A Critical Review of Empirical and Rational Strategies for Item Selection and Keying for Biographical Data Inventories

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Adam Skaja Beatty

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Paul R. Sackett, Adviser

December 2013

To my mother, Jane Ann

Abstract

Biographical data (biodata) inventories measure a person's prior experiences under the assumption that these experiences are indicative of knowledge, skills, and abilities relevant for a given purpose, typically selection. A large literature exists on methods for developing biodata inventories and how to weight biodata items, many of which are specific to the context of biodata. The current investigation reviewed the literature on the attributes of biodata inventories and compared empirical and rational methods of scale development, proposing a framework on how to conceptualize the steps involved in developing a biodata inventory. Next, using a number of large archival datasets and simulations, the effectiveness of empirical keying methods hypothesized to improve upon typically-used keying methods was tested. In archival datasets, option-keyed multiple regression was found to explain more variance in cross-validation samples than traditional alternatives, whereas a configural keying method produced results in between the two. The sample size-to-item ratio was an important factor in the extent to which option-keyed regression performed better than alternatives. The results of simulation studies indicated that in many contexts, option-keyed regression produced higher validities than traditional alternatives at 1.5 to 2 times the number of participants as items. As sample size increased, regression explained more than double the variance that traditional methods did. These findings were generally magnified by increases in the number of item response options. Practical implications and recommendations, limitations, and directions for further research are discussed.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

"The past is never where you think you left it" -- Katherine Anne Porter, *Ship of Fools*

A major and highly visible contribution of psychology to broader society is the use of psychological tests and assessments to aid in identifying individuals with the greatest potential to succeed in a given setting (Guion, 2011; Sackett, Borneman, & Connelly, 2008). In many work and higher education settings, there are large numbers of applicants for small numbers of openings. Using sound measurement tools to identify those who best fit the position in question serves both individual interests in being matched to a setting in which success is attainable and organizational interests in selecting those who will subsequently exhibit high levels of performance. Thus, developing effective techniques for personnel selection is an active research area within industrial/organizational psychology.

Many selection tools attempt to directly assess underlying attributes of an applicant. For instance, measures of knowledge and skill, such as the SAT, contain a series of questions that are designed for the purpose of obtaining an estimate of the test-taker's knowledge and skills with various cognitive problem-solving strategies. Likewise, a standard personality inventory results in an estimate of the test-taker's standing on a variety of traits, such as extraversion and conscientiousness, hypothesized to be related to performance in a particular role.

Another type of selection tool used in employment contexts is the biographical data (i.e., biodata) instrument. Used in some form since at least 1894 (Ferguson, 1961; Mumford & Stokes, 1992; Owens, 1976), biodata inventories are typically standardized paper-and-pencil measures[1] that ask applicants to report on prior experiences under the assumption that certain experiences contribute to success in a given role. Biodata inventories can have an intuitive appeal to both applicants and personnel managers because they appear to reward relevant life experience rather than performance on a test. Although initially focused largely on demographic characteristics[2], their scope has greatly broadened. For instance, a biodata inventory for a senior software engineer might ask about certifications with programming languages, the number of computer science courses taken in college, types of hobbies, or whether the applicant had ever taken on a project management role before. For large lists of broadly applicable biodata questions, see England (1971) and Farmer (2007).

Biodata inventories have been used for selection purposes within the United States Federal Government (Gandy, Dye, & MacLane, 1994), the U. S. Military (Kamp & Hough, 1986; Mumford & Stokes, 1992; Vineberg & Joyner, 1982), and a number of private companies (England, 1971). Recently, survey work has suggested that biodata inventories are not used in many organizations. For instance, Salgado, Viswesvaran, and Ones (2001) reviewed research suggesting that between .04% and 11% of U. S. firms surveyed currently used biodata in employment decisions. This number tended to be larger in other countries, with between 0% and 34% using biodata. Breaugh (2009) reported similar results based on more current studies. Part of these findings may depend on how broadly biodata is defined. Regardless of the actual incidence, use of biodata

---

[1] For exceptions, see Ployhart, Weekley, Holtz, and Kemp (2003) and Van Iddekinge, Eidson, Kudisch, and Goldblatt (2003).

[2] The 1894 study referenced above was focused on the selection of insurance salesmen and asked about marital status, amount of real estate owned, etc.

methods is usually predicated on its proposed benefits, such as ease of administration, its criterion-related validity, its potential for providing incremental validity over other common testing alternatives, and its potential for reducing adverse impact. With this in mind, I begin by reviewing the broad evidence for these benefits.[3]

Benefits of Biodata

Estimates of validity for predicting varied criteria. There is a large amount of evidence in the research literature on the criterion-related validity of biodata inventories. For brevity's sake, I will focus on and discuss the results of a selection of meta-analyses.

One of the most comprehensive and most cited references is Hunter and Hunter's (1984) classic article on "alternative" predictors of performance that could be used instead of cognitive ability tests. In addition to providing the results of their own meta-analysis, they also review the results of a number of previous meta-analyses, three of which pertain to biodata and represent largely independent samples of studies. These biodata validities range from .20 to .35.[4] Hunter and Hunter's meta-analysis includes only

---

[3] Excluding the ease of administration; given that biodata is almost always administered as a self-report paper and pencil test, this seems self-evident.

[4] There are several irregularities with the presentation of these data. One of these is Reilly and Chao (1982), who used only cross-validated observed validity coefficients, finding that the average over all criteria and occupational groups was .35. This is reported in Hunter and Hunter (1984) as .38. Another is Vineberg and Joyner (1982), which was made up of military samples, and only used cross-validated observed validities. Hunter and Hunter (1984) state that the validities were corrected for measurement error using a reliability of .60 and that the results were "derived from Vineberg and Joyner (1982)". There appear to be a large number of errors with the reporting of these results that are not explained by any systematic cause using the data reported by Vineberg and Joyner. The pattern appears to be one of aptitude tests being reported as higher than they should be, and the other tests (including biodata) being reported as lower than they should be. For instance, for the "Suitability" criterion measure, every value reported by Hunter and Hunter is the uncorrected median as reported by Vineberg & Joyner, whereas the aptitude estimate is about .05 higher than would be expected after being corrected for a reliability of .60. The only potential explanation is that Hunter and Hunter gathered all the individual studies included in Vineberg and Joyner and obtained data to calculate sample size weighted average correlations rather than the medians Vineberg and Joyner report. They don't indicate that they did this, but it is possible. Even given that, however, the previously described findings for the "Suitability" criterion seem improbable if the data were recalculated in this manner.

cross-validated coefficients, with supervisor ratings corrected for measurement error. They found that the estimated mean validity of biodata for predicting supervisor ratings was .37, with an *SD* of .10; for promotion it was .26 with an *SD* of .10; for training success, the mean was .30 with an *SD* of .11, and for tenure, the mean was .26 with an *SD* of .15.

More recently, Bliesener (1996) conducted an updated meta-analysis on observed cross-validities and analyzed a number of study characteristics as methodological and situational moderators for biodata validity. The strength of this meta-analysis is that it is able to parse apart the moderating effect of a number of study characteristics with relatively high numbers of studies per moderator level. Unforunately, it does so by aggregating across a wide range of criteria of "occupational success", which included objective performance, performance ratings, tenure, creativity, and training success, with the average observed validities for each ranging from .22 to .53. While the overall average was .30 with an *SD* of .16, for comparison purposes with Hunter and Hunter (1984), the mean for performance ratings was .32 with an *SD* of .07, for tenure it was .22 with an *SD* of .10, and for training success, it was .36 with an *SD* of .20.

Although the aggregation procedure of Bliesener (1996) does make interpretation a little difficult, a contribution is its moderator analyses by methodological decision. He finds that the vast majority of biodata validity designs reported in the literature have been concurrent (k = 135) rather than predictive (k = 24) or predictive with selection (k = 6), with concurrent cross-validities being 75% higher than predictive with selection (r = .35 vs. r = .21). Additionally, he found that cross-validity was higher when a key had been developed for that specific organization/context than when it had been developed elsewhere and applied to a new, different situation (*r* = .33 vs. *r* = .22), and that employing a double cross-validation procedure resulted in a much higher validity estimate than estimating cross-validity on a new sample (*r* = .50 vs. *r* = .28). The latter result suggests that the double cross-validation procedures where one sample is split in half may be an over-estimate of the expected validity in a different sample. These methodological characteristics have a large association with the criterion used and with each other (e.g., biodata inventories developed in a specific situation were almost always subject to cross-validation, whereas if it was a standard inventory from elsewhere, it wasn't); however, though Bliesener attempts to parse the effects apart with a meta-regression, these trends and their driving mechanisms should be regarded as preliminary.

Schmidt and Hunter (1998), in their comprehensive aggregation of meta-analyses of predictors of job performance, report the validity of biodata as .35, corrected for measurement error in the criterion and range restriction. They take this value from Rothstein, Schmidt, Erwin, Owens, and Sparks (1990), which is an estimate of the cross-validity of a key empirically-derived on a number of organizations, applied to people from 79 other organizations using the same inventory (Supervisory Profile Record) and targeted at the same job level: supervisors. Every coefficient was from a concurrent validity design. The .35 value actually appears to be the mean between the two criteria reported (a performance rating criterion, ρ = .34 and one that was a potential rating criterion, ρ = .36). Given that the study described by Rothstein et al. was specifically designed to be applicable to a wide variety of organizations, this makes it a rather unique estimate and perhaps not the best one to use if the goal is to generalize a meta-analytic estimate to the broader population of biodata inventories and contexts. That being said, the Rothstein et al. results were important because they exhibited that it was possible for a key to be widely applicable when the prevailing wisdom was that empirically-derived keys were most likely situationally-specific. Schmidt and Hunter (1998) also provided an estimate of biodata's validity for predicting training success, the .30 estimate from Hunter and Hunter (1984).

Finally, Bobko, Roth, and Potosky (1999) reported a meta-analysis of meta-analyses and large studies, calculating the mean sample size weighted average correlation and obtaining an uncorrected mean cross-validity of .28 between biodata and job performance. Their estimate included the above described studies by Hunter and Hunter (1984) and Rothstein et al. (1990), and also included an estimate of .32 from Schmitt, Gooding, Noe, and Kirsch (1984), and an estimate of .30 from Gandy et al. (1994).

In addition to the above mentioned criteria, biodata has exhibited validity for a number of others, such as leadership ratings and obtainment (Stricker & Rock, 1998), grade point average (GPA) and absenteeism in college (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Siegel, 1956), and destructive tendencies/beliefs (Gessner, O'Connor, Clifton, Connelly, & Mumford, 1993).

A critical factor for interpreting the criterion-related validity evidence for biodata is that biodata is generally considered a method rather than a construct. Similar to interviews and assessment centers, biodata inventories can be developed to tap a number of constructs (e.g., dependability, rigor of academic history, interest in sports) and are not expected to measure just one attribute (Arthur & Villado, 2008; Bobko et al., 1999; Hunter & Hunter, 1984). As a result, meta-analyses of the criterion validity of biodata should more be viewed as illustrating the range of what it could be or tends to be rather than a substantively meaningful average value. Therefore, the meta-analytic variance estimates take on particular importance as they give end-points for the expected validity of any given biodata inventory. With this strong caveat in mind, the results of the described meta-analyses indicate that biodata inventories on average can be expected to have an uncorrected cross-validity coefficient of approximately .30.

Incremental validity of biodata. The just mentioned construct-method distinction is potentially even more critical to keep in mind when trying to gauge the potential incremental validity of biodata over other selection methods. This is because it requires combining the criterion-related validity of biodata with information about its relationship with other variables. As such, it is probably more accurate to describe previous inferences about the incremental validity of biodata as inferences about the less satisfying question of the *potential* incremental validity of biodata. In other words, this would answer the question of whether it was *possible* for a biodata inventory to provide incremental validity over another measure/construct, not whether it always would. In addition, the expected incremental validity is likely to vary based on the criterion of interest. These points are important to emphasize because many times in review chapters/articles, studies of this kind are referred to as pertaining to *the* incremental validity of biodata (e.g., Breaugh, 2009; Mumford, Barrett, & Hester, 2012; Stokes &

Cooper, 2004).[5] Although this is likely due to semantic imprecision or the desire for brevity rather than a lack of awareness of the construct-method distinction, it is important to be clear on this point before attempting to generalize results. Keeping the above in mind, biodata inventories have exhibited the potential for incremental validity (either due to actual incremental validity when applied to a criterion, or simply a low correlation between the two) over cognitive ability measures (Booth, McNally, & Berry, 1978; Gandy et al., 1994; Mael & Ashforth, 1995; Mael & Hirsch, 1993; Mount, Witt, & Barrick, 2000; Pulakos & Schmitt, 1996; Schmidt, 1988; Stokes, Toth, Searcy, Stroupe, & Carter, 1999), personality (Gandy et al., 1994; McManus & Kelly, 1999; Mount et al., 2000; Pulakos & Schmitt, 1996), and interviews (Dalessio & Silverhart, 1994; Pulakos & Schmitt, 1996) in a number of empirical studies. Two studies in this domain are particularly worthy of comment.

First, Mael and Hirsch (1993) provide an interesting example of the potential for different scale construction methods to change the incremental validity of the resulting biodata measure. Specifically, they were interested in predicting four leadership criteria. They created two biodata instruments from a pool of items. One was directly keyed to the leadership criteria. The other was keyed to a number of constructs (stability, dominance, work orientation, traditional values, and energy) from the U.S. Army's Assessment of

[5] As an example, Stokes and Cooper (2004) suggest that Schmidt and Hunter (1998) "greatly underestimate the incremental validity of biodata measures" (p. 246) over cognitive ability. This is because they used a higher estimate of the correlation between cognitive ability and biodata than other researchers have found. One such study they cite was Pulakos and Schmitt (1996), who explicitly designed their biodata measure to be unrelated to cognitive ability. The correlation used by Schmidt and Hunter, on the other hand, appears to combine both biodata and judgment items (Rothstein et al., 1990; Schmidt, 1988). The point is, whatever the "true" average correlation between all biodata measures and cognitive ability is, this figure is likely not very meaningful. Schmidt and Hunter happened to use one that exhibited stronger overlap with cognitive ability.

Background and Life Experiences (ABLE; e.g., Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), and the resulting scales were termed as "biodata analogs". In other words, the goal with the analogs was to find the set of biodata items that best recreated each temperament scale. The two keying strategies differed both in the number of items they included and the weights assigned to the items. When they compared the cross-validity of these keys in a number of different samples, the incremental validity of the direct key significantly incremented over the ABLE and a "whole candidate score" (a combination of standardized test scores, high school rank, and a leadership potential score) for all criteria. However, when they examined the incremental validity of the ABLE "analog" scales over the ABLE and the whole candidate score, Mael and Hirsch (1993) found they only significantly incremented for one of the four criteria. In addition to showing that a set of biodata items has the potential to increment over cognitive ability and personality constructs (i.e., the scale directly keyed to the criteria), this study illustrates the effect that choices in the scale-development process can have on how well a biodata scale will increment (i.e., the results for the "analog" scales).

The next study of note, by Bobko et al. (1999), is a meta-analytic investigation of the intercorrelations between biodata and cognitive ability (r = .19)[6], conscientiousness (r = .51), structured interviews (r = .16), and job performance (r = .28), and is largely a re-analysis of a study by Schmitt, Rogers, Chan, Sheppard, and Jennings (1997). The

[6] There are two oddities with this value, both arising from Bobko et al.'s (1999) use of a correlation from Pulakos and Schmitt (1996). Both Bobko et al. and Schmitt et al. (1997) report this as .05, whereas the value from Pulakos and Schmitt appears to be .00. Second, Bobko et al., in discussing the Pulakos and Schmitt study, note that it did not study cognitive ability broadly, but instead focused on a verbal ability composite. As a result, they did not include it in the meta-analysis for the relationship between cognitive ability and job performance. However, they still included it in the meta-analysis for cognitive ability and biodata, which appears inconsistent.

number of studies involved in all analyses is fairly small, and to the authors' credit, they draw attention to the construct-method distinction, particularly in discussing the correlation between biodata and structured interviews (both of them being methods rather than constructs).

Although Bobko et al.'s (1999) results relating to biodata cannot be interpreted as a solid estimate for what any given biodata inventory's correlations with these constructs would be, they are potentially useful from a policy-capturing standpoint (i.e., based on inventories studied, which constructs do they overlap with?). The clear area of most overlap is with conscientiousness. This perhaps makes sense given biodata's history as a so-called "alternative" to cognitive ability for predicting job performance (Hunter & Hunter, 1984). Outside of knowledge and skills, a natural area for biodata items to tap is how dependable applicants are.

Overall, biodata inventories certainly have the potential to provide incremental predictive power over other common selection tools. However, if this is the goal, scale developers should be intentional about it from the beginning of the scale development process, as this bears on the type of items and/or scales that should be included in the inventory.

Potential for reduced subgroup differences. As just noted, part of biodata's appeal is that it is seen as an alternative to cognitive ability tests and can be expected to exhibit less adverse impact, while still being predictive of workplace outcomes (Hunter & Hunter, 1984; Pulakos & Schmitt, 1996; Reilly & Chao, 1982). Investigations of this idea have largely supported it. For instance, Terpstra, Mohamed, and Kethley (1999), in a review of legal challenges to selection methods, found no challenges to biodata inventories in the cases they reviewed. Reilly and Warech (1994) reviewed the literature on biodata and suggested that biodata has less adverse impact than cognitive tests, and "appears to meet the standard definition of fairness to minorities" (p. 145). Hough, Oswald, and Ployhart (2001) summarize the available research on biodata inventory as largely exhibiting small group differences (indexed by Cohen's *d*) between majority and minority groups.

As an example of smaller group mean differences in the biodata domain, Bobko et al. (1999), in the same meta-analytic study described above, also calculated a meta-analytic *d* for the Black-White mean difference on biodata inventories. In this case, they weighted the results of two studies to arrive at a mean *d* of .33 favoring the White group. One included study, Pulakos and Schmitt (1996) had an effect size of -.05 (n = 357), whereas Gandy et al. (1994) was reported as .35 (n = 5,758). However, the effect size from Gandy et al., referred to the initial scale. After item screening, Gandy et al.'s final operational scale had a *d* of .27 (n = 5,758). Thus, using Bobko et al.'s sample size weighting approach, the best estimate of the Black-White mean difference on an operational biodata scale based on these two studies is .25, again favoring the White group. This can be compared to an estimated Black-White mean difference of 1.10 standard deviation units for cognitive ability tests (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). The two studies included in the Bobko et al. meta-analysis also allow for the calculation of a Hispanic-White *d*-value (.08, based on a total sample size of 5,507). Additionally, the Gandy et al. study found a *d*-value of .15 in favor of females over males. Finally, Richardson, Bellows, Henry and Co. (1988), when applying a differential prediction model to their biodata inventory, reported overprediction for minorities (see Reilly & Warech, 1994).

On the other hand, there have also been some legal-related cautions about the use of biodata. For instance, Sharf (1994) and Pace and Schoenfeldt (1977) both suggest that biodata inventories, particularly specific biodata items, may be difficult to defend if they are based on blindly empirical key development procedures. This criticism will be discussed further in the section "Differences between Rational and Empirical Models" below. Along the same lines, Schmitt & Pulakos (1998) found that a number of their rationally-derived biodata subscales had no or worse validity for Black and Hispanic groups, when compared to the White group. Between 15% and 60% of the items within each scale were flagged as exhibiting differential effects by racial subgroup. They provided some post-hoc explanations for these differences and above all, suggested caution when developing biodata measures. Stokes and Cooper (2004) suggested that Schmitt and Pulakos' results point to the potential for greater racial group differences than previously thought.

In summary, although it can probably be safely assumed that most biodata inventories will exhibit lower mean differences across racial/gender groups than cognitive ability tests[7], Schmitt and Pulakos' (1998) suggestion of caution is certainly warranted. Given the potential variation in constructs tapped by biodata inventories, it

---

[7] Exceptions can certainly be thought of. For instance, any biodata item that might overlap with the physical abilities domain (e.g., "Have you ever bench-pressed over 100 pounds?") might be expected to have large Male-Female group differences, and these can sometimes be up to two times those that occur for Black-White *d*s on cognitive ability (Hough, Oswald, & Ployhart, 2001).

cannot be assumed that racial group differences will be less than .30 for every inventory. If a biodata inventory is going to be used in a battery of assessments for high-stakes selection, the operational predictive bias should be assessed according to professional guidelines (e.g., *Principles for the Validation and Use of Personnel Selection Procedures*, SIOP, 2003).

What is Biodata?

I have discussed the potential benefits to be gained from the use of biodata inventories. However, given the construct-method caveats that hinder interpretation of many of these proposed benefits, a more specific delineation of the domain of biodata is needed. Unfortunately, there is no universally accepted definition (Nickels, 1994). To illustrate the potential confusion, Stokes (1999) states, “Biodata forms tend to differ from measures of personality, interests, values, and abilities, but often capture constructs in all four domains” (p. 111). Still, I now review the attempts that have been made to sharpen the boundaries of what is and is not biodata. These take three forms: general definitions, creating taxonomies of biodata items, and comparisons with other data-collection methods.

General definitions. The first source of information about what biodata is comes from general definitions of the method. Most simply, Hough (2010) states that “...biodata are the individual’s life history, and biodata measures gather that information and score it” (p. 110). Nickels (1994) defines biodata items as those that “require people to describe behaviors and events occurring earlier in their lives” (p. 2) and states that “if there is a single characteristic of biodata measures that sets them apart from other life history measures, it is the use of a retrospective self-report to formulate a general description of a person's life history" (p. 5). Similarly, Mumford, Whetzel, Murphy, and Eubanks (2007) define biodata as "an assessment technique in which items are defined by the nature and structure of people's lives as they unfold over time" (p. 205). Mumford et al. (2012) propose three key attributes exhibited by all biodata measures: (a) a focus on behaviors/experiences that happened in an individual's past, (b) all individuals respond to the same set of questions, and (c) responses to items are provided by the individual (i.e., self-report).

In a general sense then, these definitions appear to have a high level of agreement. Biodata inventories are standardized self-report measures that ask about the past. This focus on the past tends to be agreed on as the common element to all definitions of biodata (Breaugh, 2009; Stokes et al., 1999). However, a step below this level of generality, disagreements start to appear. For instance, should only past behavior be captured, or are prior feelings and attitudes fair game (Guion, 2011)? Should items be focused on a specific instance of behavior, or can items query general trends of behavior? To further develop these ideas, researchers have proposed taxonomies of biodata items.

Taxonomies of items. There have been a number of attempts to develop taxonomies of biodata items, with the goal of deciding what should and should not be considered biodata. Because they are informative as to the types of biodata questions that can be encountered, I will describe them in depth. One of the first taxonomies was Asher (1972). He proposed that biodata items can vary on a number of dimensions, including:

- Is the subject of the item *verifiable* or *unverifiable*? An example of a verifiable item might be “How many years did you spend at the job you worked at longest?”, whereas an unverifiable item might be “What was your favorite job?”
- Is the item *historical* or *futuristic*? An example of a historical item might be “Did you have roommates in college?”, while a futuristic item might be “When do you plan to retire?”
- Does the item tap *actual behavior* or *hypothetical behavior*? An actual behavior item could be “Have you run a marathon?”, while a hypothetical behavior might be “If you trained for three months, do you think you could complete a marathon?”
- Does the item require *memory* or *conjecture*? An item that requires memory might be “Did you go on camping trips with your family when you were young?” An item that is based on conjecture might be “If your family lived in the country, do you think you would have gone on camping trips with them when you were young?”
- Is the item *factual* or *interpretive*? Asher’s (1972) example of this distinction is that a factual item might be “Do you repair mechanical things around your home such as appliances?” and an interpretive item is “If you had the training, how would you estimate your performance as an appliance repair man?” (p. 252).
- Is the item asking about a *specific* behavior, or *general* classes of behaviors? For instance, a specific item might ask whether the individual played football in high

school, whereas a general question might ask whether he/she participated in sports.

- Does the item relate to a *response* or a *response tendency*? The example given by Asher (1972) of a response is “Which of the following types of cameras do you own?” and of a response tendency is “In buying a new camera, would you most likely purchase one with automatic features?” (p. 252).
- Does the item ask about an *external event* or an *internal event*? An external event might be “How many articles have you published?”, and an internal event might be “How important do you view publishing?”

While many of these categories are potentially overlapping (e.g., all items that are verifiable are likely to be external, and internal items are likely to be unverifiable), Asher (1972) did a good job of summarizing the ambiguity that can occur when trying to develop biodata items. He concluded that there was a distinction between “hard” and “soft” biodata items, with “hard” biodata items being historical and verifiable and “soft” items being unverifiable and having responses that are “expressed in abstract value judgments rather than realistic behavior” (p. 263). He expressed a preference for the “hard” items.

Almost 20 years later, Mael (1991) continued the work of Asher (1972) and proposed a taxonomy of his own, separating biodata characteristics into three categories and highlighting contrasting poles. For each pair of opposites, he suggested that the first one he listed was either more rigorous or characteristic of what biodata should be. First, he suggested that the historical vs. future/hypothetical distinction was the main dimension that defined an item as biodata, and that it was important to focus on what the person actually did or what had actually happened rather than on some abstract idealized future state. He further stated that a focus on the historical would also preclude broad generalizations over time or situations (e.g., “Would you describe yourself as X?”), but would still allow for asking about what individuals typically have done in the past.

Mael’s next category of attributes concerns what he terms “methodological stringencies” that are supposed to aid in obtaining accurate/correct reports of behavior. These include distinctions between external and internal, objective and subjective, first-hand and second-hand, discrete and summative, and verifiable and non-verifiable items. These are similar to distinctions made by Asher (1972). The external/internal distinction is focused on if events are external or internal to the individual (i.e., behaviors vs. thoughts and feelings). The objective/subjective distinction is one between items that require only recall of an event and those that require interpretation. The first-hand/second-hand attributes refer to whether the individual is being asked to report on knowledge they have about themselves or whether they are being asked to estimate what others think or how others would respond to the item. The discrete/summative distinction refers to a specific behavior versus a summary response that requires some sort of estimation. While Mael suggests that discrete responses may be more accurately responded to as they only require memory retrieval, he acknowledges that summative items may be more appropriate when focusing on tendencies or regularly-performed behaviors that are not remarkable and would therefore not stand out and be easily recalled. The final set of attributes in this category is whether items are verifiable or non-verifiable. This refers to whether an item can be verified by some source besides the individual (e.g., records, observer testimony). He distinguishes between items that are verifiable through records and the less stringent "verification in principle". The latter definition would include all external behaviors in the presence of others, regardless of how unlikely it would be to obtain verification. He concludes that while there is no intrinsic reason to limit biodata to verifiable items, it may help with response distortion.

Mael's (1991) final set of attributes are focused on moral, fairness, and legal concerns, with the first attribute of the following pairs being judged to be most likely to avoid fairness and legal objections. As a result, rather than being proposed to more accurately define biodata, these attributes are proposed with practicality in mind. In fact, in his elaboration of each distinction, Mael seemed favorable to the second of the pair as exhibiting more biodata purity. These categories are controllable/non-controllable, equally accessible/non-equal access, job-relevant/non-job-relevant, and invasive/noninvasive. The controllable/non-controllable distinction refers to actions or behaviors the individual chooses to perform or not perform, whereas things that were done to or happened to them are non-controllable. While non-controllable items are potentially valid as indicators from the past that could explain present or future behavior (e.g., parental or demographic information), a common ethical guideline for testing is that applicants should not be penalized (or advantaged) for things they had no control over. In addition, some non-controllable items could be associated with membership in protected classes (e.g., indicators of parental socio-economic status). As a result, careful consideration of the types of non-controllable items being included in a biodata inventory was recommended by Mael (1991), though he also noted that the elimination of all non-controllable items would remove a great deal from typical biodata item pools. The equally accessible/non-equal access distinction relates to experiences and skills that are equally accessible to all individuals. For instance, asking about smartphone usage might exclude those with lower socio-economic status. Again, Mael argues that a strict interpretation of this requirement might limit the pool of external and objective items.

The job-relevant/non-job relevant distinction is based on how overtly related to job content the items are. More cautious item-developers recommend a clear point-to-point association between items and parts of the job (Mael, 1991). Mael cautions that the items most narrowly classified as job-relevant would also be those that are the easiest to fake. The final pair of item attributes is invasive/noninvasive. Although privacy protection laws are a concern, in many cases, whether an item is deemed invasive or not will be subject to individual interpretation. Mael, Connerley, and Morath (1996) conducted follow-up research and found that the broad areas seen as most invasive were items that were seen as being related to past trauma, fear of stigma, religious behavior, and intimacy (e.g., sexual behavior, having had a miscarriage).

Mael's (1991) taxonomy appears to have become an obligatory citation for overview chapters and articles on biodata (e.g., Breaugh, 2009; Mumford et al., 2012; Mumford et al., 2007; Stokes & Cooper, 2004), even though, as Nickels (1994) notes, besides the historical perspective, there is little agreement on any of the attributes. Guion (2011), for instance, agrees with almost all of the attributes Mael identifies as more characteristic of biodata, with some possible exceptions for non-controllable events because of the potential influence that events beyond a person's control can have on present behavior.

Hough (2010), on the other hand, appears to define biodata very broadly, suggesting that biodata measures can vary in terms of (a) whether quantitative or qualitative information is gathered, (b) whether typical or maximal behavior is sought, (c) whether self-evaluative or factual information is gathered, (d) the amount of structure in the response (e.g., multiple choice or free response), (e) whether the response options are continuous or categorical, (f) the source of the information (i.e., self or other), (g) how the questions are scored (i.e., rationally or empirical), and (h) whether responses are given in written form or in an interview. The distinction of Hough (2010) that may differ most from previous definitions and taxonomies is the allowance of both self-evaluative and factual information. Similarly, Mumford et al. (2007) provide their own taxonomy of types of content that could be asked about in a biodata inventory. These include situational exposure (e.g., how many times...), situational choice (e.g., decision to do something), behavior in situation, reactions to a situation, others' reactions to a situation, outcomes associated with situational exposure, life narratives, and negative life experiences. Some of these content areas veer away from the spirit of the distinctions proposed by Mael (1991) and Asher (1972) (e.g., reactions to a situation are internal, and others' reactions to a situation focus on someone else).

Outside of the broad agreement any of the above distinctions have enjoyed, there has been a small literature of empirical research that attempts to evaluate the effects that specific item types have on relevant outcomes. Although these do not add insight into the distinction between what is and is not biodata, they attempt to offer some guidance on what is preferable for validity or faking purposes. For instance, McManus and Masztal (1999) rated 255 items on each of Mael's attributes and correlated these ratings with item validity (in the prediction of survival in a sales role) and social desirability. They found that the more historical, external, objective, discrete, and verifiable an item was, the higher the validity that was obtained. There were negative relationships between validity and both job relevance and equal access. None of the attributes exhibited a significant relationship with social desirability. Lefkowitz, Gebbia, Balsam, and Dunn (1999), in a very similar design, correlated ratings of 160 biodata item attributes (largely based on Mael's taxonomy) with item validity (in the prediction of job performance ratings). They found that for items identified as valid by an empirical keying procedure, verifiability and indirectness (i.e., not first-hand, but reporting the assessments of others) were related to validity. Given the limited sampling and uncertain overlap of predictive constructs/criteria, it is difficult to reconcile the two studies' findings on the first-hand attribute.

Becker and Colquitt (1992), Graham, McDaniel, Douglas, and Snell (2002), and Harold, McFarland, and Weekley (2006), on the other hand, looked at the effect of item attributes on presumed response distortion. In general, the common thread is that biodata items can be faked (either via directed faking or by comparing applicant and incumbent samples), harming validity, and unverifiable items appear more likely to be faked than other item types.

The main difficulty with interpreting all of the empirical research comparing item attributes is that little attempt has been made to systematically control for the explanatory construct. If a certain item attribute was more heavily represented among items tapping a specific construct that was more valid than the constructs the other items in the pool tapped, this could make the higher item validities for the attribute appear to be a function of the attribute when it is really due to the underlying construct of the items. Admittedly, this is primarily a problem for attribute-validity relationships, as there doesn't seem to be a plausible reason why faking would be more or less likely depending on the construct beyond differences in facial job-relatedness. To optimally test the attribute-validity relationships, a study would have to be conducted where attributes were varied within a particular construct, the narrower the better. In any case, regardless of any conceptual reasons for preferring certain item attributes over others in a definition of biodata, it appears that there is at least some initial support that some of the attributes proposed by Mael (1991) have desirable implications for validity and resistance to faking.

Overall, taxonomies provide both illustrations of the potential for ambiguity and a means for researchers to provide examples of attributes and items that meet their personal definitions of what biodata is. Although the development of taxonomies hasn't succeeded in resolving all disagreements, they have certainly helped by providing a language through which researchers are able to talk about their differences and classify themselves on a set of continua. The final source for defining biodata comes from explicit statements of comparison with other data collection methods and constructs.

Comparisons with other data collection methods and constructs. First, it is important to differentiate this section with the above section “Incremental Validity of Biodata”. The previous section focused on (a) the difficulty in determining incremental validity due to the construct-method distinction, and (b) empirical findings of relationships between biodata inventories and other assessment tools. This section, on the other hand, will focus on researchers’ attempts to define what biodata is by contrasting it with definitions of other selection methods. The previous section, therefore, was more descriptive, whereas this section will be more prescriptive.

Biodata and standard personality inventories. The distinction between personality and biodata has probably received the most attention because personality constructs are often tapped by biodata, and stretching the definition of biodata to include subjective and internal items can lead to items that look identical to those that would be seen in a standard personality test. For instance, “Do you like to go to parties?” is a common example of a personality item that indicates extraversion, and because it is based on past behavior to some degree, could plausibly be included in a biodata inventory that allows internal items. Even if the same items are not included, Mumford and Stokes (1992) suggest that biodata items often look like variants of the type of questions in personality inventories. This confusion is exacerbated by the subtle shift toward describing biodata in construct terms (as opposed to method) to illustrate the difference between biodata and personality. Because biodata can measure personality constructs, it is important to focus on the difference between the two as a difference of typical measurement method.

A number of researchers have put forth reasons why biodata differs (or should differ) from self-report personality measures. Mumford and Owens (1987) suggested that biodata is different from personality because biodata does not focus on behavioral tendencies and instead focuses on prior behavior in specific situations. Mumford, Snell, and Reiter-Palmon (1994) proposed that there are two additional differences between the two: biodata inventories often tap content that is more similar to knowledge and skills, and biodata is often associated with choices, rather than broad preferences. Mael (1991) suggested that temperament and personality inventories may have a greater tendency to focus on subjective internal items and might avoid complex behaviors with multiple determinants (i.e., behaviors that are not pure measures of a specific construct). In another publication, Mael (1994) pointed out an interesting potential dynamic, whereby one researcher might argue that there is no difference between biodata and personality, so he or she develops a biodata inventory that has self-report personality items, then other researchers say that these hybrid measures "prove" that there is no difference between the two when the truth is more that the hybrid measures are misnamed.

Perhaps the clearest distinction between biodata and standard personality items can be made if the distinction described above between "hard" and "soft" biodata is used. It is much more difficult to equate biodata and self-report personality measures when "hard" biodata inventories are considered. External and verifiable behaviors are rarely the type of items seen in self-report personality inventories.

Biodata and training and experience measures. Training and experience (T & E) measures represent a special case in that they can primarily be viewed as a subset of biodata. More specifically, work experience is often framed in terms of tenure or age (Quinones, Ford, & Teachout, 1995; Tesluk & Jacobs, 1998), and McDaniel, Schmidt, and Hunter (1988) describe T & E methods as judgments about information provided by applicants on resumes and other application materials. This type of information can likely be assumed to fit under the objective, external, and verifiable categories of biodata items. Further, McDaniel et al. state that the primary distinguishing characteristic between T & E methods and biodata is that with T & E measures, judgment is used to weight information rather than empirical means. In other words, this is the distinction between rational and empirical keying of biodata that will be discussed further on in this section. Finally, the rationale given for why T & E measures should be related to performance is the same as that for biodata (i.e., indicators of prior knowledge, skills, and abilities are related to performance; McDaniel et al., 1988). As a result, T & E measures can be subsumed under the broad umbrella of biodata, but with a focus on a specific subset of work-relevant prior history. Indeed, Hough (2010) cites two meta-analyses of T & E measures (McDaniel et al., 1988; Quinones et al., 1995) as evidence for the predictive validity of biodata, though as noted earlier, she assumes a rather broad definition of biodata.

Biodata and other selection tools and constructs. There are a number of other constructs or methods that are differentiated from biodata in the literature, but only by a few researchers. As an example, when biodata items are typically differentiated from cognitive ability and knowledge items, it is because ability items are more focused on maximal performances rather than typical behavior and what might have occurred earlier in a person's life (e.g., Mitchell, 1994; Mumford, 1999). Traditionally, ability items also involve completing some sort of specific problem-solving task, whereas biodata items might be more centered on broad domains of knowledge obtained, like with training and experience measures. Interviews have also been distinguished from biodata in that, though both are methods of capturing life history information, biodata can be both more efficient to administer and have less potential for bias due to the lack of needing an observer (Mumford & Owens, 1987). At the extreme however, interviews and biodata could be almost totally redundant with each other if an interview consisted of reading biodata questions and response options. Overall, there are definitely rationales for separating biodata from other methods of data collection.

Does it matter? I have just reviewed literature that attempts to clarify what biodata is and is not. A pertinent question is whether it matters. One could argue that if a biodata inventory is predictive of criteria of interest, then the precise attributes exhibited by items is of little importance. In fact, all else equal, including item types from other domains (e.g., personality, interests, values) should increase predictive power, assuming each item is individually keyed to the criterion. However, even if predictive validity is the primary goal, there are a number of very practical and applied questions that are relevant to the definition of what biodata is. For instance, Mael (1991) points out that a finding that biodata measures have higher validity than temperament/personality measures is uninterpretable without knowing what biodata is and how it is different, and additionally, if they are defined to be essentially the same, then proposed benefits from using biodata over personality/temperament (e.g., resistance to social desirability) do not make sense

(Mael, 1994). In other words, deciding whether to use a biodata measure in the first place must be based on an estimate of what its benefits compared to alternatives are. Similarly, in deciding on a testing battery, it would be useful to know which predictors can be expected to be redundant and which would add incremental validity. This isn't possible if biodata is defined to simply be a medley of self-report items. Again, biodata items can be written to tap any construct, including personality constructs, but the question is how they are written and framed. Of course, these questions become much more important when attempting to explain rather than just predict.

Overall then, I concur with many others that the distinction between biodata and other measures does matter. Unfortunately, I do not have any new ideas for fine-tuning this distinction. Instead, I agree with Asher (1972), Guion (2011), and Mael (1991) that keeping biodata limited to historical, objective, and external items has the most promise for the definition of biodata as meaningfully set apart from other assessment methods. This way, it would be possible for an assessment battery to have a standard self-report personality measure, a biodata measure, and a work simulation, and have them all provide a measure of, say, conscientiousness that increments over the others. Additionally, and just as important, it would be possible to say *why* each measure provided incremental information over the others. With this definition of biodata as focused on historical external behavior, a theory for why it predicts performance can be elaborated.

Why does Biodata Work?

The one sentence rationale for why biodata works, given in practically every reference discussing biodata, is the adage that past behavior is the best predictor of future behavior. This is the principle that, for the most part, people are consistent. Mael (1991) pointed out that strict interpretation of this principle would limit the scope of biodata to previous actions of the same type (i.e., more sample than sign in the terminology of Wernimont & Campbell, 1968), so the mechanism must be elaborated. Similarly, Dean, Russell, and Muchinsky (1999) pointed out that this rationale doesn't adequately *explain* anything. As such, a number of attempts have been made to expand this rationale to be more broad and explanatory. For instance, Dean et al. stated that past behavior is "hypothesized to capture causal events or correlates of causal events that influence job candidates' behavior" (p. 24). Similarly, Nickels (1994) suggested that "biodata measures may predict performance across so many aspects of behavior as well as they do because responses to biodata items may serve to capture previous manifestations of the constructs and mechanisms that ultimately determine predictive relationships with criteria" (p. 2). In fact, this is a good summary of the ecology model, which is the most widely cited and heavily elaborated theory for why prior history is predictive (Mumford, Stokes, & Owens, 1990).

Ecology model. The foundation for the ecology model is Cronbach's (1957) call to integrate the experimental and correlational "disciplines" of the field of psychology. Owens (1968) responded to this call with what he termed the developmental-integrative model, and with his students and collaborators over the next few decades, expanded it into the ecology model (Mumford et al., 1990). The basic premise of the model is that the individual is “an active, purposeful entity, who through learning, cognition, and action seeks to maximize personal adaptation in a world of shifting environmental opportunities” (Mumford & Stokes, 1992, p. 77). Though all people start out with a set of individual differences based on hereditary and environmental resources, they soon begin the process of adaptation to their environment. An individual will make choices about which situations to enter based on the perceived valence of the outcome that is likely to be obtained. These valences are based on the needs and values of the individual. However, once these decisions are made, further adaptation and choices are needed to attain desired goals in the new situation (Mael, 1991). As a result, the ecology model is interactive; an individual’s choice about what situation to enter will affect their subsequent development, and thus, their future decisions (Breaugh, 2009). In order to satisfy all values and needs, a person will enter a wide variety of situations, and long-term adaptation is maximized to the extent that the selection of situations and developed competencies are complementary rather than competing (Mumford & Stokes, 1992). As a result of this iterative process, the individual develops a coherent and cohesive pattern of choices and differential characteristics. Thus, this model implies that a wide variety of previous behaviors can be expected to predict subsequent performance, even if they don’t facially appear that they would do so. This model, therefore, literally tries to describe why past behavior is linked to future behavior.

Some critics of the ecology model note that it places too much emphasis on the adaptive successes and not enough on failures, which can also be formative experiences (Dean et al., 1999; Farmer, 2007; Mael, 1991), and that the importance it places on choice behavior downplays the effect of environmental influences on the person. Still, this is widely regarded as the most complete theory for why biodata items and/or past behavior are predictive of future performance.

Social identity theory. The main attempt beyond the ecology model to explain why biodata is predictive is Mael's (1991) application of social identity theory, which he describes as a complement to the ecology model. Mael proposed that the emphasis of the ecology model on choice behaviors and people being active participants in their environment tended to limit the focus on environmental effects. Thus, proponents of the ecology model are not as focused on things that "happen to" a person (Farmer, 2007). In order to address this, Mael suggests that the concept of the social identity adds additional explanatory power to biodata as the combined effect of psychologically belonging to a number of perceived social categories. These social categories can include political parties, religions, fraternities, gender, and any number of other informal or formal groups. Identification with a group/category implies the perceived sharing of the characteristics of other members of that group. Applied to biodata then, social identity theory would suggest that in addition to the adaptive choices that were made, the effects of the psychological groups someone has belonged to can also affect their future behavior. In Mael's words, "biodata items encompass not only the choice-based, adaptive responses of the individual, but also the effects of all characteristics internalized through identification with the myriad psychosocial entities with whom one interacts throughout life" (p. 768). Further elaboration of social identity theory and its purported role in the organizational sciences is given in Ashforth and Mael (1989).

Implications of explanatory models. The primary practical implication of the explanatory models for why biodata works is that to ensure predictive power of a given biodata inventory, a thorough understanding of both the criterion and predictor space is needed, with special attention paid to what specific past events would be indicative of effective or ineffective performance. Building on this idea, Mumford et al. (2012), concluded that “biodata inventories will prove most effective when they capture behavior and experiences in developmentally significant situations with respect to the criterion of interest” (p. 357). They add that more recent behaviors and experiences are more likely to be predictive than more distal ones, and that biodata items can potentially draw from a wide variety of content related to previous experiences, outlined earlier in the Mumford et al. (2007) taxonomy. Related to these points, Nickels (1994) suggests that biodata items developed with specific hypotheses in mind about relationships between the predictor and criterion domain are more likely to result in inventories with high criterion-related validity. Empirical research by Reiter-Palmon and Connelly (2000) supports this idea.

Mumford, Costanza, Connelly, and Johnson (1996) further differentiate the hypotheses that can be developed as being either job-oriented or worker-oriented (referred to by other authors as direct and indirect, respectively; e.g., Mumford & Owens, 1987). Job-oriented hypotheses are in reference to the specific criterion/performance. Thus, they focus on past behaviors similar to those in the criterion domain. These could be either prior manifestations or direct antecedents of the criterion behavior of interest. For instance, items developed for a new college professor might ask about prior success when explaining or presenting information to others. Items developed in this fashion tend to have a high degree of face validity. However, job-oriented biodata items generally assume that the individual has done something very similar to the targeted performance in the past. If this is not the case or if more than one criterion is of interest, it may also be useful to take a worker-oriented approach. This means that the items are developed in order to identify an individual's standing on constructs on which performance is based (e.g., cognitive ability, conscientiousness). For instance, a worker-oriented item developed to tap conscientiousness might ask how long it has been since the respondent cleaned his/her house. While this is not obviously related to performance on most jobs, it might provide information on the respondent's standing on the orderliness facet of conscientiousness.

With the preceding in mind, there are a number of sources where content for potential content- and criterion-valid biodata items can be mined. Broadly, three proposed areas for where predictive hypotheses can emerge are the human development literature, life history interviews with incumbents, and typical factor loadings of biodata items (Mumford & Owens, 1987; Russell, 1994). More specifically, Fine and Cronshaw (1994) and Hough (2010) suggest the usefulness of functional job analysis. With its task, attribute, and behavior-based information, it can be uniquely suited for producing biodata items situated in a job-relevant context. Another potentially useful method is Flanagan's critical incidents method (1954), which Hough (1984) used to develop a biodata inventory called the Accomplishment Record for attorneys. Its development involved gathering a number of effective and ineffective job behaviors from current attorneys and categorizing the behaviors into dimensions. Then, attorneys were asked to describe their most important accomplishment related to each dimension and it was these accomplishments that were scored according to a calibrated rubric and served as the biodata "items". Finally, both Hough and Paullin (1994) and Mumford et al. (2012) endorsed construct-oriented scale construction. Briefly, this is where constructs are identified based on a job analysis or conceptual association with the criterion, and items are written to tap these constructs. This process will be described in more detail in the following section.

In summary, the models for why biodata is predictive of current/future performance are largely explanatory and provide a theoretical underpinning for the domain. However, they also imply a number of practical principles that can be applied to biodata item development to obtain items that have a direct link to criteria. As the following discussion will show, these hypotheses and methods that connect prior behavior to the criterion domain are the source of a major distinction between approaches to biodata scale development.

Models for Biodata Scale Development

With an understanding of what biodata is and why it can be expected to work, items can be written and scales can be developed. However, not all written items will exhibit empirical relationships with criteria, and even the directionality of the correlation between a specific past behavior and a criterion can be unknown. In contrast with ability items, there is not always a clear "right" answer. Thus, a number of approaches have been developed that attempt to determine which items should be used, the directionality they should be keyed in, and how much each item/item response should be weighted. The two primary approaches are the rational and empirical models.

Rational model. A purely rational perspective implies that decisions of item selection and weighting are made rationally (Hough, 2010). In other words, it proceeds entirely on the basis of expert judgment and understanding of the content domain. The rational approach is sometimes referred to as the intuitive or deductive approach, borrowing from the history of scale-development approaches in the personality domain (e.g., Burisch, 1984). Items are commonly rationally tied to constructs to create meaningful and coherent sets that can be summed to create sub-scales, which are then used to predict a desired criterion.

Common suggested benefits of the rational perspective include its interpretability as a result of specific hypotheses made by experts for why items should be predictive, greater transportability of the resulting scale to other contexts because of its focus on constructs rather than specific items/behaviors, and the ability to develop scales without item response and criterion data (Hough & Paullin, 1994; Mael & Hirsh, 1993; Mumford, 1999; Mumford & Owens, 1987). Tied to its greater interpretability is the belief that rationally-developed or screened biodata scales will have less potential for legal scrutiny (Pace & Schoenfeldt, 1977; Sharf, 1994). Some criticisms of the rational approach are that it is based on expert judgment which is potentially fallible, that items developed rationally are more likely to be transparent and hence faked, and that rationally-developed scales may exhibit lower predictive validity (Hogan, 1994; Mitchell, 1994; Mumford & Owens, 1987).

In most descriptions of the rational method, item development is also a critical feature (Hough & Paullin, 1994; Mumford & Stokes, 1992; Mumford et al., 2012) and part of what lends it its assumed interpretive superiority. Via the item-development strategies discussed earlier (i.e., job-oriented and worker-oriented) items are written to specifically target a set of constructs or behaviors thought to be related to performance. The method of scale development most often linked to the rational approach in recent years is construct-oriented scale development (e.g., Hough & Paullin, 1994; Mumford et al., 2012; Mumford et al., 1996). Hough and Paullin (1994) refer to construct-oriented scale construction as an evolution and advancement of the standard deductive method. The most recent description of this process was provided by Mumford et al. (2012). First, items are written to reflect expression of prior behaviors hypothesized to be related to relevant constructs. Then, 20 to 40 items are written per construct by trained item writers.[8] The items are then administered to 200 to 400 people, ideally as similar to those the test will be used for as possible, and items that fail to exhibit variability and acceptable item-total correlations with other items tapping the construct are removed. They suggest that somewhere between 1/3 to 1/2 of the items are removed through this process. Finally, they recommend a content review to ensure that there is still adequate construct coverage among the remaining items. As a result, construct-oriented scale construction does require item level data, just not criterion data. It is therefore a cross between intuitive and internal methods of scale development.

[8] Previous work by Mumford and his colleagues describing the item writer training process suggested that this was a major time investment, requiring up to 80 hours in training on the attributes of good and bad items, how to write items to load on a construct, and practice/feedback session (Mumford et al., 1996).

It is important to note that factorial models have sometimes been considered to belong to the rational approach (Breaugh & Dossett, 1989). This is due to the development of post-hoc hypotheses about why empirically-developed scales work by identifying underlying factors (i.e., searching for some level of rational explanation). Alternatively, factorial and rational models have sometimes both been considered to belong to the same "internal" class of scale-development methods (in contrast to the external/empirical class; Brown, 1994). For the purposes of this dissertation, I follow Hough (2010) and her distinction of deductive (rational) versus inductive (factorial) and assume that the prototypical rational model only involves the specification of *a priori* hypotheses and expert judgment on selection and weighting as described above, with the factorial approach being separate and described below.

Empirical model. In contrast to the rational model, the pure version of the empirical (sometimes called the external) model involves the selection/retention and weighting of items based on an empirical relationship with a criterion (Hough, 2010). Hogan (1994) suggests seven steps that are required in the empirical model: (a) choosing or developing the criterion, (b) identifying high and low criterion groups if needed, (c) selecting items, (d) specifying item response alternatives, (e) weighting items/response options, (f) cross-validating, and (g) developing cutoff scores. A number of these steps require elaboration.

First, empirical scaling methods require data. In fact, Hogan's (1994) seven steps should have a step added between step d and step e: the need to obtain a dataset of responses to biodata items and criterion data. Biodata items or item responses are given weights based on a metric of association with the criterion. A wide variety of methods for empirically-keying biodata items will be discussed in the last portion of the introduction. Tied to the need to obtain data is the further requirement that all empirically-keyed biodata scales be subjected to cross-validation procedures. Because they are keyed using purely data-driven methods, item weights will be most calibrated to the sample they were derived in and the observed validity will "shrink" in other samples. Therefore, in order to obtain an appropriate estimate of how well the key actually works in predicting criterion performance of new people, such as new job applicants, the key must be applied to a new sample taking the same set of items.

Second, the identification and justification of the criterion is a critical step in the development of an empirical key (England, 1971; Hogan, 1994; Mumford et al., 2012; Thayer, 1977). In contrast to the rational approach, where the criterion domain is used to develop predictive hypotheses and subsequently items, the pure empirical approach's entire process rests upon empirical relationships with a specific criterion. Therefore, it is imperative that the criterion measure be representative of the targeted performance. Any deficiencies/contamination of the criterion will be present in the resulting empirical key. Related to this, a representative and appropriate sample is needed to ensure that the key will generalize appropriately.

Broadly, the most commonly touted benefit of the empirical approach is its criterion-related validity (Hogan, 1994; Mumford & Owens, 1987); given a large enough and representative enough sample, weights derived from the empirical method should be superior to those derived from the rational method. Additionally, it has been suggested that empirical keying may uncover subtle items that are more resistant to faking (Hogan, 1994; Mitchell, 1994). Potential downsides of the empirical approach include a lack of interpretability due to only considering empirical relationships rather than coherent constructs (Hough & Paullin, 1994; Mael & Hirsch, 1993; Mumford & Owens, 1987), the need for data, and an increased potential for legal challenges if empirically identified items happen to result in subgroup differences (Pace & Schoenfeldt, 1977; Sharf, 1994).

Other models. This dissertation is primarily focused on the rational and empirical models of biodata key development, as these are typically seen as the two major competing models. However, there are two other broad classes of methods that deserve a brief mention as some research findings I discuss involve their use: the factorial and subgrouping models.

Factorial model. The factorial model of biodata scale construction (also known as inductive or internal), is the attempt to investigate the dimensionality of a biodata scale *post-hoc* (Hough, 2010). This is therefore a data driven approach, but directed internally, rather than toward an external criterion, and is most often used when there is not a set theory about the constructs that underlie a biodata scale (Mumford et al., 2007). The primary assumption is that each scale has some meaningful structure of individual difference characteristics that can be discovered using factor analysis. Beyond that, the assumptions are chiefly those of factor analysis. The internal structure of the biodata inventory would then determine the placement of items in a particular sub-scale and the direction of its keying, and factor scores are determined by summing the scores on the items that are part of a given factor. In many cases, the criterion for success is homogenous scales, and this is ensured via only including items meeting a certain threshold loading on a certain factor (e.g., .30; Chait, Carraher, & Buckley, 2000) and follow-up internal consistency analyses. Further detail on the factorial method and its applications in biodata can be found in Schoenfeldt & Mendoza (1994).

There are a number of criticisms of the factorial approach. For instance, Hough (2010) suggests that a focus on homogenous scales and deleting items with poor item-total correlations can result in less complex scales and a reduction of the variance utilized from the original item pool. For instance, Hough & Paullin (1994) reviewed a selection of studies that used the factorial approach, and the factors they retained accounted for between 19% and 49.6% of the variance in the total item pool. In a similar vein, Brown (1994) discusses the potential complications implied by the specific factor analytic approach. Using factor analysis generally assumes that the common variance among the items is the critical portion of variance to analyze (i.e., because the communalities are on the diagonal, only the common variance is being included in the dimensionality analyses). However, it may not be only the common variance that is predictive. On the other hand, using principal components analysis can create the opposite problem, where unreliable item variance could be factored. Overall, Mitchell (1994) concludes that the factorial approach "might be more akin to rationalization than to rationality" (p. 486), Guion (2011) states that the factor analysis probably only identifies constructs that could be used if one wanted to start over with a construct-oriented item development process, and Mumford et al. (2012) suggest that the rational and empirical methods should be preferred to the factorial one.

Subgrouping or clustering. The subgrouping method has been the main empirical means by which Owens and his colleagues have attempted to investigate the ecology model (Mumford et al., 1990). Technically, it is not a variable-centered approach, and instead tries to group people and identify subgroups with similar life experiences and prior behavior (Hein & Wesley, 1994). One of its proposed strengths is that rather than tying a scale and its items to a specific criterion, the subgrouping method identifies people who will tend to respond similarly to a wide variety of criteria and situations. Typically, this proceeds by factor analyzing items to reduce the number of objects that require clustering and to create profiles, entering these profiles into clustering algorithms to identify individuals with similar profiles, determining the number of clusters to retain, and finally describing those clusters (Brown, 1994; Hein & Wesley, 1994). I refer interested readers to Hein and Wesley (1994) and Mumford et al. for further information on this method.

One of the primary criticisms of the subgrouping method is that, depending on the analysis strategy, clustering solutions are either (a) not always able to classify everyone acceptably, or (b) can suggest that there are many clusters to which a person has a high probability of belonging (Brown, 1994). This creates a problem for prediction purposes, which could preclude its use in selection. However, for a method that has primarily been used for the testing of inferences from theory, this hasn't been a chief concern. In addition, there is potentially a philosophical disconnect with the idea of selecting an individual based on their similarity to one of a dozen groups (e.g., "What about the uniqueness of the individual?"). Overall, this approach appears to have been generally used by Owens and his students and to have fallen in popularity in recent years. For instance, discussion of it has been dropped from some recent biodata overviews (e.g., Hough, 2010; Stokes & Cooper, 2004), and Mumford et al. (2007) state that subgrouping techniques are still in their infancy and are rarely applied in prediction.

Commonly Proposed Differences between Rational and Empirical Models

Differences in interpretability. One of the main distinctions made between scales derived rationally and scales derived empirically is that rationally derived scales are typically more interpretable, and it is easier to communicate why a specific item is linked to the criterion (Hough, 2010; Mumford & Owens, 1987). For example, in reference to empirical keys, Reiter-Palmon and Connelly (2000) say that “it is often difficult to understand why keyed items are (or are not) related to the criterion” (p. 143), and Brown (1994) cautions that an empirically developed measure that predicts well will be of no use if its users do not accept it because the items do not ring true for them. The proposed mechanism for this distinction appears to be that empirical methods are assumed to select items strictly due to empirical relationships with criteria (Nickels, 1994) or that items are included in a pool for convenience’s sake (Korman, 1968). On the other hand, the construct-orientation of many rational scale-development approaches ensures an explanation for each item (Hough, 2010; Mumford et al., 2012). However, there are signs that this distinction is fading as proponents of empirical keying acknowledge the importance of an item pool related to the job (e.g., Hogan, 1994).

Differences in criterion-related validity. Much of the cumulative research on the criterion-related validity of biodata inventories is based on empirically keyed inventories (i.e., many of the meta-analyses of criterion-related validity and their component studies are from a time before the rational approach was heavily used). Indeed, one of the main reasons to use an empirical keying approach is the assumption of superior predictive validity (Hogan, 1994). In line with this belief, Cucina, Caputo, Thibodeaux, and MacLane (2012) recently found that it took only a very small development sample (i.e., n = ~100) for empirical methods to exhibit higher predictive validity than rational ones and that, depending on the empirical keying method, validity could more than double with sample sizes in the thousands. However, other research has challenged this assumption, with a number of results suggesting that rational methods can be just as predictive as empirical methods, if not more so (Hough & Paullin, 1994; Mumford et al., 2007; Schoenfeldt, 1999; Stokes & Searcy, 1999).

There are a number of potential problems with the claims that rational keying and empirical keying procedures are "equally effective predictors of job performance" (Mumford, Whetzel, Murphy, & Eubanks, 2007, p. 212). These largely boil down to unfair comparisons (Cooper & Richardson, 1986). For instance, Hough and Paullin (1994) in a heavily cited book chapter within the biodata literature, conducted an empirical review of studies that calculated the difference between cross-validated validity coefficients from empirical keys and validity coefficients from rational keys. They found that the sample size weighted mean difference in validities was .01, in favor of rational keys. However, no attempt was made to ensure that the comparison of methods was based on the same item pools, the same development context, or maximally predictive empirical methods. For example, the included study with the largest difference in favor of the rational method (.25) was Jackson (1975). Jackson compared the results of randomly chosen California Personality Inventory (CPI; an empirically keyed personality inventory) questions loading on sociability, social presence, and tolerance scales with similar scales from a rationally-derived personality inventory (the Jackson Personality Inventory; JPI), as well as items written by college students to tap these constructs. These scales were then used to predict self and roommate ratings of these constructs. Jackson found that both the rational scales performed better than the CPI scales, and argued that even naive item writers (college students) could beat the empirical method.

The main problem with Jackson's (1975) comparison is that the results were based on different item pools. The empirical approach for each scale was based on 16 items from one item pool, while the rationally-constructed JPI was based on 20 items from another pool, and the college student constructed inventories were 16-items that were specifically targeted to the college context. Without standardization of an initial item pool, it is unclear if empirical or rational selection/keying is superior. All else equal, it makes sense that being able to target personality items towards a specific context, like the college student item writers were able to do, may provide a method with an advantage. Also, because the criterion was simply a one-item rating on a scale of 1 to 9 based on the construct definition given by Jackson (and was specifically what the college student item writers were writing their items against), it is possible that the facets of each construct addressed by the empirical method and the rational methods were different. Finally, this implementation of the empirical method involved an inventory developed on large samples and intended to be widely predictive of the general population. It was not therefore empirically-keyed based on the specific population of interest.[9] Thus, maybe the empirical method did not perform as well because it did not have the opportunity to select the most predictive items and was not able to utilize some of its strengths.

Hough and Paullin (1994) and Jackson (1975) might argue that they were interested in comparing the broad *rational method* versus the broad *empirical method*. In other words, rather than focusing only on item keying or item selection, they were interested in how the two approaches differ with respect to item development. However, the next example I discuss from Hough and Paullin compares the performance of the methods in the same item pool. Additionally, in my estimation, adding item-writing confounds a number of largely separate issues and results in being unable to meaningfully compare the two approaches in terms of item development (i.e., an empirical approach cannot develop its own biodata items based on some sort of empirical optimization algorithm. It can only select from a set of already developed items). I will return to these distinctions later.

Another potentially unfair comparison sometimes made between the two methods is the number of items on which the empirical/rational scales are based. One such study, also included in the Hough and Paullin (1994) review, is Schoenfeldt (1999).[10] Coincidentally, the results from this study have the second highest difference in favor of the rational approach (.15) after Jackson (1975). Schoenfeldt began by developing 137

---

[9] Certainly, one could argue that this is a valid question of interest: If empirically developed keys are less generalizable, this could be a reason to prefer rationally developed scales. However, in a direct comparison between scale-development methods it seems best to put them on the same footing in terms of scope.
[10] This is cited in Hough and Paullin (1994) as Schoenfeldt (1989), which was a Division 5 Presidential address given at the American Psychological Association conference. However, all the data in Schoenfeldt (1999) is the same as that reported by Hough and Paullin.

items targeted toward 16 constructs that were found via job analysis to be predictive of performance for service employees. After converting categorical items to sets of dichotomous items, the 137 items became 240. To develop an empirical scale, the 68 items that were significantly related to one of five criteria (either based on zero-order correlations, or a significant partial correlation observed during the process of stepwise regression) were identified. Then, these 68 items were used to build a key to predict the five criteria via stepwise regression, but a maximum of 10 items per criterion were allowed. To construct the rational key, Schoenfeldt used *the same* 68-item pool of those items that were found to be significantly related to the criteria, and used *all* of them to create 16 construct scores (based on the relationship of the item with the original construct it was written for), which formed the basis for the rational key used to predict each construct. Because each of the 16 construct scores consisted of four to five individual items, it was more reliable than any of the ten individual items used for the empirical keys. Additionally, 16 total predictors were allowed instead of 10, making this a potentially unfair comparison that "stacked the deck" in favor of the rational approach. Finally, because all items were conditioned on being significantly related to criteria in the first place, this does not appear to be a very pure comparison of the empirical and rational methods. On one hand, this avoids the confound discussed in the Jackson (1975) study in that all keys were based on the same item pool, but it allows for another potential misinterpretation: basing conclusions on different numbers of items. Similar interpretational difficulties are present in Stokes and Searcy (1999)[11] and Reiter-Palmon

[11] While there are other interesting facets to Stokes and Searcy (1999), from the perspective of a rational/empirical comparison, it compares over 200 items tied to 20-47 constructs for the rational approach

and Connelly (2000).[12] In sum, many of the studies that are frequently cited as indicating the equality or superiority of the rational method for criterion-related validity are difficult to interpret.

The other problem with global statements about the equality or superiority of either approach is that it is situationally dependent. In my opinion, the research is pretty well-settled in terms of an empirical keying approach being optimal for predicting a specific criterion compared to the rational keying approach, given a large enough sample size and the same item pool.[13] This holds both with rational keying that uses unit weights instead of differential weights and with rational keying based on expert judgment, both of which are explored further in later sections. However, there will be specific settings where rational keying is the preferred or only option (e.g., no developmental sample available), and there will be settings where empirical keying is the optimal choice if one wishes to maximize predictive validity. Therefore, the primary question facing those trying to decide between scaling methods on the basis of criterion-related validity is not "which one is better?"; it is whether there is likely to be a representative and large developmental sample.

Differences in reliability. The primary distinction drawn between rational and empirical approaches with respect to reliability is in the type of reliability estimates that

---

with between 36 and 65 items for the empirical keys, and the rational scales and empirically-identified items are weighted via regression against each criterion. As a result, it bases conclusions on different numbers of items and doesn't appear to be a pure rational/empirical comparison.

[12] The rational key in Reiter-Palmon & Connelly (2000) is based on a set of items linked to constructs judged relevant to the criterion, and the empirical key is based on smaller subsets of these theory-related item pools and unit-weighted.

[13] Again, my sense is that many who argue for the superiority of the rational method do not hold that the item pool would be constant. I will address this in the section: Distinctions between Item Writing, Item Selection, and Item Weighting.

can be used. Because rational scale-development methods are oriented toward specific constructs, the use of internal consistency metrics, such as coefficient alpha, are generally recommended for the resulting scales. Mumford et al. (2007) concluded that scales created in this fashion tended to exhibit fairly high reliability (.70s to .96). On the other hand, because an empirically-derived scale can potentially be comprised of items that tap any number of constructs, internal consistency is assumed to be low (and in fact, to be inappropriate measures of reliability[14]), and test-retest reliability is most often recommended. In general, test-retest reliability has been found to be moderate to high (i.e., .60 to .96, Mumford et al., 2012; Mumford et al., 2007).

Differences in transportability/generalizability. Another area in which scales derived from rational and empirical methods have been proposed to differ is in the degree to which they can be expected to generalize to other contexts (e.g., other criteria, other jobs, other companies, other countries; Mumford, 1999). For instance, Dreher and Sackett (1983), Hunter and Hunter (1984), and Thayer (1977) have all suggested that empirically keyed biodata inventories are likely to be situationally-specific. Research by Rothstein et al. (1990) and Carlson et al. (1999) has exhibited that it is possible for empirically keyed inventories to be valid across a wide number of organizations for broad job titles like supervisors and managers. However, these procedures involved very large and diverse development samples that allowed for stringent testing of items and perhaps are not characteristic of typical biodata contexts. Therefore, they more suggest that it is *possible*

[14] Hunter and Schmidt (2004) and Schmidt and Hunter (1999) suggest that internal consistency estimates are really assessing parallel forms reliability. Coefficient alpha and split-half reliabilities work for this purpose because they assume that the items can be split into subsets that are, in effect, a parallel form of items that would measure the same construct. Because an empirically-derived key would not be able to meet this assumption, internal consistency measures are inappropriate.

for empirically keyed scales to be generalizable across organizations rather than that it can generally be assumed that they will be. It is also important to note that one of the explicit purposes of both of these projects was to develop a scale that would generalize across situations. Additionally, others have shown that it is possible for empirical keys to be generalized to other countries (Dalessio, Crosby, & McManus, 1996; Hinrichs, Haanpera, & Sonkin, 1976; Laurent, 1970), though doing so generally required modifying items to account for different cultures. On the other hand, rationally derived scales have had little investigation into their generalizability across situations, but it is generally assumed that they would transport better due to their focus on constructs as opposed to discrete behaviors as well as their avoidance of keying to specific criteria (Mumford, 1999; Stokes & Cooper, 2004).

Differences in susceptibility to faking. In terms of the rational/empirical distinction, the most common argument made is that empirical scaling may identify subtle items, whereas rationally derived items are more likely to be transparent to the respondent. Mitchell (1994) serves as an example of this belief, writing that “the nonintuitive nature of biodata scoring keys makes them highly resistant to faking and manipulation” (p. 505) and citing Pannone (1987) and Crosby and Mitchell (1988) as providing support for this belief. From a different angle, Kluger, Reilly, and Russell (1991) found that option-keyed scales (i.e., keys where different weights are developed for every response option) were more resistant to distortion than item-keyed scales. Option-keyed scales are likely more characteristic of empirical keying rather than rational due to the ease of implementation rather than having judges attempt to differentiate weights that should apply to specific item options. Alternatively, Hough and Paullin (1994) reviewed literature from the personality field comparing obvious to subtle items and concluded that subtle items may reduce the validity of scales and do not appear to be more resistant to faking.

Related to Mitchell's (1994) concern, there is another literature on biodata item faking that relates to the specific item attributes of biodata items, such as those described by the taxonomy of Mael (1991). These studies tend to find that objective and verifiable items are generally less subject to faking (e.g., Becker & Colquitt, 1992; Graham et al., 2002; Harold et al., 2006; Lautenschlager, 1994). However, these types of items can appear on both rational and empirical keys.

Differences in potential for legal action. Tying into assumed difficulties with interpreting how empirically derived scales are related to the criterion domain, many have suggested that a purely empirical process can result in using items that could be resisted for a lack of job-relevance, and at worst, directly tapping into subgroup differences (Hogan, 1994; Mumford & Owens, 1987; Pace & Schoenfeldt, 1977; Sharf, 1994). Though as of yet, there have been no successful legal challenges to biodata inventories (Terpstra et al., 1999), a commonly cited anecdote from Pace and Schoenfeldt (1977) illustrates the potential danger: In developing an empirically scaled inventory to predict theft in two companies in the Detroit area, two of the valid predictors found by Rosenbaum (1976) were "is black" and "Detroit address". Although "is Black" is obviously identified as pertaining to a protected class, "Detroit address" also implicates Black applicants as they were more likely to live within the city boundaries. Therefore, only utilizing the fact that an item is related to performance has the potential to unwittingly result in adverse impact.

At the very least, it is often recommended that a rational screen be conducted when empirical keying to eliminate items that might potentially result in adverse impact (Mumford & Owens, 1987). Going further, Sharf (1994) references an amicus brief that the American Psychological Association filed to the Supreme Court suggesting that the most defensible approach is to be able to demonstrate a logical and empirical match between each item to the knowledge and skills required to perform the job. For a dissenting view, Mitchell (1994) cautions that the rejection of items that are not face valid would result in throwing away a number of valid items. He further argues that "if a biodata item effectively predicts a job-relevant criterion measure, then that item must also be job relevant" (pp. 507-508), regardless of face validity. Many companies, however, would probably prefer to err on the side of avoiding potential negative applicant reactions and legal challenges.

There are therefore two related issues at play here: items that may unwittingly result in adverse impact, and items that may not be seen as face-valid. Empirically derived scales are implicated as being more susceptible to both, particularly due to item selection. Although rational item-development processes would likely help with job-relatedness, they are not immune to unintentional adverse impact, and prudence would suggest standard sensitivity analyses be conducted on either type of inventory.

Parallels to other areas of inquiry. Before moving on, I think it is useful to draw attention to the parallels between the rational/empirical scale development distinction in biodata and similar distinctions in other domains, as well as potential parallels to the clinical vs. statistical prediction debate. First, it is worth noting that the problem of how to key items extends more broadly than biodata. Personality scales, for instance, are one area where scale developers have wrestled with rational versus empirical approaches (e.g., Burisch, 1984; Hase & Goldberg, 1967; Jackson, 1970; Simms & Watson, 2007). In fact, Hough and Paullin (1994) and, to a lesser degree, Hogan (1994) draw on this literature in discussing the rational and empirical approaches, respectively, to biodata development. Similarly, the emerging area of situational judgment tests (SJTs) also is grappling with this issue in order to determine which responses to situations are most effective (e.g., Bergman, Drasgow, Donovan, Henning, & Juraska, 2006; Krokos, Meade, Cantwell, Pond, & Wilson, 2004; Muros, 2008). Empirical methods appear particularly similar between the SJT and biodata domains, with many of the same ones being used (e.g., mean criterion response, vertical percent method).

Second, there may be a temptation to draw parallels between the previous discussion on rational versus empirical scale development/keying and the literature comparing clinical vs. statistical prediction. For instance, Cucina et al. (2012), in a study that compared empirical and rational keying methods, found that rational approaches generally exhibited much lower validity than empirical approaches for larger sample sizes. They stated that this was "...consistent with the established notion that statistical procedures have a history of outpredicting clinical judgments" (p. 422). Although this statement is accurate, I think it is important to clarify and expand on this comparison.

The key point is that the literature on clinical vs. statistical prediction focuses largely on data combination. For instance, Meehl (1954) distinguished between a "prediction... arrived at by some straightforward application of an equation or table to the data" (p. 15) versus "judging or inferring or weighing...done by a skilled clinician" (p. 16). Meehl further clarified a mechanical approach by saying that "Once the data have been gathered from whatever source and of whatever type, the prediction itself could be turned over to a clerical worker" (1954, p. 16). Sawyer (1966) agreed with this distinction and disagreed with the notion put forward by Sarbin, Taft and Bailey (1960) that to be considered statistical or mechanical, a process had to be statistically derived as well as statistically applied. Sawyer wrote that the mechanical combination was "any set of rules whose application is objective, whatever mixture of experience and intuition their derivation involves" (p. 180). This distinction is used to guide much of the investigations of this question (e.g., Grove, Zald, Lebow, Snitz, & Nelson, 2000; Kuncel, Klieger, Connelly, & Ones, 2008), and the controversies that have resulted (e.g., Jeanneret & Silzer, 2011; Kuncel & Highhouse, 2011; Silzer & Jeanneret, 2011). Both Meehl and Grove et al. (2000) found overwhelming support for the efficacy of the mechanical method of data combination over the clinical method in their reviews of the literature.

With this framing in mind, it is clear that both the rational and empirical scale development approaches discussed in the biodata literature could be considered mechanically combined, as in both cases a scoring key is derived and consistently applied. In fact, because biodata is also collected consistently, it represents what Sawyer refers to as a "purely statistical" process (1966, p. 183). The most relevant parallels from the clinical/statistical literature between the two approaches are what are known as expert judgment-based weights (sometimes called policy capturing) or a "bootstrap" model (Dawes, 1971; Dawes & Corrigan, 1974). For expert judgment-based weights, experts are simply asked how they would weight a set of information, and the information is aggregated, and used. For a bootstrap model, previous assessor judgments are regressed on predictor values (i.e., items) to obtain a statistical representation of how important/effective/relevant the predictors are according to the assessors. This allows for elements of judgment/rationality to be involved in the development of a scoring procedure while ensuring consistency of application. Research comparing bootstrap procedures to purely empirical ones indicates that the purely empirical ones perform better (Dawes, 1979; Dawes & Corrigan, 1974; Goldberg, 1970).

Combined methods. Many recent authors have noted that the empirical and rational approaches are rarely implemented in their pure form (Breaugh, 2009; Hough, 2010; Mumford, 1999; Stokes & Reddy, 1992; Stokes, Mumford, & Owens, 1994). For instance, Stokes, Mumford and Owens (1994) wrote that "the line between these techniques is not always clear...these techniques are often combined to develop a final biodata scale" (p. 65). Although broad statements to this effect are numerous and this sort of approach is implied to be common practice, there are few explicit operationalizations in the literature of a so-called "combination" or "hybrid" procedure.

At the most basic level, items from rationally derived scales that do not predict the criterion can be removed, and items/weights from empirically derived scales that don't "make sense" can be altered or removed. Specific to weighting, Cucina et al. (2012)

state that “sometimes, researchers will create empirical scoring for weights for each response option and review the weights for consistency with theory” (p. 387), and “another approach is to combine rational and empirical weights (using unit weighting) to create a set of hybrid weights” (p. 387).

More broadly, Mael and Hirsch (1993) described two approaches that they termed the “quasi-rational” approach and “rainforest empiricism”. The quasi-rational approach, described earlier in this dissertation as using “biodata analogs”, involved empirically keying items to a set of personality constructs believed to be related to the criterion. As a result, the items were not directly keyed to the criterion (a quality of the rational approach), but were empirical in the sense that they were keyed to something external to the scale. Their other approach, rainforest empiricism (termed to be in contrast to dustbowl empiricism), was an attempt to systematically screen empirically keyed relationships to ensure that they made rational sense. Example rational tweaks involved smoothing out unlikely non-linearity in items that represent a continuum (e.g., a low weight in an otherwise monotonically increasing set of weights), combining response categories with low frequencies, and dropping items that exhibited low variance.

Although it can be debated how well either of these specific methods manage the trade-offs between the particular methods, the acknowledgement and use of combined procedures for scale development represents a shift in the theory of biodata scale development. I will argue in the following section that by being more deliberate about the design choices being made, further conceptual clarity can be achieved, and the benefits of both rational and empirical methods can be integrated.

Distinctions between Item Writing, Item Selection, and Item Weighting

I propose that much of the controversy between the rational and empirical approaches can be avoided, or at the very least clarified, by careful delineation of three phases of the scale development process: item generation/writing, item selection, and item weighting. This is not a unique idea, as the points have been made that the rational and empirical methods can be combined and that the steps of the scale development process can be distinguished (e.g., Breaugh, 2009; Brown, 1994; Hough, 2010; Hough & Paullin, 1994; Karas & West, 1999). However, there is substantial potential confusion in delimiting these steps and what specifically is called "rational" or "empirical".

Although I suspect most biodata researchers would largely agree with this proposition, the way these procedures are presented in the literature has consistently muddied the distinctions between these three scale development phases. Consider that when discussing the differences between rational and empirical scale development, many biodata references and articles highlight item development as an important characteristic of rational scale development, while not addressing where items come from for empirical scale development (e.g., Brown, 1994; Hough & Paullin, 1994; Kilcullen, White, Mumford, & Mack, 1995; Mumford, 1999; Mumford et al., 2012).[15]

Rather than confounding item development with the rational method and rational selection and keying, I contend that the scale development process can be broken up into three separate decisions: how to develop items, how to select items, and how to weight

[15] Some references have a separate "item generation" section, but still revert to describing item development as a feature of rational scaling.

items. These options can be represented in a pair of contingency tables as done in Figure 1, and fully crossed, represent 24 possible scale development combinations.

The first decision is how to develop items. Empirical item development would be somewhat of a misnomer because items cannot technically be developed empirically. What is called empirical is probably better named archival because it is usually framed as a large pool of items, out of which an empirical selection procedure will identify useful items. On the other hand, rational item development is the previously described process of writing items to target specific constructs and prior behaviors identified as job-relevant. In other words, the item pool is made up of items that were written based on hypotheses specific to the setting of the inventory.

The next decision is how to select items. The first option is that there would be no selection. Thus, every item in the original pool would be used. There are also two empirical options that select items based on data: internal and external. Empirical - internal selection would be characterized by the construct-oriented scale construction strategy, where rationally developed items are removed if they do not exhibit high enough variability or item-total correlations. Empirical - external would be if items were selected based on their relationship to a criterion (i.e., either stepwise regression, or based on some magnitude/significance cut-off of correlations or percentile weights). Often, when empirical - external selection is used, this is associated with empirical weighting because the basis for choosing items is often then used to weight the items. Finally, items could be selected rationally. This might occur if a panel of experts was choosing what they expected would be the best items from a larger pool.

The final decision to be made is how to weight the items that were selected. At its simplest, this could be done by giving each item/item response a weight of one (unit weighting). Alternatively, differential weights could be assigned, and this could be done either rationally using expert judgment, or empirically, using data to obtain an estimate of the relationship of the item/item response to the criterion.

Given this set of decisions to be made, I suggest that some options present themselves as preferable. First, it makes sense that items should be developed with at least some level of rationality. Items should be developed with a focus on the specific context and purpose for which the scale is being created. The only argument I can see against this method is that it will result in transparent items. However, if a worker-oriented or indirect approach is used, that does not have to be the case. For example, an objective biodata item that is related to extraversion wouldn't necessarily be seen as obviously related to the job. Additionally, even if some valid, subtle items were found using an archival approach to item development, chances are it would take a great deal many more items in the prospective pool to find enough that would cross-validate to fill out a scale. More items require both more time from subjects and more subjects for empirical methods that require a specific sample size-to-item ratio to be effective. Similarly, if the choice is strictly between rational assignment of weights and empirical weighting, all else equal, empirical weighting should produce a better result in terms of predictive validity. As discussed earlier, the main factors in this choice are whether there is a sample available to the researcher for the development of weights and whether there is an adequate criterion measure. However, empirical keying should be viewed as the default, only not conducted when situational constraints prohibit it. One does not have to choose “rational” for all three of the choices due to a desire to have meaningful items targeted towards the criterion/job of interest. Karas and West (1999) provided an example of using rational construct-oriented item development with empirical keying. Additionally, Reiter-Palmon and Connelly (2000) concluded that empirical keying methods from “theory-rich” item pools performed better than those from random item pools.

Still, review chapters refer to the broad classes of methods and the same lists of strengths and weaknesses that have been touted for decades (e.g., empirical keys are uninterpretable). Breaugh (2009) suggests that this is done for presentation purposes, but I think the explicit break-down of the scale-development process is much clearer.

Implications. The primary implication of adopting this framework is that many of the commonly proposed differences between the rational and empirical approaches are greatly diminished. First, by ensuring an item pool that has a rational tie to the criterion domain, this immediately aids in interpretability. Even if the items were not sorted into subscales and items were differentially weighted, the scale developer would be able to point to the theory that justified every item’s inclusion in the pool. This may even aid in interpretability; in addition to being able to argue for interpretability from the standpoint of association with a construct or job-relevant behavior, the scale developer would be able to determine which items were more or less predictive of the criterion. This could aid in theory refinement. With the increase in interpretability also comes a decrease in the potential for legal action that empirical keying has often been accused of. Finally, these delineations make clear that empirical keying and the higher predictive validity that comes with it can be conducted in any biodata scale development setting. If developers have access to a large enough sample to support empirical keying, but decline to do so due to a desire to adhere fully to “the rational approach”, they are giving up validity.

Overview of Empirical Keying Techniques and Issues in their Implementation

With the preceding set of distinctions in mind, it is clear that empirical keying techniques are broadly applicable regardless of the choices that one makes for item development or selection. The focus of this dissertation is therefore on comparing methods for how best to empirically derive a biodata scale. First, I will present a relatively new taxonomy for classifying common empirical methods. I will then discuss a set of issues and decisions that are of broad applicability across these methods before describing each method and reviewing comparisons of how these methods perform. Finally, I will review two methods that have seen little use in the empirical-keying literature and argue why they might be useful.

Taxonomy of keying techniques. Cucina and his colleagues (Cucina et al., 2012; Cucina et al., 2009) have recently proposed a taxonomy for organizing the empirical keying methods commonly used in the biodata domain that is helpful for comparing and contrasting different keying methods. They distinguish first between item- and option-level keying classes. Within the item-level class, they further distinguish between unit-weighted, correlational, and regression methods. Within the option-level class they highlight the mean criterion, percentile, rare response, and correlational families of methods, most of which have a number of variations. Because these are all what Brown

(1994) referred to as additive methods, the process common to every one of them is (a) the determining of weights, (b) multiplying each option/item by its weight, and (c) summing the resulting values to obtain a total score.

Cucina et al. (2012) also pointed out that empirical keying can combine both item and option-level keying, by first keying each option relative to a criterion and then identifying which items are more or less predictive overall. Items are often unit-weighted, which makes the implicit decision that whatever predictive differences there are at the item level, they will all be considered the same for the purpose of creating a summed total score. Alternatively, items can be correlation or regression weighted, allowing for these differences to be identified.

Initial considerations. Given the limited space in handbook chapters and overview articles on biodata and the large number of topics to cover, discussion of empirical keying techniques is usually limited to a couple pages. These might include a brief description of some common techniques, comparisons between them, and validity findings. This results in limited or no explanation of some features and decisions that need to be made that are relatively unique to empirical keying. Therefore, before providing a detailed description of the various empirical keying methods, I outline these issues in turn and highlight their relation to broader measurement concerns beyond the context of biodata.

Item-level versus option-level keying. One of the first decisions to be made after committing to empirical keying is whether an item-level or option-level keying method is going to be used. As can be inferred from the names, item-level keying is where weights are derived for every item, whereas option-level keying is where each response option of every item is evaluated and given its own weight. For item-level keying to make sense, the item's response options must be on some sort of continuum, such as a Likert scale. (Kluger et al., 1991). If the item is categorical (e.g., "Which of the following subjects in school was your favorite?"), it must be split into dichotomous items based on the response to each option if an item-level keying method is used. Because this linear, monotonic relationship is assumed, item-level keying cannot identify possible nonlinear relationships between the item and the criterion.

If nonlinear relationships are of interest or if item responses are largely not arranged on the basis of a continuum, then it will be preferable to use the option-keying method. Because each response option's weight is determined independently, nonlinearities are able to emerge. For instance, if a 5-point Likert item ended up with weights of 0, 1, 2, 1, and 0 for responses of 1, 2, 3, 4, 5, respectively, this would model a quadratic function. Another example might be weights of 0, 1, 1, 1, and 1, implying a leveling-off effect, where the only benefit gained as far as criterion performance is not responding to the lowest response. As a result, it is possible that with option-keying, a more accurate representation of the relationship between the item and the criterion can be obtained. Regardless of the level of the keying approach, Mitchell (1994) and Brown (1994) both point out that most common item- and option-keying methods result in scores which are linearly related to the criterion at the total-score level. Because correlational techniques are used to assess a scale's validity (i.e., correlating the total biodata score with the criterion), linearity is assumed.

As can be seen, these approaches are not terribly different. They largely differ based on what is done with continuum/Likert items (i.e., categorical items have to be split up into dichotomous items anyway). Practically, this decision has implications for the number of parameters that will be estimated, what is considered an "item", and the corresponding required sample size for the stability of these estimates. For example, if a scale has i = 100 Likert-style items on a 5-point scale, there are three potential numbers of items that might apply: (a) if item-level keying, there would be i items, or 100, (b) if option-level keying, there would be i * 5, or 500 items, and (c) if using regression at the option-level, then each item's responses would have to be dummy-coded, so there would be i * (5-1), or 400 items. Therefore, all else equal, option-keying is likely more susceptible to shrinkage than item-keying, as there are more parameters estimated. Additionally, if an empirical keying method is particularly sensitive to the sample size-to-item ratio, care must be taken to ensure that the right number of items is used (e.g., 100, 500, or 400) when planning a research program to develop weights. In other words, the effective number of items may differ from the number of questions that respondents answered.

There have been other proposed benefits of the option-keying approach, such as potential resistance to faking due to the nonlinear scoring keys that may emerge. Likert-style items scored at the item-level may be easier to fake because there is a high and low point; the respondent only needs to be able to discern whether the item is "good" or "bad". Alternatively, Kluger et al. (1991) provide an example option-level key for the item "How often have you had problems getting a job done because you did not have the right kind of people?" on a 1 to 5 scale (i.e., never, seldom, sometimes, often, and very often). They found that "seldom" had a weight of 1, "often" had a weight of -1, and the other three had weights of 0. Therefore, it would be more difficult for respondents to determine the best answer and to profit from distorting their responses. Kluger et al. found some support for this hypothesis in that respondents told to act like they were applying for a job were able to inflate their scores for item-keying but not for option-keying. The counter to this proposed benefit is that it may be difficult to justify non-intuitive nonlinearity in response weights. Indeed, the previously described "rainforest empiricism" strategy of Mael (1991; Mael & Hirsch, 1993) specifically tried to remove these in order to rationally revise the results of empirical keying. At the very least, if a scale developer is going to option-key and attempt to profit on nonlinearities in items, they should have a large and representative development and validation sample to be persuasive that they are not over-fitting error. Better still would be *a priori* hypotheses for items tapping constructs that are suspected to have nonlinear relationships with the criterion.

Unit-weighting versus differential-weighting of items. As is clear from Cucina et al.'s (2012) taxonomy described above, in addition to item- or option-level keying, the researcher must also choose between unit-weighting and differential-weighting of items and/or item responses. Because the typical biodata reference contains minimal treatment of this issue and some potential confusion, I provide a brief review and reference list from a selection of relevant research on this topic from the measurement, judgment and decision making, and biodata literatures.

Unit-weighting is where every response in an additive composite receives a weight of 1. This can be used either by simply weighting every item as 1 (e.g., if a scale was made up of 5-point Likert-style items, each item would simply be summed together), or where every item/item response that meets a certain threshold is given a weight of 1 (e.g., each item that correlates over .10 with the criterion gets a weight of +/- 1 depending on its sign). The main distinguishing factor is that all weights are equal. Differential weights, on the other hand, allow for items or item responses to be given different weights to reflect their stronger/weaker relationships with the criterion of interest. For instance, one way to calculate differential weights could be to use regression to determine the weights that maximize the relationship between the set of items and the criterion. Another way to apply differential weights would be to allow experts to assign different weights based on their intuitive understanding of the predictor and criterion domains. Using unit or differential weights is a decision that can be made at both the item and the response option level, and either *can* be used with both the empirical and rational scale development strategies.

Opinions diverge, however, on whether one *should* use differential weights. The primary argument is one of cross-validity and shrinkage: weights that are heavily dependent on a specific sample (i.e., empirically derived differential weights), all else equal, are more subject to shrinkage than weights that are minimally dependent on any specific sample (i.e., unit weights). Advocates of the rational approach are more likely to propose explanations for why unit weights should be preferred over differential weights, or at the very least that the decision doesn't matter. Mumford et al. (2007), for instance, suggest that unit weights result in scales that have similar validity to those using differential weighting, citing Dawes (1971). Mumford and Stokes (1992), building on an argument by Guilford (1954), suggest that “little is usually gained from the application of more complex weighting schemes” (p. 106). Similarly, Hough and Paullin (1994) and Reiter-Palmon and Connelly (2000) cite a number of measurement texts and articles in support of this idea.

Conversely, at an intuitive level, it makes sense that differential weights should have some value, as not all predictors are equally powerful and some predictors overlap considerably, so they don’t provide much extra information above one another (Wang & Stanley, 1970). Regression weights are called “optimal” for a reason; if one item has a correlation of .40 with the criterion of interest, and another has a correlation of .10, should they really both be weighted equally? Additionally, asking redundant questions is less efficient and can potentially end up weighting certain constructs more than they “should” be in a biodata composite. For example, consider a scale that purports to be a measure of deviancy. If the scale has a question about theft, a question about arson, and three questions about the pattern of absence in different contexts, applying unit weights (assuming variances were equal) would mean that the scale scores would be determined largely by the participants’ tendencies towards absence. In other words then, the question is not whether unit weights will result in valid or invalid weighting systems (they should be valid), but whether there is a better option, and if so, when it should be used.

Much of the confident rhetoric regarding the preferable status of unit weights comes from Wilks’ theorem (1938). Wilks showed that, in general, as the number of variables increases, as the variability in weights decreases, and as the average inter-variable correlation increases, it is expected that the correlation between linear composites derived from any two sets of weights would increase. Ree, Carretta, and Earles (1998) therefore discuss the finding that unit weights perform similarly to differential/regression weights as a "mathematical inevitability" (p. 409). Similarly, Lawshe and Schucker (1959) conclude that "...nothing is to be gained by differentially weighting the items in a long test" (p. 104). Many of those who indicate a preference for unit weights (or indicate that it doesn't matter) from the biodata realm cite either Wilks or a psychometric textbook/article that uses Wilks as part of the basis for their recommendations (e.g., Hough & Paullin, 1994; Mumford & Stokes, 1992; Reiter-Palmon & Connelly, 2000). However, the assumptions of Wilks' theorem are rarely discussed. Specifically, Gulliksen (1950) notes that:

> ...if the standard deviation of the set of weights is very large in comparison to the mean, changes in weights used can produce great changes in scores regardless of the number of variables to be combined, and regardless of their intercorrelation. For example, if both positive and negative weights are permitted, the mean of the distribution of weights will be near zero, while the standard deviation will be very large. (p. 315)

This implies that if using raw empirically derived weights (e.g., raw point-biserial correlation weights), which allow for positive and negative values, composites could potentially be less correlated with unit weights than might be assumed by Wilks' theorem when there are a lot of items.[16] This is not to say that they certainly will, or that differential weights will be more correlated with a criterion than unit weights, just that it is mathematically possible that with large numbers of items, all composites are not practically identical in terms of rank order.

Outside of arguments based on Wilks' theorem, there is a related literature in the judgment and decision making and psychometrics literatures that allows for conclusions based on empirical and simulation studies that unit weights and regression weights perform similarly in many settings (e.g., Dawes, 1971; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975). Similarly, others have found that regression weighting can underperform unit weighting in settings with small samples (e.g., Dorans & Drasgow, 1978; Einhorn & Hogarth, 1975; Raju, Bilgic, Edwards, & Fleer, 1999). For a current review of findings with a fairly comprehensive reference list, see Bobko, Roth, and Buster (2007).

Placing this controversy in a broader context, Einhorn (1986) presents arguments contrasting the belief in the possibility for perfect prediction (clinical approach) with the acceptance of error (statistical approach). Whatever the underlying belief for error (e.g., inherently uncertain world, a belief that knowledge of the world will always be incomplete, or similarly, that the use of any algorithm or equation will never fully encapsulate the complexity of the phenomenon under study), he argues that if one can accept some small amount of error in using unit weights, it might outweigh the potential

[16] As an aside, this assumption may have been more tenable in the past as many weighting strategies involved adding constants or transforming data to not allow for negative weights (e.g., England, 1971; Guion, 1965).

risk that differential relative weights will depart heavily from their actual weights.[17] Hence the name of his paper *Accepting Error to Make Less Error*, and he proposes that this strategy is “likely to be a safer and more accurate strategy over a wide range of practical situations” (p. 394). Dana and Dawes (2004) make similar arguments.

Whatever the rationale in support of unit weights, the broad conclusion appears to be that with small sample sizes, unit weights can be expected to be equal to or even be superior to differential weights, whereas at increasing sample sizes differential weights can be expected to perform increasingly better than unit weights. The sample sizes where differential weights are expected to provide an advantage are relatively modest by some estimates. For instance, Bobko et al. (2007) suggest based on their literature review that this point might be between 120 and 200. In the biodata domain, Hough (2010) proposes the use of unit weights when developmental sample sizes are fewer than 150. In a more recent and comprehensive simulation study based on real data in a biodata context, Cucina et al. (2012) investigated unit weighting and differential weighting at the option-level. At the option-level, differential weights (i.e., raw weights) generally performed better with respect to cross-validity than unit-weights (i.e., weights are assigned to 1, 0, or -1 based on magnitude or significance of the raw weight) at every derivation sample size, the smallest of which was 150. The differences between the two methods, however, tended to be fairly small (i.e., difference in cross-validity of .01 to .02) throughout the

[17] For example, if V1 and V2 truly should be weighted at a ratio of 2:1, respectively, then using equal weights ensures that the data won't result in V2 being weighted higher than V1 due to sampling error.

sample sizes tested, up to $N = 3{,}515$.[18] The findings from one biodata inventory therefore suggest that validity isn't greatly affected by using unit weights over differential weights. However, transforming raw weights to unit weights adds another step to the process; if empirically option keying anyway, the Cucina et al. results suggest that it probably makes sense to use raw differential weights.

Keying methods that require the development of extreme groups. A procedure that is required for some empirical keying techniques when there is a continuous criterion is the development of upper and lower extreme criterion groups. The purpose of this is two-fold. First, before computers, it was necessary to employ computational shortcuts for efficiency. Although a correlation between a biodata item response and a continuous criterion might be a good measure of their relationship, it is time-consuming to calculate by hand for every item. Second, if it is given that the criterion will be split to accommodate computation, it is desirable that the resulting groups are as significantly different as possible. This is the logic behind choosing groups on either extreme of the criterion range.

Within the biodata literature, there are two approaches to assuring the significance of the difference between extreme criterion groups. The first is characterized by England (1971) who highlighted criterion selection and the development of criterion groups as the most important steps in the Vertical Percent Method (VPM), suggesting that "without adequate criterion group development, the entire study will be a waste of time" (p. 15). His primary recommendation appeared to be to apply a criterion-referenced viewpoint (as

[18] They also investigated unit-weighting versus stepwise-regression at the item-level, but this confounds item weighting and selection and should increase sample size requirements above those required for regular multiple regression.

in the distinction between norm-referenced and criterion-referenced), whereby consideration of what constitutes effective criterion performance and what constitutes ineffective criterion performance is taken into account. For instance, is the top half of the sample “effective”, and only the bottom 10% considered truly “ineffective” based on performance standards? England notes that any group size is possible as long as the groups are large enough to create stable weights and “as long as the desirable group truly represents the type of employee desired in the future and the undesirable group represents that type one wishes to avoid” (p. 14). He concludes that adherence to simple rules of thumb may not apply in some cases.[19]

On the other hand, other researchers, while still endorsing the importance of the “appropriate identification of the criterion groups” (Hogan, 1994, p.77), have instead advocated rules of thumb, such as Kelley’s (1939) criterion of 27% in each group (Hogan, 1994) or the top and bottom thirds (Devlin, Abrahams, & Edwards, 1992; Hough, 2010; Stead, Shartle, & Associates, 1940). Hough (2010) stated that this top and bottom third split was “ideal”. In addition, in potential conflict with the advice offered by England (1971), current standard advice is that the two groups be approximately equal (Hogan, 1994; Hough, 2010; Stokes & Cooper, 2004). These sorts of rules of thumb appear to be the most used in the biodata domain.

This dichotomization to create extreme upper and lower groups has a broader context within the psychometrics literature that is potentially informative (e.g., Alf & Abrahams, 1975; Feldt; 1961; Kelley, 1939; Preacher, Rucker, MacCallum, & Nicewander, 2005). Before beginning, however, some potentially confusing terms and

[19] See Aamodt and Pierce (1987) for an example of this criterion-referenced approach.

distinctions must be discussed. First, a recent article by Preacher et al. (2005) has differentiated between an “extreme groups” design, and a “post hoc subgrouping” design. An extreme groups design, according to this article, is one in which researchers wish to investigate the relationship between two variables, but due to one constraint or another (e.g., cost, time), the ability to administer one of the measures is prohibitive. In this case, some researchers decide to administer the less-constrained variable first as a screening tool, then only administer the constrained one to the most extreme groups. The key distinction is that data for every subject under consideration exists for the less-constrained variable, but sampling of the constrained variable is limited. As Butts and Ng (2009), DeCoster, Iselin, and Gallucci (2009), and Preacher et al. (2005) noted in a set of recent articles, creating so-called “extreme groups” increases the power of tests relating a continuous IV to a DV and that it has been a way to select participants to maximize the likelihood of finding significant relationships among the variables; all else equal, creating extreme groups will usually improve power and more extreme groups results in further increases in power (Preacher et al., 2005). Therefore, this is a means to lower the sample size necessary to identify an effect without reducing statistical power. This procedure of removing the middle of a distribution is also given as an example of how artificially increasing variance, or “range enhancement”, can increase a correlation beyond its “true” value in the full distribution (Sackett & Yang, 2000).

Preacher et al. (2005) concluded that certain applications of extreme groups analysis are a useful way to achieve cost-efficiency. This is contrasted with “post hoc subgrouping”, which Preacher et al. recommend strongly against. This is characterized by “obtaining variables x and y for all individuals in a sample, but analyzing data on y only for those individuals scoring at the extremes on variable x” (p. 190). In other words, this, with the addition of dichotomizing the criterion, is the situation encountered when empirically keying biodata with upper and lower criterion groups. Their rationale for avoiding splitting the sample in this case is that the power for the test of a null hypothesis that the correlation between x and y = 0 is at a maximum when all of the cases are being used (citing Alf and Abrahams, 1975, p. 571).

In a pair of related articles, Kelley (1939) and Feldt (1961) analytically prove that if upper and lower criterion groups are developed for the purpose of determining which variables (e.g., biodata items) are most related to the criterion, the optimal proportion in each group should be approximately 27%. This may seem to contradict the finding described in the previous paragraph that power to detect correlations significantly different from zero is at a maximum when all cases are being used. The key is in the statistics being used: Preacher et al. (2005) are talking about a correlation, whereas Kelley and Feldt are referring to the difference in means between the top and bottom group on the criterion. Kelley notes that even though the idea of removing up to 48% of the data to improve statistical power may appear counter-intuitive, it is predicated on the fact that the decision has already been made to split the sample into the upper and lower groups (i.e., already not using full information).

There are several implications of the previous discussion for practice in biodata when using extreme groups methods. First, when using percentile methods that require extreme criterion groups and the difference in percents (or equivalently means) is the underlying basis for weighting, if there is a compelling theory or performance-based rationale to determine success or failure on the criterion measure, it should probably be prioritized over statistical rules of thumb. If this does not apply or if it would create unacceptably small groups, Kelley's (1939) and Feldt's (1961) recommendation that upper and lower groups of approximately 27% of the sample in which weights are derived should be followed. Second, when using a weighting technique that assigns weights based on correlations, it is not necessary to split the sample into extreme groups, and doing so could reduce the statistical power to find an effect and/or artificially inflate the effect size.

Cross-validation procedures. As described earlier, one of the key features of empirical keying is the need to cross-validate derived keys on other samples. This is because empirically derived weights capitalize on sample-specific idiosyncrasies that are not expected to generalize (Cureton, 1950; Mosier, 1951). There are a number of different cross-validation procedures, however, each with their purported strengths and weaknesses. Recent authors have made an important distinction regarding the parameter of interest in cross-validation designs. It is possible to be interested in estimating the population variance explained (i.e., ρ^2; this is what the standard Wherry, 1931, adjusted R^2 formula attempts to estimate) or the estimated variance explained in a different sample drawn from that population (ρ_c^2; Cascio & Aguinis, 2005; Guion, 2011; Raju, Bilgic, Edwards, & Fleer, 1997; St. John & Roth, 1999). For estimating predictive efficacy in the future, it is argued that an estimate of ρ_c^2 is conceptually what is desired, and this is what both empirical and formula cross-validity procedures attempt to estimate.

Empirical cross-validation refers to any procedure that applies weights derived on one sample to another sample to estimate shrinkage in validity. These procedures vary in design based on whether multiple samples are drawn and whether multiple equations are used as follows:

- Single-sample and single-equation: One sample is split into two smaller subsamples. A common split is 2/3 in one subsample and 1/3 in the other. Weights are derived in one subsample (usually the larger, for stability of weights) and applied to the other subsample. The variance explained in the other subsample is taken as the estimate of ρ_c^2. This is probably the most criticized procedure due to its lack of independent samples (e.g., Guion, 2011; Murphy, 1984). For the same reason, it is also a very common cross-validation procedure because it only requires one sample, making it easier to implement.
- Multiple samples and single-equation: In this procedure, two independent samples are drawn from the population of interest. One set of weights is derived in one sample and applied to the other sample. The variance explained in the other sample is taken as the estimate of ρ_c^2. The only difference between this approach and the previous is that independent samples are used rather than one sample split into subsamples. This is viewed as preferable to the single-sample/single-equation approach due to its reliance on independent samples (e.g., Guion, 2011; Mitchell & Klimoski, 1986; Murphy, 1983, 1984). However, I question the extent to which it matters unless there is a specific shift in sampling strategy being pursued with an attempt to be more veridical with respect to the predictive context (e.g.,

weights were derived on an incumbent sample and cross-validity is going to be estimated on a separate applicant sample). If samples are going to be drawn identically (e.g., same pool of employees in the same company) and exposed to more or less the same testing conditions, instructions, and forms, then the distinction between "multiple" and "single" sample designs is largely semantic.

- Multiple samples and multiple equations: This procedure requires that two or more independent random samples are drawn from the population. Weights are derived in each sample, and applied to the other sample(s). The average cross-validity of the equations serves as the estimate of ρ_c^2. Mosier (1951) is credited with suggesting this procedure applied to two samples, referred to as double cross-validation. Some suggest that with this particular procedure, it is no longer clear what the cross-validity statistic is estimating. It's no longer answering the question "How well does this set of weights perform in another sample?" and instead answers the more muddy question of "What is the average performance of two/three sets of weights in different samples?", where the average doesn't necessarily refer to any of the sets of weights (Murphy, 1984).
- Single-sample and multiple equations: While not acknowledged in some taxonomies (e.g., St. John & Roth, 1999), this type of empirical cross-validation procedure involves utilizing double/triple cross-validation on subsamples drawn from a single sample (Raju et al., 1997). These procedures have been extended to bootstrap models (i.e., multiple subsamples drawn; Efron & Gong, 1983; Efron & Tibshirani, 1998) and jackknife models (i.e., weights derived on every subject

except one, then those weights used to predict the criterion for that subject, conducted on every subject; Efron & Gong, 1983). If using double/triple cross-validation, it doesn't make as much sense to split the subsamples into unequal groups, as derived weights should be stable for all groups.

In addition to calculating an estimate of ρ_c^2, it is also important to consider which weights to use. In both of the single-equation procedures, this isn't an issue; there is only one set of weights. In the multiple-equation scenarios, such as with double/triple cross-validation, it is more uncertain as there are multiple sets of weights. It is possible to average the weights, analogous to a meta-analytic mean. Additionally, Mosier (1951) suggested the use of the full sample to obtain final weights as these would be the most stable weights. Guion (2011), Hogan (1994), and Mitchell and Klimoski (1986) concurred with this strategy. Although this will probably produce the "best" possible estimate of weights given the available data, it makes the estimate of ρ_c^2 inaccurate and likely a conservative underestimate of how the actual weights used will perform in new samples (i.e., ρ_c^2 still has to be estimated as the average of the performance of the weights based on smaller samples rather than the actual weights based on the full sample).

There are two common criticisms of all empirical cross-validity estimates. The first is that it "wastes" data.[20] This has been illustrated as a cost-benefit trade-off, with the idea that the best estimate of empirical weights (i.e., weights with the lowest variance due to sampling error) will be derived by using the maximum possible sample size (e.g., Mosier, 1951; Murphy, 1984). Therefore, splitting the sample for an estimate of cross-validity will paradoxically lower the likelihood that the weights will cross-validate as

[20] The one exception is procedures where final weights are still derived on the full sample.

well and subject them to greater shrinkage. This observation has led some to suggest that empirical cross-validation, especially single-sample, is rarely justified (e.g., Murphy, 1984). The second common criticism of empirical estimates of cross-validity is that they are highly dependent on the representativeness of the samples used, specifically the cross-validation sample (Murphy, 1984). Given a large enough sample size, two samples with the same characteristics should exhibit zero shrinkage. However, if the derivation and cross-validation samples are both similar and biased, this may result in an over-estimate of the cross-validity of the weights. This is often suggested to be a larger concern for single-sample cross-validation designs as they are arguably not independent. Thus, a single-sample approach can account only for random error and not systematic error like a multiple-sample design *possibly* could (Cascio & Aguinis, 2005).

An illustration of concerns with the generalizability of empirically-derived weights and usefulness of cross-validity analyses is that almost all biodata validity analyses are based on concurrent validation designs (e.g., utilizing incumbents). For instance, Bliesener (1996) found that 82% of the studies in his meta-analysis used this design. Even some procedures specifically designed to develop keys to be generalizable across organizations (e.g., Carlson et al., 1999; Rothstein et al., 1990) base their keying on incumbent samples. To the extent that incumbents respond differently from applicants, any cross-validation analyses done solely with incumbents may be over-estimates. Therefore, biodata scale developers should pay close attention to how they structure their cross-validation procedures and the inferences they make from them.

An alternative to empirical cross-validity estimates is to use formulas rather than data to estimate cross-validity. This allows weights to be derived using all available data. Unfortunately, these formulas are only available for regression-based procedures, and not for regression-based procedures involving variable-selection (Cattin, 1980; Dorans & Drasgow, 1980; Murphy, 1984; St. John & Roth, 1999). This includes stepwise regression, which is probably the most common regression-based procedure used in biodata keying. Although Schmitt and his colleagues have suggested that formulas for estimating ρ_c^2 can be used with stepwise regression if the initial number of predictors before selection is substituted for the number of the final chosen set of predictors, results on the operation of this procedure have been mixed and may rely on large samples sizes relative to the initial set of predictors (e.g., Mitchell & Klimoski, 1986; Schmitt & Ployhart, 1999). As a result, formula cross-validation doesn't appear to be used very often in biodata contexts, as ordinary linear regression is used very rarely. For more information on formula methods of cross-validation, see the work of Raju and colleagues (Raju et al., 1997; Raju et al., 1999).

Empirical Keying Methods

Option-level Empirical Keying Methods.

Percentile methods. There are two broad methods that fit into the Percentile family: the Vertical Percent Method (VPM; sometimes referred to as the Weighted Application Blank), and the Horizontal Percent Method (HPM). Within these, a number of different variations exist. Credit for the VPM is given to England (1961, 1971), Stead et al. (1940), or Strong (1926), depending on the author, and credit for the HPM is given to Stead et al. Both methods start by creating extreme criterion groups as described above. Again, common splits are the top and bottom 27%, and the top and bottom third. Then, the VPM proceeds by calculating the percent in each of the two groups that chose each response option. Finally, the percent in the bottom group is subtracted from the percent in the top group to obtain a weight for that response option. Positive weights mean that the response option was endorsed more in the high criterion group, whereas negative weights mean it was endorsed more in the low criterion group. The results of this procedure are referred to as "raw" weights.

Most variations of the VPM appear to be for the purposes of computational simplicity or to reduce capitalization on chance by using more coarse weighting strategies. The trade-off, of course, is that with large sample sizes, using coarser weights could potentially limit validity. In any case, other procedures for using the VPM can involve some function of the magnitude or significance of the raw weight. For instance, a scale developer could decide to assign a weight of +1 if the raw difference was between the percents was 5% or higher, and -1 if it was -5% or lower (e.g., Cucina et al., 2012; Devlin et al., 1992). The same procedure could be conducted, with 10% as a cutoff on either side. Similarly, Guion (1965) describes a number of procedures involving significance, such as a test of the significance of the difference of the two percentages, then using the nearest whole number of t as the weight. Finally, VPM raw weights can be turned into "net weights", or "assigned weights". Net weights were introduced by Strong (1926) and take into account the percentage in each criterion group that selected a response, and the percentage difference between the groups, and range from -28 to 28.

Tables for these conversions are provided by England (1961, 1971), Guion, Stead et al. (1940), and Strong. To simplify the scoring process from net weights and to rely less on chance differences, England (1961, 1971) recommended that they be transformed into assigned weights that range from 0 to 2. After determining all of the response option weights, a total score for an individual is obtained by adding all the weights assigned to his/her response options. Simulation studies have generally shown that all the VPM variations perform very similarly as far as validity in a cross-validation sample is concerned, with England's assigned weights working a little less well than the others, and both Strong's net weights and raw weights performing a little better than the others, presumably due to the coarseness effect (Cucina et al., 2012; Devlin et al., 1992).[21]

The HPM, on the other hand, after creating high and low criterion groups, proceeds by dividing the number of people who endorsed the response option in the high group by the number who endorsed the response in both the high and the low group. This figure is then turned into a percent, and divided by 10. A weight of 5 indicates that both groups responded in the same manner, whereas weights between 5 and 10 and weights between 0 and 5 mean that the high criterion group had a higher or lower endorsement rate, respectively. Similar to simplification steps in the VPM, the suggestion to divide by 10 appears to only be for the sake of not having to deal with manipulating large numbers (i.e., the validity of the composite will be the same either way). The HPM is used much less frequently than the VPM, so there are less variations described, and most studies appear to use this raw weight. Guion (1965), however, did suggest alternatives based on

[21] One potential practical benefit of using coarser weight strategies is that items with a weight of 0 can be dropped and the scale can be shortened.

magnitude and significance; this time the magnitude and significance was based on the difference between the HPM weight and a criterion reference point.[22] After calculation of weights, total scores are calculated the same as for the VPM; all weights given to the individual's responses are summed.

Although the HPM is included in most biodata references, it is rarely given much more detail than calling it a variation of the VPM and a couple sentence description of how to implement it. Nobody appears to have evaluated the processes and whether any method inherently makes sense over the other. Table 1 shows the results of applying the VPM and HPM to a set of response situations. All of them assume a sample of 60 people, with the top 20 and bottom 20 being selected for criterion groups. From these, I vary the number of endorsers in each group from 0 to 20 by increments of 5. For the HPM, when either group's number of endorsers is 0, the weight doesn't vary, no matter how many people are in the other group. Ideally, if there are 20 out of 20 instead of 1 out of 20 in the high group, the situations wouldn't be given equal weight. These extreme values

---

[22] Surprisingly, there appears to be some disagreement in how to implement the HPM. Most innocuous are imprecise explanations. Brown (1994) and Stokes and Cooper (2004), for example, say to take the total number of high-criterion group members who endorsed the option and divide it by the total number who selected that response option, *no matter the group they belong to* (i.e., it could be assumed that this is inclusive of the dropped middle group as well). A little further away, Hough (2010) and Mumford and Owens (1987) say that it is the number of subjects in the high criterion group who endorsed the item divided by the *total number of people in the high and low* groups (i.e., doesn't specify whether they endorsed or not), and Dean et al. (1999) indicated the denominator was the *total number of subjects* (i.e., not just the ones that selected the response option, and not just in the high and low group). Finally, Ramsay (2002) suggested that it is the percent of people in the high criterion group who endorsed that option divided by the total percentage of people who chose the option, then multiplying by 10. This is the same procedure as described by Stead et al. (1940), but uses percents in the division rather than the actual numbers. It is unclear what the effect of most of these modifications would be on predictive validity, but the effects of the Brown, Stokes and Cooper, and Ramsay modifications should change the output fairly minimally. The effects of the Hough, Mumford and Owens, and Dean et al. alterations on the actual operation of the procedure is to limit the flexibility and range of calculations. Because the denominator is constant, the weight produced will not vary based on the number of people who endorsed the option in the low group. For example, five high-criterion respondents endorsing an option will result in that option being given the same weight no matter how many respondents endorse the option from the low group.

illustrate the main driver of the HPM: its focus on the percent of the high criterion group that endorsed the response option. The correlation between the HPM and VPM methods is .90 using the set of conditions in Table 1. If cases are removed where one of the groups is 0, the correlation increases to .98, suggesting that most of the difference between the two methods in terms of rank order was due to this lack of variance.[23] Given the lack of sensitivity to these situations, it seems ill-advised to use the HPM over the VPM on logical grounds alone. As discussed below, the HPM shows no advantages in simulation studies either.

Correlational methods. The correlational family is a set of methods that weight response options based on the correlation between that response option and the criterion. They can vary based on the type of correlation coefficient used, whether the criterion is continuous or dichotomous, and whether differential or unit weights are used. Perhaps the most common correlational method is the point-biserial method, which is used when the criterion is continuous (i.e., continuous criterion, dichotomous response option data). Either the raw correlation can be used as a weight, or response options can be unit-weighted based on significance or magnitude (e.g., if the correlation is significant at the .05 level, it is given a weight of either +1 or -1 depending on the directionality of the effect size). It may be better to use magnitude when the sample size is large, as significance may not be a very strong differentiator in these cases. Cucina and his colleagues endorse this method due to its ease of use in standard statistical packages and

[23] If this simulation is extended to include every combination of sample sizes in each group without the step restriction (i.e., 1 to 20, increments of 1), the corresponding correlations are .91 and .95, which are closer due to the relatively lower number of affected values. Before, one out of five were affected; in the increment-by-1 example, it's one out of every 20.

similar performance to other relatively well-performing methods such as the VPM (Cucina et al., 2012; Cucina et al., 2009). As an extension of the point-biserial method, Dean and Russell have conducted research indicating that due to different endorsement rates across response options, the standard error of the point-biserial correlation can differ (see Dean et al., 1999). As a result, response options with an identical point-biserial value may be subject to more or less sampling error. Adjusting for this effect (i.e., dividing the point-biserial estimate by a bootstrapped estimate of the standard error) resulted in slightly better predictive validities.

In addition to the concern about the standard error, it should also be noted that the point-biserial correlation does not range from -1 to 1 like a standard Pearson Product-Moment correlation. Due to the dichotomous predictor, the maximum possible correlation obtained is also a function of the response frequencies. If the dichotomy is an equal 50/50 split, the maximum point-biserial correlation obtainable is .798 (Cohen, 1983; MacCallum, Zhang, Preacher, & Rucker, 2002). This decreases as the splits become more unequal. This effect can be corrected by using the biserial correlation (e.g., Taylor & Ellison, 1967), which estimates what the correlation coefficient would be if the dichotomous variable were continuous. This method appears to be rarely used, however.

When the criterion is dichotomous as well (either naturally, or dichotomized[24]), the correlational method used is the phi coefficient. Lecznar and Dailey (1950) appear to be the first to apply this method to biodata. Similar to the point-biserial correlation, either raw weights or unit weights based on magnitude or significance can be used. Cucina et al.

---

[24] I remind the reader of the previously-described concerns with using extreme groups with the correlational methods.

(2012) make two points worth mentioning. First, the phi coefficient does not attempt to estimate what the correlation would be if the two dichotomous variables were continuous; that is what the tetrachoric correlation does. Second, both the phi coefficient and point-biserial correlations can be computed by using the standard Pearson Product-Moment formula. The phi coefficient and point-biserial methods are computational shortcuts from the pre-personal computer era and now largely exist as a way of conceptually differentiating the coarseness of the variables involved in the correlation.

Mean criterion method. The mean criterion method is a fairly simple option-level weighting method where the weight assigned to each response option is the mean score obtained on the criterion for those who endorsed that response. Accredited to Guttman (1941), it doesn't appear to have been used in many applied settings (for one example, see Karas & West, 1999), and instead is usually only included in recent lists of empirical keying methods or simulation studies (e.g., Cucina et al., 2012; Dean et al., 1999; Devlin et al., 1992; Ramsay, 2002). This method is referred to as the mean standardized criterion method when the criterion is standardized before assigning weights, thereby putting the weights on a standard scale metric. Weights are derived on all data rather than extreme groups. Regardless of which version is used, a strength of this method is the ease of explanation/interpretation of what the weights mean.

Rare response method. The conceptual basis of the rare response method is that odd or unusual responses have unique informational value on the respondent's standing on the characteristic of interest. Though it is often grouped with empirical keying methods, it is not criterion-based. Instead, weights are derived based solely on the predictor/biodata items, and points are given for responses that occur infrequently. Telenson, Alexander, and Barrett (1983) were the first to bring the rare response method out of psychopathology/clinical psychology to the biodata domain. They suggested that if an option was responded to by 30% or more of respondents, it be given a weight of 0, if an option was endorsed by 15-30%, it got a weight of 1, and the weight was 2 if it was endorsed by less than 15% of applicants. Note that weights were derived using the whole derivation sample, not extreme groups. Rather than making a choice of directionality based on data, it was made logically (i.e., if it seemed like it would be positively or negatively related to the criterion). Beyond the original article suggesting its use (Telenson et al., 1983), it has performed poorly in the simulation and empirical studies it has been examined in (e.g., Aamodt & Pierce, 1987; Devlin et al., 1992).

In general, the rare response method's assumption that someone who tended to answer questions differently would be those who would perform well may be untenable. However, even though some suggest it may therefore be only of use in clinical settings (e.g., Cucina et al., 2012; Devlin et al., 1992), given its basis in identifying unconventionality, it could potentially be of use for predicting work criteria like creativity or counterproductive work behaviors. If used for the latter purpose though, scale developers would need to be careful about violating *Americans with Disabilities Act* regulations, similar to personality tests that measure abnormal personality.

Deviant response method. Another rarely discussed or used empirical keying alternative is the deviant response method (Neidt & Malloy, 1954). This particular method was not included in the taxonomy presented by Cucina et al. (2012). It uses a correlational or VPM procedure to produce weights for extreme criterion groups scaled based on the deviation from the regression line after first partialling out the variance due to an initial predictor set.

For instance, Webb (1960) keyed a biographical inventory of 200 items. He compared a typical VPM approach (top and bottom 27% keyed to first year grades) to an approach where college GPA was regressed onto high school GPA, the American Council on Education Psychological Examination, and a locally constructed math achievement test, and high and low groups were constructed (again, 27%) based on the deviation of their predicted score from their actual score using standard regression. In both cases, items were kept if they had a correlation of more than .09 or less than -.09 with the criterion. He found that the VPM did much better in terms of zero-order cross-validity, and that although the deviant technique provided .03 to .07 additional validity points over and above the original predictors, both methods were practically identical in incremental validity in cross-validation. Taken as a whole, these findings suggest the common sense conclusion that the method performed as intended: to identify and include only items related to the portion of variance left over after accounting for the initial predictor set, thus lowering the potential stand-alone validity when compared to the VPM, while still providing similar incremental explanatory power. In contrast to Webb, Malloy (1955) found that application of the deviant method resulted in additional incremental validity over the VPM method even in a cross-validation sample (.02 to .07 additional validity points depending on the sample).

From the very small literature on this method as applied to biodata, three conclusions seem warranted. First, it is geared toward the more limited set of circumstances where the examination of incremental validity is of interest. Second, it seems that its main usefulness would be in item selection rather than weighting (i.e., not wanting to choose items that are redundant with what is already in the predictive model). As such, the main benefit of the deviant technique appears to be in trying to create a smaller biodata scale from a pool of items for a very specific context. The downside to this is that the items would now be selected in reference to one specific criterion and one specific set of other predictor variables and would not be easily generalizable to use either on their own or with any other set of predictors. As a weighting/keying technique, using the standard VPM approach is much more flexible. Finally, the meaningfulness of each weight derived by the deviant method is much more opaque than typical empirical weighting procedures. In reference to the above study by Webb (1960), a positive weight would essentially mean: "Students who endorsed item X tended to be those who obtained a higher first-year college GPA than would be predicted by previous grades and cognitively-loaded tests". Malloy (1955) reflected this interpretation by referring to those in the high deviation group as "overachievers" and those in the low deviation group as "underachievers". Brown (1994) concluded that deviant keying is a "rather convoluted and logically deficient approach to criterion score development" (p. 210).

With the above in mind, use of the deviant technique seems viable only in the limited set of circumstances when (a) a prediction battery already exists and can be expected to be stable for some time (e.g., SAT and high school GPA in predicting expected college GPA), (b) it is acceptable for the inventory to be constructed empirically (i.e., predictive power is the main interest rather than trying to adequately measure specific constructs or adhere to an item blueprint), (c) the inventory will never be used on its own, and (d) it is an acceptable tradeoff to justify every item weight in reference to the predictor battery.

Item-level empirical keying techniques. Item-level empirical keying procedures are given much less attention in the biodata literature than option-keying methods. As a result, these descriptions will be much more brief. The earlier noted characteristics of item-level keying apply to these methods (e.g., assumed linear relationship between item and criterion).

Unit item-weighting. This can be viewed as the default for item-level keying. Rather than attempting to differentially weight at the item level, the responses to each item are simply summed to obtain a total score or subscale score. This is therefore the default regardless of whether the items are dichotomous (i.e., 0/1 endorsement), Likert (e.g., 1-5 or 1-7 scale), or if the response options were already option-keyed.

Correlational item-weighting. Also referred to as pattern-of-response keying by Lecznar and Dailey (1950), this involves simply weighting each item by its correlation with the criterion. Thus, it is like the correlational methods at the option-level described above, but used at the item-level. Like the option-level methods, either raw correlation weights or unit weights based on the magnitude or significance of the correlation can be used. Brief descriptions of the correlational method at the item-level are given in Hogan

(1994) and Hough (2010), and a recent empirical example of its use is provided by Stokes and Searcy (1999).

Multiple regression item-weighting. Although it is possible to use multiple regression to weight items, this appears to be very rarely done by biodata scale developers. The only instance I could find was Steinhaus and Waters (1991), who used it as an illustrative temporary step between unit-weighting and stepwise regression. Similarly, Cucina et al. (2012), state that it is very rarely used, and stepwise regression tends to be used instead. This is somewhat surprising given the potential benefits of regression (i.e., capturing redundancy) and issues identified with stepwise regression. I return to this thought in the below section: "Option-level regression".

Stepwise regression item-weighting. As described earlier, stepwise regression is a procedure that combines empirical item selection and empirical item weighting by selecting and weighting items only if they add a statistically significant amount to prediction beyond the items already selected. This procedure is usually referred to as differential regression in the biodata literature (e.g., Hogan, 1994; Hough, 2010; Stokes & Cooper, 2004), with Malone (1977) most often cited as the originating reference. Stepwise regression tends to be viewed with some degree of suspicion due to its likelihood of capitalizing on chance, underestimation of confidence intervals, and poor performance when multicollinearity is present (e.g., Cohen, Cohen, West, & Aiken, 2003; Thompson, 1995). In light of the potential to capitalize on chance, Cohen et al. (2003) suggested that the ratio of the sample size to the original number of predictors before selection be 40:1, whereas Hough (2010) simply suggested that sample sizes in the thousands are required for stable results. Cucina et al. (2012), Gandy et al. (1994), and Schoenfeldt (1999) are all recent examples of studies that utilized stepwise regression. Cucina et al. compared the unit-weighting and stepwise regression methods at the item-level, finding that stepwise regression performed better than unit-weighting in cross-validation only at a sample size approximately 13 times greater than the number of items (N = 1,850), and required even larger sample sizes to outperform unit-weighting when items were already empirically keyed at the option-level.

Comparison of validity of empirical keying techniques. For a long time, there were few comparisons between different empirical keying approaches. When there were comparisons, they were usually in reference to a proposed new method, such as the rare response method (e.g., Aamodt & Pierce, 1987; Telenson et al. 1983) and the deviant response method (e.g., Neidt & Malloy, 1954; Webb, 1960), and didn't vary the sample size-to-item ratio. To fill this gap, the two most comprehensive comparisons have been conducted by Devlin et al. (1992) and Cucina et al. (2012).

Devlin et al. (1992) used five cohorts of people who attended the U.S. Naval Academy to test key-development strategies on groups with varying sample sizes (75, 150, 300, 600, and 1,200). They compared five VPM methods (raw, 5% and 10% magnitude of percent difference unit-weighting, Strong's (1926) net weights, and England's assigned weights), the HPM, the phi coefficient method, the mean criterion method, and the rare response method. They found that the variations of the VPM generally had the highest cross-validities; the variation that performed better varied based on the sample size, with the two best generally being the raw difference and the Strong net weights. The mean criterion method had the highest validities in the derivation samples but exhibited the greatest shrinkage. At the highest sample sizes, however, its cross-validity was comparable to the others. The rare response method exhibited almost no cross-validity.

Cucina et al. (2012) went a step further than Devlin et al. (1992), comparing rational, empirical, and hybrid (using information from rational and empirical keying) methods. With respect to the option-level empirical keying methods, they sampled from 3,515 people to develop keys, with 1,757 in the cross-validation sample. Keys were developed on developmental groups ranging from $N = 150$ to $N = 3{,}515$. They compared 12 different methods: biserial correlations, raw phi coefficients, phi coefficients unit-weighted if significant, raw point-biserial correlations, point-biserial correlations unit-weighted if significant, the mean criterion method, the HPM, the VPM unit-weighted if the difference was at least 5%, the VPM unit-weighted if the difference was at least 10%, the VPM with England's assigned weights, the VPM with Strong's net weights, and raw VPM weights. In general, they concluded that it does not matter much which option-level procedure is chosen, though they recommended the point-biserial method for ease of implementation. Additionally, after weighting each item's response options, they either unit weighted or used stepwise regression at the item-level. In general, they found that stepwise regression had lower validity at lower sample sizes, but higher validity at larger sample sizes. They then proposed a decision-tree based on their simulation results.

Two common limitations are present in the Devlin et al. (1992) and Cucina et al. (2012) articles. First, they do not vary the number of items or discuss findings in terms of the sample size-to-item ratio. Although the absolute sample size is useful for the interpretation of a given study's findings, to be able to generalize simulation comparison results to other settings with other biodata scales, the number of items relative to the sample size is a critical piece of information. Although this statistic can be calculated from their articles, without varying the number of items, the picture is not complete. It is possible that for some methods, the sample size-to-item ratio required may also be dependent on the number of items, particularly for regression-based methods, like those used in Cucina et al. (2012). In other words, Cucina et al.'s (2012) decision-tree could only be accurate for their specific inventory (or inventories of similar length). The second limitation is that neither simulation study is exhaustive of potential methods (nor claims to be). Two potential areas empirical keying can be extended to are option-level regression and configural methods.

Option-level regression. One option that is not mentioned in biodata references is the possibility of dummy-coding the entire set of response options and using multiple regression on the resulting dataset. Conceptually, it is plausible that multiple regression could outperform other option-level keying methods by taking into account redundancies among variables in addition to being able to account for non-linearity. However, it has two sample size related disadvantages that may preclude its operational use in a number of settings. First, a sample size greater than the number of items is needed. Second, in addition to overcoming sampling variability in the predictor-criterion relationships, regression must also overcome sampling variability in the predictor intercorrelations. Weiss (1976) suggested that the vertical percent method may be more stable in cross-validation than regression for this very reason as it uses less information about the specific derivation sample. Therefore, regression may outperform other option-level keying methods, but only with large sample sizes. To wit, a common rule of thumb for multivariate methods is a need for a sample 10 times larger than the number of variables. Using this metric for a 300-item biodata questionnaire, a sample of 3,000 would be needed. In many settings this would be untenable, suggesting that other methods, even if somewhat less accurate, might be preferred.

Along these lines, Monte Carlo work (Campbell, 1974) has recently been cited in a number of references on biodata (e.g., Hough, 2010; Mumford & Stokes, 1992; Stokes & Cooper, 2004; Stokes & Reddy, 1992) for its recommendations on sample size requirements for using regression procedures versus zero-order correlations and unit-weighting. Campbell (1974) reported that multivariate procedures should not be conducted with sample sizes below 150, and multiple regression probably should not be attempted unless the sample size is above 250 and the multiple *R* of the predictors is relatively large (> .50). However, Campbell also noted that these recommendations were based on sets of data where there were either four or eight relatively reliable predictors, which likely does not generalize to cases where many items are being examined in an item pool. In addition, because the number of predictors was low and limited to those expected to be highly relevant to the criteria, the average validities and intercorrelations for the items were higher than one might expect in the average biodata inventory (e.g., mean validity = .26, *SD* = .14, and mean intercorrelation = .41, *SD* = .25). Given these restrictions, these recommendations seem to warrant further investigation before being applied to the typical biodata inventory.

There is a final potential problem with using regression at the option-keying level. If an item has more than two response options, dummy-coding creates sets of variables that are not independent (i.e., if respondents can only endorse one response option, the other response options by definition cannot be endorsed). This violates the assumption of independence among the predictors, which can make the interpretation of regression weights difficult (Dean et al., 1999). As a result, if regression does perform better than other option-keying methods, this may only apply to dichotomous items (i.e., items with only one resulting variable when dummy-coded), and perform more poorly when there are more response options.

Configural methods. Regression and other previously described empirical approaches are typically used to investigate main effects. As such, they don't capitalize on the potential predictive value of higher-order interactions and/or configurations of predictor scores. In many areas of study, it is quite possible that the real predictive power lies in configurations or patterns. Configural methods have long been proposed to have the potential to provide great increases in predictive power (e.g., Gaier & Lee, 1953; Meehl, 1950)

Meehl was one of the first and most vocal supporters of applying a configural approach to psychology (1950). His particular focus was on clinical psychology, where he argued that it was common procedure for clinicians to use combinations of predictor scores (or less formally, their observations) to diagnose patients. However, these combinations were usually made on an individual basis and could vary from clinician to clinician. Meehl argued that it was important to investigate these relationships empirically, rather than relying on anecdotal evidence or intuition, and even proposed the development of "cookbooks" of common and/or generalizable configurations, backed up by data (Meehl, 1956).

In order to substantiate the potential importance of investigating configural approaches, Meehl (1950) gave the provocative example of two dichotomous items designed to differentiate "schizophrenic" from "normal". With the stipulation that each of the items has 50% difficulty within each category (i.e., half of the schizophrenics and half of the normals endorse each of the items as true, and half of both groups endorse each item as "false"), he argued that typical methods of dichotomous item analysis, such as the significance of difference of proportions or phi coefficients, would find no validity for the items and they would therefore be removed from a scale. To illustrate his point, he added the extreme stipulation that the inter-item correlation among the items be 1 for the normal group, and -1 for the schizophrenic group. In other words, this intercorrelation pattern implies the response table presented in Figure 2. Although this is an extreme case, Meehl proceeded to demonstrate that because all people either answer both items the same way ("normal" group) or always give opposite answers ("schizophrenic" group), if the information from these items is combined, perfect classification is obtained, as shown in Figure 3. The combination of these items is a basic example of what is meant by taking a configural approach. Meehl went on to discuss how this concept can be extended to less extreme inter-item correlations, more than two items, and cases where the items actually do have validity. This pattern of results became known as "Meehl's paradox", and its underlying mathematics were expanded by Horst (1954, 1968), who demonstrated that it was a special case of the properties of the linear combinations of products of item scores (i.e., the sum of items and their interactions).

Extending his research beyond an example into empirical research, Meehl took initial steps in examining the predictive efficacy of configural approaches, focusing on the Minnesota Multiphasic Personality Inventory (MMPI; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and its ability to diagnose psychopathology. Both Meehl (1959) and Meehl and Dahlstrom (1960) found support for the idea that techniques that are built to capitalize on configural relationships can aid in prediction. Specifically, Meehl (1959) found that in a comparison of five statistical methods and clinical judgment on the cross-validated predictive power of identifying psychotics, the four statistical methods designed to detect configural relationships all exhibited higher validity than the judgment and remaining statistical method (linear discriminant function). Similarly, in a related study, Meehl and Dahlstrom further examined the usefulness of a procedure similar to the one found most effective in Meehl (1959) and again determined that it was more effective than clinical judgments. Taken together, these studies were viewed as evidence that at least in some domains, benefits can be obtained for both explanation and prediction by investigating configural relationships. Today, applied usage of empirically-backed configurations of personality scales can still be found in so-called "cookbooks" (e.g., McAllister, 1996) and the computerized version of the MMPI (Meehl, 1989).

Similar to the proposed efficacy of configurations in prediction and explanation in clinical psychology, some in the biodata field have suggested configural approaches as a way of improving the usefulness of biodata (e.g., Brown, 1994). One simple way of adding configural relationships to predictive systems is to calculate the interaction between two items or response options and incorporate that interaction as part of the empirical keying procedure. However, the number of potential combinations in even small inventories can quickly become unwieldy and overwhelm the power of the available sample size to detect interactions that are significant. This approach could potentially be useful for investigating a few potentially critical interactions suggested by theory.

Another approach for solving the "number of interactions" problem is provided by more advanced configural methods. One example of such an attempt is a set of methods known as Automatic Interaction Detection (AID) or sometimes referred to as classification and regression trees. This is what McQuitty (1957) referred to as a cumulative approach. The basic idea is to partition a dataset of subjects into a set of mutually exclusive and exhaustive groups (i.e., all subjects are classified, and each subject is classified into only one group) based on the statistical relationships found with a dependent variable (Magidson, 1994). This process is often depicted as growing a tree as individuals are ultimately nested within mutually-exclusive branches based on their responses.

For instance, consider a dataset with ten independent variables and one dependent variable. In the first step, the procedure would find the independent variable that had the strongest relationship with the dependent variable and calculate the optimal splits in that independent variable that would maximize a statistical criterion (e.g., largest chi-square statistic, most homogeneity). This would create a certain number of resultant nodes into which the subjects in the dataset would be classified. The next step would examine those new nodes, look for the next most predictive variable and best splits in that variable and continue in the same fashion. Conceptually, partitioning the dataset in such a manner is creating a set of interactions. While there are a number of specific procedures that implement this type of process, they largely differ only in how they calculate splits, how many splits they allow, and what types of variables they allow (e.g., ordinal, interval; Kass, 1980). AID methods are only one type of configural method (for many more see Hastie, Tibshirani, & Friedman, 2009; Witten, Frank, & Hall, 2011), but they are one of the only configural methods that have been applied to biodata.

Most applications of configural techniques to biodata inventories to this point have provided uninterpretable or disappointing results. For instance, Gandy et al. (1994) chose a series of two-way and three-way interactions that were thought to be theoretically relevant in predicting job performance of federal employees. However, they provided only an example of a significant interaction rather than giving an omnibus increment of adding every possible interaction term into their regression model. Tanofsky, Shepps, and O'Neill (1969) used AID procedures to examine six previously predictive items for insurance salesmen's sales production. They found six significant subdivisions, but no attempt was made to cross-validate these findings or to compare them to any other method. Finally, Grauer (2006) applied neural networking (i.e., a machine learning procedure conceptually related to the AID family) and regression to the prediction of a number of performance and turnover criteria. Neither technique performed well with the seven predictors used, but regression results tended to perform better and exhibit less shrinkage than neural networking. From the biodata literature reviewed, there does not appear to have been direct comparisons of configural methods to other empirical keying methods beyond regression, and most previous efforts seem to have been more along the lines of a proof of concept investigation of a new method. Dean et al. (1999) concluded that “it remains to be seen whether procedures using configural predictor combinations incrementally increase prediction or understanding at a level that justifies their complexity and cost” (p.18).

Recapitulation and Purpose

In the preceding, I have circled around a number of related themes that I now wish to emphasize before turning to the details of the present set of studies. First, empirical keying is not incompatible with a rigorous item-development strategy targeted at specific constructs or job-related behaviors. Its use or the use of rational keying should be predicated on sample size concerns rather than full adherence to an “approach” or “philosophy”. Second, instead of a specific sample size threshold, the sample size-to-item ratio is likely to be most informative regarding sample size requirements. Third, although research that compares many empirical keying procedures exists, it is by no means comprehensive, and the relative effectiveness of rarely-used methods such as option-level regression and configural methods is unknown. With this in mind, this dissertation will attempt to answer the following research questions:

Research Question 1: Do the option-level regression and configural methods (specifically those from the AID family) provide enhanced predictive power when compared to more traditional empirical keying procedures such as the vertical percent method and point biserial correlation method?

Research Question 2: If any benefits are obtained from using option-level regression and configural methods, to what sample sizes (in terms of the sample size-to-item ratio) will these benefits generalize?

Research Question 3: Are required sample size-to-item ratio findings dependent on the number of response options items have?

STUDY 1

The first study focused on whether the use of option-keyed regression and/or AID methods result in any benefits over using more traditional keying methods like the vertical percent method and the point-biserial correlation method (i.e., *Research Question 1*). I investigated this question using a very large archival dataset that contained responses to an inventory with biodata-style questions. Given that the dataset was large enough to fairly safely ignore the effects of sampling error, these analyses can be viewed as addressing whether it is *possible* for either option-keyed regression or AID methods to outperform traditional keying methods. Later studies will replicate findings and investigate the boundaries under which the findings hold.

Method

Sample. The College Board collected data from 150,374 students in the 2006 entering cohort at 110 universities and colleges in the United States. The universities and colleges were specifically selected to increase diversity across a number of categories, including geographic location, school size, public versus private schools, and school selectivity. The sampling plan and subsequently obtained sample size give this dataset a large degree of generalizability among college students. See Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008) for prior research using this dataset. Although this sample is from an educational context, the properties of the biodata inventory used are analogous to what one would expect in an employment setting, and the wide variability among schools sampled limits the chance that idiosyncrasies related to one specific school setting would skew results in any way.

Measures.

Freshman grade point average (FGPA). FGPA was provided by the universities/colleges from student records after a student's first year at that institution.

Student Descriptive Questionnaire (SDQ). The SDQ is a 331-item self-report measure students completed at the time of taking the SAT. To facilitate analyses, some items that involved students choosing from a list of possible responses and were presented as categorical variables with a large number of possible values were transformed into binary items that corresponded to endorsement of each of the possibilities[25]. This resulted in 365 base items. The scoring methods that were applied to the SDQ in this study involved assigning weights to individual response options, and thus each variable was converted into a series of dichotomous Yes/No variables indicating whether a particular response option was endorsed. An "item" in this scheme is based on the total number of response options per item. This has implications for the sample size needed and subsequent interpretations of study results (i.e., in this inventory, this procedure resulted in 948 "items").

Although the SDQ was constructed to gather data about SAT takers for research purposes, it taps a variety of constructs broadly relevant to the prediction of academic achievement. The SDQ includes items dealing with academic rigor (e.g., number of courses taken, honors and advanced placement courses taken), academic achievement (e.g., average course grade in various disciplines, class rank), academic and social

[25] Specifically, this involved sports participation and projected major choice. For example, sports participation was represented in the dataset as 6 variables, each of which indicating whether one of 36 sports had been selected by the respondent. Rather than 6 items with 36 possible responses each, these items were converted to 36 binary items each of which indicated whether the sport had been selected by the respondent in any of the original 6 items.

engagement (e.g., participation in a wide variety of extracurricular and sports activities), and level of aspiration (e.g., degree goal and intended major) as well as demographic information (e.g., father's and mother's education and family income, ethnic and gender identification, and ESL status). While not all of these reflect past behavior (e.g., college intentions are more values/belief-based), all of these variables have the potential to exhibit a relationship with FGPA by tapping engagement and motivational constructs.[26] This reflects one common attribute of biodata, namely, a multifaceted item pool tapping a number of constructs.

Procedures and data analysis plan. Infrequently endorsed options on the SDQ were pared down in accordance with recommended practice in dealing with empirically keyed biodata (Hogan, 1994). Specifically, when a response option had fewer than 200 subjects endorsing it, I combined its data with the adjacent response option. For instance, if there were eight response options for the number of semesters of English a student took in high school (i.e., 1-8), and fewer than 200 students took only one semester of English, those students who took one semester of English were merged with those who took two semesters. Though these recommendations are primarily for ensuring there are no response options with very small sample sizes, I felt that it might also produce more meaningful and interpretable results in the weighting procedures, while still retaining the purpose and the bulk of the variability of the initial questions. This was only done in situations where the response set was arguably continuous as in the above example. Subsequent to the response option reduction procedure, I conducted four procedures to

[26] A copy of the current SDQ in use by the College Board can be downloaded at http://www.collegeboard.com/sss/repository/SAT%20Questionnaire.pdf.

develop weights for the SDQ responses: the VPM, the point-biserial correlation method (PBC), option-keyed multiple regression, and the Chi-squared Automatic Interaction Detector (CHAID).

Vertical percent method (VPM). The VPM is a traditional empirical biodata keying procedure described by England (1961, 1971) and based on a weighting scheme proposed by Strong (1926). The VPM has considerable evidence supporting its validity in comparison to both other empirical keying procedures and rational keying procedures, even after cross-validation (Devlin et al., 1992; Hogan, 1994; Mitchell & Klimoski, 1982; Mumford & Owens, 1987).

First, high and low criterion groups were chosen. When dealing with dichotomous criteria such as turnover, this is fairly simple, with those who stayed being coded as the high criterion group, and those who left coded as the low criterion group. When using continuous criteria, there are more options. Hogan (1994) suggested that the most common procedure was Kelley's (1939) recommendation to set the high criterion group as the upper 27% and the low criterion group as the bottom 27% in order to maximize differences while having large sample sizes in both groups. Additionally, Devlin et al. (1992) compared a number of different splits (e.g., top 5%, top 10%, etc.) in predicting a GPA analogue for the U.S. Naval Academy, and the results were fairly stable across the various operationalizations of the VPM. For the FGPA criterion, I used Kelley's recommendation and selected the top 27% for the high criterion group (FGPA above 3.47) and the bottom 27% for the low criterion group (FGPA below 2.63). I tested multiple other splits (1%, 5%, 10%, 15%, 20%, 27%) that cover a wide range of extreme group cut points and, as shown in Table 2, composites based on these different splits all correlated above $r = .99$ with one another. Thus the split utilized to generate the high and low criterion groups did not affect conclusions.

After choosing criterion groups, the next step was to compute the percent within each of the criterion groups who chose each item alternative. Then, the percent from the low criterion group was subtracted from the high criterion group. Because other studies have shown that typical transformations of these weights often perform worse (e.g., Cucina et al., 2012; Devlin et al., 1992), this raw difference was the weight that was used. Finally, each of these weights was applied to the subjects' item responses, and these weights were summed to create a total biodata score for each individual.

Empirically keyed biodata weights have a risk of exhibiting shrinkage when weights are applied to a sample other than the one in which the weights were derived (Hogan, 1994; Mumford & Owens, 1987). Indeed, any procedure that uses purely empirical bases to generate a weight will likely capitalize on sample-specific idiosyncrasies. As such, even though there was an enormous and arguably representative sample, I used a cross-validation procedure to test the generalizability of the weights derived from the VPM. I randomly selected 2/3 of the entire sample of students to be in the weight-development group, and used the remaining 1/3 of the sample to be the cross-validation group. This resulted in sample sizes of 100,898 and 49,476 for the weight-derivation and cross-validation samples, respectively. Using these samples, weights derived via the VPM in the weight-development group were applied to both groups. The resulting composites of weighted responses in both groups were then used to predict FGPA.

Point-biserial correlation (PBC). Similar to the VPM, analyses using the PBC took place at the response option level. Each response option, represented as a binary 0/1, was correlated with the criterion, and the resulting correlation was used as the weight for that response option. The weights for the endorsed response options were summed, and this composite was used as the biodata score. The same procedure for cross-validation was used as was used with the VPM, using the same random draw of subjects into the weight-derivation and cross-validation groups.

Option-keyed multiple regression. The VPM is specifically used in biodata contexts, and much of the detail of its implementation, though not complex, is tedious. It would be useful to know if more traditional, standard methods of data analysis besides the PBC method work similarly or better. One likely reason why option-keyed multiple regression could be expected to do better than traditional option-keyed methods is that regression takes into account the intercorrelations between variables when deriving weights, whereas the VPM and PBC examine each item individually. This could lead to over/under-weighting measurements of the same construct when using the VPM or PBC method because the weight each construct received would be a function of how many items there were that measured it. With regression, redundancies among items would be eliminated. As such, I conducted multiple regressions using the same data and response options from the VPM and PBC analyses. I then converted each variable's response options into dichotomous Yes/No dummy variables indicating whether that particular option was endorsed. For instance, following traditional dummy-coding procedures and using the previous example of number of semesters of English, the eight possible semesters were converted to seven Yes/No (1/0) variables, with all entries being 0 indicating the student was in a reference group (whichever group wasn't included as a dummy variable). When the questions only had two response options (i.e., were already dichotomous; "Did you take an Honors English course"), I left them as they were. I then regressed FGPA on these variables. This causes a shift in the meaning of what an item is. The number of items in this scheme is now the number of response options minus one for each item. In this case, the result was 583 response options. This has implications for the sample size needed to use regression; specifically, the weight derivation group must have at least as many people as dummy-coded response options plus one.[27] The same cross-validation procedure was used as described previously, and the samples of subjects used in each group were based on the same random draw as for the above two methods.

Chi-squared automatic interaction detector. The above three procedures are focused on capturing main effects. It is possible that there are interactions between variables that would allow for enhanced prediction over simpler linear relationships. To investigate that possibility, I used a procedure from the family of analyses known as Classification and Regression Trees or machine learning called the Exhaustive Chi-squared Automatic Interaction Detector (CHAID) (Biggs, De Ville, & Suen, 1991; Kass, 1980).

[27] To reduce confusion, carrying forward, the original items will be referred to as "base items". The set of transformed variables that include a binary variable for every response option will be referred to as "total response options". Finally, the set of transformed variables that is used with regression will be referred to as "dummy-coded response options".

CHAID starts by selecting the independent variable with the strongest relationship with the dependent variable. It then splits subjects into separate groups based on their responses to that independent variable and finds the next independent variable that is most strongly associated with each of the resultant groups through brute-force empiricism, using chi-square tests with adjusted Bonferonni significance values. The algorithm then continues until it reaches its specified limits (depth of the tree, and minimal size of nodes for splitting) or there are no longer significant splits to be made. To conduct the analysis, I utilized the Decision Trees package from SPSS. I applied this algorithm to the FGPA criterion, and used the same weight-derivation and cross-validation samples as described for the previous three procedures.

Because CHAID involves finding a series of multi-way interactions, there is a strong possibility that multi-way interactions at deeper levels of the tree are highly dependent on the sample on which the weights are being derived. This made it important to investigate the optimal settings for parameters in the CHAID algorithm. To do so, I systematically varied the depth of the tree (from 2 to 14 levels) and the significance level for splitting nodes (p-values from .05 to .00000001). It was expected that the deeper the depth of the tree, the higher the R^2 in the weight-derivation sample, but the worse the tree would perform in a cross-validation sample. Similar results should hold for more lenient alpha levels. I found the combination that maximized variance explained in the cross-validation sample and used that as my comparison to the other methods.

For all methods, weights were derived on the weight-derivation sample across all schools, applied to all data, composites were created, and then the performance of the method was calculated separately within the weight-derivation and cross-validation samples of each of the schools in order to ensure I was not conflating between-school and within-school effects. The estimates for each school's samples were then averaged across all schools weighting by sample size. This helped to balance schools with differing sample sizes in the final average, and it also allowed examination of the variability of results between schools.

Results and Discussion

Sample size weighted means, standard deviations, and correlations between study variables for the weight-derivation and cross-validation samples are presented in Table 3 for composites used in the prediction of FGPA. Although correlations in the cross-validated samples tended to be slightly lower than those in the weight-derivation samples, the amount of shrinkage was quite small. This is expected given the large sample size in this dataset. In addition, although there is some indication of variability among the correlations across schools (as shown by the *SD* rhos in Table 3), it doesn't appear to be of a magnitude to radically alter conclusions based on an individual school (i.e., FGPA tends to have relatively robust predictability). The traditional methods are correlated extremely highly with each other (above .99), whereas the correlations between the other empirical keys are lower, indicating differentiation on what variables they are focusing on or how they are weighted.

Table 4 reports the results of the investigation of the efficacy of methods of creating biodata weights. As these results are also based on sample-size weighting, they will not be exactly equal to the squared correlations from Table 3. Traditional composites result in an R^2 of approximately .10 in the cross-validation sample, with virtually no shrinkage exhibited when compared to results in the weight-derivation sample. In contrast, the option-keyed multiple regression composite resulted in an R^2 of .20 in the cross-validation sample, but exhibited a little bit more shrinkage. In this sample of subjects and items then, regression weights explained approximately twice as much variance as the traditional methods. This is a dramatic increase in predictive power. However, even with such a large sample size, regression appeared to pick up on non-replicable sample-specific idiosyncrasies resulting in shrinkage from the weight-derivation validity, which perhaps indicates caution for use in smaller samples.

The results from CHAID are more difficult to clearly interpret. As discussed in the method section, using CHAID requires specifying a number of parameters for algorithm operation. Using the specific parameters that maximized the cross-valid variance explained, the resulting composite explained 16.5% of the variance in the weight-derivation sample, and 14.6% in the cross-validation sample. Based on this result, one might conclude that CHAID split the difference between the traditional methods and multiple regression.

However, the parameters for CHAID were chosen after testing 104 combinations of tree depths and significance for splitting nodes to determine which combination maximized the variance explained in the cross-validation sample. The tables in Appendix A contain the results of these trials. In general, the more lenient the alpha level required for node-splitting and the deeper the depth of the tree, the higher the weight-derivation variance explained, but the greater the keys exhibited shrinkage in the cross-validation samples. Figure 4 provides an example of the effects of varying tree depth at a node splitting significance level of .05. It suggests that there is a maximum R^2 in the cross-validation sample at a certain point, after which a sharp drop-off occurs with deeper trees. The weight-derivation variance explained, on the other hand, increases the deeper the tree gets. The keys that maximized variance explained in the cross-validation sample tended to be much less complex than the keys that maximized the variance explained in the weight-derivation sample. For instance, the key that maximized variance explained in the weight-derivation sample had 13,612 nodes, whereas the key that maximized variance in the cross-validation sample had only 244 nodes. In total, out of the 104 combinations, the mean cross-valid variance explained was .13, and 16% of the combinations produced results inferior to the traditional methods. Even in this large of a dataset, this procedure leading to specific algorithm parameters chosen could be open to the charge of capitalizing on chance. For a smaller, more typical, dataset this procedure becomes even more tenuous. Overall, in spite of the added complexity involved in conducting analyses with CHAID, it produced results inferior to those from using option-keyed regression. Given this, and the specialized nature of the technique and hefty computational time requirements limiting its regular use, I removed CHAID from further consideration.

In sum, this study used a 331-item biodata questionnaire in an educational sample of 150,374 college students from a wide variety of schools. Keying based on multiple regression explained roughly twice the variance explained in cross-validation samples than the traditional methods. In contrast, the use of CHAID produced results inferior to multiple regression and was removed from analyses in subsequent studies. The results of this study were based on a large dataset, and should be firm with respect to this biodata instrument (SDQ) and this population (college freshmen). However, other instruments and populations could operate differently. As a result, Study 2 will attempt to replicate the predictive superiority of option-keyed multiple regression to traditional keying alternatives in two additional archival datasets.

STUDY 2

Study 2 continues the investigation of *Research Question 1* (Does the option-level regression keying method provide enhanced predictive power when compared to more traditional empirical keying procedures?) by replicating the findings in Study 1 on two separate datasets. One dataset is based in an educational context, but focuses on items purported to measure a specific construct to predict FGPA. The second dataset is based in an employment context and focuses on the prediction of job performance ratings. The finding of similar results would bolster the conclusion that option-keyed regression will outperform traditional methods and that the results in Study 1 were not due to a sample- or inventory-specific aberration.

Method

The procedures in Study 1 were replicated on two additional archival datasets with different qualities in order to test the generalizability of findings. Table 5 compares a number of relevant features of the datasets in Studies 1 and 2 and will be discussed in more depth later.

Academic Rigor Index (ARI). First, whereas the SDQ arguably tapped a wide number of constructs, one replication was conducted using a set of questions focused on one specific construct: the academic rigor of an entering college student's high school curriculum. These data also came from the College Board and were gathered under a similar sampling plan as the SDQ. Specifically, the data were gathered by the College Board in collaboration with a set of diverse and broadly representative colleges and universities. The sample includes 67,644 students from the 2007 entering class of 110 schools who completed all measures in the SAT Admitted Class Evaluation Service™ (ACES™) dataset. Academic rigor was measured by gauging the difficulty and breadth of students' high school curricula in English, mathematics, natural science, social science/history, and foreign and classical languages. Item information included general course titles and grade level when taken as well as Advanced Placement, honors, or dual-enrollment participation. Each question required a simple Yes/No based on whether a student participated in a given course during a given grade, and there were 395 items in total. Because these were all binary items, 395 is also the total number of response options and the number of dummy-coded response options. As with the SDQ dataset in Study 1, the criterion used with this dataset was FGPA. The use of the ACES™ dataset to develop the ARI is described in more detail by Wiley, Wyatt, and Camara (2010).

Individual Achievement Record (IAR). The second replication extended analyses to a different context. Rather than focusing on the educational domain, it used data gathered from a work context. Data from 5,277 employees were gathered to develop and validate the IAR, a biodata inventory designed for selection in the U. S. federal government across a wide variety of jobs (Gandy et al., 1994). The IAR has 139 items that were developed by investigating job-analysis information on federal occupations, extant taxonomies of past behavior items, and screening on acceptability for public-sector use. This included questions tapping educational experience, work history/skills, and interpersonal relations. Example questions include the number of previous jobs held, grade point average in college major, and the number of clubs/activities participated in while in high school. When option-coded, the 139 base items resulted in 695 total response options and 556 dummy-coded response options. The criterion in this dataset was the mean on a performance rating scale filled out by a supervisor. The criterion items loaded heavily on a general factor, with 71% of the variance due to the first unrotated component.

Data analysis procedure. With each of these datasets/inventories, procedures from Study 1 were largely replicated. All items were converted to option-level data and rare response options for item representing a continuum were collapsed. The participants were split into a 2/3 weight-derivation group and a 1/3 cross-validation group. The VPM (with 27% extreme groups), PBC, and option-keyed regression were applied to the weight-derivation group to obtain weights. These weights were then applied to all data, and the variance explained by the resulting composites was calculated for both the weight-derivation and cross-validation groups to compare to results from Study 1. In the ARI dataset, empirical key results were calculated at the school level then sample-size weighted. In the IAR dataset, results were calculated across the whole dataset.

Results and Discussion

Table 6 is the mirror of Table 3 and contains descriptive statistics for the ARI dataset, including means, standard deviations, and correlations between the empirical keys and FGPA for both the weight-derivation and cross-validation groups. Many findings from Study 1 were replicated with this dataset. As with the SDQ dataset, there was little shrinkage when comparing correlations in the weight-derivation group to the cross-validation group, but the degree of shrinkage was slightly larger in this dataset. Additionally, there was a very high correlation between the VPM and PBC methods (again, it was over .99 in this dataset), and lower, but still high, correlations between option-keyed regression and the traditional methods.

One difference between the results presented in Table 6 and those presented in Table 3 is that the *SD* rhos indicate a greater potential for overlap among empirical keys. Specifically, in the SDQ dataset, a school would have to be fairly aberrant in both the performance of the traditional methods (higher than the mean) and regression (lower than the mean), approximately a combination of two standard deviations, in order for there to be overlap. In the ARI dataset, the traditional methods could outperform regression if one of them was one standard deviation higher or lower, respectively, than the average. This suggests that regression's predictive superiority is less definite in this dataset.

Table 7 compares empirical key performance in aggregated variance explained terms and is comparable to the SDQ results in Table 4. Coinciding with the smaller sample size in the ARI dataset, the degree of shrinkage is a little higher than was in the SDQ dataset, and traditional methods explained the same amount of variance as each other. There was still an advantage in predictive power for the regression method, but it was not as large as it was in the SDQ dataset. In Study 1, regression explained double the variance as the traditional methods. In the ARI results, regression explained only about 44% more variance than the traditional methods did.

Turning to results from the IAR dataset, the descriptive statistics are presented in Table 8. Again, the correlation between the keys derived via the traditional methods was above .99, and the correlation between regression and the traditional methods was lower, but still high (.538 in the cross-validation sample). The main difference in this table compared to previous ones is that shrinkage is higher, particularly with the regression key. Validity for this key just about halved in the cross-validation sample. Although the validity for regression was still higher than for the traditional methods, the difference is smaller than with the previous datasets. This, combined with the large amount of shrinkage, suggests the possibility of boundary conditions for the superiority of the regression method.

In Table 9, the variance explained by the three empirical keys is presented. In line with the results from Table 8, regression has an edge over the traditional methods, but it is smaller than it was in the ARI and SDQ datasets, explaining 26% more than the traditional methods. Additionally, the transformation to variance explained makes the degree of shrinkage for the regression composite even more stark, with the cross-validation variance explained being slightly higher than one-third of the size of the variance explained for the weight-derivation sample.

Overall then, many of the findings from Study 1 were replicated in Study 2. The traditional methods performed almost identically, and regression was the superior empirical keying method in all three datasets. What did vary, however, was regression's edge over the traditional methods. It went from 100% more variance explained than traditional methods to 44%, to 26%. There were a number of potential differences in the datasets that could have been explanatory factors in these results. For instance, biodata item content and the number of these items differed across these datasets, as did the criteria, the context, and the sample size.

Two of the potential explanatory factors listed seem less likely than the others to play a role in the variation of empirical key performance. Neither FGPA nor rated job performance is a criterion that is often thought of as particularly difficult to predict (as opposed to retention and turnover, for instance) and there doesn't seem to be evidence that either would be more suited than the other for measurement with biodata-style questions. Similarly, it seems unlikely that biodata keying in general or a specific method would lend itself particularly well to an educational or an occupational context. Additionally, variation in results occurred both within and across criteria and contexts (i.e., results differ within educational contexts with FGPA as a criterion and they both differ from the employment context with performance as a criterion). Item content and sample size concerns therefore seem more likely to explain the variation in how the keying methods performed relative to one another.

With respect to the item content, if one inventory contained items that were "better" or "worse" in terms of predictiveness than the other inventories, or had items with more or less redundancy, this could result in differences in how the empirical keys operate. A window into this is presented in Table 10, which contains the average validity and intercorrelation for the dummy-coded response options for the predictors in each of the datasets investigated in Study 1 and Study 2. The results from this table suggest that the item content is similar in the datasets in terms of how predictive the response options tend to be and how related they are to one another. Therefore, although response option validity and intercorrelation may play a role in how these keying methods operate relative to one another, they do not appear likely to be a major role in the difference in results between Study 1 and Study 2.

What do appear to be likely contributors to the variation of results, however, are the number of items and the sample size in the datasets. Table 5 provides a comparison of the number of items and respondents in each dataset. First, the number of dummy-coded response options is correlated with the cross-valid variance explained by regression. This isn't surprising, as all else equal, the more items with the same validity, the higher the potential for prediction. Also of note is the last column, the ratio of the sample size in the weight-derivation group to the number of dummy-coded response options. This ratio correlates perfectly with the drop in the advantage of regression over the traditional keying methods in variance explained and coincides with the increasing shrinkage as this ratio decreases as well. This indicates that regression might need a large sample size to provide maximum benefit from its use. However, even six times the number of participants to items can be a prohibitive requirement for keying a large biodata inventory, much less 100 times.

It would be informative to investigate further to determine whether there is a point at which regression will not outperform traditional keying methods. As the weight-derivation sample size nears the point at which it is barely sufficient for regression to run (i.e., as it approaches the number of dummy-coded response options), it seems reasonable to expect that it would perform worse relative to methods that do not have this method-imposed requirement. Fortunately, some of the archival data already examined provide a straightforward approach to estimating the performance of the empirical keys at smaller sample size levels, due to the hierarchical structure of students grouped within schools. Study 3 will use this data to begin an investigation into the required sample size for regression to perform well relative to traditional empirical keying methods.

STUDY 3

Study 3 is a partial investigation of *Research Question 2* (If any benefits are obtained from using option-level regression and configural methods, to what sample size-to-item ratios will these benefits generalize?). Option-keyed multiple regression resulted in more optimal cross-valid weights than using the VPM and/or PBC methods in three archival datasets, but this advantage may have been due only to the large sample size-to-item ratio in the datasets (e.g., approximately 100,000 students in the weight-derivation sample to approximately 950 item responses for the SDQ dataset) as this advantage dropped as these ratios decreased. Given that the majority of biodata development contexts will not have the luxury of sample sizes in the thousands, the question of how well any benefits of using option-keyed multiple regression will transfer to settings with smaller sample sizes is of considerable interest. One avenue for testing this generalizability is simply to use the data from Study 1, but with smaller subsets of students. Luckily, the SDQ dataset facilitates testing of this hypothesis via the pre-existing subgroupings of students based on the colleges/universities they attended. Because these schools cover a wide range of sample size-to-item ratios, it allows the investigation of how useful these methods are at smaller sample sizes and whether there are any specific sample size-to-item ratios at which certain methods are more preferable than others.

Method

The SDQ inventory described in Study 1 was used to investigate the generalizability of findings to settings with smaller sample sizes. For each of the 47 colleges/universities with sample size large enough such that a 2/3 weight-derivation sample would be larger than the 583 dummy-coded response options in the SDQ, the students were split into a 2/3 weight-derivation sample and a 1/3 cross-validation sample. The VPM (with 27% extreme groups), PBC, and option-keyed regression methods were then applied to the weight-derivation sample of each school to develop school-specific keys. These keys were then applied to the weight-derivation and cross-validation samples for that school to obtain estimates of validity and cross-validity for each method within each school. Considering that as sample sizes get smaller, the random sample split can have a much larger impact on results, the above procedure was repeated 25 times for each school (i.e., 25 random sample splits creating 25 keys for each keying method), and the mean of these trials was calculated. The extent to which regression or traditional methods exhibited higher cross-validity was compared to the sample size-to-item ratio for each school to obtain rough descriptive evidence of the ratios that need to be available before profiting from the use of option-keyed regression over the VPM or PBC methods.

Results and Discussion

Table 11 presents results for the aggregated cross-validation empirical key performance for each of the 47 schools with a large enough weight-derivation sample size. Sample size-to-item ratios ranged from 1.07 to 6.52 (i.e., weight-derivation sample sizes of 622 to 3,800) and there was a wide range in the effectiveness of the school-specific keys, both in absolute terms and in relative terms. For instance, the cross-valid variance explained using the VPM ranged from .03 to .15, and the variance explained for the PBC ranged from .04 to .16. More interesting however, is that there were many instances where regression did not work well. Specifically, there were regression keys for 17 schools that had a cross-valid variance explained of .03 or less (i.e., at or below the minimum for the traditional methods) when using regression. Given that situations where regression performed poorly seemed to overlap with those where the sample size-to-item ratio are lower, this indicates that there are indeed sample size limits to the performance of option-keyed regression as an effective empirical keying method.

Turning to the comparative results, there were only eight schools where a regression key tied or exceeded both the VPM and PBC in cross-valid variance explained. In order to better visualize the relationship between sample size and the relative advantage resulting from use of the option-keyed regression method, I plotted the difference between the cross-valid variance explained by the PBC method (which often explained slightly more variance than the VPM) and regression against the ratio of weight-development sample size to dummy-coded response options in Figure 5. Though the number of schools at a sample size ratio of 3 and higher is small, examination of the figure suggests that somewhere around a sample size-to-item ratio of 3 or 4, the multiple regression method starts to gain a predictive advantage over the traditional methods. Below this ratio, regression performs worse, and often much worse.

Broadly, visual analysis of the plotted relationship between the sample size-to-item ratio and the relative advantage of option-keyed regression over over methods suggests that it is strong and positive. In line with this observation, the correlation between the sample size ratio and the regression-PBC difference is .79. It is also worth noting that when looking at the largest sample size ratios, there is also variance in terms of how much variance multiple regression explains over the traditional methods. In School 37, for instance, there is a sample size-to-item ratio of 6.52, and regression explains about half again the variance as the traditional keying methods did. This is comparable to the findings from the ARI dataset. On the other hand, in School 20, there was a sample size-to-item ratio of 6.21, and regression explained only about 10% more variance than the traditional methods did, which is worse than the findings from the IAR dataset.

Given the patterns of results from Studies 1, 2, and 3, it appears that sample size plays a large role in how effective regression is relative to traditional keying methods. When the sample size is small relative to the number of dummy-coded response options, regression performs very poorly. On the other hand, when the sample size is large, keys developed using regression can result in very large gains in variance explained. Clearly, however, sample size is not the only factor in how well regression performs relative to the VPM and PBC methods, given the variance in results from similar sample size-to-item ratios. For instance, one factor that could play an important role is the average intercorrelation among the set of dummy-coded response options. Because this directly relates to the amount of redundancy in the predictor set, it could impact the performance of regression relative to traditional methods that don't take redundancy into account. Another factor is the number of items. All else equal, with more items, there is more opportunity for validity in predicting criteria. It is uncertain how the number of items would affect the relative performance of regression and the traditional methods. Simply put, though analysis of the archival datasets was very useful, they are still only three inventories and three sample draws. In order to investigate the properties of the regression method in a more fine-grained manner, it would be informative to expand analyses to datasets with a wider variety of parameters. Therefore, Studies 4 and 5 will focus on simulation work to try to investigate boundary conditions of the absolute and relative effectiveness of regression.

STUDY 4

Study 4 continues the investigation of *Research Question 2*. Study 3 focused on using archival data from a set of 47 schools. However, these results were based on one inventory and, although the schools represent a range of sample size-to-item ratios, each school's results represented only one random sample of students at that ratio. In order to draw more firm conclusions about optimal sample sizes for using the VPM, PBC, and option-keyed regression methods, I therefore conducted a simulation study to model a range of parameters and evaluate the outcome of repeating each set of parameters a large number of times.

Method

Simulation details. I used the open-source statistical programming language R (R Core Team, 2013) to conduct this simulation study. Each trial in the simulation consisted of generating a correlation matrix and generating raw data corresponding to that matrix. I generated correlation matrices that approximated the validities and intercorrelations of the items from the datasets used in Study 1 and Study 2 (as specified in Table 10) and also investigated reasonable extensions of these parameters for datasets involving 48, 120, and 204 response options (hereafter, referred to as "items" or "base items"). These item numbers were modeled after a report of the average length of biodata inventories being 100 to 200 items (Reiter-Palmon & Connelly, 2000). I chose 50, 125, and 200 as a range of item numbers and then rounded to the nearest multiple of 12 in order to match conditions required in Study 5.[28] For each number of items, I randomly sampled values for item intercorrelations from normal distributions with fully-crossed combinations of means of .00, .10, and .20, and standard deviations of .05, .10, and .15. Standard deviations for the item validities were tied to those of the intercorrelations, but the mean was held constant at .04 to correspond to the value observed in the SDQ dataset, and to what is likely common when applying these methods to the first item-screening pass in the development of an empirically keyed biodata inventory (i.e., mean validities close to zero). Additionally, adding further simulation conditions to account for different mean validities, separating the standard deviation of validities from the standard deviation of intercorrelations, and increasing sample sizes would have dramatically increased computing time. Because the correlation matrices were randomly generated, they occasionally violated properties of real data and needed to be corrected to ensure that they were positive definite (Knol & Berger, 1991).

After correlation matrices were generated, I created raw data. The sample sizes varied as a ratio of the weight-derivation sample size to the number of items, with sample size equal to the number of items + 2 as a lower bound, and then with sample size-to-item ratios of 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. For the 48 item conditions then, this resulted in sample sizes of 50, 60, 72, 84, 96, 144, 192, 240, 288, 480, 1,200, 2,400, and 4,800. Fully crossing all conditions (number of items, mean predictor intercorrelation, standard deviation of intercorrelations and validities, and sample size ratios) resulted in 351 simulation conditions.

[28] In order to simulate the same number of items for 3, 4, and 5-option items to create as close a comparison as possible, I needed values that were divisible by 2, 3, and 4 (i.e., the number of dummy-coded items that are derived from 3, 4, and 5-option items). The smallest number that meets this specification is 12.

The sample size in each condition was used for the weight-derivation group and the cross-validation group was one-half the size of the weight-derivation sample size. This corresponds to a situation where the sample is split into 2/3 weight-derivation and 1/3 cross-validation. After generating sample data from a normal distribution (mean = 0, *SD* = 1) that corresponded to the given correlation matrix, I then dichotomized each of the "item" variables to simulate a dichotomous Yes/No biodata item. The split for each item was also randomly generated (from a uniform distribution [-1.25 *SD*s, 1.25 *SD*s]). The dependent variable was left continuous. Because the data for the "items" was dichotomized, the parameters defined to construct the correlation matrices were reduced in magnitude in the raw data. Specifically, the mean validities were reduced on average by a factor of approximately .74, and the mean intercorrelations were reduced on average by a factor of approximately .55 using the double-multiplication approximation presented in Hunter and Schmidt (1990). As such, the simulation investigated modest levels of mean intercorrelations of items.

Once there were two samples of dichotomized data, I derived weights using the VPM, PBC, and option-keyed regression methods in the larger sample (i.e., the weight-derivation sample), and then applied the weights from each of these methods to both samples and calculated the variance explained by the resulting keys. This process was repeated 1,000 times for each condition.

Results and Discussion

Full results for all the simulation conditions are presented in Appendix B. However, it is admittedly difficult to integrate the findings of 27 tables, so I created two summary tables. The first, Table 12, provides the sample size-to-item ratio for every combination of the other varied parameters where regression ties or exceeds the VPM and PBC methods to illustrate the factors that contribute to regression performing relatively better or worse than traditional methods. The second summary table, Table 13, is constructed similarly to Table 12, but it instead illustrates the cross-valid variance explained for the maximum sample size condition (i.e., 100:1). This allows the identification of which simulation parameters lend themselves to the maximum advantage for regression.

With respect to Table 12, there are a number of useful findings. This table shows that as the standard deviation increases, the number of items increases, and the mean predictor intercorrelation increases, the sample size-to-item ratio needed before regression ties or exceeds traditional methods decreases (the one exception being in the smallest item condition where the *SD* = .05). Roughly then, as the table goes from top-left to bottom-right, the required sample size ratio decreases. The required sample size ratio for regression to tie or exceed the PBC is always the same or higher than the ratio needed to tie or exceed the VPM. Figure 6 through Figure 8 display this pattern of results graphically using a log transformation on sample size to increase interpretability.

These results largely coincide with the hypothesized reasons why regression would out-perform traditional methods in the first place. As the number of items increases, the amount of redundancy increases in the item pool, as does the amount of validity. Similarly, increases in the mean predictor intercorrelation represent direct increases in the degree of redundancy in the items. Finally, increased standard deviations increase the variance (and range) of validities, and they do the same for intercorrelations and therefore the redundancy among the items. At a high level, from these results it appears that if a weight-derivation sample size is 2 to 3 times the number of dummy-coded response options, regression will tie or equal traditional option-keying methods in terms of cross-valid variance explained in most cases.

It's worth pointing out that if one examines the full results in the appendices, the difference between regression and the traditional methods is sometimes quite small. For instance, particularly in the simulation conditions in the top-left of Table 12, the sample size-to-item ratio where regression "wins" is often little different from the next few closest ratios on either side, with differences in the thousandth place. Additionally, in these same conditions, the standard deviations of the cross-valid variance explained by the keying methods can overlap. This suggests that there are many contexts where it is uncertain which method will perform better than the others at lower sample sizes, even when regression might perform better on average.

As stated earlier, Table 13 reports the cross-valid variance explained for each of the keying methods at maximum studied sample sizes. There are two patterns of interest: those related to absolute maximum values and those related to relative maximum values. With respect to absolute maximum values, they appear to increase as the number of items increase and as the standard deviations increase. This makes sense as there is more validity in the item pool to take advantage of. However, the absolute validity for all three methods tends to decrease as the mean intercorrelation increases. This also makes sense as it effectively offsets some of the validity in the item pool by increasing redundancy.

The interesting finding is that this effect seems to impact the traditional methods much more than regression. In extreme cases, the variance explained can be reduced from approximately .32 to .04 when holding everything else constant but moving from a mean intercorrelation of .00 to a mean intercorrelation of .20. The comparable regression results for these conditions are dropping from .37 to .26.

Turning to findings of the relative maximum performance of keying strategies from Table 13, there were similar patterns. As the number of items increased, the relative edge in favor of regression generally increased. Similarly, as the mean intercorrelation increased, the relative superiority of regression also increased. However, as the standard deviation of validities and intercorrelations increased, the relative edge of regression generally decreased. This is likely due to the accompanying increased validities for all methods to identify and capitalize on with large sample sizes in these conditions.

Overall, the results of Study 4 suggest a number of interesting and practically useful conclusions. First, the weight-derivation sample size required for regression to tie or exceed the variance explained by the VPM and PBC can often be modest, as low as 1.5 times the number of dummy-coded response options. Additionally, the predictive advantage of multiple regression over the traditional methods can be massive: over 6 times the variance explained in the most extreme cases. The extent to which these extreme settings would actually occur, however, is an open question. For instance, the results in Table 10 suggested that the parameters in the datasets investigated in Study 1 and Study 2 were more characteristic of the simulation conditions where the mean intercorrelation was .00 and the standard deviation of validities and intercorrelations was between .05 and .10. These are conditions where simulation results showed that regression gains an edge over traditional methods with a small sample size-to-item ratio (1.75 to 2), but doesn't far outweigh them in terms of variance explained at maximum sample sizes. One important difference between the archival datasets and the simulation conditions, however, is that the simulations didn't include item numbers as large as were in the archival datasets. Also pertaining to the generalizability of the most extreme simulation findings, it seems unlikely that the mean intercorrelation between items would reach as high as .20 except in investigations of specific constructs, in which case, there probably wouldn't be hundreds of items.

Nevertheless, the simulation results seem supportive of the idea that regression should be considered a viable keying option that can begin to have advantages over traditional methods at reasonable sample sizes. Traditional methods that consider items individually don't do an optimal job of capturing the totality of the information about items in an inventory.

One potential limitation of the simulation findings, however, is that they pertain only to dichotomous items. Items with more response categories introduce dependency among sets of dummy-coded response options, as only one of them can typically be selected. It is unclear how this dependency will affect the performance of option-keyed regression. This is important for generalizability as most inventories will likely contain items that have more than two response options. Study 5 will therefore expand the simulation to items with 3, 4, and 5 response options.

STUDY 5

Study 5 addressed *Research Question 3* (Are required sample size-to-item ratio findings dependent on the number of response options items have?). The simulation in Study 4 focused only on dichotomous items and revealed that, depending on the condition, at weight-derivation sample sizes as low as 1.5 times the number of items, there were advantages in predictive validity from using option-keyed regression over the VPM and PBC methods. However, the benefits of option-keyed regression at modest sample sizes may be exaggerated in that set of findings. This is because option-keying dichotomous items maintains a single item, whereas dummy-coding, with more response options per item, creates dependent sets of items. If one of the dummy items from a parent item is selected, the other dummies cannot be (assuming a typical self-report scale where only one response is allowed per item). This may make it more difficult for option-keyed regression to accurately assign weights and maintain validity advantages at smaller sample sizes.

On the other hand, it is possible that adding these dependent sets of items could actually advantage the regression method. This is because the process of dummy-coding items has a set of predictable effects because the resulting item sets are partially ipsative (Hicks, 1970). A set of items is ipsative when the sum of the scores on those items is equal for every participant. An example of this is when an item is decomposed into a full set of binary items for every response option possibility. Assuming the item is one where only one response option can be chosen, the sum over this set of items for every respondent should be equal to 1. When a set of items is ipsative, there are a number of mathematical results that must be true. One of these is that, assuming the variances are equal, the average intercorrelation will be equal to -1/(*nr*-1) where *nr* is the number of response options. For example, if there were three response options, the average intercorrelation among them would be -.50; if there were four, it would be -.33. In other words, this dependency manifests itself as a negative correlation. Another mathematical result, also assuming the variances are equal, is that the sum of the validity coefficients will be equal to zero.

Option-keyed regression, however, does not operate on the full set of binary response options. It instead operates on the traditional dummy-coded set, where one of the binary items is dropped per item. This can result in what is often termed “partial ipsativity”, because it does not adhere to the exact definition of ipsativity, but shares many of the same qualities. For instance, there should still be a trend of average negative intercorrelations within a dummy-coded item set, and the average of the validities should still trend towards zero. This is why regression may actually be advantaged by the dependency among the item sets; it has the ability to pick up the pattern of negative relationships, whereas the traditional methods do not. It is therefore an open question as to whether the results from Study 4 will generalize to inventories with larger numbers of response options, and if results were different, whether they would be more or less in favor of regression.

This question is partially addressed by Study 1 and Study 3. The SDQ has items with a number of different response options, from 2 to 13. See Table 14 for the frequencies of the items with each number of response options. A little more than half the dummy-coded response options in the SDQ are derived from dichotomous items, whereas the rest are derived from items with more than two response options. With this mixture of items, regression still performed relatively well relative to the traditional methods, explaining double the cross-valid variance explained in Study 1, and tying/exceeding them at a sample size-to-item ratio of around three in Study 3. However, these results apply only to one inventory and set of participants, and regardless of the mixture of items with different numbers of response options, still rely heavily on dichotomous items. Therefore, in order to isolate the effect, if any, of more response options on the performance of option-keyed regression, I conducted a series of simulation studies replicating the procedure from Study 4 but extending it to include items with more response options.

Method

Simulation details. The simulation procedures for this study were very similar to those for Study 4. Numbers of items, mean inter-item correlation and validity starting parameters, and sample size-to-item ratios were the same. This was so that the results would be comparable to those from Study 4 as much as reasonably possible. There are two ways to interpret the goal of having the same number of items in the simulations in Study 4 and Study 5. First, this could mean that the number of dummy-coded response options is the same. This would result in fewer base items. For instance, in the simulation conditions with 48 3-option items, in order to result in 48 dummy-coded response options, the number of base items would need to be 24 (i.e., each 3-option item contributes two dummy-coded items). The second way of implementing the goal of keeping the same number of items is that the number of base items would be the same, but these would then be split into the number of dummy-coded response options for that condition. Taking the same example of 48 3-option items, this would result in 96 dummy-coded items (i.e., 48 items each divided into two dummy-coded items).

There are conceptual positives and negatives to each of these procedures. With the first option, it seems to be a more direct answer to the question of what would happen if there were the same number of items, but some were part of dependent sets. However, given that there are fewer base items, there is less initial validity in the item pool when using the same item generation parameters, which could obscure comparisons to Study 4. This effect would be somewhat offset by higher effective validities for items due to having larger numbers of response options (i.e., instead of dichotomizing items, they are split into 3, 4, or 5 response options, which doesn't reduce the resulting effective correlation as much). The reverse is true for the second interpretation. It would be based on the same amount of base items/validity in the item pool, but would result in a much larger number of dummy-coded response options than were investigated in Study 4, which also makes comparisons less interpretable. As a result, I conducted the simulation studies both ways. Table 15 provides a comparison of the number of base items, total responses, and dummy-coded responses involved in each of the simulation conditions.

Despite efforts to make the results of these simulations comparable to those from Study 4, they will differ in a few ways in addition to differences with the number of response options/ base items. The first is that, in line with the study's purpose, rather than items being dichotomized, they were split into three, four, and five response options, and then dummy-coded. This had two effects. First, as described above, it resulted in modified validity and intercorrelation parameters in the observed data prior to dummy-coding (i.e., correlations were reduced less than they were when the items were dichotomized). The second effect of adding more than two response options per item occurred as a result of dummy-coding. Rather than the validity and intercorrelation parameters being the effective parameters in the dataset that was subjected to keying methods as they were with dichotomous items, the validity and intercorrelation parameters applied before dummy-coding and were modified as a result of ipsativity.

The second difference between the simulation studies is in the cut scores that defined the frequency assigned to each response option. For this study, given the additional complexity of managing the random generation of multiple cut-points[29], rather than randomly generating response frequencies as in Study 4, I developed a set of response distributions for each number of response options (i.e., three, four, and five). These distributions are presented in Table 16, Table 17, and Table 18. For each generated item, I randomly chose one of the response distributions to apply to the item.

The final difference between the simulation studies is that in some simulation parameter combinations I did not simulate every sample size-to-item ratio. For conditions where the item condition represents the number of base items, for item conditions of 120 and 204, I did not investigate sample size-to-item ratios of 25, 50, and 100. This is because the size of the data matrices involved dramatically increased the length of time a simulation trial took. For example, in the 204 dichotomous item conditions, the

[29] If two cut points are randomly generated very close together, there may be very few or no people in one of the response options, which creates programming difficulties.

maximum data matrix size for a weight-derivation group was 20,400 rows by 204 columns. This stayed the same in conditions where the number of dummy-coded options equaled the item condition. However, in conditions where the number of base items is the item condition, this increased to 40,800 rows by 408 columns for 3 response options, 61,200 rows by 612 columns for 4 response options, and 81,600 rows by 816 columns for 5 response options. Conducting these additional simulation trials would have increased computing time by months with very little added benefit.

After datasets were created in the manner described, I used the VPM, PBC, and option-keyed regression methods in the two-thirds weight-derivation sample to develop weights, and then applied the weights from each of these methods to both samples. This process was repeated 1,000 times for each condition and results within condition were averaged.

Results and Discussion

Results from this simulation study are even more difficult to integrate and summarize than those for Study 4, given that it involved almost six times the number of simulation conditions (i.e., all the same conditions for three response options and two ways of operationalizing them). Full simulation results are presented in Appendices C through H. As before, I will focus primarily on summary tables.

Table 19 through Table 24 are mirrors of Table 12 from Study 4 and display the sample size-to-item ratios where regression ties or exceeds the traditional methods for each parameter combination. Across all numbers of response options and both ways of conceptualizing a number-of-items match to the specifications from Study 4, patterns from Study 4 remain the same. In general, as the number of items increases, the mean intercorrelation increases, and the standard deviation of the validities and intercorrelations increases, the sample size-to-item ratio needed for regression to tie or exceed the traditional methods decreases. Similar to the dichotomous item results, the main exception to this trend is in the 48 item condition with a standard deviation of .05 as the mean intercorrelation increases. These trends can be observed graphically by looking at Figure 9 through Figure 26. There are two additional conclusions that can be drawn from this set of simulations, however.

First, as the number of response options per item increases, the advantages from regression appear at smaller sample size-to-item ratios. This holds even in the conditions where the number of base items is quite small (i.e., the number of 5-option base items in conditions where the sum of the dummy-coded items equals the item condition is 12 for the 48 item conditions). In fact, in the conditions where the number of base items is equal to the item condition, the vast majority of sample size-to-item ratios where regression starts to gain an edge are at 1.25 times the number of dummy-coded response options. This is the smallest studied ratio where it is reasonable to expect an edge for regression in a cross-validation sample, because the only smaller ratio is one that is essentially at the limit where regression will run. The exception to this trend is with the same set of oddly-performing conditions described above (i.e., 48 items and .05 standard deviation of validities and intercorrelations). In this limited set of conditions, the more response options per item, the larger the sample size-to-item ratio needs to be in the conditions where the mean intercorrelation is equal to .10.

The second additional conclusion that can be drawn from Table 19 through Table 24 is that, compared within the same number of response options per item, the benefits from regression over traditional methods are reached at smaller sample sizes in the set of conditions where the base items are equal to the item parameter (i.e., Table 20, Table 22, and Table 24), rather than the conditions where the dummy-coded items are equal to the item parameter (i.e., Table 19, Table 21, and Table 23). In retrospect, this dovetails with the findings that the benefits of regression are obtained at smaller sample size ratios with more items. This is because, minus some degree of overlap depending on how many response options per item there are, the conditions where the base items are equal to the item parameter are investigating the same effect as those where the dummy-coded items are equal to the item parameter; they simply involve more base items.

Table 25 through Table 30 are the analogue to Table 13 from Study 4 and report the variance explained in the cross-validation sample in the maximum sample size-to-item ratio for each condition. First, the patterns observed in Study 4 for this analysis are generally replicated. Again, as the number of items increases and the standard deviation of validities and intercorrelations increases, the maximum cross-valid variance explained increases, and as the mean predictor intercorrelation increases, the maximum variance explained decreases for all keying methods. However, in conditions where the number of base items equals the item parameter (i.e., Table 26, Table 28, and Table 30), maximum cross-validation validities were generally larger than those for dichotomous items for all methods, not just regression. Holding other parameters constant, as the number of response options per item increased, so did the maximum validity. This is true even considering the lower maximum sample size ratios in the conditions where the number of items equals 120 and 204.

This effect seems likely to be due to the increasing signal of criterion standing that a response option provides, as there are more per item. With a binary item, a response option only indicates which of two groups a person belongs to. With five response options, much more information can be conveyed. All else equal, given a linear relationship of an item to a criterion, as the number of response options increases, the weights for endorsing any particular response option, particularly those at either the minimum or the maximum, should increase. Add to this a very large sample size (i.e., the maximum sample size-to-item ratio studied) so the directionality of the item is correctly established, and this finding seems reasonable.

The conditions where increasing the number of response options per item generally reduced maximum keying method effectiveness in cross-validation samples were cases where the dummy-coded options were equal to the number of response options (i.e., Table 25, Table 27, and Table 29). In particular, this happened when the number of items parameter was smaller and the standard deviations of validities and intercorrelations was lower (i.e., the range of validity was smaller and there were fewer items with large validities). Therefore, for these conditions increasing the number of response options was confounded with decreasing the number of base items, which could certainly play a role in reducing the maximum variance explained. As a result, the conditions where the number of dummy-coded response options equal the number of items parameter seem poorly suited for identifying the effects of varying the number of response options per item on maximum variance explained. Even so, there were still some parameter combinations where increasing the number of response options increased the maximum variance explained in spite of fewer base items.

Overall then, the results from this study suggest that when items are continuous and have linear relationships with criteria, holding base items constant, increasing the number of response options per item generally increases the performance of both the traditional empirical keying methods that consider items independently and option-keyed regression. It also tends to increase the relative performance of regression to the traditional methods. Even when not holding base items constant, adding dependency in the form of dummy-coded response options doesn't seem to harm the effectiveness of option-keyed regression, and in many cases lowered the sample size-to-item ratio at which regression performed better than the VPM and PBC methods. This therefore further extends the generalizability of the potential usefulness of option-keyed regression.

GENERAL DISCUSSION

I began this investigation by comparing empirical and rational methods of biodata inventory development and suggesting that these were not necessarily mutually exclusive strategies. I proposed a framework for conceptualizing the biodata inventory development process and argued that the rational approach is ideal for item development, but the empirical approach is likely the optimal option for determining keying weights (assuming a sample is obtainable). I then reviewed empirical keying methods and research comparing these methods, pointing out that two of the most thorough comparisons (Cucina et al., 2012; Devlin et al., 1992) had two common deficits: neither varied the number of items or took the sample size-to-item ratios into account, nor were they exhaustive of keying methods. With these deficits in mind, the current study focused on the usefulness of two empirical keying methods, option-keyed multiple regression and CHAID, in comparison to commonly-used alternatives, the VPM and PBC methods.

Initial tests with three archival datasets from different contexts and keyed against different criteria indicated that option-keyed regression resulted in superior prediction in cross-validation samples when compared to the traditional alternatives. CHAID exhibited improved prediction over the traditional alternatives, but was still less predictive than option-keyed regression. This, combined with the complexity involved in its use, resulted in the decision to drop CHAID from further consideration. Muddying interpretation, the three archival datasets were generally very large (weight-derivation $n \approx 3{,}000$ to 100,000) and it was possible that option-keyed regression provided an advantage only because of the sample size. The investigation of keys derived in naturally-occurring subsets (i.e., schools) of one of the datasets suggested that regression provided advantages with sample sizes as small as around 1,800 (i.e., in reference to the biodata inventory, a sample size-to-item ratio of about 3).

Because all of the just described results were based on three specific large inventories, I conducted a series of simulation studies in an attempt to investigate the necessary sample size needed for option-keyed regression to exceed the prediction offered by traditional methods. To build on previous research, I focused on the sample size-to-item ratio as an estimate of the likelihood of shrinkage, as it takes into account both the size of the sample and the length of the inventory. Results with dichotomous items across a range of simulation parameters suggested that as the standard deviation of predictor validities and intercorrelations increased, the number of items increased, and the mean predictor intercorrelation increased, the average required weight-derivation sample size-to-item ratio for regression to exceed the traditional methods decreased. In terms of absolute performance, maximum cross-validities for all methods increased as the number of items increased and the standard deviations of the validities and intercorrelations increased, but decreased as the mean predictor intercorrelation increased. The general pattern observed was that at the largest sample sizes studied, regression had the largest absolute cross-validities, followed by the PBC and VPM methods. In terms of relative performance, as the number of items and mean predictor intercorrelation increased, the maximum benefit of using option-keyed regression over traditional alternatives increased. As the standard deviation of validities and intercorrelations increased, however, the maximum relative benefit decreased.

The dichotomous results had the unique property that the number of items was the same as the number of dummy-coded items. As such, the dummy-coding process for these items avoided creating dependent sets of items. It was unclear what effect using items with more response items, where dummy-coding would result in this dependency, would have on the advantages of using option-keyed regression over the traditional alternatives. Therefore, additional simulations were conducted with items with three, four, and five response options. The results of these simulations showed that the advantages from regression generally appeared at smaller sample size-to-item ratios as the number of response options per item increased. Similarly, in the majority of simulation conditions, the maximum cross-validity increased for all methods as the number of response options increased.

Taken in total, the results suggest that the sample size needed for option-keyed regression to be a preferred choice to traditional methods could often be modest and was more a function of the number of items to be keyed than an absolute need for a sample of a specific size. Additionally, as sample sizes increased relative to the number of items, it was common for there to be a large benefit to using regression in terms of prediction.

Though I have already covered results with respect to answering the three research questions, there are a number of points related to interpretation and practical application that I have not discussed. I address each of these in turn.

Practical Considerations for the Application of Option-Keyed Regression

This set of studies focused on a series of questions with direct relevance for how biodata inventories are developed and keyed. Many of the high-level ideas and findings have already been discussed. For example, inventories can be made up of items that are rationally-derived to focus on relevant criteria, but empirically keyed to maximize prediction. Another general conclusion is that option-keyed regression can provide benefits in prediction over commonly-used alternatives in a number of circumstances, particularly those with large sample sizes. However, the actual implementation of these findings entails a number of issues to be aware of.

Issues with the interpretation and use of weights. Although the validity of a biodata key in predicting valued criteria is one important factor in the evaluation and use of a biodata inventory, there are other considerations. Among these are the face validity to test-takers and any organizational stakeholders, and the defensibility against legal action. Although any empirical keying procedure can result in unintuitive weights, at least most traditional methods are based only on one bivariate relationship. In other words, they only reflect the relationship between an item or a response option and the criterion. Option-keyed regression, on the other hand, would likely be much more difficult to explain, justify, and defend in an easily understandable manner. This is because every weight is in reference to every other weight in the composite. To the extent that the justification for using weight is opaque, it could reduce stakeholder buy-in and increase the chance of legal challenge.

The use of option-keyed regression also raises complications in terms of scale revision. If an item is deleted (or many items are deleted) because of a rational screen of weights that go against theory and the understanding of the construct space, decisions would have to be made about whether to re-derive weights or leave them the same. Both approaches have concerns associated with them. If weights are re-derived based on the same sample it could result in a further capitalization on sample-specific variance because of item selection based on odd empirical results. This could necessitate the need for a new sample. If weights are left the same after dropping items, it creates awkwardness in the explanation and justification for using those weights because they are based in relation to items that are no longer part of the inventory (or still included, but no longer part of the scored key).

Issues with the results of prediction. Related to concerns with interpretation are the results when weights are applied to actual test-takers' data. Both traditional keying methods that consider each item/response option independently and option-keyed regression have the potential for inconvenient outcomes. Traditional methods can over or under-weight certain constructs relative to their importance in the criterion space due to having many or few items measuring that construct, relative to items measuring other constructs. Similarly, regression can, in many cases, maximize prediction for the selection system by managing redundancy, but it introduces the possibility that certain test-takers will not get the credit they "should" for particular life experiences. For instance, if two responses are correlated .80, but one has a slight edge in validity, the results of a regression will give much more weight to the response with higher validity. Therefore, if someone endorses the response with less validity, but not the one with more validity, they may end up disadvantaged compared to bivariate methods because they happened to have a rare response combination (i.e., one, but not the other). Again, although this outcome is "optimal" from a regression standpoint when focused on the viewpoint of the selection system, from the viewpoint of the individual, this could appear unfair. In order to avoid this circumstance, biodata key developers could consider creating composites from items/response options that have an intercorrelation above a certain threshold. Paradoxically, this would reduce some of the potential advantage for using the regression method in the first place by removing redundancy in the item pool.

Cross-validation versus estimating shrinkage. Unlike the previous sections, this section doesn't discuss a problem or warning with applying option-keyed regression, but instead discusses a possible advantage. The need for a cross-validation sample has been cited as a necessity when using empirical-keying methods, and this additional sample size requirement has been framed as an advantage in favor of using rational-keying methods (Hough, 2010; Hough & Paullin, 1994). Indeed, if one were planning to use a cross-validation sample, then the sample-size ratio benchmark of 1.5 to 2 only reflects the weight-derivation sample. With this sample size, it would be expected that the weights are superior to those obtained from the VPM or PBC, but there would not be a cross-validated estimate of their validity. The Monte Carlo simulation in Studies 4 and 5 specified a cross-validation sample size of 1/2 the sample size of the weight-derivation group. Using this study design, a sample size of 2.25 to 3 times the number of items would be required to estimate the validity of the derived weights in a new sample (though 1.5 or 2 would still be applicable if one were just interested in deriving weights).

However, with the move to a regression framework, a variety of shrinkage formulas are available to estimate validity without the need for a cross-validation sample. Using separate cross-validation samples in lieu of shrinkage estimates has even been framed as being potentially wasteful of data that could be used to gain more stable weight estimates when derived using the full sample (e.g., Browne, 1975; Cotter & Raju, 1982; Mosier, 1951). Therefore when using regression to obtain biodata weights in applied settings, a profitable approach may be investigating the derivation of weights on the full sample and applying a shrinkage formula to estimate validity in a new sample. See Raju and colleagues (1997) for an accessible review of shrinkage estimates.

It is worth noting that there is no analogue to shrinkage formulas when using the VPM or PBC methods. This could result in regression providing superior results to those methods at even smaller sample size-to-item ratios than reported in the current study. Specifically, if one were to use the same study design as in Studies 4 and 5, then the sample size ratios specify a scenario where 1.5 times the sample size is being used to validate the key for the VPM/PBC methods than would be required when using the regression method with a shrinkage formula.

Of course, the preceding discussion refers only to the practice of splitting one sample into a weight-derivation and cross-validation sample. There would still be utility in obtaining cross-validity estimates from entirely different samples and other cross-validation variations described in the “Cross-validation procedures” section of the introduction.

Sample size requirements. I demonstrated through the analysis of archival data and simulation studies that sample sizes in the thousands are not necessarily a requirement for a benefit to be obtained from using option-keyed regression over traditional alternatives. However, it could be suggested that a sample size of even double the number of dummy-coded items is still out of reach in many settings, and thus applicability may be limited to either small inventories or very large organizations and testing companies. This is especially true when considering that using option-keyed regression on categorical items with more than two response options can drastically increase the number of effective items. It is certainly true that there will be settings where the needed sample is not attainable. However, in the settings where it is attainable, option-keyed regression offers a preferable alternative to the VPM and PBC methods. It is also worth noting that in many cases the limit to sample size is not the size of the organization, but rather the nature of the criterion. It is not uncommon to use only a small sample of employees in a validation study in which rated performance is the criterion, as training supervisors on rating processes and obtaining ratings is costly. Alternatively, there are many settings where the criterion to be predicted by a biodata inventory is regularly collected by human resource management systems (e.g., turnover, absence, workplace injuries) and thus criterion measures are readily available for all employees. In these cases the sample size burden would not be nearly as great.

Flexibility of keying with regression. As a final note, it is worth pointing out that the use of option-keyed regression as a keying method allows for more flexibility than other option-keying methods. Specifically, regression allows for the derivation of weights for continuous items and dummy-coded items at the same time. The VPM requires that continuous items be split into manageable groups/levels to determine the percent endorsement of that group. The PBC method shares this advantage with regression (i.e., rather than point-biserial correlations, Pearson product-moment correlations are calculated). This added flexibility of the PBC and option-keyed regression, depending on whether accounting for redundancy is desired, is a conceptual reason to prefer these methods to the VPM, in addition to their validity advantages.

The Utility of Configural Keying Methods

The results concerning CHAID, and their bearing on the applicability of configural methods to keying biodata inventories deserve further comment. Although CHAID exhibited a respectable validity in the weight-derivation sample in Study 1, this validity shrunk in the cross-validation sample. Because its cross-valid variance explained was less than regression's and not appreciably more than the traditional methods, it was dropped from further investigation as it is a much more complex and rarely applied method. However, it is important to note that the specific validity used for CHAID was the most cross-valid out of 104 trees created by varying the tree construction parameters. There is little guidance in the literature on what the values of these parameters should be set to by default, and the optimal parameters would almost certainly have to be determined via some sort of cross-fold procedure as with other data mining techniques (Hastie et al., 2009).

Although the results of this study may suggest that CHAID and other configural techniques provide disappointing results when applied to biodata measures, CHAID was only applied to one dataset in this investigation. It is possible that the SDQ just didn't have much in terms of interactions for CHAID to pick up on. Even beyond this possibility, there are certainly cases where configural methods may prove useful. Of note, they could be used in a more limited and exploratory capacity to identify the most predictive and meaningful interactions. These interactions could then be used with more traditional regression methods. Additionally, more advanced data mining methods, such as elastic net regression, random forests, and least angle regression could be investigated. However, regardless of the specific method, use of configural methods to develop an entire key that is generalizable would likely require a large sample size because of base rate concerns with multi-way interactions.

Generalizability of Findings and Limitations

This investigation had a number of strengths with respect to the generalizability of its findings. First, the same pattern of results was found in all three archival datasets. Given that these were from two different contexts (educational vs. employment) and used different criteria (FGPA vs. performance) and item selection assumptions (broad vs. targeted construct), this allows for the conclusion that the general pattern of results is not idiosyncratic or isolated. Second, the simulation results were based on parameters that were largely shared by the three archival datasets. This therefore grounds the simulation results in actual data. Indeed, the mean validities and intercorrelations used are a likely consequence of the typical application of biodata keying. Because a prospective inventory contains a wide number of items, only some of which will show relationships with a criterion, and there being no reason why one would expect intercorrelations between them to be high or to exhibit any particular directionality, a mean validity and intercorrelation of close to .00 seems reasonable.

There are, however, a few caveats and limitations to the conclusions and generalizations that can be drawn. Although the simulations were based on parameters that were observed in actual datasets, the simulated items were not exhaustive of patterns that might occur in real data. Specifically, the items simulated were continuous and exhibited linear relationships with criteria. The lack of non-linearity in the dataset shouldn't be a large problem in the interpretability of results comparing the methods, as all the methods studied should have been able to identify and take advantage of non-linearity because they all operated at the response option level. However, it is possible that results could have been different if items with non-linear relationships with criteria were included. For instance, if non-linearity in the item pool makes it more difficult for regression and/or the traditional alternatives to identify proper weights, it could have an effect on the extent to which the methods perform in both an absolute and a relative sense.

Similarly, all the items simulated were based on continuous distributions. As a result, dummy-coding the response options resulted in sets of items where a response to one of the dummy-coded response options conveyed information on the relative standing on that set of items (e.g., if the weight is negative, a respondent who endorses that response option is relatively low on the continuum of the item on which that set of dummy-coded response options is based). However, if items were nominal and only one response option was typically chosen (e.g., "Which of the following religions do you belong to?"), these response options would be dependent to the extent that only one could be selected, but independent in terms of their relationship to the criterion. As with non-linearity, the effects of this different pattern of dependence are an open question.

In either case of adding item types, conducting a simulation would require different item generation parameters and assumptions. For instance, the average and standard deviation of item validities and intercorrelations would not be sufficient to model non-linearity, and different forms of non-linearity would need to be defined and operationalized (e.g., quadratic functions, leveling off), preferably with some grounding in real data. By the same token, to generate nominal items, the item generation process would need to be reversed, with mutually exclusive response options simulated then packaged together. This would increase the complexity in generating response data because the response options would need to correspond to particular correlational patterns, but also paired with response options with no overlap in endorsement. The implementation of both extensions is therefore not entirely straightforward.

Ultimately, the generalizability of these results should be seen as most applicable to cases where there is a large pool of items with an uncertain relationship to a criterion, as this was the context of both the archival datasets and the simulations (which were based on parameters from the archival datasets). Further, the simulation conclusions were based on continuous items with a linear relationship to the criteria. Extensions of the simulations to conditions beyond those observed in the archival datasets allow for the speculation on the likely outcome when option-keyed regression and/or traditional methods are applied to other datasets, but the results are still deficient to the extent that the simulation conditions did not vary all relevant variables and include all item types. While fully investigating every important variable is infeasible, the conditions most likely to be relevant that were not simulated were those where the mean validity varied and those where the standard deviation of validities and intercorrelations were separately defined.

It is worth pointing out again that many of the most extreme instances of regression's superior performance in the simulation studies, both in an absolute sense and relative to other keying techniques, were the result of applying parameters that may not be typical. Specifically, many of the more extreme examples were in cases with larger numbers of items, a large standard deviation of validities and intercorrelations, and, depending on the particular finding, higher mean intercorrelations. The archival datasets, however, were fairly uniform with the standard deviation and mean intercorrelations being in the lower range of conditions studied. As a result, it is reasonable to question the generalizability of the conditions where regression exhibited the largest advantage to real datasets. For instance, an average mean intercorrelation of .20 is quite high. As a comparison, this approaches the mean intercorrelation of some personality scales that are ostensibly measuring a single construct (Goldberg, 1999). The likelihood of someone using a 204-item scale (i.e., the highest simulated number of items condition in this study) with an average intercorrelation that high seems unlikely given that acceptable reliability would have been reached with far fewer items and testing time is often at a premium. Nonetheless, the simulation conditions studied are useful for illuminating the general principles that govern when option-keyed regression can be expected to do better or worse when compared to traditional empirical keying methods.

Directions for Future Research

The above discussion identifies a number of opportunities for further research to flesh out the findings from this set of studies. First, option-keyed regression and the traditional approaches could be applied to additional biodata inventories. Although the datasets already used provide a fair amount of generalizability, further replication with different criteria, such as more difficult-to-predict dichotomous outcomes like retention, or with inventories further along in the scale development process would provide additional support for recommendations to use option-keyed regression. For instance, it would be interesting to see the extent to which regression has an advantage in deriving a key for an established biodata inventory where only the predictive items remain and/or where redundant items were removed. Additionally, further simulation work could be conducted to investigate the effect of previously discussed changes to conditions, such as the use of different mean validities, separating the standard deviation of validities and intercorrelations, and adding nominal items and non-linearity. All of these would serve to increase understanding of the extent to which the findings with respect to option-keyed regression generalize to other contexts and inventories.

Another avenue for further research is the application of different configural and/or data mining techniques to biodata inventories. Although there have already been some investigations of configural approaches to keying biodata inventories (e.g., Gandy et al., 1994; Grauer, 2006; Tanofsky et al., 1969), these, much like CHAID, have generally not been the most advanced and/or modern options. Modern alternatives, such as least angle regression, elastic net regression, random forests, and multivariate adaptive regression splines can have a number of attractive properties that combine and expand on the advantages of traditional keying methods. For instance, many of them, in addition to accounting for redundancy, non-linearity, and interactions, also automatically correct for over-fitting (i.e., optimizing the complexity/parsimony tradeoff that caused problems for CHAID in this study), and account for model uncertainty through model averaging. Although larger sample sizes may still be required, it would be fascinating to see whether the promise of configural methods can be realized with these modern alternatives.

Finally, the use of shrinkage formulas in lieu of an empirical cross-validation estimate was proposed earlier as a possible advantage of using option-keyed regression. An extension of the current work that evaluates the performance of shrinkage formulas in comparison to the empirical cross-validation estimates in both the archival datasets and simulation work would be informative for further understanding the viability of this suggestion. In such a study, the sample size demands would be recast in terms of total sample size needed for an estimate of cross-validation. Assuming a shrinkage formula produced results similar to the empirical findings, viewing the sample size requirements in these terms should make option-keyed regression preferable to traditional keying methods at even smaller sample sizes.

Conclusion

In summary, when developing biodata inventories, even though items are likely best developed rationally and targeted toward a specific criterion and context, empirical methods of weight development will almost certainly result in stronger validities given a large enough sample. The current study thoroughly investigated the behavior of option-keyed regression, an alternative to traditional empirical keying methods rarely, if ever, used in biodata contexts, across a variety of settings and found that it often explained more variance than the results of often-used alternatives that do not take item redundancy into account. These results suggest that if a weight-derivation sample is larger than the number of dummy-coded response options in a biodata inventory, option-keyed regression should be attempted along with other keying methods. In addition to improving prediction, keys with more validity should also help shed light on the relationship of previous experiences to valued criteria. Future research outlined in the present study should serve to move both of these goals forward even further.

REFERENCES

Aamodt, M. G., & Pierce, W. L., Jr. (1987). Comparison of the rare response and vertical percent methods for scoring the biographical information blank. *Educational and Psychological Measurement, 47*, 505-511.

Alf, E. F., Jr., & Abrahams, N. M. (1975). The use of extreme groups in assessing relationships. *Psychometrika, 40*, 563-572.

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435-442.

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology, 25*, 251-269.

Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *The Academy of Management Review, 14*, 20-39.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45*, 389-406.

Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14*, 223-235.

Biggs, D., De Ville, B., & Suen, E. (1991). A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics, 18*, 49-62.

Bliesener, T. (1996). Methodological moderators in validation biographical data in personnel selection. *Journal of Occupational and Organizational Psychology, 69*, 107-120.

Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods, 10*, 689-709.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561-589.

Booth, R. F., McNally, M. S., & Berry, N. H. (1978). Predicting performance effectiveness in paramedical occupations. *Personnel Psychology, 31*, 581-593.

Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review, 19*, 219-231.

Breaugh, J. A., & Dossett, D. L. (1989). Rethinking the use of personal history information: The value of theory-based biodata for predicting turnover. *Journal of Business and Psychology, 3*, 371-385.

Browne, M. W. (1975). Predictive validity of a linear regression equation. *British Journal of Mathematical and Statistical Psychology, 28,* 78-87.

Brown, S. H. (1994). Validating biodata. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 199-236). Palo Alto, CA: Consulting Psychologists Press.

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39*, 214-227.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butts, M. M., & Ng, T. W. H. (2009). Chopped liver? OK. Chopped data? Not OK. In C. E. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 361-386). New York, NY: Routledge.

Campbell, J. P. (1974). *A monte carlo approach to some problems inherent in multivariate prediction: With special reference to multiple regression* (Rep. No. 2002). Arlington, VA: Personnel and Training Research Programs, Office of Naval Research.

Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H., & Erwin, F. (1999). Generalizable biographical data validity can be achieved without multi-organizational development and keying. *Personnel Psychology, 52*, 731-755.

Cascio, W. F., & Aguinis, H. (2005). Test development and use: New twists on old questions. *Human Resource Management, 44,* 219-235.

Cattin, P. (1980). Estimation of the predictive power of a regression model. *Journal of Applied Psychology, 65*, 407-414.

Chait, H. N., Carraher, S. M., Buckley, M. R. (2000). Measuring service orientation with biodata. *Journal of Managerial Issues, 12*, 109-120.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249-253.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* New York, NY: Routledge.

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*, 179-184.

Cotter, K. L., & Raju, N. S. (1982). An evaluation of formula-based population squared cross-validity estimates and factor score estimates in prediction. *Educational and Psychological Measurement, 42*, 493-519.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Crosby, M. M., & Mitchell, T. W. (1988, April). The obscurity of biodata predictor-criterion relationships: A blessing in disguise? In T. W. Mitchell (Chair), *Advancing the theory and method of biodata*. Symposium conducted at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.

Cucina, J. M., Caputo, P. M., Thibodeaux, H. F., & MacLane, C. N. (2012). Unlocking the key to biodata scoring: A comparison of empirical, rational, and hybrid approaches at different sample sizes. *Personnel Psychology, 65*, 385-428.

Cucina, J. M., Bayless, J. M., Thibodeaux, H. F., Busciglio, H. H., & MacLane, C. N. (2009, April). *How to score biodata measures: A master tutorial*. Master tutorial presented at the 24th annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Cureton, E. E. (1950). Validity, reliability, and baloney. *Educational and Psychological Measurement, 10,* 94-96.

Dalessio, A. T., & Silverhart, T. A. (1994). Combining biodata test and interview information: Predicting decisions and performance criteria. *Personnel Psychology, 47*, 303-315.

Dalessio, A. T., Crosby, M. M., & McManus, M. A. (1996). Stability of biodata keys and dimensions across English-speaking countries: A test of the cross-situational hypothesis. *Journal of Business and Psychology, 10*, 289-296.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics, 29*, 317-331.

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist, 26*, 180-188.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

Dean, M. A., Russell, C. J., & Muchinsky, P. M. (1999). Life experiences and performance prediction: Toward a theory of biodata. In G. Ferris (Ed.), *Research* *in Personnel and Human Resources Management (*pp. 245-281). Greenwich, CT: JAI Press.

DeCoster, J., Iselin, A.R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods, 14*, 349-366.

Devlin, S. E., Abrahams, N. M., & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. *Military Psychology, 4*, 119-136.

Dorans, N., & Drasgow, F. (1978). Alternative weighting schemes for linear prediction. *Organizational Behavior and Human Performance, 21*, 316-345.

Dorans, N. J., & Drasgow, F. (1980). A note on cross-validating prediction equations. *Journal of Applied Psychology, 65*, 728-730.

Dreher, G. F., & Sackett, P. R. (1983). *Perspectives on staffing and selection*. Homewood, IL: Irwin.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*, 36-48.

Efron, B., & Tibshirani, R. J. (1998). *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*, 387-395.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13*, 171-192.

England, G. W. (1961). *Development and use of weighted application blanks*. Dubuque, IA: W. C. Brown.

England, G. W. (1971). *Development and use of weighted application blanks* (Bulletin No. 55), Minneapolis, MN: Industrial Relations Center, University of Minnesota.

Farmer, W. L. (2007). *A brief review of biodata history, research, and applications* (Report No. NPRST-TN-07-3). Millington, TN: Bureau of Naval Personnel.

Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika, 26*, 307-316.

Ferguson, L. W. (1961). The development of industrial psychology. In B. H. Gilmer (Ed.), *Industrial Psychology* (pp. 18-37). New York, NY: McGraw-Hill.

Fine, S. A., & Cronshaw, S. (1994). The role of job analysis in establishing the validity of biodata. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 39-63). Palo Alto, CA: Consulting Psychologists Press.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Gaier, E. L., & Lee, M. C. (1953). Pattern analysis: The configural approach to predictive measurement. *Psychological Bulletin, 50*, 140-148.

Gandy, J. A., Dye, D. A., & MacLane, C. N. (1994). Federal government selection: The individual achievement record. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information* *in selection and performance prediction* (pp. 275-309). Palo Alto, CA: Consulting Psychologists Press.

Gessner, T. E., O'Connor, J. A., Clifton, T. C., Connelly, M. S., & Mumford, M. D. (1993). The development of moral beliefs: A retrospective study. *Current Psychology, 12*, 236-254.

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*, 422-432.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In. I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe, Vol. 7* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology, 16*, 573-592.

Grauer, E. (2006). *Applying neural networking techniques to improve performance and turnover prediction* (Doctoral dissertation). Bowling Green State University, Bowling Green, OH.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Guilford, J. P. (1954). *Psychometric methods*. New York, NY: McGraw-Hill.

Guion, R. M. (1965). *Personnel testing*. New York, NY: McGraw-Hill.

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions.* New York, NY: Routledge.

Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: John Wiley & Sons.

Guttman, L. (1941). An outline of the statistical theory of prediction. In P. Horst (Ed.), *The prediction of personal adjustment: A survey of logical problems and research techniques, with illustrative application to problems of vocational selection, school success, marriage, and crime* (pp. 253-311). New York, NY: Social Science Research Council.

Harold, C. M., McFarland, L. A., & Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment, 14*, 336-346.

Hase, H. D., & Goldberg, L. R. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin, 67*, 231-248.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.

Hein, M., & Wesley, S. (1994). Scaling biodata through subgrouping. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 171-196). Palo Alto, CA: Consulting Psychologists Press.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.

Hinrichs, J. R., Haanpera, S., & Sonkin, L. (1976). Validity of a biographical information blank across national boundaries. *Personnel Psychology, 29*, 417-421.

Hogan, J. B. (1994). Empirical keying of background data measures. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 69-107). Palo Alto, CA: Consulting Psychologists Press.

Horst, P. (1954). Pattern analysis and configural scoring. *Journal of Clinical Psychology, 10*, 3-11.

Horst, P. (1968). Configural analysis and pattern recognition. *Journal of Clinical Psychology, 24*, 383-405.

Hough, L. M. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology, 69*, 135-146.

Hough, L. M. (2010). Assessment of background and life experience: The past as prologue. In J. C. Scott, & D. H. Reynolds (Eds.), *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent* (pp. 109-139). San Francisco, CA: Jossey-Bass.

Hough, L. M., & Paullin, C. (1994). Construct-oriented scale reconstruction: The rational approach. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 109-145). Palo Alto, CA: Consulting Psychologists Press.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*, 152-194.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*, 334-349.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd edition). Thousand Oaks, CA: Sage.

Jackson, D. N. (1970). A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2, pp. 61-96). New York, NY: Academic Press.

Jackson, D. N. (1975). The relative validity of scales prepared by naive item writers and those based on empirical methods of personality scale construction. *Educational and Psychological Measurement, 35*, 361-370.

Jeanneret, R., & Silzer, R. (2011). Individual psychological assessment: A core competency for industrial-organizational psychology. *Industrial and Organizational Psychology, 4*, 342-351.

Kamp, J. D., & Hough, L. M. (1986). Utility of biographical data for predicting job performance. In L. M. Hough (Ed.), *Literature review: Utility of temperament, biodata, and interest assessment for predicting job performance* (ARI Research Note No. 88-02, pp. 91-130). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Karas, M., & West, J. (1999). Construct-oriented biodata development for selection to a differentiated performance domain. *International Journal of Selection and Assessment, 7*, 86-96.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of Applied Statistics, 29*, 119-127.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*, 17-24.

Kilcullen, R. N., White, L. A., Mumford, M. D., & Mack, H. (1995). Assessing the construct validity of rational biodata scales. *Military Psychology, 7*, 17-28.

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76*, 889-896.

Knol, D. K., & Berger, M. P. F. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research, 26*, 457-477.

Kobrin, J. K., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. Research Report #2008-5. New York, NY: The College Board.

Korman, A. K. (1968). The prediction of managerial performance: A review. *Personnel Psychology, 21*, 295-322.

Krokos, K. J., Meade, A. W., Cantwell, A. R., Pond, S. B., III, & Wilson, M. A. (2004, April). *Empirical keying of situational judgment tests: Rationale and some examples.* Paper presented at the 19th annual meeting of the Society for Industrial and Organizational Psychology, Chicago, IL.

Kuncel, N. R., & Highhouse, S. (2011). Complex predictions and assessor mystique. *Industrial and Organizational Psychology, 4*, 302-306.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2008). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. In I. L. Kwaske (Chair). *Individual assessment: Does the research support the practice?* Symposium presented at the 23rd annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Laurent, H. (1970). Cross-cultural cross-validation of empirically validated tests. *Journal of Applied Psychology, 54*, 417-423.

Lautenschlager, G. J. (1994). Accuracy and faking of background data. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 391-419). Palo Alto, CA: Consulting Psychologists Press.

Lawshe, C. H., & Schucker, R. E. (1959). The relative efficiency of four test weighting methods in multiple prediction. *Educational and Psychological Measurement, 19*, 103-114.

Lecznar, W. B., & Dailey, J. T. (1950). Keying biographical inventories in classification test batteries. *American Psychologist, 5*, 279.

Lefkowitz, J., Gebbia, M. I., Balsam, T., & Dunn, L. (1999). Dimensions of biodata items and their relationships to item validity. *Journal of Occupational and Organizational Psychology, 72*, 331-350.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44*, 763-792.

Mael, F. A. (1994). If past behavior really predicts future, so should biodata's. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 273-291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mael, F. A., & Ashforth, B. E. (1995). Loyal from day one: Biodata, organizational identification, and turnover among newcomers. *Personnel Psychology, 48*, 309-333.

Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. *Personnel Psychology, 46*, 719-738.

Mael, F. A., Connerley, M., & Morath, R. A. (1996). None of your business: Parameters of biodata invasiveness. *Personnel Psychology, 49*, 613-650.

Magidson, J. (1994). The CHAID approach to segmentation modeling: Chi-squared automatic interaction detection. In R. P. Bagozzi (Ed.), *Advanced methods of market research* (pp. 118-157). Cambridge, MA: Blackwell Publishers.

Malloy, J. (1955). The prediction of college achievement with the life experience inventory. *Educational and Psychological Measurement, 15*, 170-180.

Malone, M. P. (1977). *Predictive efficiency and discriminatory impact of verifiable biographical data as a function of data analysis procedure* (Doctoral dissertation). Illinois Institute of Technology, Chicago, IL.

McAllister, L. (1996). *A practical guide to CPI interpretation.* Palo Alto, CA: Consulting Psychological Press.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283-314.

McManus, M. A., & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. *Personnel Psychology, 52*, 137-148.

McManus, M. A., & Masztal, J. J. (1999). The impact of biodata item attributes on validity and socially desirable responding. *Journal of Business and Psychology, 13*, 437-446.

McQuitty, L. L. (1957). Isolating predictor patterns associated with major criterion patterns. *Educational and Psychological Measurement, 17*, 3-42.

Meehl, P. E. (1950). Configural scoring. *Journal of Consulting Psychology, 14*, 165-171.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1956). Wanted - A good cookbook. *American Psychologist, 11*, 263-272.

Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology, 6*, 102-109.

Meehl, P. E. (1989). Paul E. Meehl. In G. Lindzey (Ed.), *A history of psychology in autobiography* (Vol. 8, pp. 337-389). Stanford, CA: Stanford University Press.

Meehl, P. E., & Dahlstrom, W. G. (1960). Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *Journal of Consulting Psychology, 24*, 375-387.

Mitchell, T. W. (1994). The utility of biodata. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 485-516). Palo Alto, CA: Consulting Psychologists Press.

Mitchell, T. W., & Klimoski, R. J. (1986). Estimating the validity of cross-validity estimation. *Journal of Applied Psychology, 71*, 311-317.

Mosier, C. I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement, 11*, 5-11.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53*, 299-323.

Mumford, M. D. (1999). Construct validity and background data: Issues, abuses, and future directions. *Human Resource Management Review, 9*, 117-145.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11*, 1-31.

Mumford, M. D., & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background measures. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology: Vol. 3* (2nd ed., pp. 61-138). Palo Alto, CA: Consulting Psychologists Press, Inc.

Mumford, M. D., Barrett, J. D., & Hester, K. S. (2012). Background data: Use of experiential knowledge in personnel selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 353-382). Oxford: Oxford University Press.

Mumford, M. D., Snell, A. F., & Reiter-Palmon, R. (1994). Personality and background data: Life history and self-concepts in an ecological system. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 583-625). Palo Alto, CA: Consulting Psychologists Press.

Mumford, M. D., Stokes, G. S., & Owens, W. A. (1990). *Patterns of life history: The ecology of human individuality*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mumford, M. D., Costanza, D. P., Connelly, M. S., & Johnson, J. F. (1996). Item generation procedures and background data scales: Implications for construct and criterion-related validity. *Personnel Psychology, 49*, 361-398.

Mumford, M. D., Whetzel, D. L., Murphy, S. T., & Eubanks, D. L. (2007). Background data. In D. L. Whetzel, & G. R. Wheaton (Eds.), *Applied measurement: Industrial psychology in human resources management* (pp. 201-233). New York:, NY: Psychology Press.

Muros, J. P. (2008). *Know the score: An exploration of keying and scoring approaches for situational judgment tests* (Doctoral dissertation). University of Minnesota, Minneapolis, MN.

Murphy, K. R. (1983). Fooling yourself with cross-validation: Single sample designs. *Personnel Psychology, 36*, 111-118.

Murphy, K. R. (1984). Cost-benefit considerations in choosing among cross-validation methods. *Personnel Psychology, 37*, 15-22.

Neidt, C. O., & Malloy, J. P. (1954). A technique for keying items of an inventory to be added to an existing test battery. *Journal of Applied Psychology, 38*, 308-312.

Nickels, B. J. (1994). The nature of biodata. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 1-16). Palo Alto, CA: Consulting Psychologists Press.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187-207.

Owens, W. A. (1968). Toward one discipline of scientific psychology. *American Psychologist, 23*, 782-785.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 609-644). Chicago, IL: Rand McNally.

Pace, L. A., & Schoenfeldt, L. F. (1977). Legal concerns in the use of weighted applications. *Personnel Psychology, 30,* 159-166.

Pannone, R. D. (1987, August). Effects of faking on biodata validity coefficients. In L. J. Stricker (Chair), *Problems of biodata distortion in personnel selection systems.* Symposium conducted at the annual meeting of the American Psychological Association, New York.

Ployhart, R. E., Weekley, J. A., Holtz, B. C., & Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology, 56*, 733-752.

Porter, K. A. (1962). *Ship of fools*. New York, NY: Little, Brown and Company.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*, 178-192.

Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*, 241-258.

Quinones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology, 48*, 887-910.

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement, 21*, 291-305.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1999). Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement, 23*, 99-115.

Ramsay, M. J. (2002). *Comparing five empirical biodata scoring methods for personnel selection* (Master's thesis). University of North Texas, Denton, TX.

Ree, M. J., Carretta, T. R., & Earles, J. A. (1998). In top-down decisions, weighting variables does not matter: A consequence of Wilks' theorem. *Organizational Research Methods, 1*, 407-420.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35*, 1-62.

Reilly, R. R., & Warech, M. A. (1994). The validity and fairness of alternatives to cognitive tests. In L. C. Wing, & B. R. Gifford (Eds.), *Policy issues in employment testing* (pp. 131-224). Boston, MA: Kluwer Academic Publishers.

Reiter-Palmon, R., & Connelly, M. S. (2000). Item selection counts: A comparison of empirical key and rational scale validities in theory-based and non-theory-based item pools. *Journal of Applied Psychology, 85*, 143-151.

Richardson, Bellows, Henry and Co., Inc. (1988). *Technical reports: The law enforcement candidate record*. Washington, D. C.: Author.

Rosenbaum, R. W. (1976). Predictability of employee theft using weighted application blanks. *Journal of Applied Psychology, 61*, 94-98.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297-330.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75*, 175-184.

Russell, C. J. (1994). Generation procedures for biodata items: A point of departure. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 17-38). Palo Alto, CA: Consulting Psychologists Press.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85,* 112-118.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*, 215-227.

Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D.

S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 165-199). London, England: Sage.

Sarbin, T. R., Taft, R., & Bailey, D. E. (1960). *Clinical inference and cognitive theory.* New York, NY: Holt, Rinehart & Winston.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178-200.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 33*, 272-292.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27*, 183-198.

Schmitt, N., & Ployhart, R. E. (1999). Estimates of cross-validity for stepwise regression and with predictor selection. *Journal of Applied Psychology, 84*, 50-57.

Schmitt, N., & Pulakos, E. D. (1998). Biodata and differential prediction: Some reservations. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 167-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719-730.

Schoenfeldt, L. F. (1999). From dust bowl empiricism to rational constructs in biographical data. *Human Resource Management Review, 9*, 147-167.

Schoenfeldt, L. F., & Mendoza, J. L. (1994). Developing and using factorially derived biographical scales. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 147-169). Palo Alto, CA: Consulting Psychologists Press.

Sharf, J. C. (1994). The impact of legal and equal employment opportunity issues on personal history inquiries. In G. S. Stokes, M. D., Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 351-390). Palo Alto, CA: Consulting Psychologists Press.

Siegel, L. (1956). A biographical inventory for students: II. Validation of the instrument. *Journal of Applied Psychology, 40*, 122-126.

Silzer, R., & Jeanneret, R. (2011). Individual psychological assessment: A practice and science in search of common ground. *Industrial and Organizational Psychology, 4*, 270-296.

Simms, L. L., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Frayley, & R. F. Kruger (Eds.), *Handbook of research methods in personality psychology* (pp. 240-258). New York, NY: Guilford.

Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the validation and use of personnel selection procedures* (4th Ed.). Bowling Green, OH.

St. John, C. H., & Roth, P. L. (1999). The impact of cross-validation adjustments on estimates of effect size in business policy and strategy research. *Organizational Research Methods, 2*, 157-174.

Stead, W. H., Shartle, C. L., & Associates (1940). *Occupational counseling techniques: Their development and application.* New York, NY: American Book Company.

Steinhaus, S. D., & Waters, B. K. (1991). Biodata and the application of a psychometric perspective. *Military Psychology, 3*, 1-23.

Stokes, G. S. (1999). Introduction to special issue: The next one hundred years of biodata. *Human Resource Management Review, 9*, 111-116.

Stokes, G. S., & Cooper, L. A. (2004). Biodata. In M. Hersen (Ed.-in-Chief) & J. C. Thomas (Vol. Ed.), *Comprehensive handbook of psychological assessment: Vol. 4. Industrial and organizational assessment* (pp. 243-268). Hoboken, NJ: John Wiley.

Stokes, G. S., & Reddy, S. (1992). Use of background data in organizational decisions. In C. L. Cooper, & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 7, pp. 285-321). Chichester, England: John Wiley & Sons.

Stokes, G. S., & Searcy, C. A. (1999). Specification of scales in biodata form development: Rational vs. empirical and global vs. specific. *International Journal of Selection and Assessment, 7*, 72-85.

Stokes, G. S., Mumford, M. D., & Owens, W. A. (1994). *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction*. Palo Alto, CA: Consulting Psychologists Press.

Stokes, G. S., Toth, C. S., Searcy, C. A., Stroupe, J. P., & Carter, G. W. (1999). Construct/rational biodata dimensions to predict salesperson performance: Report on the U.S. Department of Labor sales study. *Human Resource Management Review, 9*, 185-218.

Stricker, L. J., & Rock, D. A. (1998). Assessing leadership potential with a biographical measure of personality traits. *International Journal of Selection and Assessment, 6*, 164-184.

Strong, E. K. (1926). An interest test for personnel managers. *Journal of Personnel Research, 5*, 194-203.

Tanofsky, R., Shepps, R. R., & O'Neill, P. J. (1969). Pattern analysis of biographical predictors of success as an insurance salesman. *Journal of Applied Psychology, 53*, 136-139.

Taylor, C. W., & Ellison, R. L. (1967). Biographical predictors of scientific performance. *Science, 155*, 1075-1080.

Telenson, P. A., Alexander, R. A., & Barrett, G. V. (1983). Scoring the biographical information blank: A comparison of three weighting techniques. *Applied Psychological Measurement, 7*, 73-80.

Terpstra, D. E., Mohamed, A. A., & Kethley, R. B. (1999). An analysis of federal court cases involving nine selection devices. *International Journal of Selection and Assessment, 7*, 26-34.

Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personnel Psychology, 51*, 321-355.

Thayer, P. W. (1977). Somethings old, somethings new. *Personnel Psychology, 30*, 513-524.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55*, 525-534.

Van Iddekinge, C. H., Eidson, C. E., Jr., Kudisch, J. D., & Goldblatt, A. M. (2003). A biodata inventory administered via interactive voice response (IVR) technology: Predictive validity, utility, and subgroup differences. *Journal of Business and Psychology, 18*, 145-156.

Vineberg, R., & Joyner, J. N. (1982). *Prediction of job performance: Review of military studies.* Alexandria, VA: Human Resources Research Organization.

Wang, M. W., & Stanley, J. C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40*, 663-705.

Webb, S. C. (1960). The comparative validity of two biographical inventory keys. *Journal of Applied Psychology, 44*, 177-183.

Weiss, D. J. (1976). Multivariate procedures. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 327-362). Chicago, IL: Rand McNally.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376.

Wherry, R. J., Sr. (1931). A new formula for predicting the shrinkage of multiple correlation. *Annals of Mathematical Statistics, 2*, 440-457.

Wiley, A., Wyatt, J., & Camara, W. J. (2010). *The development of a multidimensional index of college readiness for SAT students*. (College Board Research Report 2010-3). New York, NY: The College Board.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika, 3*, 23-40.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann Publishers.

Table 1

Comparison of the Application of the Vertical Percent Method (VPM) and Horizontal Percent Method (HPM) to a Selection of Response Situations

N in High Criterion Group	*N* in Low Criterion Group	Percent of High Criterion Group	Percent of Low Criterion Group	VPM	HPM
0	0	0%	0%	0	N/A
0	5	0%	25%	-25	0
0	10	0%	50%	-50	0
0	15	0%	75%	-75	0
0	20	0%	100%	-100	0
5	0	25%	0%	25	10
5	5	25%	25%	0	5
5	10	25%	50%	-25	3.33
5	15	25%	75%	-50	2.5
5	20	25%	100%	-75	2
10	0	50%	0%	50	10
10	5	50%	25%	25	6.67
10	10	50%	50%	0	5
10	15	50%	75%	-25	4
10	20	50%	100%	-50	3.33
15	0	75%	0%	75	10
15	5	75%	25%	50	7.5
15	10	75%	50%	25	6
15	15	75%	75%	0	5
15	20	75%	100%	-25	4.29
20	0	100%	0%	100	10
20	5	100%	25%	75	8
20	10	100%	50%	50	6.67
20	15	100%	75%	25	5.71
20	20	100%	100%	0	5

Note. Values in this table assume that both the high and low criterion groups are comprised of 20 people.

Table 2

Correlations between Vertical Percent Method Composites Using Different Extreme Group Splits in the SDQ Dataset

	Low Cut	High Cut	1	2	3	4	5	6
1. VPM 27% Extreme Groups	2.63	3.47		1.000	1.000	0.999	0.997	0.993
2. VPM 20% Extreme Groups	2.43	3.60	1.000		1.000	1.000	0.998	0.994
3. VPM 15% Extreme Groups	2.26	3.69	1.000	1.000		1.000	0.999	0.995
4. VPM 10% Extreme Groups	2.04	3.79	0.999	1.000	1.000		0.999	0.996
5. VPM 5% Extreme Groups	1.65	3.90	0.997	0.998	0.999	0.999		0.997
6. VPM 1% Extreme Groups	0.78	4.00	0.993	0.994	0.995	0.996	0.997	

Note. VPM = Vertical Percent Method. Each composite was created by applying the VPM with the specified percent of the weight-derivation sample (N = 100,898) in the high and low extreme groups. The resulting weights were then applied to the entire sample. Low Cut = the cut-off on the FGPA criterion that was used to create the low extreme group. High Cut = the cut-off on the FGPA criterion that was used to create the high extreme group. Correlations above the diagonal are correlations from the weight-development sample. Correlations below the diagonal are from the cross-validation sample. Correlations are sample-weighted means across schools.

Table 3

Descriptive Statistics and Correlations among Empirical Keying Methods in the Student Descriptive Questionnaire Dataset

	Mean[a]	*SD*[a]	1[b]	2[b]	3[b]	4[b]	5[b]
1. Vertical Percent Composite Keyed on FGPA	-3.50 (-3.48)	3.98 (3.96)		.997 (.001)	.636 (.056)	.523 (.060)	.306 (.066)
2. Point-Biserial Composite Keyed on FGPA	-4.16 (-4.14)	3.47 (3.46)	.997 (.001)		.646 (.053)	.525 (.060)	.310 (.066)
3. Multiple Regression Composite Keyed on FGPA	-0.26 (-0.26)	0.31 (0.31)	.636 (.054)	.646 (.051)		.785 (.026)	.462 (.047)
4. CHAID Predicted Value Keyed on FGPA	2.97 (2.97)	0.29 (0.29)	.519 (.059)	.521 (.059)	.779 (.025)		.403 (.050)
5. FGPA	2.97 (2.97)	0.65 (0.66)	.300 (.069)	.304 (.058)	.444 (.053)	.377 (.051)	

Note. FGPA = first-year grade point average. Correlations above the diagonal are correlations from the weight-derivation sample. Correlations below the diagonal are from the cross-validation sample. All correlations are significant at $p < .001$.

[a]First value is for the weight-derivation sample. Values in parentheses are for the cross-validation sample. [b]First value is the mean sample size-weighted correlation across schools. Values in parentheses are the standard deviations of the correlations corrected for sampling error (*SD* rho).

Table 4

Comparison of Vertical Percent Method, Point-Biserial Correlation Method, Option-Keyed Multiple Regression, and CHAID Empirical Keying Methods in the Student Descriptive Questionnaire Dataset

Empirical Keying Method	Variance Explained
Vertical Percent Composite Keyed on FGPA	.099 (.097)
Point-Biserial Composite Keyed on FGPA	.101 (.099)
Multiple Regression Composite Keyed on FGPA	.216 (.201)
CHAID Predicted Value Keyed on FGPA	.165 (.146)

Note. FGPA = first-year grade point average. First value is variance explained in the weight-derivation sample. Values in parentheses are variance explained in the cross-validation sample. Results are based on sample-size weighting of variance explained from each school.

Table 5

Comparison of Number of Items, Response Options, and Sample Sizes for Databases Used in Study 1 and Study 2

Dataset	Base Items	Full Response Options	Dummy-coded Response Options	Weight-Derivation Sample Size	Cross-Validation Sample Size	Ratio of WD N to Response Options
2006 College Board	365	948	583	100,898	49,476	173:1
Academic Rigor Index	395	395	395	44,701	22,943	114:1
Individual Achievement Record	139	695	556	3,483	1,794	6:1

Note. Base Items = the number of items presented to participants; Full Response Options = the number of base items each split into binary items, one for each response option; Dummy-coded Response Options = the number of full response options with one response option dropped per base item; Ratio of WD *N* to Response Options = the ratio of the weight-derivation sample size to the number of dummy-coded response options.

Table 6

Descriptive Statistics and Correlations among Empirical Keying Methods in the Academic Rigor Index Dataset

	Mean[a]	*SD*[a]	1[b]	2[b]	3[b]	4[b]
1. Vertical Percent Composite Keyed on FGPA	0.70 (0.71)	1.42 (1.41)		.993 (.001)	.739 (.056)	.233 (.062)
2. Point-Biserial Composite Keyed on FGPA	0.60 (0.61)	1.30 (1.30)	.993 (.001)		.749 (.052)	.233 (.063)
3. Multiple Regression Composite Keyed on FGPA	0.37 (0.37)	0.25 (0.25)	.738 (.059)	.748 (.054)		.299 (.069)
4. FGPA	2.93 (2.93)	0.67 (0.66)	.215 (.061)	.214 (.061)	.264 (.060)	

Note. FGPA = first-year grade point average. Correlations above the diagonal are correlations from the weight-derivation sample. Correlations below the diagonal are from the cross-validation sample. All correlations are significant at $p < .001$.

[a]First value is for the weight-derivation sample. Values in parentheses are for the cross-validation sample. [b]First value is the mean sample size-weighted correlation across schools. Values in parentheses are the standard deviations of the correlations corrected for sampling error (*SD* rho).

Table 7

Comparison of Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression Empirical Keying Methods in the Academic Rigor Index Dataset

Empirical Keying Method	Variance Explained
Vertical Percent Composite Keyed on FGPA	.060 (.054)
Point-Biserial Composite Keyed on FGPA	.061 (.054)
Multiple Regression Composite Keyed on FGPA	.096 (.078)

Note. FGPA = first-year grade point average. First value is variance explained in the weight-derivation sample. Values in parentheses are variance explained in the cross-validation sample. Results are based on sample-size weighting of variance explained from each school.

Table 8

Descriptive Statistics and Correlations among Empirical Keying Methods in the Individual Achievement Record Dataset

	Mean[a]	*SD*[a]	1	2	3	4
1. Vertical Percent Composite Keyed on Performance	0.97 (1.03)	2.16 (2.14)		.995	.582	.335
2. Point-Biserial Composite Keyed on Performance	0.82 (0.88)	2.14 (2.12)	.995		.602	.346
3. Multiple Regression Composite Keyed on Performance	0.26 (0.26)	0.39 (0.42)	.538	.557		.576
4. Performance	3.53 (3.54)	0.68 (0.70)	.313	.322	.362	

Note. Correlations above the diagonal are correlations from the weight-derivation sample. Correlations below the diagonal are from the cross-validation sample. All correlations are significant at $p < .001$.

[a]First value is for the weight-derivation sample. Values in parentheses are for the cross-validation sample.

Table 9

Comparison of Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression Empirical Keying Methods in the Individual Achievement Record Dataset

Empirical Keying Method	Variance Explained
Vertical Percent Composite Keyed on Performance	.112 (.098)
Point-Biserial Composite Keyed on Performance	.120 (.104)
Multiple Regression Composite Keyed on Performance	.332 (.131)

Note. First value is variance explained in the weight-derivation sample. Values in parentheses are variance explained in the cross-validation sample.

Table 10

Mean Validities and Predictor Intercorrelations in Dummy-coded Archival Datasets

	Validity	Predictor Intercorrelation
Student Descriptive Questionnaire	.04 (.04)	.04 (.07)
Academic Rigor Index	.04 (.04)	.02 (.07)
Individual Achievement Record	.03 (.05)	.03 (.06)

Note. First value is the mean and the value in parentheses is the standard deviation.

Table 11

School-level Key and Sample Size Comparisons in the Student Descriptive Questionnaire Dataset

School	Full *N*	WD *N*	*N* Ratio	VPM	PBC	MR	VPM Diff	PBC Diff	MR Tie or Beat
School 1	1,836	1,212	2.08	.05	.08	.05	.00	-.03	
School 2	3,193	2,107	3.61	.06	.07	.07	.01	.00	X
School 3	2,303	1,520	2.61	.08	.10	.08	.00	-.02	
School 4	1,008	665	1.14	.06	.08	.01	-.05	-.07	
School 5	1,361	898	1.54	.09	.10	.04	-.05	-.06	
School 6	2,341	1,545	2.65	.08	.10	.06	-.02	-.04	
School 7	1,024	676	1.16	.08	.09	.01	-.07	-.08	
School 8	1,136	750	1.29	.09	.11	.03	-.06	-.08	
School 9	943	622	1.07	.07	.09	.01	-.06	-.08	
School 10	1,971	1,301	2.23	.10	.11	.07	-.03	-.04	
School 11	1,486	981	1.68	.09	.11	.02	-.07	-.09	
School 12	1,278	843	1.45	.10	.12	.02	-.08	-.10	
School 13	1,458	962	1.65	.04	.05	.04	.00	-.01	
School 14	2,841	1,875	3.22	.07	.08	.08	.01	.00	X
School 15	3,731	2,462	4.22	.12	.12	.15	.03	.03	X
School 16	1,495	987	1.69	.10	.11	.03	-.07	-.08	
School 17	1,787	1,179	2.02	.07	.07	.03	-.04	-.04	
School 18	2,381	1,571	2.70	.14	.13	.09	-.05	-.04	
School 19	2,854	1,884	3.23	.06	.07	.03	-.03	-.04	
School 20	5,483	3,619	6.21	.10	.10	.11	.01	.01	X
School 21	2,002	1,321	2.27	.04	.05	.04	.00	-.01	
School 22	2,520	1,663	2.85	.09	.10	.07	-.02	-.03	
School 23	4,178	2,757	4.73	.08	.10	.13	.05	.03	X
School 24	1,309	864	1.48	.06	.06	.01	-.05	-.05	
School 25	976	644	1.10	.05	.08	.00	-.05	-.08	
School 26	2,283	1,507	2.58	.12	.13	.08	-.04	-.05	
School 27	1,362	899	1.54	.15	.16	.07	-.08	-.09	
School 28	1,013	669	1.15	.13	.15	.01	-.12	-.14	
School 29	3,320	2,191	3.76	.09	.10	.07	-.02	-.03	
School 30	1,247	823	1.41	.03	.05	.01	-.02	-.04	
School 31	2,306	1,522	2.61	.08	.09	.05	-.03	-.04	
School 32	1,495	987	1.69	.05	.07	.04	-.01	-.03	
School 33	2,217	1,463	2.51	.13	.15	.09	-.04	-.06	
School 34	2,418	1,596	2.74	.08	.10	.09	.01	-.01	
School 35	2,213	1,461	2.51	.12	.12	.08	-.04	-.04	
School 36	2,086	1,377	2.36	.07	.09	.08	.01	-.01	
School 37	5,758	3,800	6.52	.11	.12	.17	.06	.05	X
School 38	3,294	2,174	3.73	.12	.12	.12	.00	.00	X
School 39	4,192	2,767	4.75	.09	.10	.11	.02	.01	X
School 40	1,385	914	1.57	.09	.12	.04	-.05	-.08	
School 41	1,194	788	1.35	.03	.04	.02	-.01	-.02	
School 42	1,393	919	1.58	.09	.10	.04	-.05	-.06	

(Table 11 continues)

(Table 11 continued)

School	Full *N*	WD *N*	*N* Ratio	VPM	PBC	MR	VPM Diff	PBC Diff	MR Tie or Beat
School 43	1,287	849	1.46	.10	.12	.02	-.08	-.10	
School 44	1,303	860	1.48	.11	.12	.03	-.08	-.09	
School 45	2,423	1,599	2.74	.12	.13	.10	-.02	-.03	
School 46	1,729	1,141	1.96	.09	.10	.03	-.06	-.07	
School 47	951	628	1.08	.04	.06	.01	-.03	-.05	

Note. Full *N* = full sample size; WD *N* = weight-derivation sample size; *N* Ratio = ratio of weight-derivation sample size to the number of dummy-coded response options in the 2006 College Board dataset (i.e., 583); VPM = variance explained for the vertical percent method in the cross-validation sample; PBC = variance explained for the point-biserial correlation method in the cross-validation sample; MR = variance explained for option-keyed regression in the cross-validation sample; VPM Diff = MR – VPM; PBC Diff = MR – PBC; MR Tie or Beat = whether MR ties or beats both VPM and PBC. Key validity results are the aggregation of 25 weight-derivation/cross-validation sample splits for key development. For the difference variables, a positive value means multiple regression has a higher cross-validity and a negative value means the other keying method has a higher cross-validity.

Table 12

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Dichotomous Items

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	3	2	1.75	2	1.75	1.75	1.75	1.75	1.75
Point-Biserial Correlation	5	3	3	3	2	2	2	2	2
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	6	1.75	1.75	3	1.5	1.75	1.75	1.5	1.75
Point-Biserial Correlation	10	3	2	3	1.75	2	2	1.75	1.75
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	4	1.75	1.75	1.75	1.5	1.5	1.5	1.5	1.5
Point-Biserial Correlation	5	1.75	1.75	1.75	1.75	1.75	1.5	1.5	1.5

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dichotomous items. The decision on whether the methods are tied or exceed one another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values, and the effective values will be smaller due to dichotomization.

Table 13

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Dichotomous Items

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.090	.230	.312	.228	.337	.386	.310	.379	.416
Point-Biserial Correlation	.096	.241	.325	.239	.350	.399	.323	.393	.430
Option-Keyed Multiple Regression	.101	.276	.374	.270	.404	.462	.367	.453	.498
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.050	.181	.277	.073	.205	.305	.074	.194	.297
Point-Biserial Correlation	.052	.186	.286	.074	.210	.313	.074	.197	.303
Option-Keyed Multiple Regression	.074	.254	.361	.186	.373	.447	.267	.429	.486
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.037	.146	.234	.043	.141	.239	.040	.119	.212
Point-Biserial Correlation	.037	.149	.241	.043	.143	.244	.040	.120	.216
Option-Keyed Multiple Regression	.074	.248	.346	.181	.360	.435	.258	.414	.474

Note. The maximum sample size-to-item ratio investigated in this set of simulations was 100:1. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values, and the effective values will be smaller due to dichotomization.

Table 14

Frequency Distribution of Items with Given Numbers of Response Options Used with the Student Descriptive Questionnaire Dataset

Number of Response Options in Item	Item Frequency
2	299
3	11
4	17
5	16
6	7
7	7
8	3
9	1
10	2
11	1
12	0
13	0
14	1

Note. The number of response options is the full number of response options. The frequency distribution of items with a given number of dummy-coded response options would be obtained by subtracting one from the first column.

Table 15

Items and Response Options for the Each Set of Simulation Conditions

Simulation Set	Base Items			Full Responses			Dummy-coded Responses		
	48	120	204	48	120	204	48	120	204
Dichotomous	48	120	204	96	240	408	48	120	204
Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition	24	60	102	72	180	306	48	120	204
Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options	48	120	204	144	360	612	96	240	408
Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition	16	40	68	64	160	272	48	120	204
Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options	48	120	204	192	480	816	144	360	612
Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition	12	30	51	60	150	255	48	120	204
Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options	48	120	204	240	600	1,020	192	480	816

Table 16

Response Patterns Simulated for Simulation Conditions Involving Items with Three Response Options

	Proportion Endorsing First Option	Proportion Endorsing Second Option	Proportion Endorsing Third Option
Response Pattern 1	.33	.33	.33
Response Pattern 2	.25	.25	.50
Response Pattern 3	.25	.50	.25
Response Pattern 4	.50	.25	.25
Response Pattern 5	.60	.20	.20
Response Pattern 6	.20	.60	.20
Response Pattern 7	.20	.20	.60
Response Pattern 8	.40	.40	.20
Response Pattern 9	.20	.40	.40
Response Pattern 10	.40	.20	.40
Response Pattern 11	.15	.15	.70
Response Pattern 12	.15	.70	.15
Response Pattern 13	.70	.15	.15

Note. Percentages not summing to one are due to rounding and were specified as exact in simulation trials. First option through third option corresponds to 1 through 3, respectively, on a 3-point scale.

Table 17

Response Patterns Simulated for Simulation Conditions Involving Items with Four Response Options

	Proportion Endorsing First Option	Proportion Endorsing Second Option	Proportion Endorsing Third Option	Proportion Endorsing Fourth Option
Response Pattern 1	.25	.25	.25	.25
Response Pattern 2	.15	.20	.30	.35
Response Pattern 3	.35	.30	.20	.15
Response Pattern 4	.15	.15	.20	.50
Response Pattern 5	.50	.20	.15	.15
Response Pattern 6	.20	.20	.20	.40
Response Pattern 7	.40	.20	.20	.20
Response Pattern 8	.15	.35	.35	.15
Response Pattern 9	.20	.30	.30	.20
Response Pattern 10	.15	.15	.55	.15
Response Pattern 11	.15	.55	.15	.15

Note. First option through fourth option corresponds to 1 through 4, respectively, on a 4-point scale.

Table 18

Response Patterns Simulated for Simulation Conditions Involving Items with Five Response Options

	Proportion Endorsing First Option	Proportion Endorsing Second Option	Proportion Endorsing Third Option	Proportion Endorsing Fourth Option	Proportion Endorsing Fifth Option
Response Pattern 1	.20	.20	.20	.20	.20
Response Pattern 2	.40	.15	.15	.15	.15
Response Pattern 3	.15	.40	.15	.15	.15
Response Pattern 4	.15	.15	.40	.15	.15
Response Pattern 5	.15	.15	.15	.40	.15
Response Pattern 6	.15	.15	.15	.15	.40
Response Pattern 7	.15	.15	.15	.25	.30
Response Pattern 8	.15	.15	.25	.30	.15
Response Pattern 9	.15	.15	.30	.25	.15
Response Pattern 10	.15	.30	.25	.15	.15
Response Pattern 11	.30	.25	.15	.15	.15

Note. First option through fifth option corresponds to 1 through 5, respectively, on a 5-point scale.

Table 19

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	3	1.75	1.5	2	1.5	1.5	1.5	1.5	1.5
Point-Biserial Correlation	5	2	1.5	3	1.5	1.5	1.75	1.5	1.5
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	6	1.5	1.5	3	1.5	1.5	1.75	1.5	1.5
Point-Biserial Correlation	25	2	1.5	4	1.5	1.5	1.75	1.5	1.5
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	5	1.25	1.5	1.75	1.25	1.25	1.5	1.25	1.25
Point-Biserial Correlation	6	1.5	1.5	2	1.5	1.5	1.5	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into three response options.

Table 20

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	3	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Point-Biserial Correlation	3	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	4	1.5	1.5	1.5	1.25	1.5	1.25	1.25	1.25
Point-Biserial Correlation	5	1.5	1.5	1.5	1.5	1.5	1.25	1.25	1.5
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	3	1.5	1.25	1.5	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	3	1.5	1.5	1.5	1.25	1.25	1.25	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. Sample size ratios of 25, 50, and 100 were only investigated for conditions where the number of items is 48. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into three response options.

Table 21

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	4	2	1.25	3	1.25	1.25	1.75	1.25	1.25
Point-Biserial Correlation	5	3	1.5	3	1.5	1.25	2	1.25	1.25
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	6	1.75	1.25	4	1.25	1.25	2	1.25	1.25
Point-Biserial Correlation	25	2	1.5	5	1.5	1.25	2	1.25	1.25
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	3	1.5	1.25	3	1.25	1.25	1.5	1.25	1.25
Point-Biserial Correlation	10	1.5	1.25	3	1.25	1.25	1.5	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into four response options.

Table 22

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	2	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	3	1.25	1.25	1.5	1.25	1.25	1.25	1.25	1.25
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	3	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	4	1.25	1.25	1.5	1.25	1.25	1.25	1.25	1.25
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	1.75	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	2	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. Sample size ratios of 25, 50, and 100 were only investigated for conditions where the number of items is 48. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into four response options.

Table 23

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	1.5	2	1.25	3	1.25	1.25	2	1.25	1.25
Point-Biserial Correlation	5	3	1.5	3	1.25	1.25	3	1.25	1.25
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	6	3	1.25	4	1.25	1.25	3	1.25	1.25
Point-Biserial Correlation	50	3	1.25	10	1.25	1.25	3	1.25	1.25
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	4	1.5	1.25	3	1.25	1.25	1.75	1.25	1.25
Point-Biserial Correlation	10	2	1.25	3	1.25	1.25	1.75	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into five response options.

Table 24

Ratios of Sample Size-to-Items at which Multiple Regression Cross-Validation Variance Explained Ties or Exceeds the Traditional Alternatives for All Simulation Conditions Involving Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	1.75	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	3	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	3	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	3	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	1.75	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25
Point-Biserial Correlation	2	1.25	1.25	1.25	1.25	1.25	1.25	1.25	1.25

Note. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. Sample size-to-item ratios are the weight-derivation sample size relative to the number of dummy-coded response options for this set of simulation conditions as outlined in Table 15. The decision on whether the methods are tied or exceed on another is based on rounding to the third decimal place. Sample size ratios investigated include 1+2, 1.25, 1.5, 1.75, 2, 3, 4, 5, 6, 10, 25, 50, and 100. Sample size ratios of 25, 50, and 100 were only investigated for conditions where the number of items is 48. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into five response options.

Table 25

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.059	.182	.319	.160	.355	.436	.277	.422	.481
Point-Biserial Correlation	.062	.188	.329	.166	.366	.447	.286	.434	.492
Option-Keyed Multiple Regression	.067	.231	.432	.191	.483	.586	.360	.570	.640
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.040	.154	.291	.066	.243	.362	.078	.244	.370
Point-Biserial Correlation	.041	.159	.299	.067	.248	.369	.079	.247	.376
Option-Keyed Multiple Regression	.051	.225	.423	.137	.448	.567	.252	.533	.624
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.031	.136	.254	.043	.178	.297	.045	.164	.281
Point-Biserial Correlation	.031	.140	.262	.043	.181	.302	.045	.165	.284
Option-Keyed Multiple Regression	.051	.241	.413	.148	.441	.557	.261	.524	.614

Note. The maximum sample size-to-item ratio investigated in this set of simulations was 100:1. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into three response options.

Table 26

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.127	.315	.415	.246	.378	.438	.347	.432	.477
Point-Biserial Correlation	.132	.325	.426	.265	.398	.457	.368	.452	.495
Option-Keyed Multiple Regression	.149	.432	.560	.374	.553	.619	.509	.613	.659
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.059	.230	.351	.076	.224	.334	.077	.208	.324
Point-Biserial Correlation	.060	.235	.360	.077	.230	.346	.078	.212	.333
Option-Keyed Multiple Regression	.106	.401	.538	.237	.516	.603	.370	.580	.644
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.039	.175	.295	.047	.152	.268	.042	.131	.235
Point-Biserial Correlation	.040	.178	.302	.047	.155	.272	.042	.132	.238
Option-Keyed Multiple Regression	.111	.394	.528	.246	.503	.595	.364	.568	.631

Note. The maximum sample size-to-item ratio investigated in conditions where the number of items is 48 was 100:1; it was 10:1 for other conditions. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into three response options.

Table 27

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.044	.134	.268	.115	.312	.429	.201	.406	.485
Point-Biserial Correlation	.047	.142	.283	.123	.328	.448	.213	.425	.505
Option-Keyed Multiple Regression	.051	.170	.401	.139	.479	.646	.260	.619	.713
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.031	.119	.254	.057	.228	.367	.071	.258	.397
Point-Biserial Correlation	.033	.125	.267	.059	.237	.382	.073	.266	.410
Option-Keyed Multiple Regression	.038	.165	.406	.103	.451	.625	.190	.584	.699
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.025	.107	.231	.039	.178	.311	.044	.182	.309
Point-Biserial Correlation	.027	.112	.242	.040	.184	.322	.044	.187	.318
Option-Keyed Multiple Regression	.038	.183	.411	.111	.456	.620	.215	.578	.687

Note. The maximum sample size-to-item ratio investigated in this set of simulations was 100:1. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into four response options.

Table 28

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.141	.350	.449	.281	.422	.483	.387	.475	.514
Point-Biserial Correlation	.150	.368	.468	.306	.449	.509	.417	.502	.539
Option-Keyed Multiple Regression	.174	.543	.669	.485	.678	.738	.634	.729	.766
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.063	.242	.378	.078	.237	.359	.078	.214	.345
Point-Biserial Correlation	.064	.251	.392	.079	.245	.372	.079	.218	.354
Option-Keyed Multiple Regression	.126	.501	.654	.322	.641	.721	.484	.700	.755
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.041	.185	.315	.045	.160	.275	.041	.132	.240
Point-Biserial Correlation	.042	.191	.326	.046	.164	.285	.041	.134	.245
Option-Keyed Multiple Regression	.137	.503	.645	.341	.633	.711	.483	.688	.747

Note. The maximum sample size-to-item ratio investigated in conditions where the number of items is 48 was 100:1; it was 10:1 for other conditions. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into four response options.

Table 29

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.033	.104	.231	.092	.273	.411	.162	.384	.484
Point-Biserial Correlation	.036	.111	.245	.098	.289	.432	.172	.405	.505
Option-Keyed Multiple Regression	.038	.127	.336	.109	.428	.664	.203	.633	.753
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.026	.095	.219	.051	.211	.361	.066	.259	.396
Point-Biserial Correlation	.028	.101	.231	.053	.220	.377	.067	.269	.412
Option-Keyed Multiple Regression	.031	.126	.343	.080	.413	.652	.146	.596	.734
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.023	.092	.205	.037	.175	.311	.042	.191	.333
Point-Biserial Correlation	.024	.097	.217	.038	.182	.325	.043	.197	.344
Option-Keyed Multiple Regression	.031	.140	.370	.086	.441	.643	.162	.600	.733

Note. The maximum sample size-to-item ratio investigated in this set of simulations was 100:1. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into five response options.

Table 30

Cross-Valid Variance Explained for all Keying Methods at Maximum Sample Size-to-Item Ratio for All Simulation Conditions Involving Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

	Number of Items = 48			Number of Items = 120			Number of Items = 204		
	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15	sd = .05	sd = .10	sd = .15
Mean Predictor Intercorrelations = .00									
Vertical Percent Method	.151	.374	.476	.299	.445	.506	.414	.500	.542
Point-Biserial Correlation	.161	.394	.498	.329	.475	.535	.447	.529	.569
Option-Keyed Multiple Regression	.188	.614	.745	.555	.751	.804	.713	.797	.829
Mean Predictor Intercorrelations = .10									
Vertical Percent Method	.064	.257	.398	.079	.246	.371	.078	.221	.353
Point-Biserial Correlation	.066	.266	.413	.081	.253	.386	.079	.225	.364
Option-Keyed Multiple Regression	.137	.576	.728	.385	.723	.788	.569	.772	.819
Mean Predictor Intercorrelations = .20									
Vertical Percent Method	.041	.188	.324	.046	.161	.277	.041	.131	.248
Point-Biserial Correlation	.042	.193	.334	.046	.164	.285	.041	.132	.254
Option-Keyed Multiple Regression	.150	.579	.718	.413	.715	.783	.570	.765	.812

Note. The maximum sample size-to-item ratio investigated in conditions where the number of items is 48 was 100:1; it was 10:1 for other conditions. sd parameter is the standard deviation of the generated validities and intercorrelations around their mean. All simulation conditions assume a mean predictor validity of .04. All specified correlational values are the simulated continuous values and the effective values will be smaller due to splitting predictors into five response options.

Rational Item Development

		Item Weighting		
		Differential Weights - Rational	Differential Weights - External	Unit-Weighting
	Rational			
Item Selection	Empirical - Internal			
	Empirical - External			
	All Items			

Archival Item Development

		Item Weighting		
		Differential Weights - Rational	Differential Weights - External	Unit-Weighting
	Rational			
Item Selection	Empirical - Internal			
	Empirical - External			
	All Items			

Figure 1. An Illustrative Model of Scale Development Choices

Normal					Schizophrenic				
		Item 2					**Item 2**		
		T	F				T	F	
Item 1	T	50	0	50	**Item 1**	T	0	50	50
	F	0	50	50		F	50	0	50
		50	50	100			50	50	100

Figure 2. Recreation of Meehl's (1950) Paradox Example: Part 1

		Normal	**Schizophrenic**	
Response Pattern	TT or FF	100	0	100
	TF or FT	0	100	100
		100	100	200

Figure 3. Recreation of Meehl's (1950) Paradox Example: Part 2

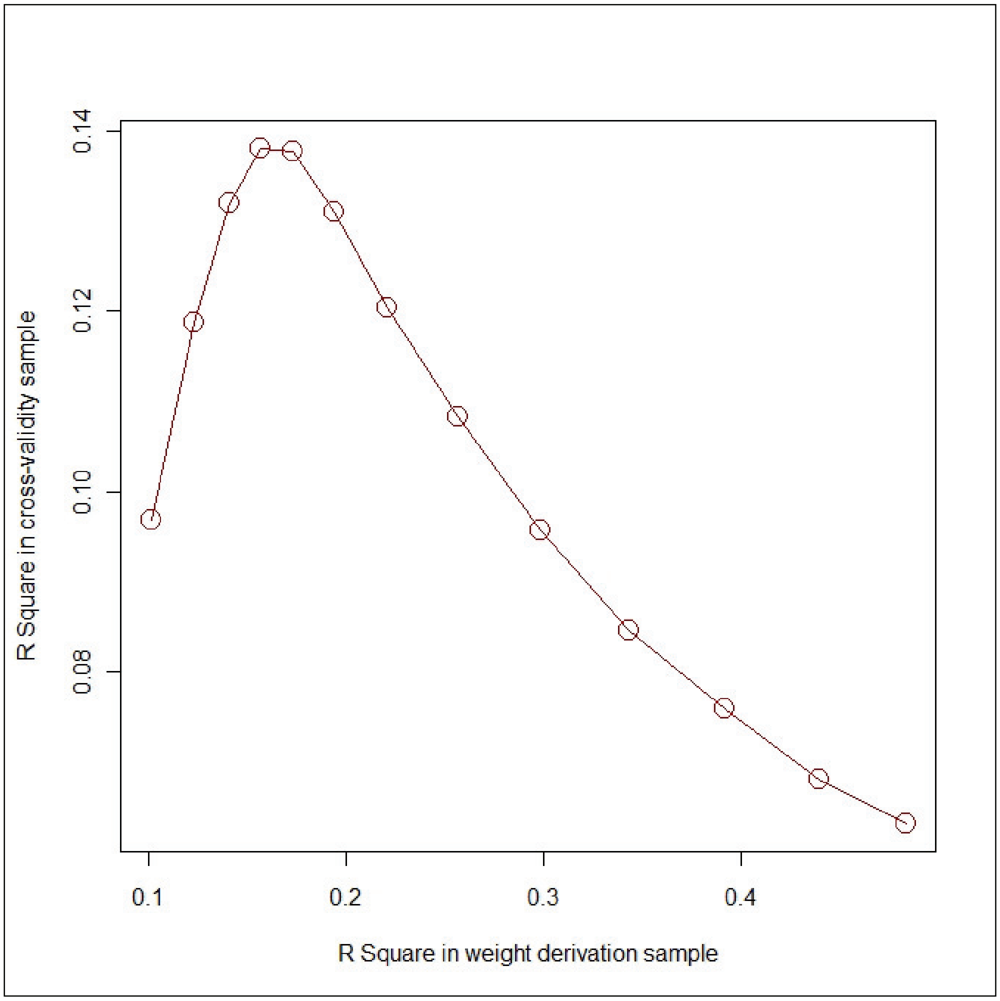

Figure 4. Visual Illustration of the Relationship between Variance Explained in the Weight-Derivation Sample and the Cross-validation Sample Using a Series of CHAID Maximum Depths to Predict FGPA with a Node Splitting Alpha of .05. Each point represents a different maximum depth for the CHAID tree from 2 to 14 (left to right).

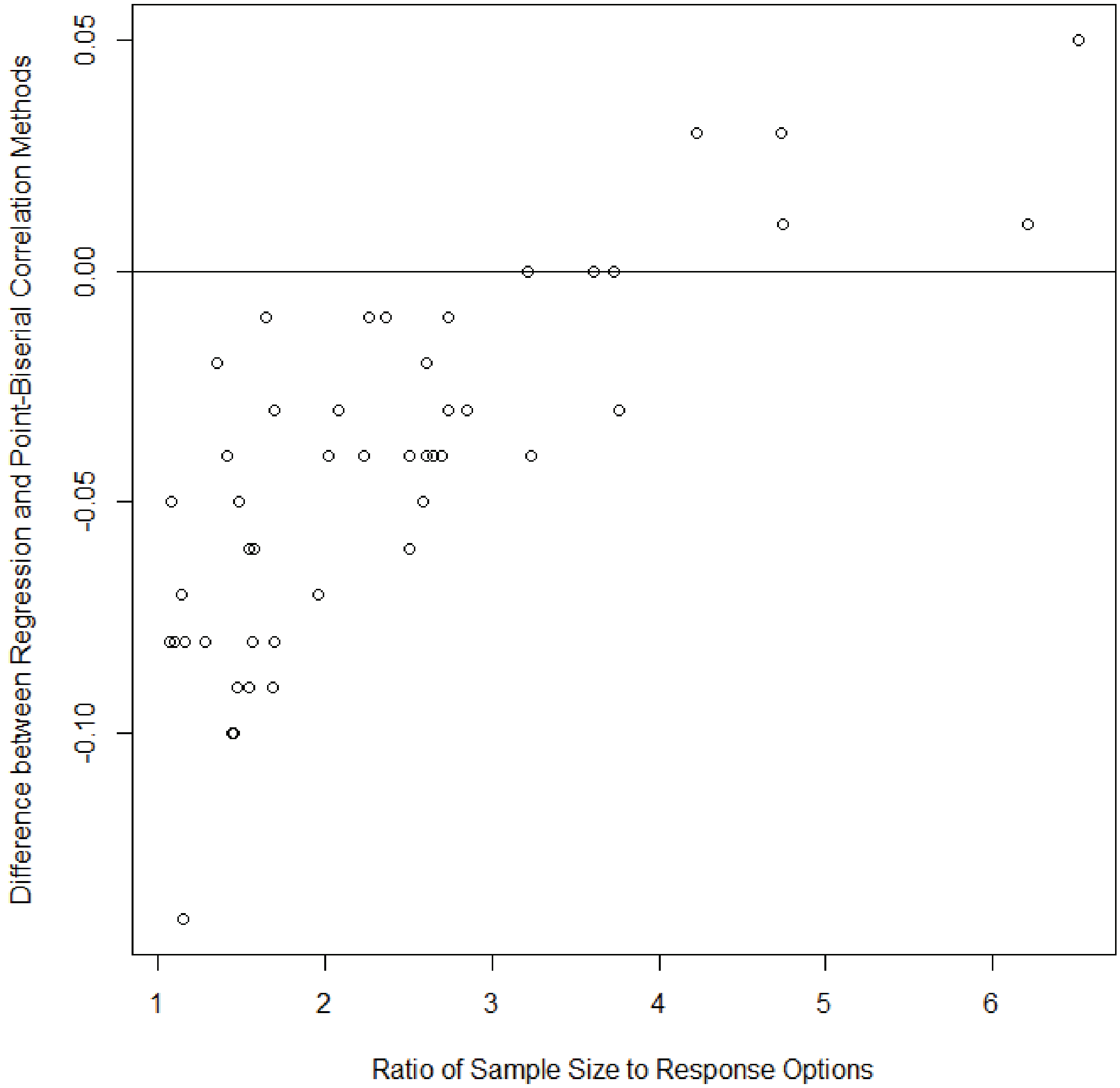

Figure 5. School-Level Differences between the Option-Keyed Multiple Regression and Point-Biserial Correlation Methods Performance in the Cross-Validation Sample Plotted Against the Ratio of the Weight-Development Sample Size to the Number of Dummy-coded Response Options. Keys are developed within the weight-development sample of each school of the Student Descriptive Questionnaire dataset and applied to that school's cross-validation sample. Each plot point is the average cross-valid variance explained of 25 keys in 25 unique weight-derivation/cross-validation sample splits.

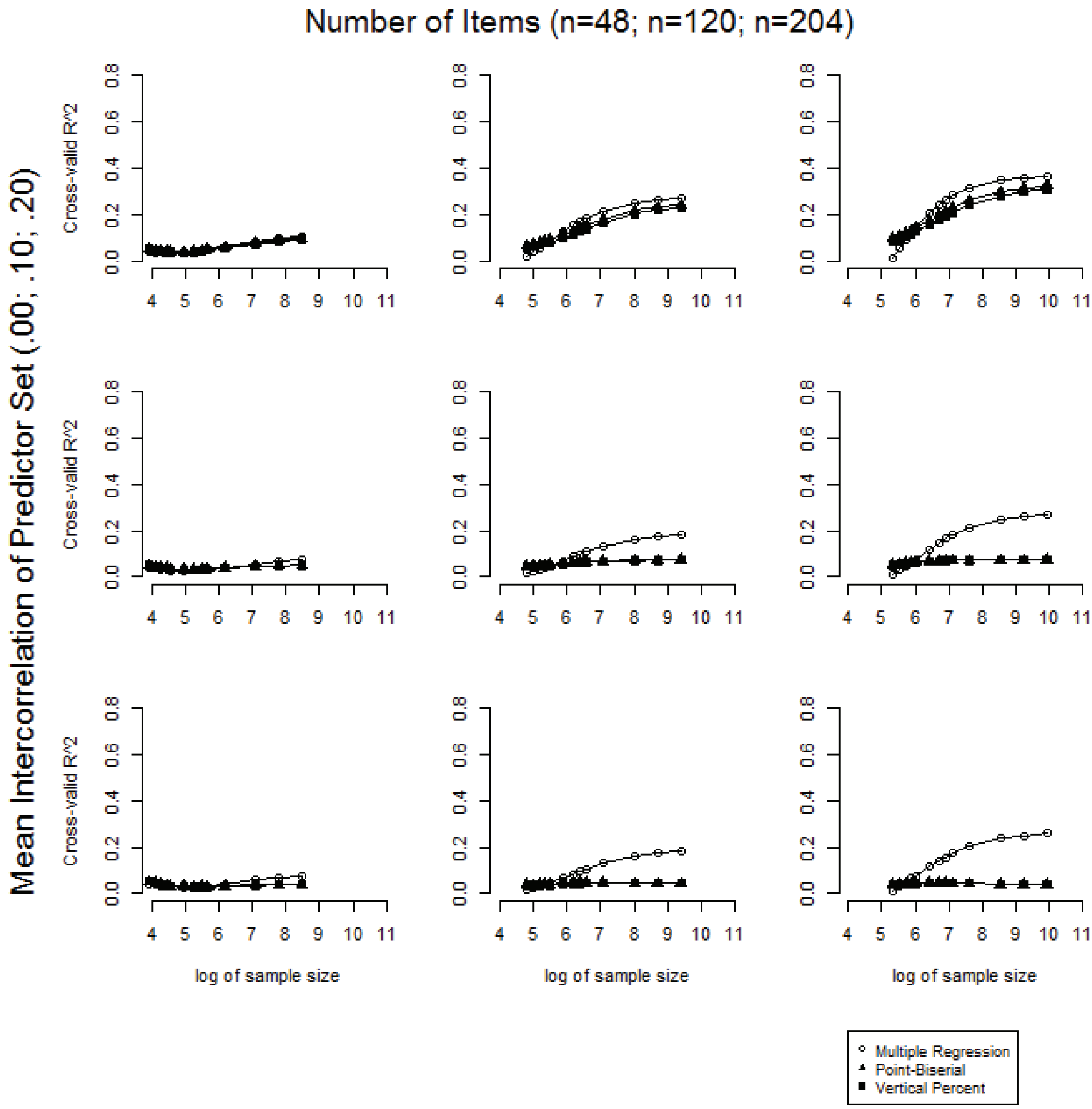

Figure 6. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Two Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

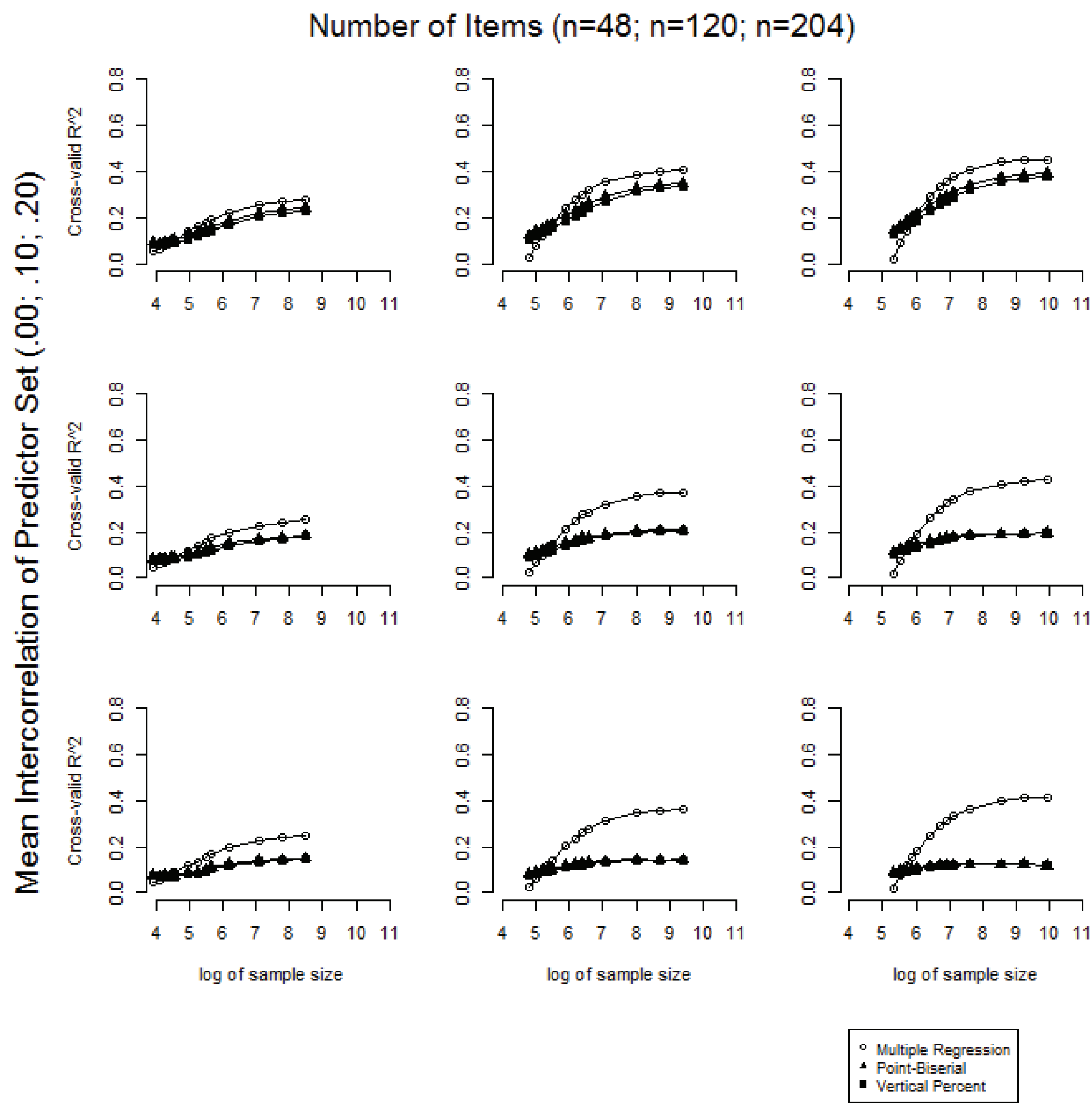

Figure 7. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Two Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

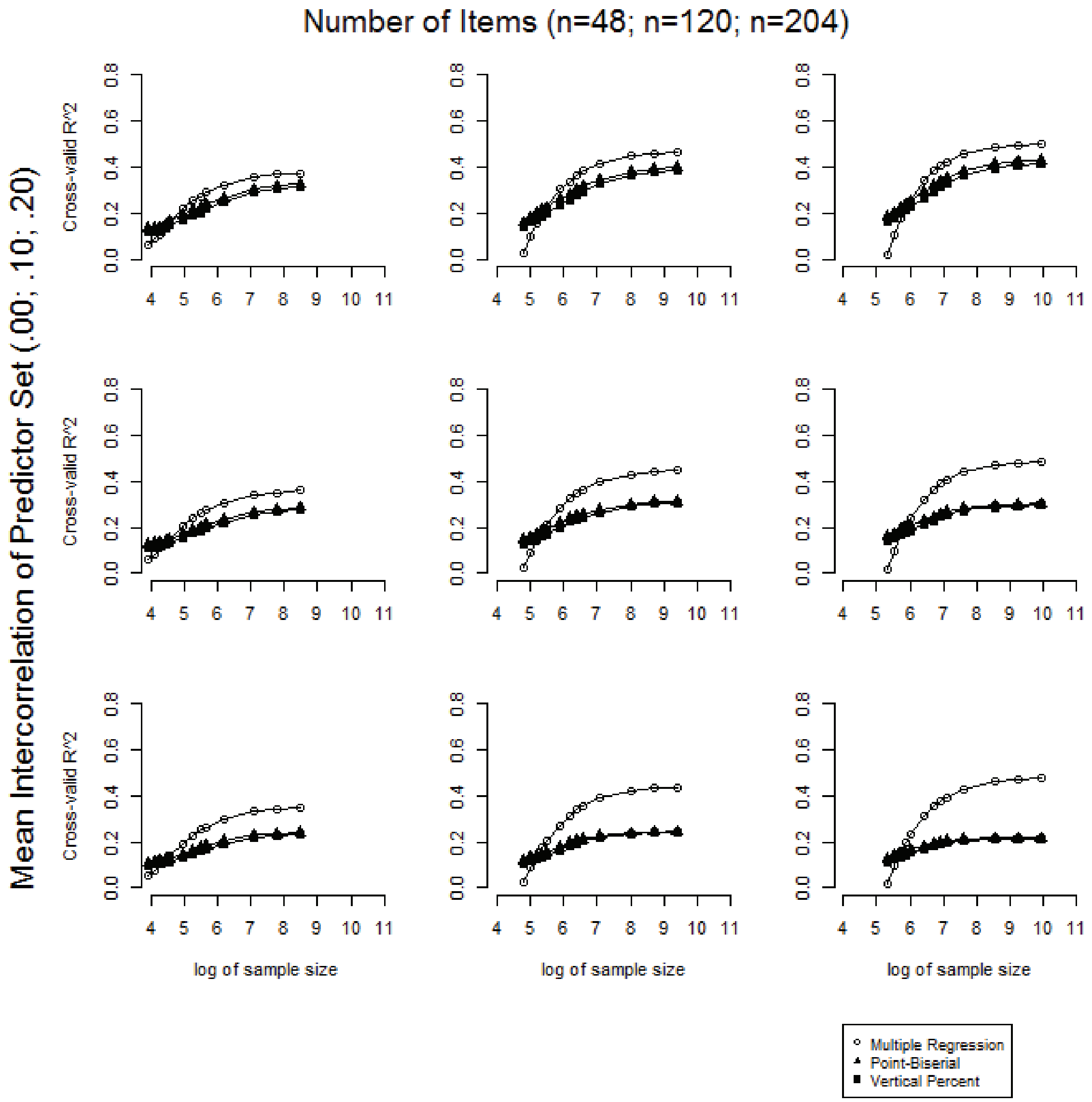

Figure 8. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Two Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

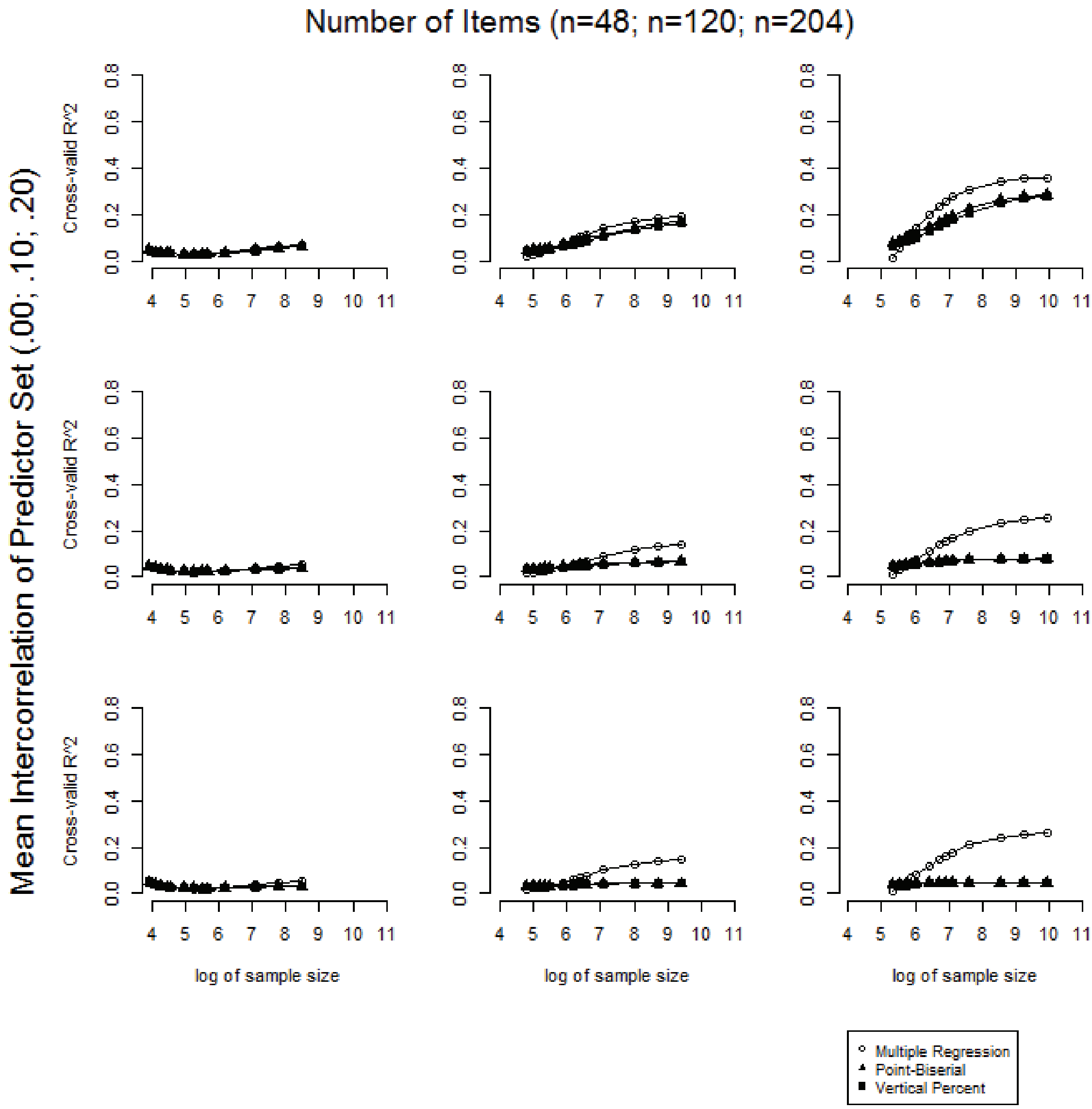

Figure 9. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

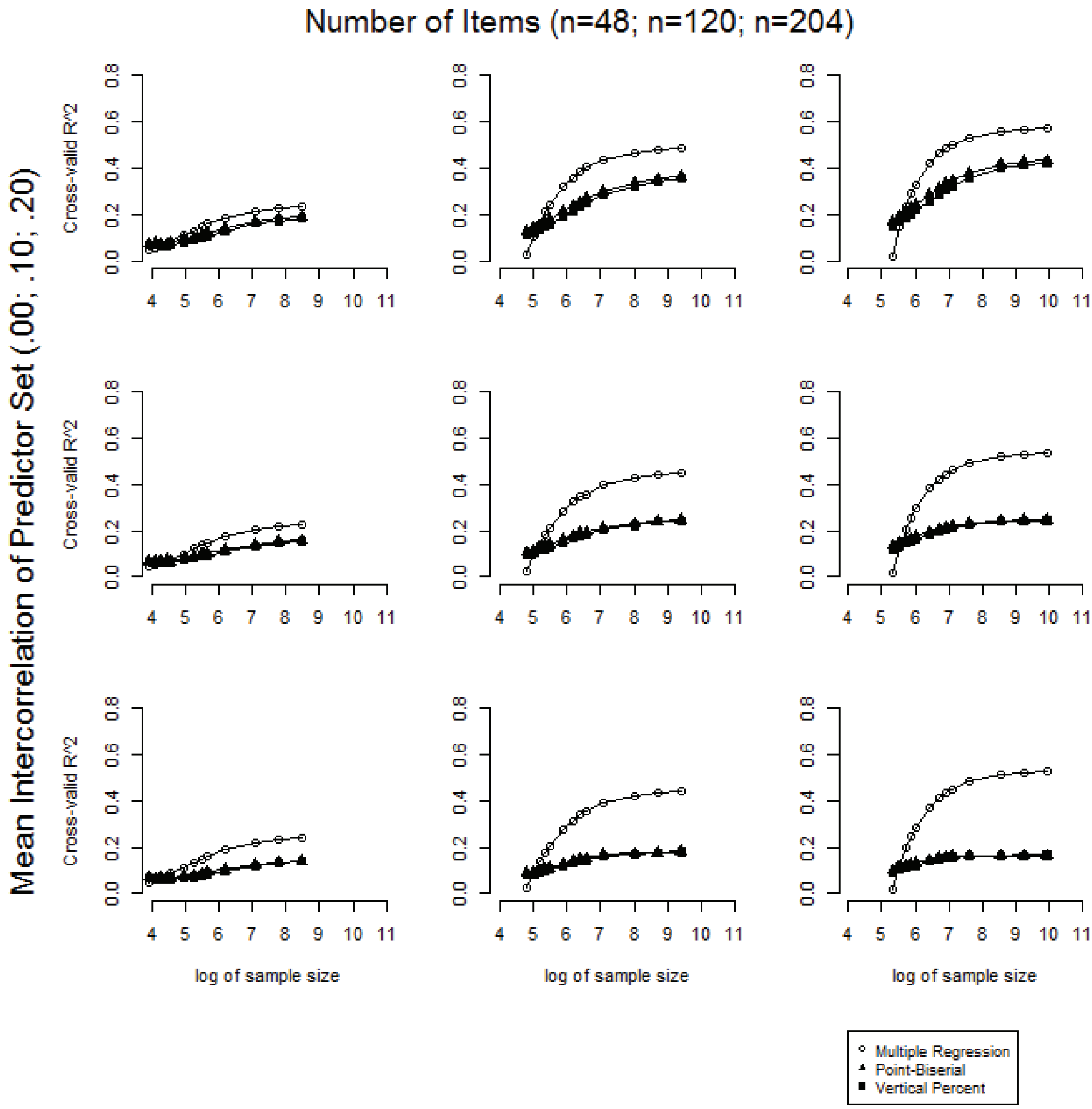

Figure 10. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

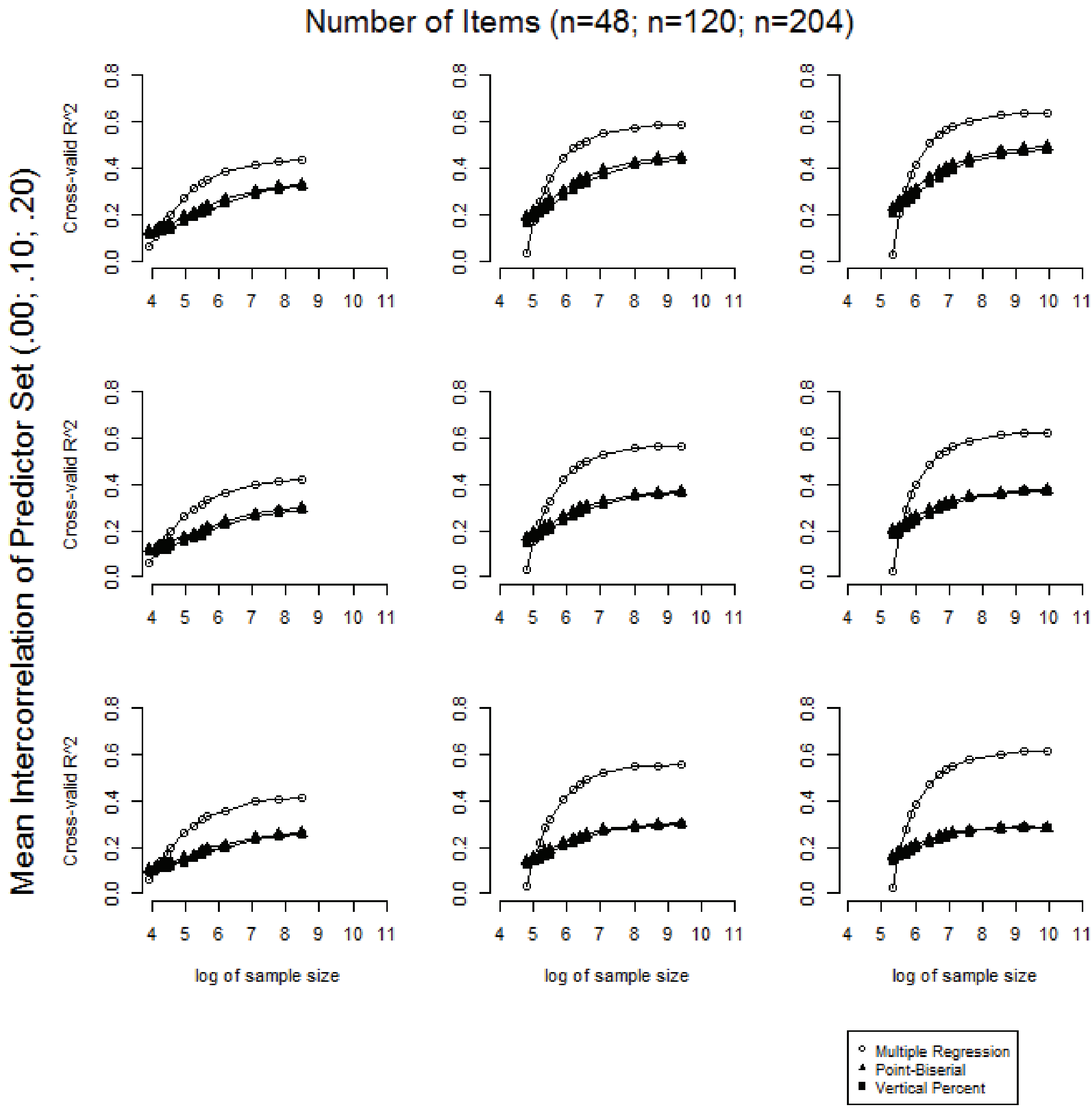

Figure 11. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

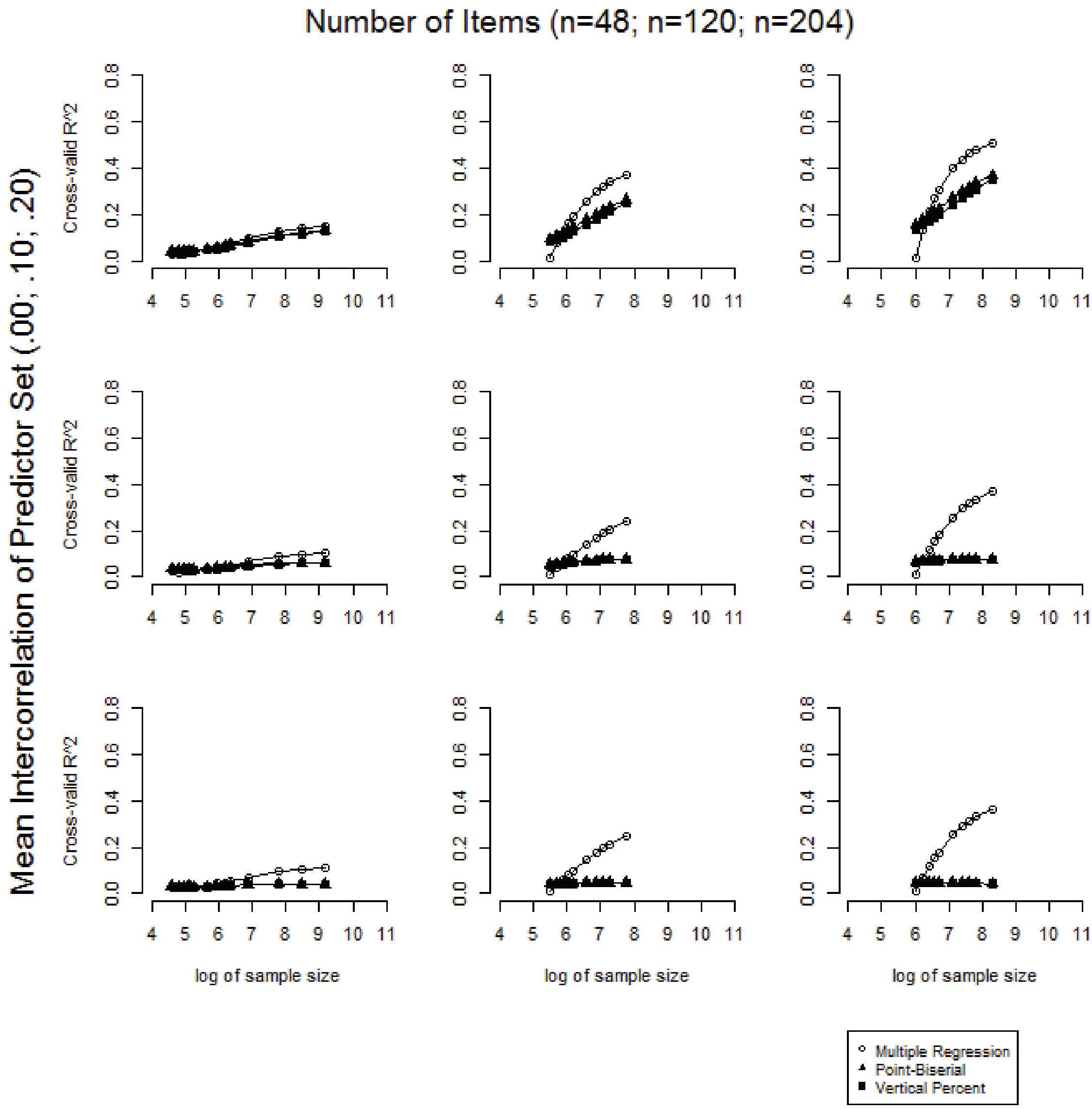

Figure 12. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

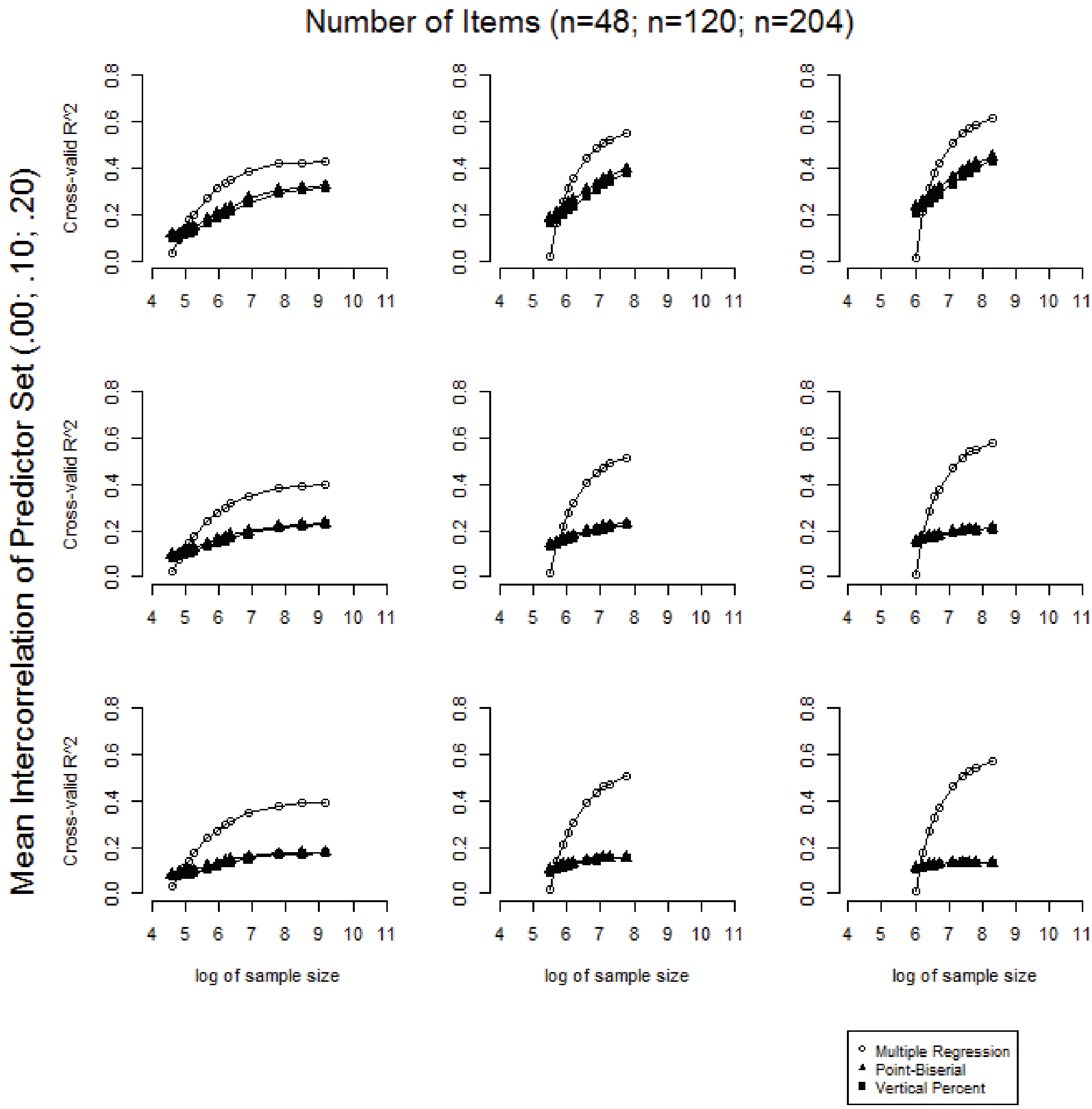

Figure 13. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

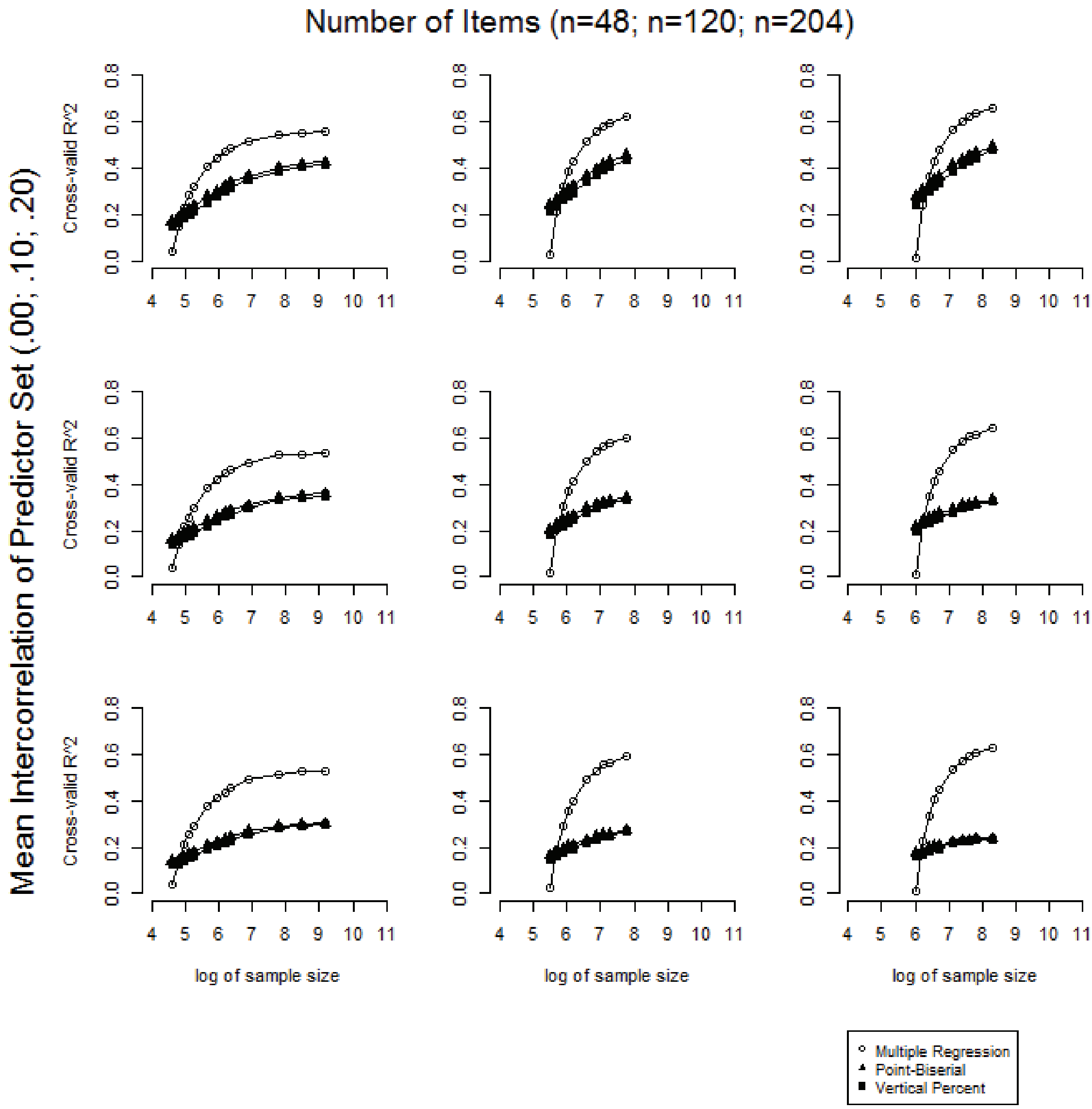

Figure 14. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

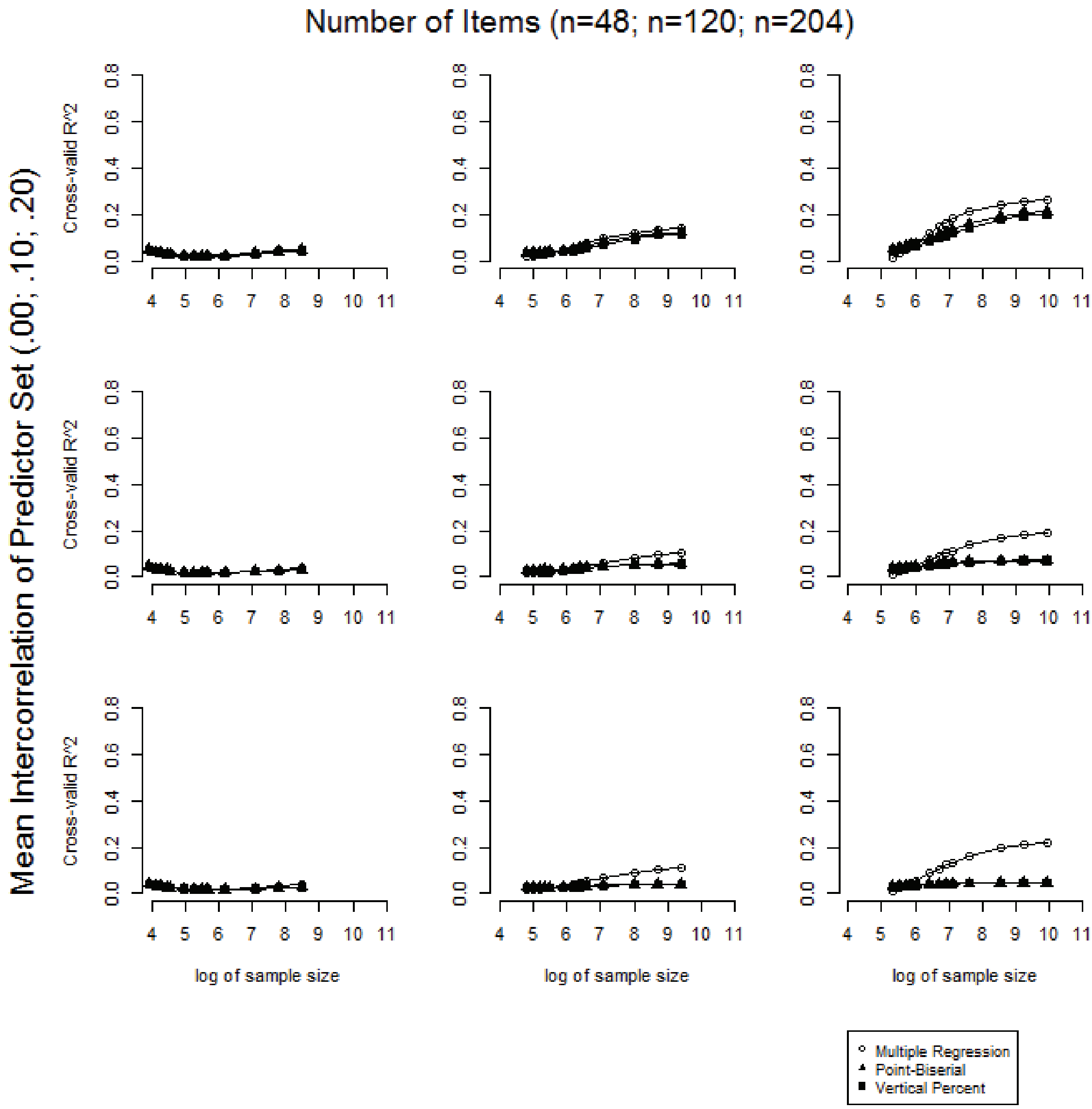

Figure 15. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

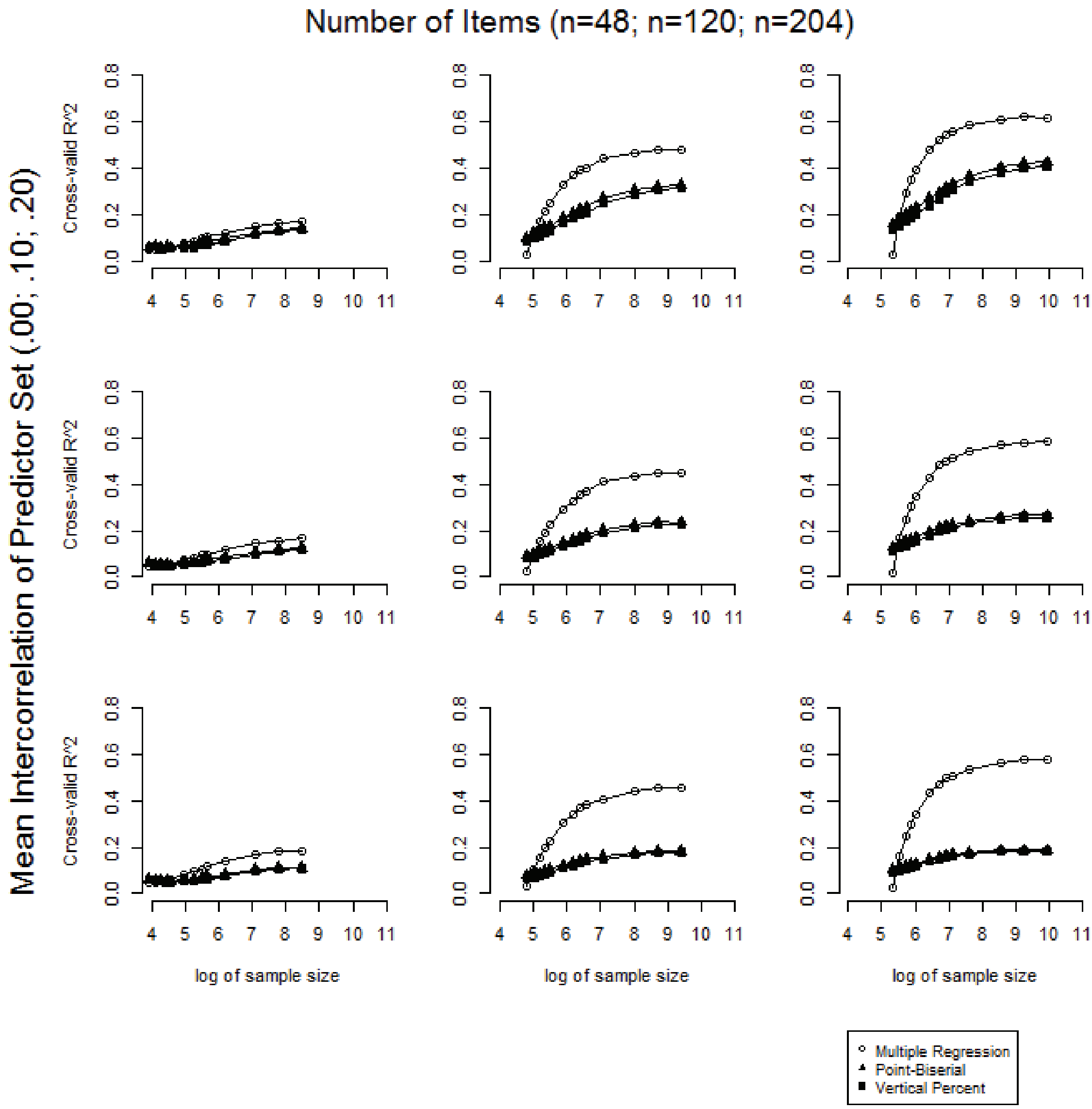

Figure 16. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

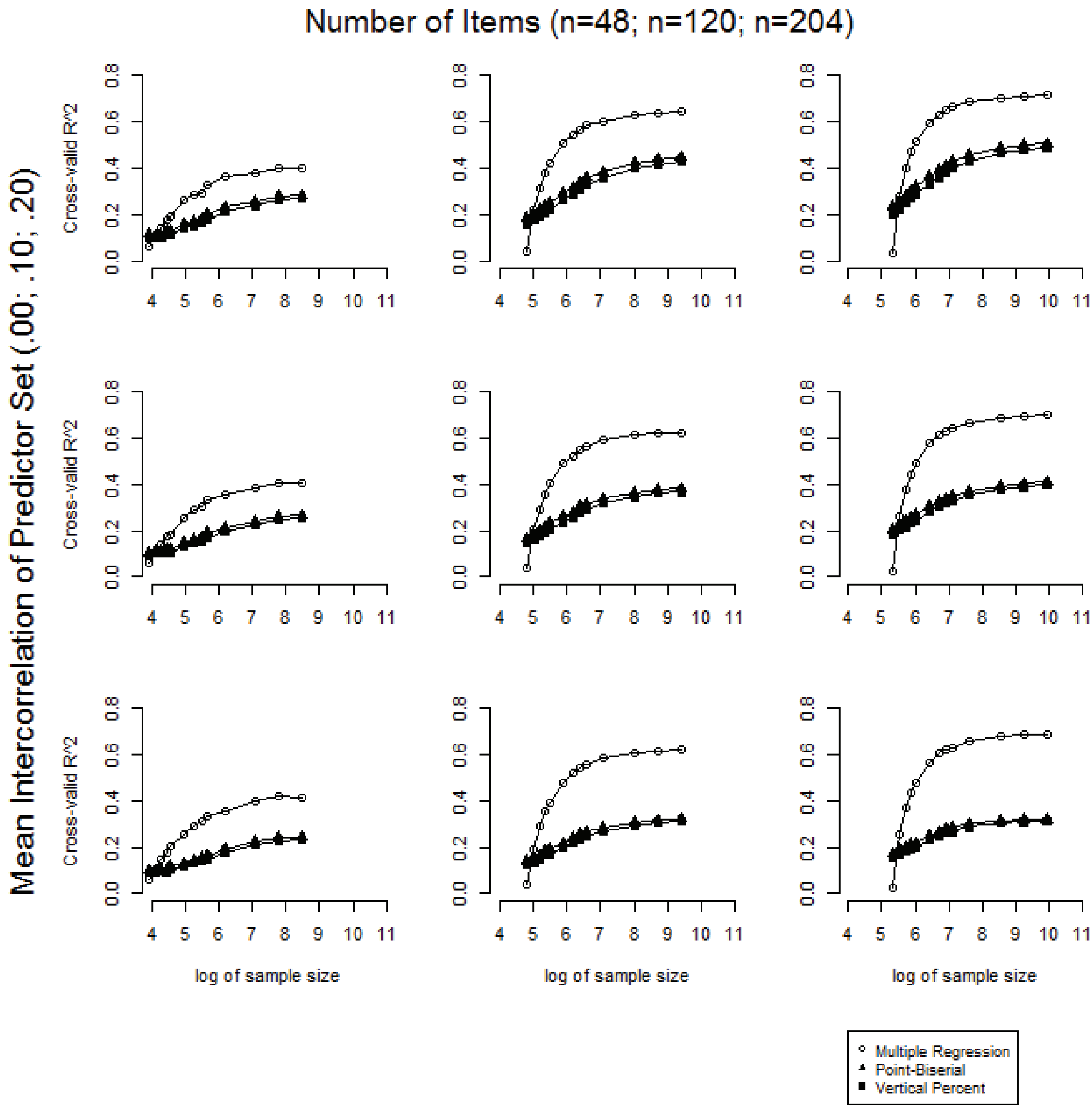

Figure 17. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

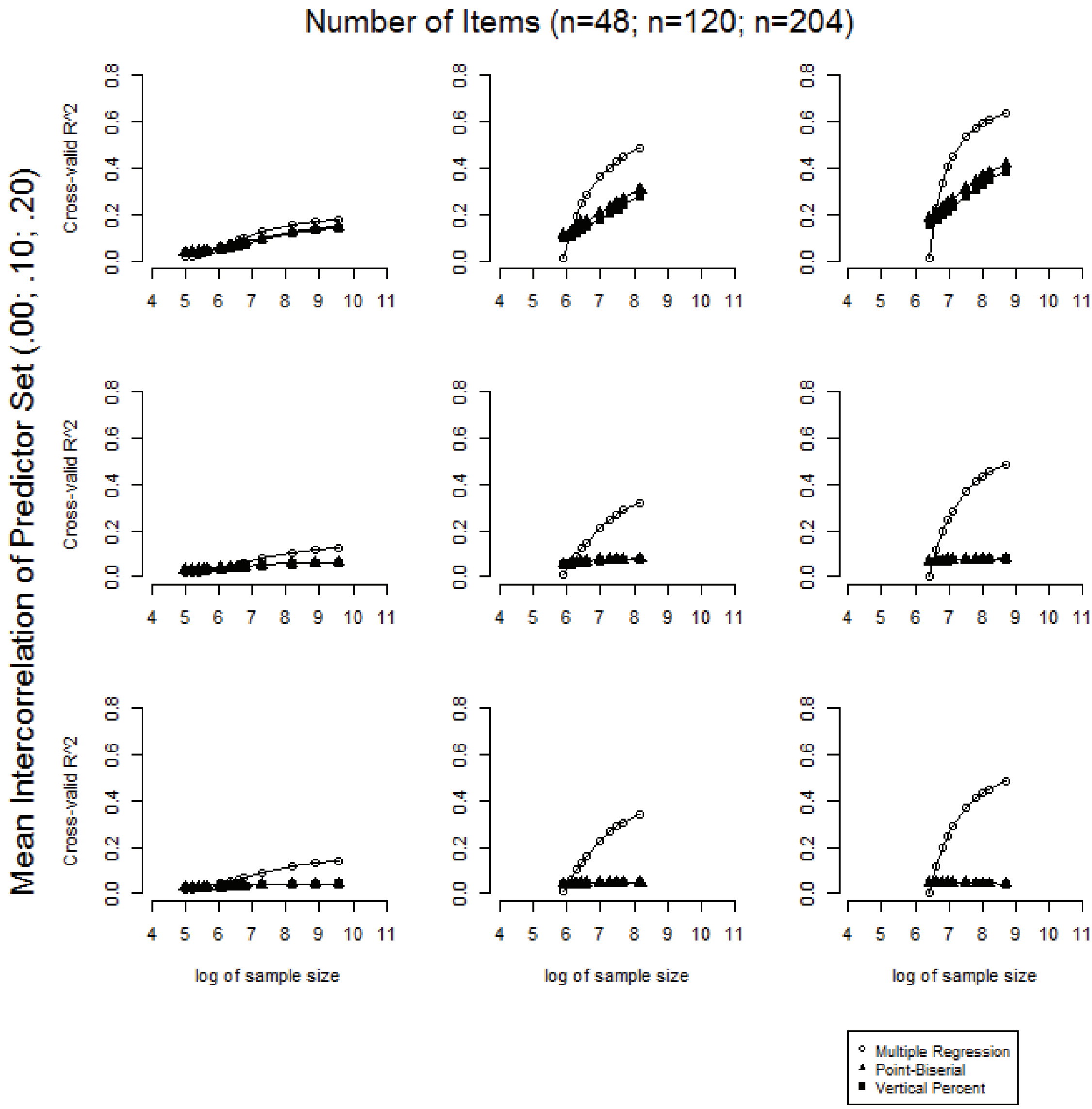

Figure 18. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

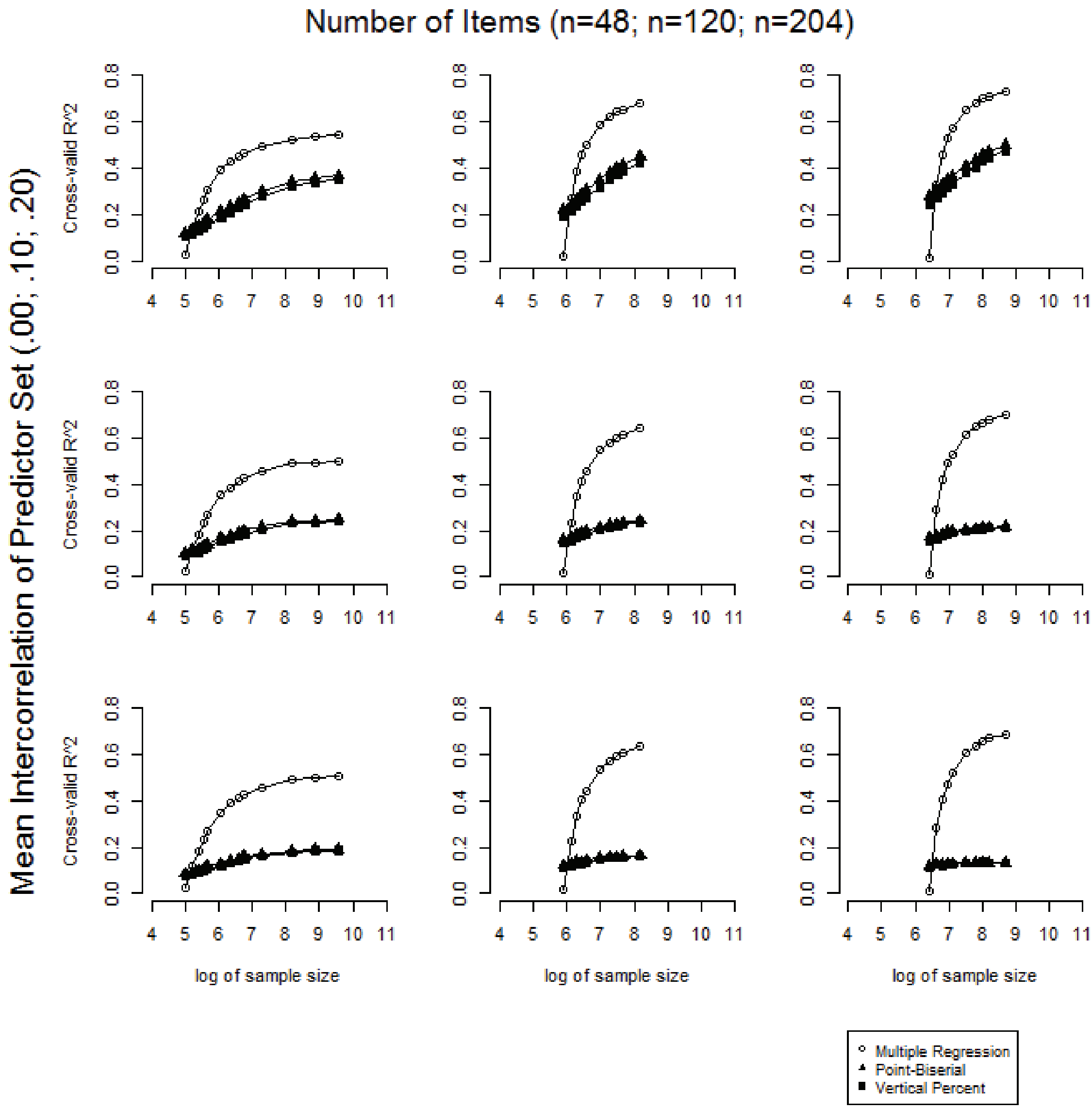

Figure 19. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

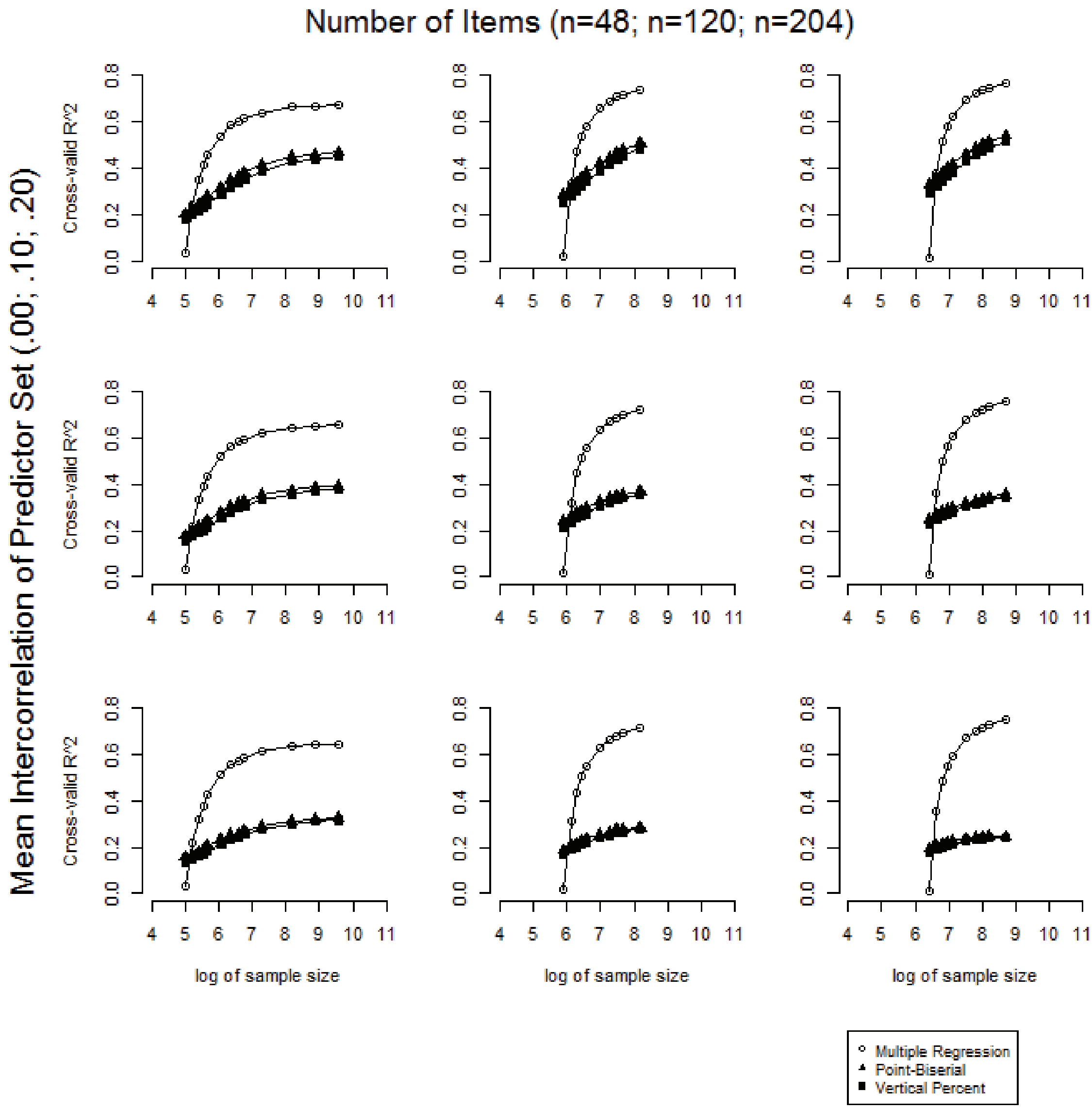

Figure 20. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

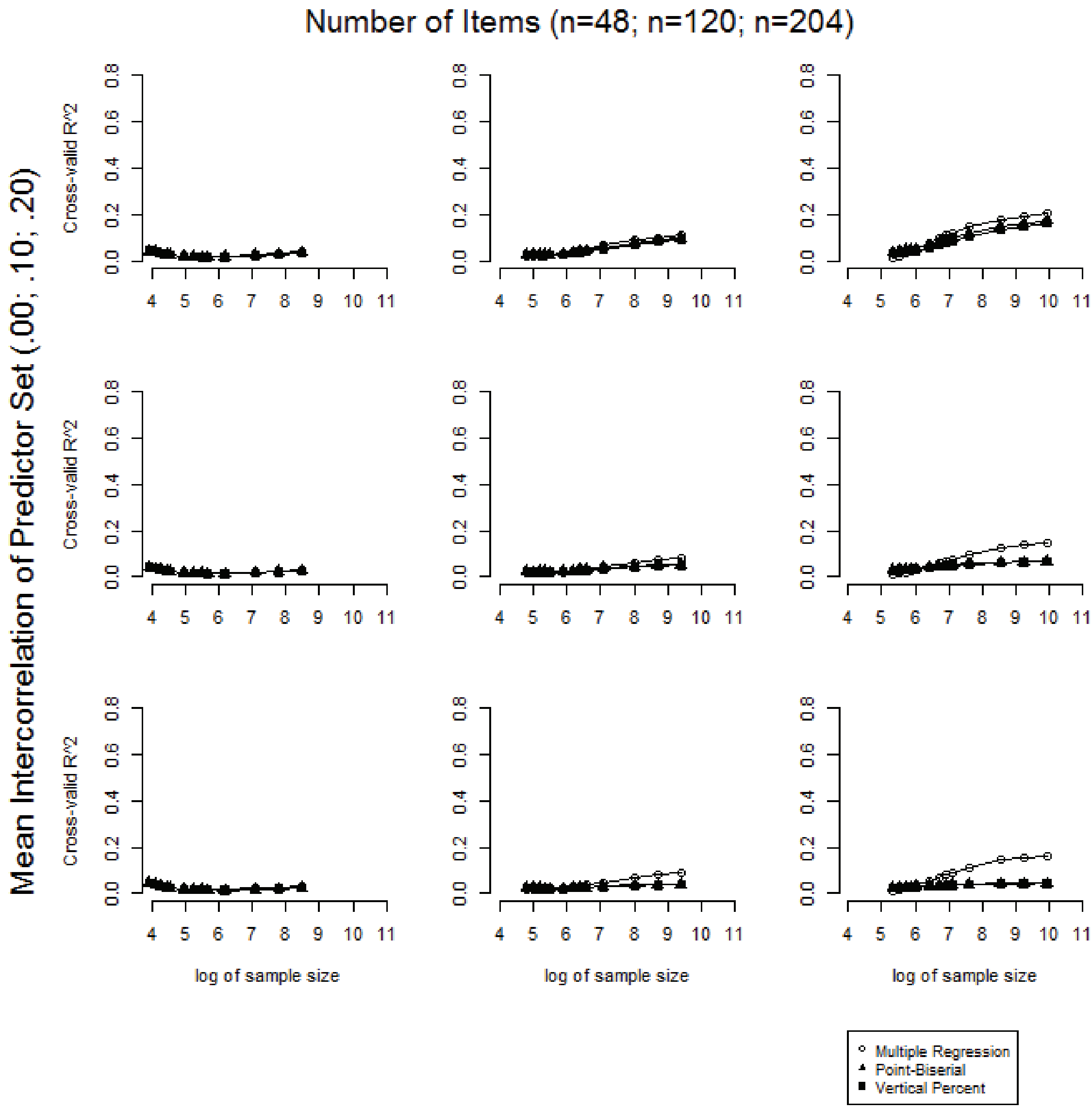

Figure 21. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

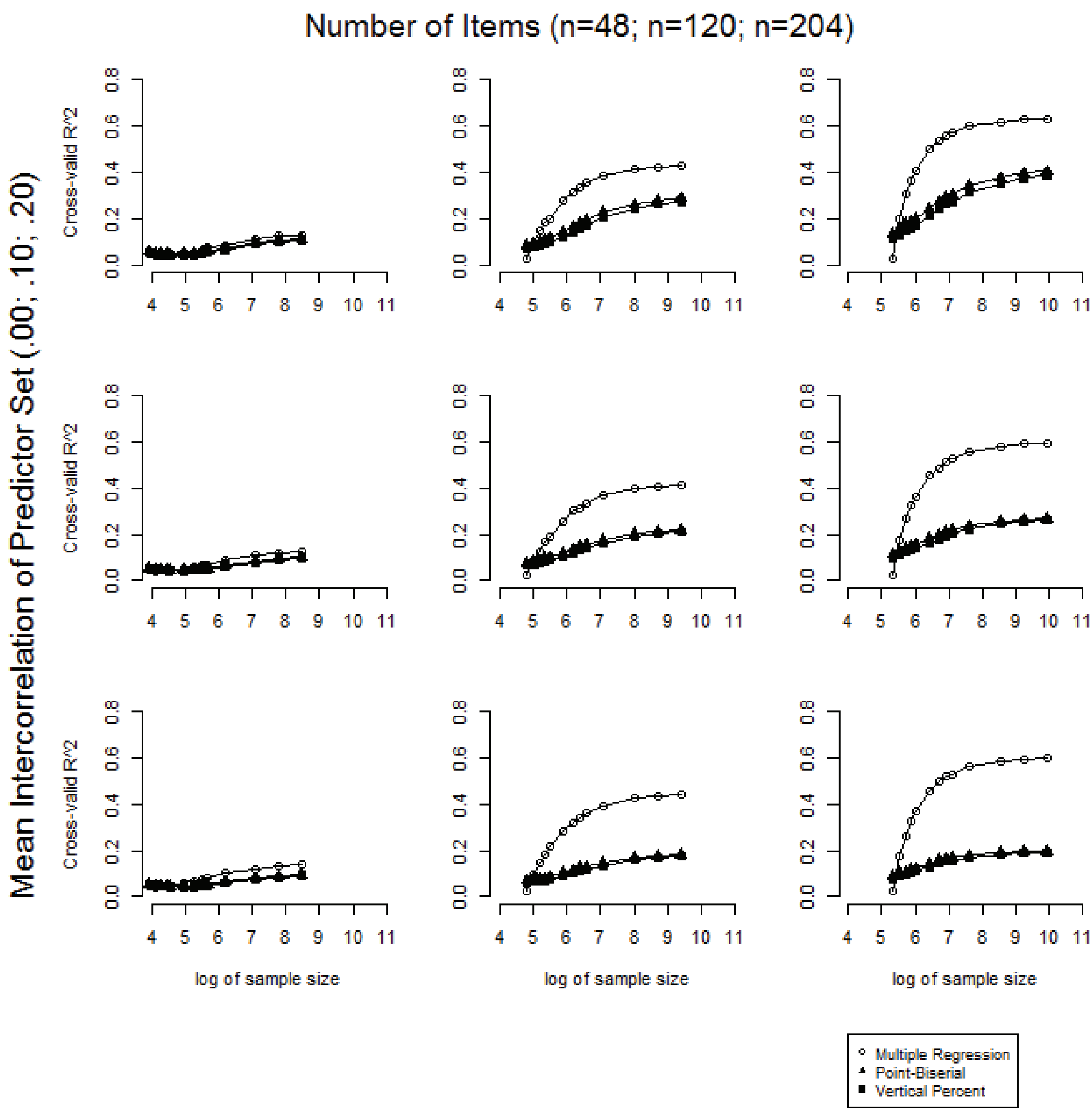

Figure 22. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

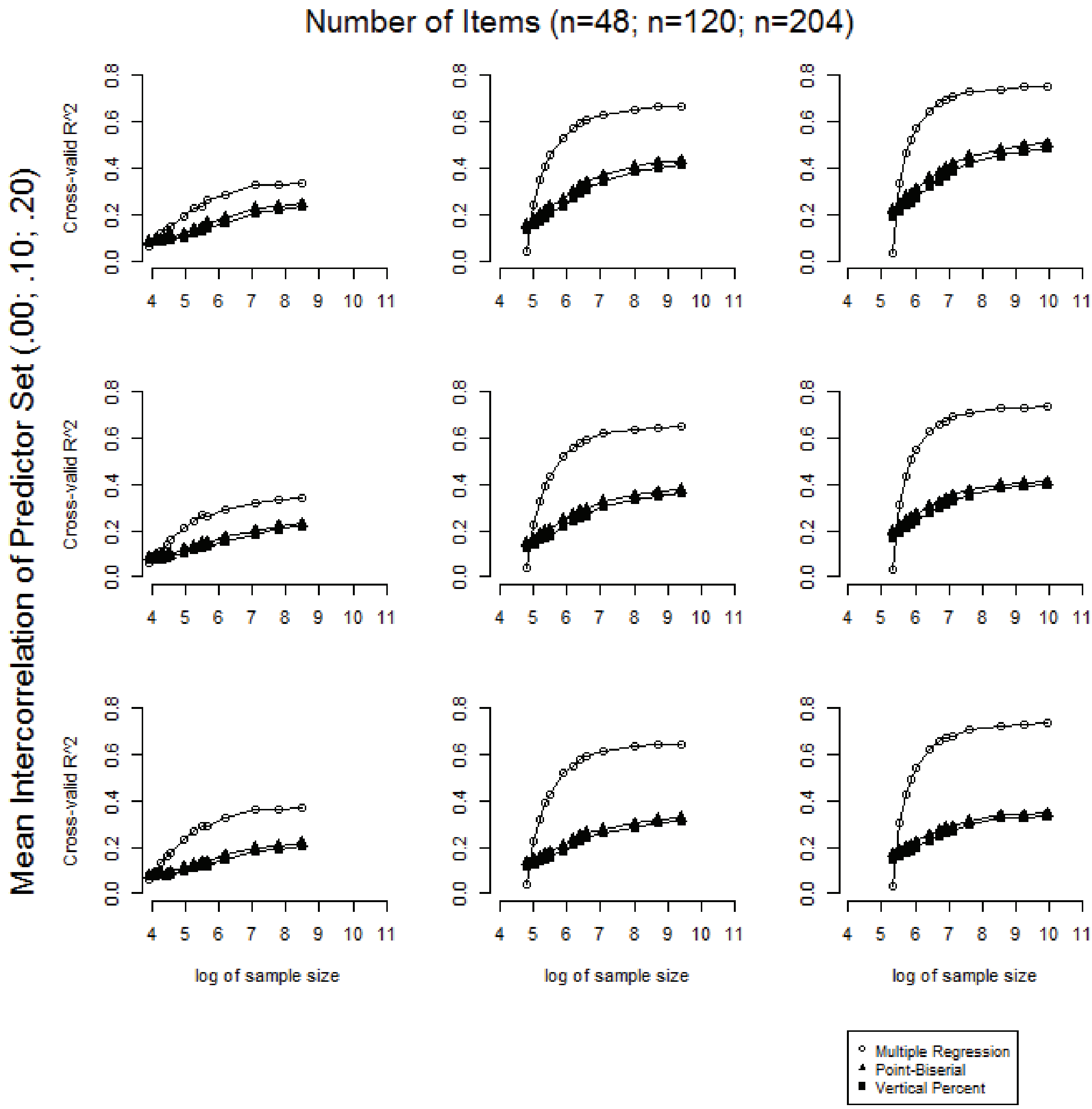

Figure 23. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

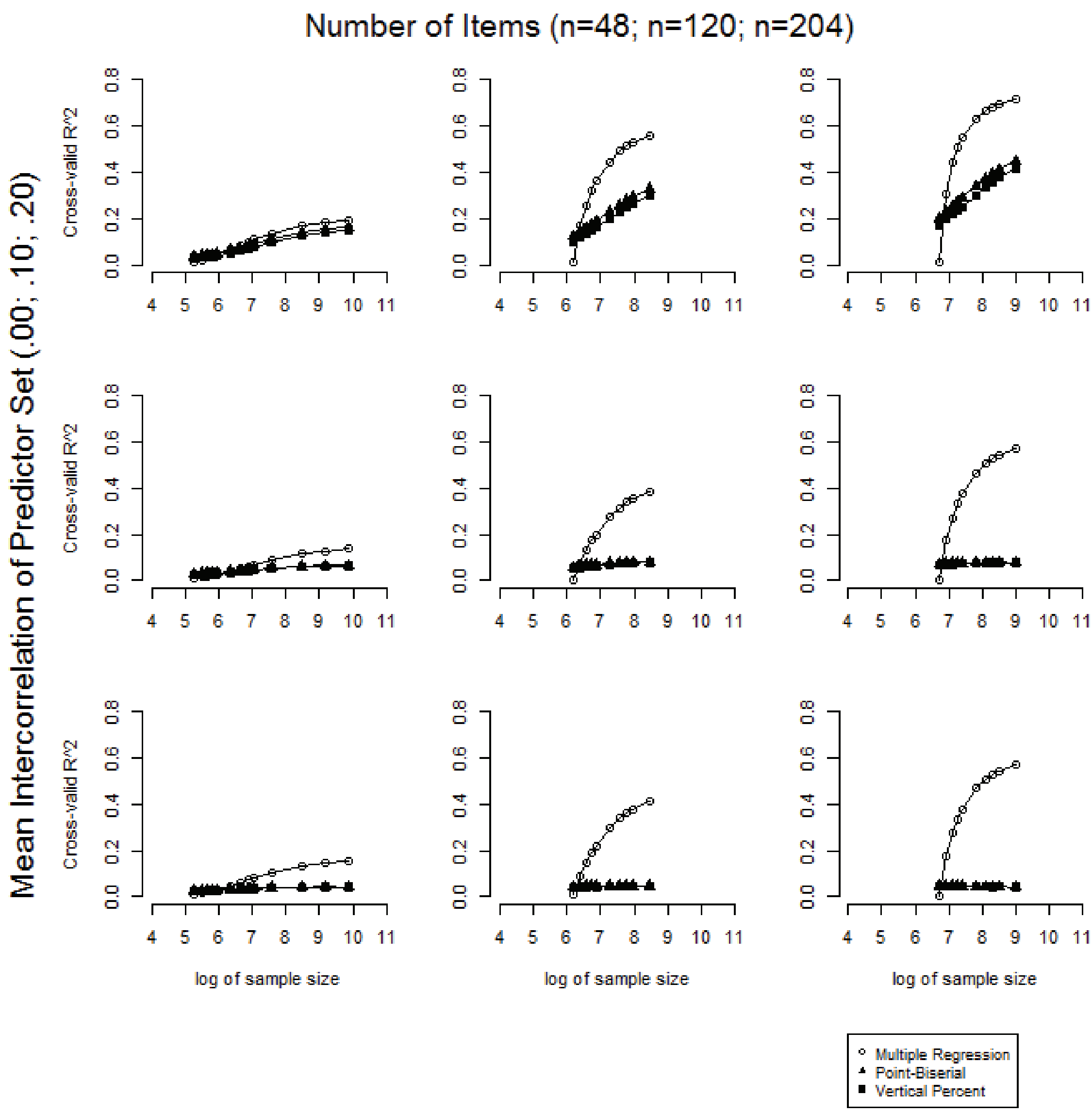

Figure 24. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

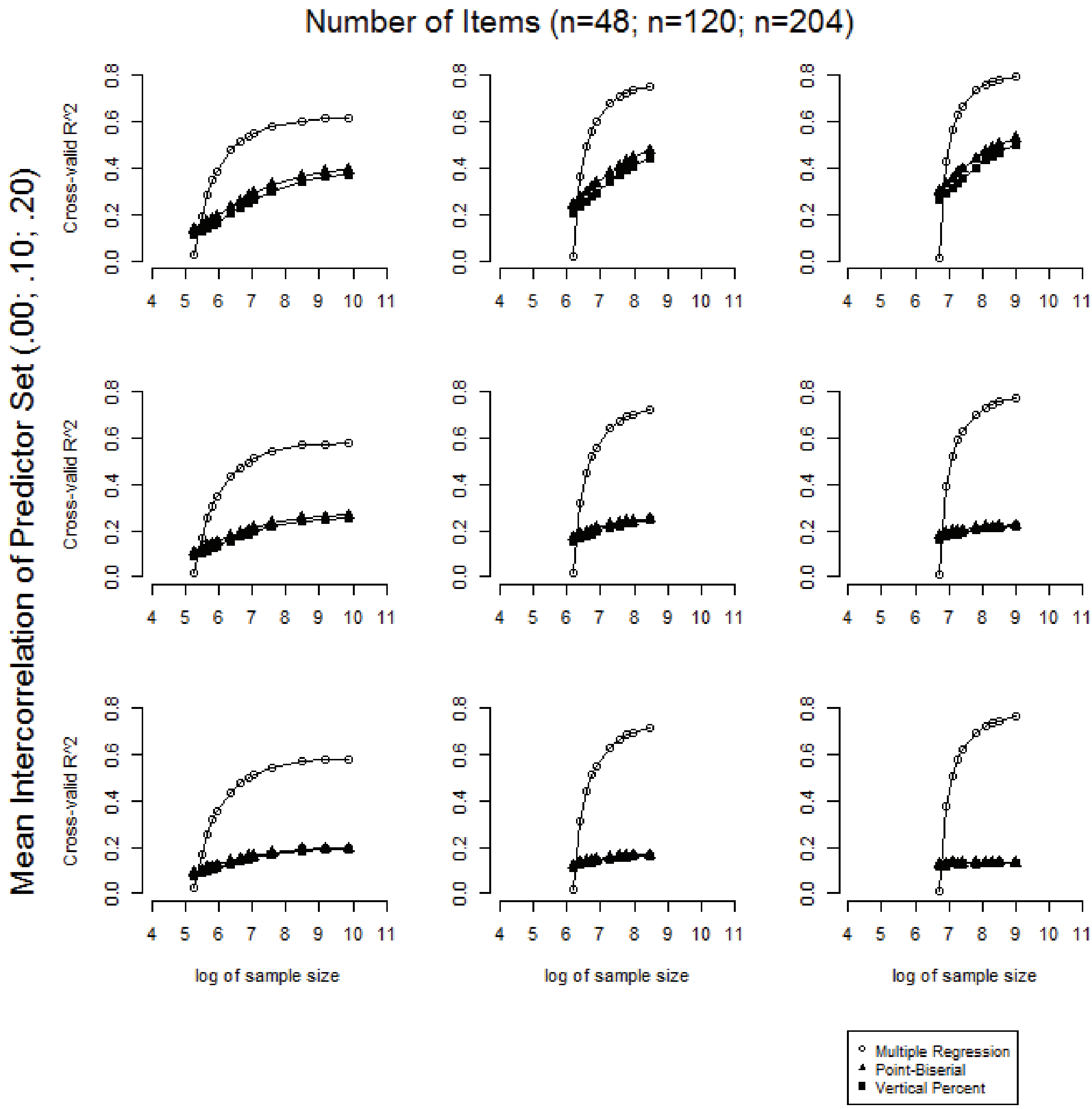

Figure 25. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

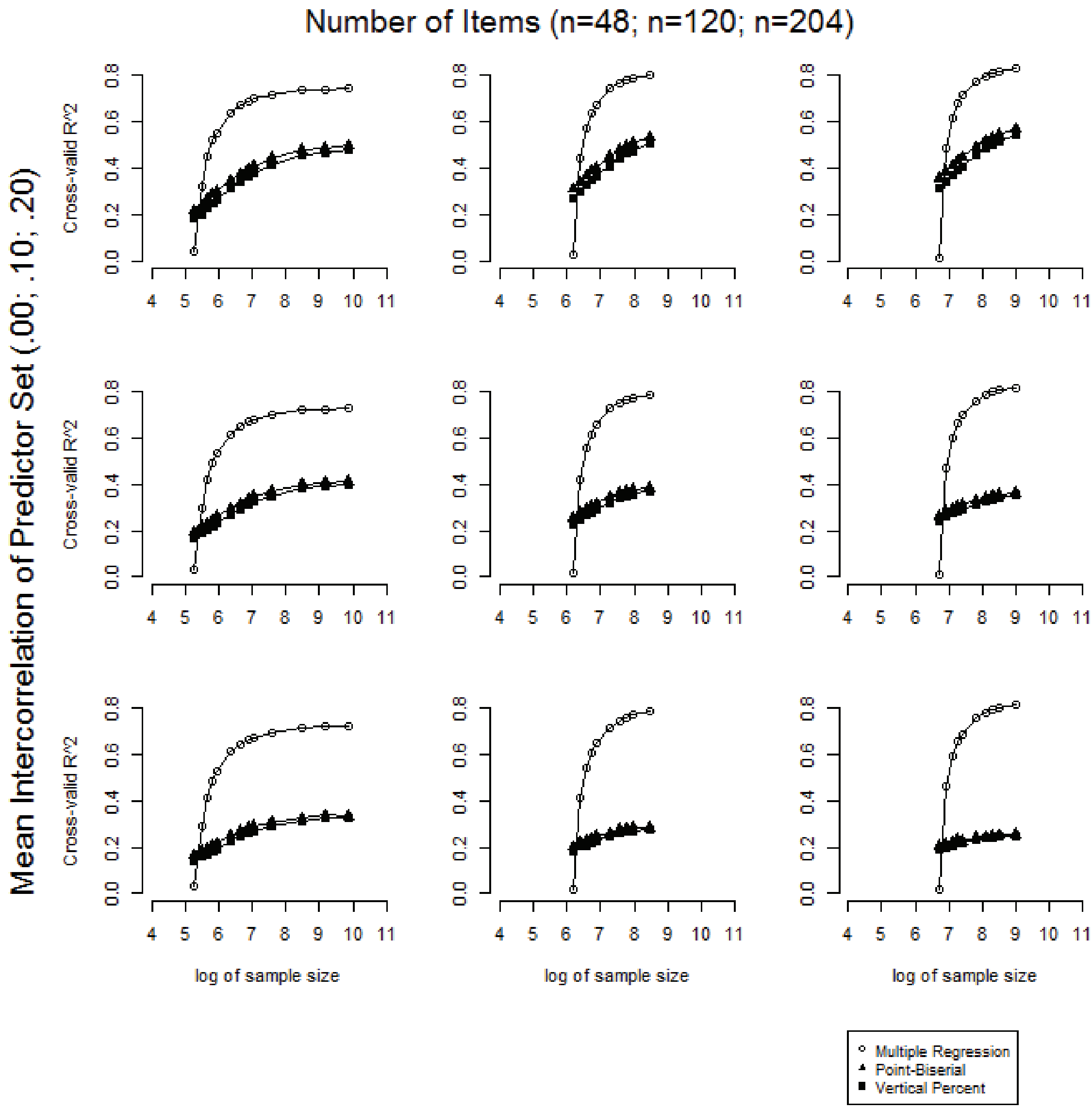

Figure 26. Effects of Varying the Mean Inter-Item Correlations and the Number of Items on the Performance of the Multiple Regression, Point-Biserial Correlation, and Vertical Percent Methods in Simulation Conditions where the Standard Deviation of the Validities and Inter-Item Correlations is .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options. The simulations all assume a mean continuous validity of .04. The condition labels are represented from top-left (number of items = 48, mean inter-item correlation = .00) to bottom-right (number of items = 204, mean inter-item correlation = .20). For interpretive purposes, log sample sizes of 4, 7, and 10 correspond to sample sizes of 55, 1,097, and 22,026, respectively.

APPENDICES

Appendix A: CHAID Algorithm Performance in the Student Descriptive Questionnaire Dataset at Varying Tree Depths and Node Splitting Alpha Levels

Table A1

CHAID Performance at an Alpha Level of .05

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	79	54	.141	.132
5	5	202	124	.157	.138
6	6	439	252	.173	.138
7	7	878	480	.194	.131
8	8	1,621	862	.221	.120
9	9	2,744	1,441	.257	.108
10	10	4,282	2,225	.298	.096
11	11	6,179	3,189	.343	.085
12	12	8,437	4,333	.392	.076
13	13	10,984	5,623	.439	.068
14	14	13,612	6,956	.483	.063

Note. All results are for a node splitting alpha level of .05. This means that the chi-square test in the algorithm needed to be significant at a .05 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A2

CHAID Performance at an Alpha Level of .01

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	79	54	.141	.132
5	5	198	122	.157	.138
6	6	419	242	.173	.138
7	7	812	447	.192	.132
8	8	1,413	758	.216	.123
9	9	2,194	1,163	.244	.113
10	10	3,102	1,628	.273	.104
11	11	4,100	2,136	.300	.097
12	12	5,112	2,653	.327	.090
13	13	6,004	3,102	.347	.086
14	14	6,796	3,505	.363	.084

Note. All results are for a node splitting alpha level of .01. This means that the chi-square test in the algorithm needed to be significant at a .01 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A3

CHAID Performance at an Alpha Level of .001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	77	53	.141	.132
5	5	175	110	.156	.138
6	6	319	191	.169	.140
7	7	502	290	.181	.139
8	8	696	394	.191	.136
9	9	854	477	.199	.132
10	10	992	549	.205	.130
11	11	1,069	589	.208	.129
12	12	1,101	605	.209	.128
13	13	1,113	611	.209	.128
14	14	1,119	614	.210	.128

Note. All results are for a node splitting alpha level of .001. This means that the chi-square test in the algorithm needed to be significant at a .001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A4

CHAID Performance at an Alpha Level of .0001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	73	50	.141	.132
5	5	154	98	.155	.139
6	6	258	158	.166	.142
7	7	365	218	.174	.143
8	8	437	258	.178	.142
9	9	470	276	.180	.142
10	10	488	285	.181	.142
11	11	492	287	.181	.142
12	12	494	288	.181	.142
13	12	494	288	.181	.142
14	12	494	288	.181	.142

Note. All results are for a node splitting alpha level of .0001. This means that the chi-square test in the algorithm needed to be significant at a .0001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A5

CHAID Performance at an Alpha Level of .00001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	69	48	.141	.132
5	5	132	87	.153	.140
6	6	210	133	.163	.143
7	7	270	167	.168	.146
8	8	303	186	.170	.146
9	9	325	198	.172	.145
10	10	327	199	.172	.145
11	10	327	199	.172	.145
12	10	327	199	.172	.145
13	10	327	199	.172	.145
14	10	327	199	.172	.145

Note. All results are for a node splitting alpha level of .00001. This means that the chi-square test in the algorithm needed to be significant at a .00001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A6

CHAID Performance at an Alpha Level of .000001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	67	47	.140	.132
5	5	126	84	.153	.140
6	6	191	123	.161	.144
7	7	224	143	.164	.146
8	8	244	155	.165	.146
9	9	254	161	.166	.146
10	10	256	162	.166	.146
11	10	256	162	.166	.146
12	10	256	162	.166	.146
13	10	256	162	.166	.146
14	10	256	162	.166	.146

Note. All results are for a node splitting alpha level of .000001. This means that the chi-square test in the algorithm needed to be significant at a .000001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A7

CHAID Performance at an Alpha Level of .0000001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	67	47	.140	.132
5	5	120	81	.152	.140
6	6	173	114	.159	.143
7	7	193	126	.161	.144
8	8	204	133	.162	.144
9	8	204	133	.162	.144
10	8	204	133	.162	.144
11	8	204	133	.162	.144
12	8	204	133	.162	.144
13	8	204	133	.162	.144
14	8	204	133	.162	.144

Note. All results are for a node splitting alpha level of .0000001. This means that the chi-square test in the algorithm needed to be significant at a .0000001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Table A8

CHAID Performance at an Alpha Level of .00000001

Max Depth	Actual Depth	Number of Nodes	Number of Terminal Nodes	R^2 Weight-Derivation	R^2 Cross-validation
2	2	8	5	.102	.097
3	3	25	17	.123	.119
4	4	67	47	.140	.132
5	5	112	77	.151	.139
6	6	149	100	.156	.143
7	7	165	110	.158	.144
8	8	171	114	.159	.144
9	8	171	114	.159	.144
10	8	171	114	.159	.144
11	8	171	114	.159	.144
12	8	171	114	.159	.144
13	8	171	114	.159	.144
14	8	171	114	.159	.144

Note. All results are for a node splitting alpha level of .00000001. This means that the chi-square test in the algorithm needed to be significant at a .00000001 level for a node to be split. Max Depth is a parameter set for the CHAID algorithm that specifies how many levels deep the tree can be built, whereas the Actual Depth is how deep the tree went in this sample. The Number of Nodes is how many nodes there are in the entire tree and the Number of Terminal Nodes is the number of endpoints that were not split into a deeper interaction by the algorithm. The weight-derivation sample was the sample of 100,898 used for every analysis in Study 1, whereas the cross-validation sample was the sample of 49,476.

Appendix B: Simulation Results for Items with Two Response Options

Table B1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.428 (.085)	.051 (.061)	.527 (.072)	.050 (.062)	.981 (.026)	.042 (.055)
1.25	60	48	.389 (.079)	.044 (.058)	.479 (.068)	.046 (.059)	.835 (.065)	.036 (.049)
1.50	72	48	.359 (.071)	.042 (.053)	.440 (.065)	.044 (.056)	.711 (.075)	.038 (.050)
1.75	84	48	.332 (.065)	.037 (.047)	.404 (.060)	.040 (.051)	.621 (.073)	.035 (.045)
2	96	48	.310 (.063)	.037 (.046)	.378 (.059)	.040 (.048)	.556 (.072)	.036 (.046)
3	144	48	.254 (.052)	.035 (.039)	.307 (.050)	.039 (.040)	.409 (.061)	.036 (.040)
4	192	48	.220 (.046)	.037 (.037)	.262 (.045)	.041 (.040)	.331 (.053)	.039 (.039)
5	**240**	**48**	**.197 (.043)**	**.040 (.035)**	**.233 (.042)**	**.045 (.037)**	**.284 (.049)**	**.045 (.040)**
6	288	48	.185 (.037)	.045 (.036)	.217 (.037)	.050 (.038)	.258 (.042)	.051 (.039)
10	480	48	.152 (.031)	.052 (.032)	.174 (.031)	.058 (.033)	.197 (.035)	.062 (.034)
25	1,200	48	.120 (.025)	.070 (.027)	.132 (.025)	.076 (.028)	.144 (.028)	.082 (.030)
50	2,400	48	.109 (.022)	.082 (.024)	.118 (.022)	.088 (.025)	.126 (.024)	.094 (.026)
100	4,800	48	.105 (.021)	.090 (.023)	.111 (.022)	.096 (.024)	.118 (.023)	.101 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.462 (.084)	.086 (.099)	.552 (.073)	.094 (.105)	.985 (.021)	.055 (.072)
1.25	60	48	.432 (.079)	.081 (.086)	.516 (.068)	.087 (.089)	.867 (.055)	.061 (.075)
1.50	72	48	.412 (.074)	.084 (.080)	.487 (.067)	.095 (.087)	.769 (.064)	.077 (.083)
1.75	84	48	.388 (.077)	.094 (.084)	.456 (.070)	.102 (.086)	.693 (.068)	.087 (.080)
2	96	48	.373 (.069)	.091 (.080)	.437 (.062)	.104 (.084)	.644 (.070)	.100 (.083)
3	**144**	**48**	**.332 (.062)**	**.107 (.076)**	**.383 (.058)**	**.122 (.080)**	**.520 (.067)**	**.141 (.083)**
4	192	48	.310 (.056)	.121 (.071)	.353 (.054)	.137 (.075)	.462 (.061)	.164 (.084)
5	240	48	.295 (.053)	.132 (.070)	.331 (.052)	.147 (.073)	.424 (.059)	.177 (.079)
6	288	48	.287 (.053)	.145 (.067)	.321 (.051)	.159 (.067)	.403 (.057)	.194 (.074)
10	480	48	.268 (.048)	.168 (.061)	.294 (.048)	.183 (.062)	.355 (.054)	.223 (.067)
25	1,200	48	.249 (.045)	.204 (.053)	.266 (.045)	.217 (.053)	.310 (.050)	.255 (.057)
50	2,400	48	.246 (.044)	.223 (.048)	.259 (.044)	.235 (.049)	.298 (.048)	.272 (.052)
100	4,800	48	.243 (.045)	.230 (.046)	.254 (.045)	.241 (.047)	.290 (.049)	.276 (.050)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.485 (.084)	.121 (.118)	.568 (.069)	.136 (.124)	.987 (.019)	.058 (.076)
1.25	60	48	.464 (.079)	.121 (.108)	.543 (.068)	.136 (.113)	.883 (.050)	.091 (.093)
1.50	72	48	.441 (.075)	.123 (.099)	.511 (.065)	.138 (.103)	.796 (.059)	.108 (.096)
1.75	84	48	.429 (.075)	.133 (.099)	.491 (.068)	.151 (.106)	.733 (.064)	.138 (.103)
2	96	48	.412 (.070)	.147 (.100)	.473 (.064)	.166 (.103)	.691 (.063)	.162 (.104)
3	**144**	**48**	**.380 (.064)**	**.169 (.094)**	**.426 (.059)**	**.191 (.097)**	**.583 (.063)**	**.222 (.104)**
4	192	48	.368 (.062)	.195 (.092)	.408 (.059)	.216 (.095)	.536 (.062)	.256 (.098)
5	240	48	.355 (.059)	.202 (.082)	.391 (.056)	.221 (.082)	.501 (.059)	.272 (.089)
6	288	48	.350 (.056)	.221 (.078)	.383 (.054)	.239 (.079)	.482 (.058)	.291 (.084)
10	480	48	.334 (.053)	.246 (.069)	.360 (.052)	.265 (.069)	.439 (.055)	.323 (.072)
25	1,200	48	.328 (.049)	.289 (.059)	.345 (.049)	.304 (.059)	.406 (.050)	.357 (.059)
50	2,400	48	.325 (.051)	.304 (.056)	.339 (.050)	.317 (.056)	.394 (.052)	.368 (.056)
100	4,800	48	.322 (.050)	.312 (.052)	.335 (.050)	.325 (.052)	.385 (.051)	.374 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.395 (.083)	.051 (.064)	.490 (.076)	.051 (.066)	.983 (.023)	.039 (.051)
1.25	60	48	.362 (.077)	.043 (.058)	.442 (.071)	.045 (.059)	.827 (.066)	.039 (.049)
1.50	72	48	.328 (.073)	.039 (.051)	.400 (.067)	.041 (.054)	.705 (.074)	.035 (.045)
1.75	84	48	.301 (.066)	.037 (.046)	.365 (.062)	.038 (.046)	.615 (.075)	.032 (.042)
2	96	48	.280 (.062)	.032 (.040)	.337 (.060)	.036 (.044)	.546 (.073)	.028 (.037)
3	144	48	.215 (.051)	.032 (.037)	.256 (.050)	.034 (.038)	.389 (.061)	.026 (.032)
4	192	48	.182 (.042)	.031 (.034)	.214 (.041)	.034 (.035)	.311 (.050)	.029 (.032)
5	240	48	.161 (.038)	.033 (.033)	.186 (.038)	.036 (.034)	.266 (.046)	.031 (.031)
6	288	48	.144 (.033)	.032 (.030)	.166 (.033)	.036 (.031)	.234 (.040)	.032 (.031)
10	**480**	**48**	**.111 (.025)**	**.038 (.025)**	**.123 (.025)**	**.041 (.025)**	**.173 (.030)**	**.042 (.028)**
25	1,200	48	.076 (.016)	.044 (.019)	.081 (.016)	.046 (.020)	.118 (.021)	.057 (.024)
50	2,400	48	.064 (.013)	.048 (.017)	.067 (.013)	.050 (.017)	.099 (.018)	.066 (.020)
100	4,800	48	.059 (.013)	.050 (.014)	.061 (.013)	.052 (.014)	.091 (.018)	.074 (.019)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.433 (.088)	.078 (.089)	.519 (.076)	.084 (.094)	.984 (.024)	.045 (.059)
1.25	60	48	.407 (.080)	.079 (.085)	.485 (.071)	.085 (.088)	.862 (.056)	.058 (.071)
1.50	72	48	.383 (.077)	.075 (.078)	.451 (.071)	.087 (.085)	.763 (.066)	.067 (.073)
1.75	84	48	.358 (.072)	.080 (.075)	.419 (.067)	.090 (.081)	.684 (.071)	.083 (.082)
2	96	48	.348 (.071)	.081 (.075)	.401 (.066)	.089 (.079)	.635 (.069)	.087 (.078)
3	**144**	**48**	**.298 (.060)**	**.093 (.071)**	**.341 (.057)**	**.103 (.075)**	**.505 (.064)**	**.118 (.082)**
4	192	48	.274 (.053)	.105 (.068)	.309 (.053)	.116 (.071)	.445 (.061)	.141 (.077)
5	240	48	.256 (.052)	.110 (.061)	.287 (.052)	.122 (.064)	.403 (.058)	.155 (.075)
6	288	48	.248 (.049)	.118 (.058)	.275 (.050)	.129 (.060)	.383 (.056)	.172 (.069)
10	480	48	.222 (.048)	.139 (.056)	.241 (.047)	.151 (.057)	.331 (.053)	.201 (.065)
25	1,200	48	.199 (.042)	.162 (.048)	.210 (.042)	.169 (.048)	.285 (.047)	.229 (.054)
50	2,400	48	.192 (.040)	.170 (.044)	.200 (.040)	.177 (.044)	.272 (.045)	.243 (.049)
100	4,800	48	.190 (.041)	.181 (.043)	.196 (.041)	.186 (.043)	.268 (.046)	.254 (.049)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.468 (.086)	.114 (.107)	.549 (.074)	.126 (.114)	.986 (.020)	.059 (.075)
1.25	60	48	.442 (.082)	.120 (.108)	.516 (.072)	.132 (.115)	.880 (.049)	.080 (.090)
1.50	72	48	.424 (.077)	.121 (.103)	.490 (.068)	.138 (.107)	.795 (.058)	.109 (.098)
1.75	84	48	.406 (.077)	.127 (.102)	.465 (.069)	.140 (.106)	.729 (.066)	.133 (.106)
2	**96**	**48**	**.394 (.071)**	**.132 (.095)**	**.449 (.065)**	**.147 (.098)**	**.687 (.065)**	**.147 (.103)**
3	144	48	.357 (.064)	.154 (.089)	.399 (.061)	.171 (.093)	.574 (.063)	.201 (.097)
4	192	48	.343 (.063)	.175 (.083)	.380 (.059)	.192 (.085)	.527 (.062)	.241 (.091)
5	240	48	.330 (.057)	.184 (.080)	.363 (.055)	.202 (.081)	.494 (.059)	.259 (.088)
6	288	48	.321 (.056)	.195 (.076)	.349 (.054)	.212 (.077)	.469 (.058)	.273 (.082)
10	480	48	.301 (.053)	.219 (.066)	.323 (.052)	.236 (.066)	.424 (.053)	.305 (.071)
25	1,200	48	.290 (.052)	.255 (.059)	.304 (.051)	.266 (.059)	.390 (.052)	.341 (.060)
50	2,400	48	.285 (.049)	.267 (.054)	.296 (.049)	.277 (.054)	.377 (.051)	.352 (.056)
100	4,800	48	.285 (.049)	.277 (.051)	.295 (.049)	.286 (.052)	.372 (.049)	.361 (.051)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.357 (.094)	.050 (.066)	.436 (.093)	.051 (.065)	.982 (.024)	.041 (.055)
1.25	60	48	.319 (.085)	.043 (.056)	.392 (.088)	.044 (.058)	.827 (.065)	.039 (.054)
1.50	72	48	.290 (.080)	.036 (.048)	.350 (.082)	.038 (.049)	.703 (.074)	.033 (.044)
1.75	84	48	.258 (.072)	.033 (.044)	.314 (.076)	.034 (.045)	.608 (.073)	.028 (.037)
2	96	48	.239 (.069)	.032 (.040)	.292 (.072)	.033 (.040)	.543 (.073)	.029 (.037)
3	144	48	.184 (.052)	.029 (.034)	.219 (.057)	.031 (.035)	.389 (.058)	.026 (.032)
4	192	48	.152 (.046)	.027 (.030)	.178 (.051)	.029 (.030)	.310 (.049)	.028 (.032)
5	**240**	**48**	**.133 (.040)**	**.028 (.029)**	**.152 (.043)**	**.030 (.030)**	**.264 (.042)**	**.030 (.030)**
6	288	48	.118 (.036)	.026 (.025)	.133 (.039)	.028 (.027)	.235 (.042)	.032 (.029)
10	480	48	.089 (.029)	.030 (.024)	.098 (.030)	.033 (.025)	.173 (.032)	.041 (.029)
25	1,200	48	.058 (.017)	.034 (.018)	.061 (.018)	.036 (.018)	.116 (.021)	.056 (.024)
50	2,400	48	.047 (.013)	.036 (.015)	.048 (.012)	.036 (.015)	.100 (.020)	.067 (.021)
100	4,800	48	.042 (.011)	.037 (.013)	.043 (.010)	.037 (.012)	.091 (.018)	.074 (.019)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.395 (.095)	.072 (.084)	.476 (.090)	.074 (.085)	.983 (.022)	.046 (.063)
1.25	60	48	.362 (.089)	.065 (.071)	.435 (.086)	.070 (.077)	.858 (.060)	.051 (.062)
1.50	72	48	.339 (.083)	.068 (.071)	.401 (.082)	.074 (.074)	.754 (.066)	.066 (.075)
1.75	**84**	**48**	**.320 (.080)**	**.071 (.073)**	**.375 (.083)**	**.079 (.080)**	**.680 (.071)**	**.079 (.080)**
2	96	48	.300 (.075)	.069 (.067)	.350 (.076)	.076 (.070)	.620 (.071)	.086 (.079)
3	144	48	.263 (.071)	.081 (.066)	.299 (.071)	.087 (.066)	.503 (.065)	.115 (.077)
4	192	48	.233 (.061)	.082 (.057)	.263 (.066)	.090 (.060)	.435 (.061)	.133 (.074)
5	240	48	.221 (.060)	.092 (.059)	.247 (.062)	.100 (.061)	.401 (.059)	.151 (.071)
6	288	48	.212 (.059)	.101 (.060)	.235 (.061)	.110 (.062)	.375 (.056)	.165 (.071)
10	480	48	.191 (.052)	.116 (.054)	.205 (.053)	.124 (.055)	.328 (.051)	.193 (.060)
25	1,200	48	.166 (.048)	.135 (.051)	.173 (.049)	.140 (.052)	.282 (.046)	.228 (.053)
50	2,400	48	.155 (.044)	.139 (.046)	.161 (.045)	.144 (.047)	.268 (.046)	.241 (.049)
100	4,800	48	.154 (.042)	.146 (.043)	.158 (.042)	.149 (.044)	.261 (.044)	.248 (.046)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	48	.432 (.093)	.097 (.101)	.511 (.086)	.108 (.107)	.987 (.019)	.056 (.070)
1.25	60	48	.410 (.092)	.105 (.098)	.479 (.086)	.114 (.103)	.876 (.051)	.073 (.081)
1.50	72	48	.382 (.085)	.106 (.094)	.446 (.083)	.120 (.099)	.790 (.061)	.105 (.097)
1.75	**84**	**48**	**.366 (.082)**	**.112 (.094)**	**.426 (.078)**	**.124 (.098)**	**.725 (.062)**	**.125 (.096)**
2	96	48	.354 (.082)	.111 (.087)	.407 (.077)	.127 (.093)	.682 (.065)	.138 (.101)
3	144	48	.318 (.072)	.130 (.081)	.358 (.070)	.143 (.083)	.567 (.062)	.190 (.096)
4	192	48	.299 (.068)	.147 (.081)	.332 (.068)	.161 (.084)	.514 (.059)	.228 (.090)
5	240	48	.292 (.068)	.160 (.077)	.320 (.069)	.174 (.079)	.480 (.061)	.251 (.085)
6	288	48	.283 (.066)	.169 (.075)	.311 (.065)	.184 (.077)	.462 (.059)	.263 (.081)
10	480	48	.266 (.062)	.191 (.073)	.285 (.063)	.204 (.073)	.417 (.054)	.296 (.073)
25	1,200	48	.253 (.058)	.220 (.063)	.264 (.059)	.229 (.064)	.379 (.052)	.330 (.059)
50	2,400	48	.246 (.057)	.228 (.058)	.254 (.057)	.235 (.059)	.365 (.048)	.339 (.052)
100	4,800	48	.243 (.057)	.234 (.059)	.251 (.058)	.241 (.060)	.359 (.049)	.346 (.052)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.460 (.051)	.057 (.056)	.551 (.046)	.064 (.059)	.994 (.008)	.020 (.027)
1.25	150	120	.434 (.051)	.062 (.052)	.516 (.043)	.070 (.057)	.861 (.036)	.039 (.042)
1.50	180	120	.410 (.048)	.069 (.054)	.483 (.043)	.079 (.057)	.762 (.041)	.057 (.048)
1.75	210	120	.388 (.046)	.074 (.049)	.457 (.042)	.084 (.052)	.692 (.044)	.073 (.051)
2	240	120	.374 (.042)	.078 (.048)	.437 (.039)	.090 (.051)	.641 (.043)	.084 (.053)
3	**360**	**120**	**.333 (.040)**	**.097 (.048)**	**.383 (.037)**	**.112 (.050)**	**.517 (.042)**	**.126 (.056)**
4	480	120	.310 (.037)	.114 (.046)	.353 (.036)	.130 (.050)	.459 (.041)	.155 (.056)
5	600	120	.295 (.033)	.124 (.041)	.333 (.032)	.140 (.043)	.421 (.037)	.170 (.048)
6	720	120	.287 (.035)	.137 (.044)	.321 (.034)	.152 (.045)	.398 (.039)	.185 (.050)
10	1,200	120	.266 (.031)	.161 (.038)	.292 (.031)	.177 (.039)	.349 (.034)	.215 (.043)
25	3,000	120	.251 (.030)	.202 (.035)	.268 (.030)	.216 (.036)	.308 (.034)	.252 (.040)
50	6,000	120	.244 (.029)	.219 (.032)	.258 (.029)	.231 (.032)	.292 (.032)	.264 (.035)
100	12,000	120	.241 (.029)	.228 (.030)	.253 (.030)	.239 (.031)	.284 (.033)	.270 (.034)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.488 (.053)	.108 (.077)	.571 (.045)	.124 (.080)	.995 (.007)	.024 (.033)
1.25	150	120	.463 (.053)	.122 (.072)	.539 (.045)	.141 (.076)	.884 (.032)	.073 (.059)
1.50	180	120	.445 (.049)	.132 (.068)	.513 (.043)	.152 (.074)	.803 (.036)	.118 (.070)
1.75	210	120	.435 (.047)	.142 (.070)	.497 (.042)	.162 (.072)	.748 (.039)	.144 (.072)
2	**240**	**120**	**.424 (.047)**	**.154 (.066)**	**.483 (.042)**	**.175 (.067)**	**.706 (.040)**	**.177 (.069)**
3	360	120	.397 (.042)	.187 (.064)	.444 (.039)	.210 (.065)	.606 (.039)	.241 (.069)
4	480	120	.380 (.039)	.207 (.056)	.421 (.038)	.230 (.058)	.554 (.039)	.278 (.062)
5	600	120	.372 (.039)	.224 (.054)	.408 (.037)	.247 (.053)	.526 (.038)	.301 (.056)
6	720	120	.367 (.037)	.240 (.051)	.400 (.035)	.263 (.051)	.506 (.037)	.322 (.051)
10	1,200	120	.356 (.036)	.272 (.048)	.382 (.035)	.293 (.048)	.467 (.036)	.353 (.047)
25	3,000	120	.350 (.033)	.312 (.038)	.368 (.032)	.328 (.037)	.433 (.032)	.387 (.037)
50	6,000	120	.348 (.032)	.328 (.035)	.362 (.032)	.342 (.035)	.421 (.032)	.398 (.035)
100	12,000	120	.347 (.032)	.337 (.033)	.360 (.031)	.350 (.033)	.415 (.031)	.404 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.507 (.054)	.140 (.086)	.585 (.045)	.160 (.090)	.996 (.006)	.027 (.036)
1.25	150	120	.488 (.053)	.161 (.082)	.558 (.045)	.182 (.084)	.896 (.028)	.098 (.070)
1.50	180	120	.473 (.051)	.174 (.078)	.535 (.043)	.197 (.081)	.825 (.033)	.157 (.080)
1.75	210	120	.459 (.048)	.187 (.077)	.518 (.043)	.211 (.079)	.771 (.037)	.193 (.078)
2	**240**	**120**	**.451 (.046)**	**.200 (.074)**	**.506 (.042)**	**.225 (.075)**	**.735 (.037)**	**.230 (.080)**
3	360	120	.431 (.044)	.235 (.070)	.475 (.039)	.261 (.070)	.646 (.038)	.305 (.069)
4	480	120	.417 (.042)	.255 (.062)	.455 (.039)	.280 (.063)	.598 (.037)	.338 (.062)
5	600	120	.413 (.040)	.278 (.058)	.447 (.037)	.304 (.059)	.575 (.036)	.366 (.057)
6	720	120	.409 (.038)	.294 (.053)	.440 (.036)	.317 (.053)	.557 (.037)	.383 (.053)
10	1,200	120	.401 (.036)	.325 (.047)	.425 (.035)	.345 (.048)	.521 (.034)	.414 (.046)
25	3,000	120	.395 (.034)	.364 (.040)	.413 (.033)	.381 (.039)	.489 (.032)	.448 (.037)
50	6,000	120	.395 (.034)	.380 (.037)	.410 (.033)	.394 (.036)	.479 (.033)	.459 (.035)
100	12,000	120	.394 (.034)	.386 (.035)	.408 (.033)	.399 (.034)	.473 (.032)	.462 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.373 (.071)	.044 (.047)	.440 (.075)	.047 (.047)	.994 (.009)	.019 (.027)
1.25	150	120	.338 (.064)	.043 (.045)	.394 (.069)	.048 (.048)	.841 (.040)	.026 (.031)
1.50	180	120	.303 (.064)	.046 (.043)	.350 (.070)	.050 (.045)	.733 (.046)	.033 (.035)
1.75	210	120	.276 (.059)	.047 (.039)	.315 (.063)	.051 (.041)	.656 (.045)	.040 (.038)
2	240	120	.260 (.056)	.048 (.039)	.296 (.062)	.053 (.040)	.597 (.044)	.045 (.039)
3	**360**	**120**	**.206 (.046)**	**.054 (.033)**	**.230 (.051)**	**.058 (.034)**	**.462 (.041)**	**.068 (.039)**
4	480	120	.179 (.042)	.059 (.033)	.193 (.044)	.063 (.034)	.396 (.038)	.086 (.040)
5	600	120	.160 (.037)	.060 (.030)	.171 (.037)	.064 (.031)	.356 (.035)	.097 (.038)
6	720	120	.147 (.033)	.065 (.031)	.155 (.033)	.067 (.031)	.328 (.035)	.110 (.038)
10	1,200	120	.117 (.024)	.066 (.025)	.121 (.022)	.069 (.025)	.274 (.030)	.134 (.034)
25	3,000	120	.091 (.016)	.071 (.019)	.093 (.015)	.073 (.019)	.226 (.028)	.165 (.030)
50	6,000	120	.082 (.013)	.072 (.015)	.083 (.012)	.073 (.015)	.208 (.025)	.177 (.026)
100	12,000	120	.078 (.012)	.073 (.014)	.079 (.011)	.074 (.013)	.202 (.024)	.186 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.433 (.067)	.093 (.071)	.504 (.065)	.105 (.075)	.995 (.008)	.024 (.035)
1.25	150	120	.409 (.065)	.104 (.069)	.470 (.065)	.115 (.073)	.877 (.033)	.066 (.057)
1.50	180	120	.384 (.063)	.109 (.063)	.439 (.065)	.123 (.065)	.794 (.036)	.100 (.065)
1.75	**210**	**120**	**.365 (.061)**	**.115 (.064)**	**.413 (.064)**	**.128 (.066)**	**.735 (.040)**	**.129 (.068)**
2	240	120	.352 (.060)	.122 (.061)	.394 (.063)	.135 (.065)	.692 (.040)	.149 (.067)
3	360	120	.315 (.058)	.142 (.060)	.346 (.060)	.155 (.063)	.588 (.040)	.214 (.063)
4	480	120	.295 (.053)	.153 (.054)	.319 (.054)	.166 (.056)	.534 (.039)	.246 (.061)
5	600	120	.280 (.053)	.165 (.054)	.300 (.054)	.176 (.057)	.504 (.037)	.276 (.054)
6	720	120	.270 (.051)	.168 (.052)	.285 (.051)	.177 (.053)	.484 (.037)	.286 (.053)
10	1,200	120	.243 (.044)	.182 (.048)	.255 (.043)	.190 (.048)	.439 (.035)	.321 (.046)
25	3,000	120	.225 (.041)	.200 (.044)	.232 (.040)	.205 (.043)	.404 (.033)	.355 (.038)
50	6,000	120	.218 (.037)	.204 (.039)	.223 (.036)	.209 (.038)	.394 (.031)	.369 (.034)
100	12,000	120	.213 (.035)	.205 (.036)	.217 (.035)	.210 (.036)	.386 (.031)	.373 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.473 (.063)	.130 (.084)	.544 (.058)	.146 (.088)	.996 (.006)	.026 (.035)
1.25	150	120	.451 (.060)	.140 (.078)	.511 (.057)	.157 (.082)	.893 (.027)	.091 (.069)
1.50	180	120	.431 (.059)	.150 (.075)	.487 (.059)	.168 (.078)	.818 (.034)	.139 (.078)
1.75	210	120	.418 (.058)	.162 (.076)	.469 (.058)	.183 (.077)	.769 (.037)	.182 (.077)
2	**240**	**120**	**.406 (.055)**	**.173 (.073)**	**.451 (.054)**	**.191 (.075)**	**.725 (.037)**	**.210 (.080)**
3	360	120	.381 (.054)	.200 (.066)	.417 (.054)	.220 (.067)	.635 (.039)	.283 (.067)
4	480	120	.365 (.051)	.225 (.062)	.394 (.052)	.242 (.062)	.591 (.038)	.327 (.061)
5	600	120	.354 (.052)	.236 (.063)	.380 (.051)	.252 (.062)	.562 (.038)	.350 (.059)
6	720	120	.347 (.051)	.244 (.061)	.369 (.051)	.258 (.061)	.544 (.036)	.366 (.055)
10	1,200	120	.331 (.049)	.265 (.055)	.348 (.049)	.278 (.055)	.507 (.035)	.400 (.045)
25	3,000	120	.318 (.046)	.289 (.050)	.328 (.046)	.299 (.050)	.473 (.034)	.430 (.039)
50	6,000	120	.317 (.045)	.304 (.047)	.326 (.045)	.312 (.047)	.465 (.033)	.444 (.036)
100	12,000	120	.312 (.045)	.305 (.046)	.320 (.045)	.313 (.046)	.458 (.032)	.447 (.034)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.298 (.097)	.034 (.038)	.360 (.112)	.038 (.042)	.993 (.010)	.019 (.026)
1.25	150	120	.272 (.091)	.034 (.038)	.315 (.105)	.037 (.039)	.842 (.038)	.024 (.028)
1.50	180	120	.240 (.085)	.035 (.037)	.279 (.097)	.039 (.039)	.733 (.044)	.034 (.036)
1.75	**210**	**120**	**.222 (.081)**	**.037 (.032)**	**.252 (.094)**	**.038 (.035)**	**.654 (.046)**	**.039 (.037)**
2	240	120	.201 (.076)	.035 (.033)	.231 (.089)	.038 (.033)	.595 (.044)	.045 (.038)
3	360	120	.158 (.064)	.039 (.030)	.175 (.071)	.042 (.032)	.461 (.039)	.068 (.040)
4	480	120	.132 (.056)	.042 (.030)	.141 (.059)	.044 (.031)	.392 (.038)	.083 (.040)
5	600	120	.115 (.051)	.043 (.029)	.121 (.052)	.045 (.030)	.352 (.035)	.094 (.038)
6	720	120	.104 (.046)	.045 (.028)	.110 (.048)	.046 (.029)	.323 (.034)	.104 (.037)
10	1,200	120	.077 (.030)	.044 (.024)	.079 (.028)	.044 (.023)	.270 (.028)	.130 (.035)
25	3,000	120	.055 (.016)	.043 (.016)	.055 (.014)	.043 (.015)	.222 (.025)	.162 (.029)
50	6,000	120	.048 (.010)	.043 (.013)	.049 (.009)	.043 (.013)	.205 (.024)	.174 (.025)
100	12,000	120	.045 (.008)	.043 (.010)	.045 (.007)	.043 (.010)	.197 (.024)	.181 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.373 (.095)	.074 (.065)	.433 (.103)	.083 (.069)	.995 (.007)	.022 (.030)
1.25	150	120	.341 (.092)	.079 (.061)	.391 (.098)	.088 (.067)	.878 (.033)	.061 (.055)
1.50	180	120	.318 (.090)	.086 (.063)	.365 (.097)	.095 (.065)	.790 (.039)	.091 (.062)
1.75	**210**	**120**	**.307 (.088)**	**.092 (.064)**	**.346 (.097)**	**.101 (.065)**	**.730 (.041)**	**.118 (.064)**
2	240	120	.291 (.086)	.093 (.059)	.326 (.091)	.104 (.063)	.686 (.039)	.140 (.066)
3	360	120	.253 (.080)	.108 (.059)	.276 (.086)	.117 (.063)	.578 (.038)	.202 (.062)
4	480	120	.230 (.074)	.116 (.058)	.248 (.080)	.124 (.062)	.523 (.040)	.236 (.058)
5	600	120	.216 (.073)	.122 (.057)	.230 (.076)	.128 (.059)	.493 (.037)	.260 (.055)
6	720	120	.207 (.068)	.126 (.058)	.219 (.072)	.133 (.060)	.472 (.036)	.274 (.054)
10	1,200	120	.183 (.060)	.134 (.056)	.188 (.059)	.137 (.056)	.429 (.034)	.308 (.046)
25	3,000	120	.162 (.049)	.141 (.050)	.165 (.047)	.144 (.049)	.394 (.033)	.344 (.038)
50	6,000	120	.150 (.041)	.140 (.043)	.151 (.039)	.142 (.041)	.379 (.032)	.354 (.035)
100	12,000	120	.145 (.038)	.141 (.039)	.147 (.037)	.143 (.038)	.372 (.031)	.360 (.032)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	120	.420 (.087)	.105 (.079)	.483 (.092)	.119 (.085)	.995 (.007)	.025 (.035)
1.25	150	120	.399 (.083)	.118 (.077)	.458 (.089)	.133 (.081)	.890 (.029)	.085 (.066)
1.50	180	120	.375 (.082)	.125 (.075)	.426 (.086)	.141 (.079)	.814 (.037)	.132 (.072)
1.75	**210**	**120**	**.362 (.084)**	**.130 (.073)**	**.408 (.088)**	**.147 (.079)**	**.761 (.039)**	**.172 (.077)**
2	240	120	.350 (.081)	.141 (.074)	.389 (.085)	.155 (.079)	.722 (.037)	.202 (.076)
3	360	120	.322 (.080)	.162 (.071)	.353 (.084)	.177 (.074)	.626 (.039)	.272 (.066)
4	480	120	.308 (.076)	.182 (.071)	.333 (.081)	.196 (.075)	.582 (.037)	.311 (.061)
5	600	120	.292 (.075)	.190 (.073)	.314 (.076)	.202 (.074)	.551 (.036)	.337 (.057)
6	720	120	.293 (.073)	.202 (.069)	.309 (.075)	.213 (.072)	.535 (.036)	.354 (.053)
10	1,200	120	.273 (.071)	.218 (.070)	.285 (.073)	.227 (.073)	.498 (.035)	.391 (.046)
25	3,000	120	.256 (.065)	.232 (.067)	.264 (.065)	.238 (.068)	.464 (.032)	.421 (.037)
50	6,000	120	.251 (.063)	.238 (.064)	.256 (.063)	.243 (.064)	.454 (.033)	.432 (.035)
100	12,000	120	.245 (.060)	.239 (.060)	.250 (.060)	.244 (.061)	.446 (.032)	.435 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.485 (.042)	.084 (.054)	.571 (.036)	.098 (.059)	.997 (.004)	.014 (.018)
1.25	255	204	.461 (.039)	.093 (.052)	.538 (.034)	.108 (.055)	.877 (.025)	.057 (.042)
1.50	306	204	.439 (.037)	.102 (.050)	.510 (.033)	.118 (.053)	.791 (.029)	.089 (.048)
1.75	357	204	.423 (.035)	.115 (.051)	.488 (.031)	.132 (.053)	.731 (.030)	.119 (.050)
2	**408**	**204**	**.413 (.035)**	**.125 (.046)**	**.473 (.031)**	**.144 (.048)**	**.687 (.032)**	**.145 (.051)**
3	612	204	.382 (.031)	.153 (.045)	.431 (.029)	.172 (.047)	.583 (.030)	.202 (.050)
4	816	204	.363 (.029)	.177 (.041)	.406 (.027)	.199 (.042)	.528 (.029)	.238 (.045)
5	1,020	204	.353 (.028)	.193 (.039)	.391 (.027)	.214 (.040)	.498 (.029)	.262 (.043)
6	1,224	204	.348 (.027)	.208 (.038)	.382 (.026)	.230 (.039)	.477 (.028)	.281 (.040)
10	2,040	204	.334 (.025)	.240 (.034)	.361 (.024)	.260 (.034)	.435 (.026)	.314 (.034)
25	5,100	204	.323 (.024)	.280 (.028)	.342 (.024)	.297 (.028)	.396 (.025)	.346 (.029)
50	10,200	204	.321 (.023)	.298 (.026)	.337 (.023)	.313 (.026)	.384 (.024)	.359 (.026)
100	20,400	204	.321 (.023)	.310 (.024)	.335 (.023)	.323 (.024)	.379 (.023)	.367 (.024)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.502 (.042)	.127 (.064)	.581 (.035)	.147 (.069)	.997 (.004)	.015 (.022)
1.25	255	204	.485 (.041)	.146 (.063)	.555 (.036)	.167 (.065)	.895 (.021)	.089 (.054)
1.50	306	204	.469 (.038)	.162 (.061)	.533 (.033)	.184 (.063)	.821 (.026)	.142 (.059)
1.75	357	204	.456 (.036)	.179 (.060)	.516 (.032)	.203 (.061)	.769 (.027)	.187 (.057)
2	**408**	**204**	**.447 (.035)**	**.188 (.056)**	**.502 (.031)**	**.212 (.057)**	**.731 (.027)**	**.217 (.059)**
3	612	204	.425 (.032)	.224 (.049)	.470 (.030)	.249 (.049)	.640 (.029)	.289 (.053)
4	816	204	.413 (.030)	.255 (.047)	.452 (.029)	.281 (.048)	.594 (.028)	.337 (.048)
5	1,020	204	.406 (.030)	.270 (.043)	.441 (.028)	.293 (.043)	.568 (.028)	.354 (.044)
6	1,224	204	.405 (.030)	.288 (.043)	.436 (.028)	.311 (.042)	.551 (.028)	.376 (.041)
10	2,040	204	.394 (.028)	.319 (.038)	.419 (.026)	.340 (.037)	.513 (.026)	.408 (.035)
25	5,100	204	.391 (.025)	.358 (.029)	.409 (.025)	.375 (.029)	.482 (.024)	.440 (.028)
50	10,200	204	.390 (.026)	.373 (.028)	.405 (.026)	.388 (.028)	.471 (.025)	.450 (.027)
100	20,400	204	.388 (.026)	.379 (.027)	.402 (.025)	.393 (.027)	.464 (.025)	.453 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.520 (.042)	.162 (.070)	.595 (.035)	.186 (.072)	.998 (.003)	.017 (.021)
1.25	255	204	.500 (.041)	.178 (.064)	.567 (.034)	.203 (.068)	.903 (.021)	.107 (.059)
1.50	306	204	.486 (.040)	.198 (.066)	.548 (.034)	.223 (.068)	.836 (.025)	.176 (.066)
1.75	357	204	.478 (.037)	.212 (.065)	.534 (.033)	.238 (.065)	.788 (.026)	.220 (.064)
2	**408**	**204**	**.468 (.036)**	**.226 (.060)**	**.521 (.032)**	**.253 (.061)**	**.753 (.027)**	**.255 (.061)**
3	612	204	.451 (.033)	.267 (.054)	.493 (.030)	.293 (.055)	.670 (.027)	.339 (.053)
4	816	204	.443 (.033)	.296 (.049)	.479 (.031)	.321 (.049)	.628 (.028)	.383 (.048)
5	1,020	204	.438 (.031)	.313 (.044)	.470 (.029)	.338 (.044)	.604 (.028)	.406 (.043)
6	1,224	204	.434 (.031)	.329 (.044)	.464 (.030)	.352 (.043)	.587 (.028)	.423 (.041)
10	2,040	204	.429 (.029)	.362 (.038)	.453 (.028)	.383 (.037)	.554 (.026)	.456 (.034)
25	5,100	204	.426 (.027)	.397 (.032)	.444 (.026)	.413 (.031)	.524 (.024)	.484 (.028)
50	10,200	204	.425 (.026)	.410 (.028)	.440 (.026)	.425 (.028)	.514 (.024)	.494 (.026)
100	20,400	204	.424 (.026)	.416 (.027)	.438 (.026)	.430 (.027)	.508 (.024)	.498 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.341 (.078)	.050 (.039)	.386 (.089)	.055 (.041)	.996 (.005)	.013 (.018)
1.25	255	204	.301 (.073)	.055 (.040)	.335 (.082)	.059 (.042)	.857 (.028)	.031 (.031)
1.50	306	204	.271 (.071)	.057 (.038)	.302 (.078)	.062 (.040)	.758 (.032)	.049 (.034)
1.75	357	204	.250 (.066)	.062 (.038)	.273 (.071)	.066 (.039)	.689 (.034)	.065 (.039)
2	**408**	**204**	**.230 (.060)**	**.061 (.034)**	**.248 (.062)**	**.065 (.035)**	**.638 (.033)**	**.079 (.040)**
3	612	204	.182 (.050)	.067 (.033)	.191 (.048)	.070 (.034)	.515 (.032)	.120 (.040)
4	816	204	.153 (.037)	.069 (.031)	.161 (.036)	.071 (.031)	.456 (.029)	.146 (.040)
5	1,020	204	.139 (.034)	.071 (.028)	.144 (.031)	.073 (.028)	.420 (.029)	.165 (.037)
6	1,224	204	.130 (.031)	.072 (.027)	.132 (.028)	.073 (.027)	.395 (.029)	.180 (.037)
10	2,040	204	.105 (.019)	.072 (.022)	.107 (.017)	.073 (.022)	.346 (.026)	.211 (.033)
25	5,100	204	.086 (.012)	.073 (.016)	.086 (.010)	.074 (.015)	.302 (.024)	.246 (.027)
50	10,200	204	.079 (.009)	.073 (.012)	.080 (.009)	.074 (.012)	.287 (.023)	.259 (.025)
100	20,400	204	.077 (.008)	.074 (.011)	.077 (.008)	.074 (.010)	.281 (.022)	.267 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.418 (.070)	.104 (.060)	.475 (.073)	.116 (.063)	.997 (.004)	.016 (.021)
1.25	255	204	.394 (.069)	.116 (.058)	.441 (.074)	.129 (.061)	.888 (.022)	.079 (.049)
1.50	306	204	.366 (.067)	.123 (.057)	.407 (.071)	.135 (.059)	.811 (.026)	.126 (.058)
1.75	**357**	**204**	**.350 (.065)**	**.130 (.056)**	**.386 (.070)**	**.142 (.058)**	**.757 (.029)**	**.161 (.056)**
2	408	204	.334 (.066)	.135 (.056)	.367 (.069)	.147 (.058)	.715 (.031)	.188 (.059)
3	612	204	.299 (.060)	.151 (.054)	.321 (.063)	.161 (.057)	.621 (.030)	.263 (.052)
4	816	204	.278 (.058)	.162 (.052)	.295 (.057)	.170 (.053)	.575 (.029)	.302 (.046)
5	1,020	204	.261 (.056)	.170 (.051)	.274 (.056)	.178 (.052)	.545 (.028)	.328 (.043)
6	1,224	204	.253 (.050)	.173 (.050)	.265 (.049)	.180 (.050)	.528 (.028)	.344 (.041)
10	2,040	204	.231 (.047)	.182 (.047)	.240 (.045)	.188 (.046)	.489 (.026)	.376 (.034)
25	5,100	204	.207 (.035)	.187 (.037)	.212 (.034)	.191 (.036)	.454 (.025)	.409 (.028)
50	10,200	204	.200 (.031)	.191 (.034)	.204 (.031)	.194 (.033)	.443 (.025)	.420 (.027)
100	20,400	204	.199 (.030)	.194 (.031)	.202 (.030)	.197 (.031)	.439 (.024)	.429 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.470 (.059)	.143 (.068)	.531 (.059)	.161 (.071)	.997 (.004)	.016 (.023)
1.25	255	204	.445 (.059)	.156 (.067)	.499 (.062)	.173 (.071)	.900 (.021)	.100 (.055)
1.50	306	204	.428 (.059)	.170 (.064)	.479 (.061)	.191 (.068)	.832 (.025)	.166 (.062)
1.75	**357**	**204**	**.412 (.056)**	**.179 (.063)**	**.456 (.058)**	**.199 (.065)**	**.782 (.027)**	**.210 (.062)**
2	408	204	.402 (.058)	.187 (.062)	.443 (.061)	.206 (.064)	.746 (.028)	.241 (.063)
3	612	204	.376 (.055)	.214 (.059)	.406 (.056)	.230 (.061)	.659 (.029)	.321 (.054)
4	816	204	.355 (.055)	.230 (.058)	.379 (.056)	.245 (.059)	.617 (.028)	.364 (.048)
5	1,020	204	.350 (.054)	.247 (.056)	.370 (.054)	.260 (.057)	.594 (.027)	.392 (.042)
6	1,224	204	.345 (.055)	.256 (.057)	.363 (.055)	.269 (.058)	.577 (.026)	.409 (.041)
10	2,040	204	.330 (.052)	.273 (.054)	.342 (.052)	.282 (.053)	.542 (.027)	.442 (.036)
25	5,100	204	.311 (.046)	.287 (.048)	.319 (.045)	.295 (.048)	.511 (.025)	.471 (.028)
50	10,200	204	.303 (.045)	.292 (.046)	.311 (.045)	.298 (.046)	.500 (.024)	.480 (.027)
100	20,400	204	.303 (.042)	.297 (.043)	.310 (.043)	.303 (.043)	.496 (.024)	.486 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.258 (.105)	.037 (.035)	.294 (.124)	.038 (.036)	.997 (.005)	.013 (.018)
1.25	255	204	.226 (.098)	.039 (.037)	.255 (.113)	.042 (.038)	.856 (.027)	.031 (.030)
1.50	**306**	**204**	**.201 (.092)**	**.039 (.034)**	**.223 (.105)**	**.042 (.035)**	**.755 (.031)**	**.047 (.037)**
1.75	357	204	.183 (.088)	.041 (.031)	.201 (.099)	.044 (.033)	.686 (.034)	.065 (.040)
2	408	204	.168 (.082)	.042 (.033)	.180 (.090)	.044 (.034)	.633 (.034)	.074 (.038)
3	612	204	.125 (.064)	.044 (.030)	.129 (.065)	.045 (.031)	.510 (.033)	.114 (.039)
4	816	204	.108 (.054)	.046 (.030)	.109 (.053)	.047 (.030)	.448 (.029)	.141 (.037)
5	1,020	204	.091 (.044)	.045 (.028)	.090 (.039)	.044 (.027)	.410 (.029)	.157 (.038)
6	1,224	204	.081 (.038)	.043 (.025)	.080 (.033)	.043 (.024)	.387 (.028)	.172 (.034)
10	2,040	204	.063 (.021)	.044 (.021)	.062 (.018)	.043 (.020)	.337 (.025)	.204 (.032)
25	5,100	204	.048 (.009)	.041 (.013)	.048 (.007)	.041 (.012)	.294 (.023)	.237 (.026)
50	10,200	204	.043 (.006)	.040 (.009)	.043 (.005)	.040 (.009)	.278 (.022)	.250 (.024)
100	20,400	204	.041 (.005)	.040 (.007)	.041 (.005)	.040 (.007)	.272 (.022)	.258 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.346 (.106)	.080 (.058)	.398 (.116)	.090 (.062)	.997 (.004)	.014 (.019)
1.25	255	204	.317 (.101)	.085 (.056)	.359 (.114)	.095 (.061)	.885 (.023)	.074 (.048)
1.50	**306**	**204**	**.292 (.101)**	**.087 (.055)**	**.323 (.110)**	**.096 (.059)**	**.806 (.028)**	**.115 (.054)**
1.75	357	204	.280 (.099)	.097 (.056)	.305 (.104)	.103 (.058)	.750 (.030)	.150 (.054)
2	408	204	.266 (.097)	.100 (.057)	.288 (.102)	.108 (.061)	.709 (.031)	.179 (.056)
3	612	204	.226 (.086)	.110 (.059)	.241 (.090)	.116 (.062)	.612 (.030)	.248 (.050)
4	816	204	.205 (.080)	.116 (.057)	.216 (.083)	.122 (.061)	.564 (.029)	.288 (.046)
5	1,020	204	.191 (.073)	.118 (.056)	.199 (.073)	.122 (.057)	.535 (.029)	.311 (.044)
6	1,224	204	.180 (.071)	.120 (.060)	.186 (.069)	.123 (.059)	.516 (.029)	.331 (.040)
10	2,040	204	.159 (.059)	.122 (.055)	.161 (.055)	.123 (.053)	.477 (.026)	.362 (.034)
25	5,100	204	.135 (.041)	.122 (.041)	.136 (.037)	.122 (.038)	.442 (.026)	.397 (.029)
50	10,200	204	.129 (.034)	.123 (.036)	.130 (.033)	.124 (.035)	.432 (.025)	.409 (.027)
100	20,400	204	.122 (.026)	.119 (.027)	.123 (.025)	.120 (.026)	.426 (.024)	.414 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table B27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Two Response Options

Ratio of Sample Size to Items	Actual Sample Size	Items	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	204	.406 (.094)	.110 (.066)	.462 (.102)	.125 (.071)	.997 (.004)	.016 (.022)
1.25	255	204	.384 (.094)	.125 (.069)	.428 (.101)	.138 (.074)	.897 (.021)	.097 (.056)
1.50	**306**	**204**	**.365 (.090)**	**.135 (.066)**	**.410 (.097)**	**.150 (.070)**	**.828 (.025)**	**.158 (.059)**
1.75	357	204	.347 (.090)	.141 (.067)	.386 (.097)	.155 (.070)	.777 (.027)	.199 (.062)
2	408	204	.336 (.090)	.151 (.071)	.370 (.096)	.165 (.075)	.739 (.028)	.235 (.064)
3	612	204	.304 (.088)	.169 (.071)	.325 (.091)	.180 (.074)	.653 (.029)	.312 (.052)
4	816	204	.293 (.085)	.185 (.074)	.310 (.088)	.194 (.076)	.611 (.029)	.357 (.048)
5	1,020	204	.280 (.081)	.189 (.070)	.294 (.083)	.197 (.072)	.584 (.028)	.378 (.043)
6	1,224	204	.271 (.077)	.194 (.071)	.284 (.077)	.203 (.072)	.565 (.028)	.392 (.041)
10	2,040	204	.252 (.073)	.205 (.070)	.261 (.073)	.212 (.071)	.529 (.026)	.427 (.035)
25	5,100	204	.231 (.062)	.212 (.062)	.236 (.061)	.216 (.062)	.501 (.025)	.460 (.028)
50	10,200	204	.223 (.057)	.213 (.058)	.227 (.056)	.216 (.057)	.489 (.025)	.468 (.027)
100	20,400	204	.217 (.052)	.212 (.052)	.221 (.052)	.216 (.052)	.484 (.024)	.474 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix C: Simulation Results for Items with Three Response Options – Dummy-coded Response Options Summed Equals Item Condition

Table C1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.418 (.083)	.047 (.062)	.518 (.072)	.049 (.065)	.981 (.026)	.043 (.057)
1.25	60	24	72	48	.384 (.078)	.036 (.048)	.473 (.069)	.038 (.050)	.830 (.067)	.036 (.050)
1.50	72	24	72	48	.349 (.070)	.034 (.046)	.427 (.064)	.035 (.046)	.701 (.073)	.031 (.043)
1.75	84	24	72	48	.323 (.066)	.031 (.040)	.394 (.061)	.032 (.043)	.611 (.074)	.030 (.040)
2	96	24	72	48	.298 (.065)	.031 (.044)	.364 (.059)	.033 (.045)	.541 (.072)	.030 (.039)
3	144	24	72	48	.235 (.049)	.025 (.031)	.285 (.047)	.027 (.032)	.384 (.058)	.025 (.032)
4	192	24	72	48	.200 (.041)	.027 (.031)	.241 (.040)	.029 (.032)	.306 (.048)	.028 (.031)
5	**240**	**24**	**72**	**48**	**.177 (.041)**	**.025 (.028)**	**.211 (.041)**	**.029 (.030)**	**.260 (.048)**	**.029 (.028)**
6	288	24	72	48	.161 (.035)	.028 (.028)	.190 (.035)	.031 (.030)	.229 (.041)	.031 (.030)
10	480	24	72	48	.127 (.028)	.031 (.025)	.146 (.027)	.034 (.027)	.168 (.031)	.036 (.028)
25	1,200	24	72	48	.091 (.023)	.043 (.024)	.101 (.023)	.048 (.025)	.111 (.026)	.051 (.027)
50	2,400	24	72	48	.080 (.020)	.053 (.022)	.086 (.021)	.057 (.022)	.093 (.023)	.061 (.024)
100	4,800	24	72	48	.073 (.019)	.059 (.021)	.077 (.019)	.062 (.021)	.083 (.021)	.067 (.022)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.451 (.086)	.071 (.087)	.543 (.072)	.073 (.088)	.984 (.024)	.046 (.063)
1.25	60	24	72	48	.424 (.080)	.070 (.077)	.505 (.071)	.076 (.087)	.855 (.058)	.056 (.072)
1.50	72	24	72	48	.395 (.075)	.065 (.070)	.469 (.068)	.074 (.076)	.749 (.070)	.064 (.071)
1.75	84	24	72	48	.376 (.072)	.064 (.067)	.442 (.065)	.071 (.072)	.678 (.073)	.070 (.077)
2	**96**	**24**	**72**	**48**	**.355 (.071)**	**.071 (.072)**	**.416 (.064)**	**.081 (.078)**	**.621 (.076)**	**.085 (.084)**
3	144	24	72	48	.302 (.065)	.079 (.066)	.352 (.062)	.089 (.072)	.488 (.077)	.110 (.089)
4	192	24	72	48	.279 (.061)	.089 (.064)	.319 (.059)	.100 (.069)	.426 (.075)	.128 (.085)
5	240	24	72	48	.265 (.060)	.100 (.069)	.299 (.060)	.113 (.074)	.391 (.078)	.147 (.092)
6	288	24	72	48	.252 (.059)	.108 (.067)	.284 (.059)	.119 (.070)	.367 (.076)	.159 (.088)
10	480	24	72	48	.230 (.059)	.128 (.063)	.253 (.059)	.140 (.067)	.317 (.075)	.184 (.085)
25	1,200	24	72	48	.208 (.056)	.160 (.062)	.221 (.057)	.170 (.063)	.271 (.074)	.216 (.081)
50	2,400	24	72	48	.200 (.058)	.173 (.061)	.209 (.059)	.181 (.062)	.254 (.075)	.224 (.079)
100	4,800	24	72	48	.196 (.058)	.182 (.060)	.203 (.060)	.188 (.062)	.246 (.075)	.231 (.078)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.497 (.088)	.115 (.112)	.578 (.075)	.129 (.121)	.989 (.016)	.062 (.081)
1.25	60	24	72	48	.476 (.081)	.118 (.111)	.554 (.070)	.132 (.115)	.893 (.049)	.102 (.103)
1.50	**72**	**24**	**72**	**48**	**.451 (.080)**	**.125 (.106)**	**.522 (.071)**	**.141 (.113)**	**.817 (.058)**	**.152 (.121)**
1.75	84	24	72	48	.433 (.076)	.133 (.101)	.495 (.069)	.149 (.108)	.757 (.066)	.180 (.128)
2	96	24	72	48	.416 (.076)	.137 (.100)	.476 (.070)	.157 (.107)	.714 (.072)	.200 (.129)
3	144	24	72	48	.386 (.076)	.170 (.099)	.433 (.072)	.191 (.106)	.619 (.080)	.273 (.128)
4	192	24	72	48	.372 (.072)	.188 (.099)	.411 (.071)	.209 (.104)	.573 (.084)	.310 (.126)
5	240	24	72	48	.362 (.072)	.206 (.098)	.396 (.071)	.224 (.100)	.546 (.085)	.332 (.125)
6	288	24	72	48	.352 (.075)	.214 (.093)	.384 (.074)	.234 (.096)	.525 (.087)	.346 (.115)
10	480	24	72	48	.340 (.072)	.250 (.088)	.364 (.073)	.268 (.091)	.491 (.087)	.384 (.112)
25	1,200	24	72	48	.331 (.079)	.288 (.088)	.347 (.079)	.302 (.089)	.458 (.091)	.412 (.102)
50	2,400	24	72	48	.331 (.081)	.310 (.085)	.343 (.082)	.322 (.086)	.449 (.093)	.428 (.097)
100	4,800	24	72	48	.331 (.079)	.319 (.082)	.341 (.080)	.329 (.083)	.444 (.091)	.432 (.094)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.405 (.082)	.046 (.061)	.497 (.073)	.048 (.065)	.980 (.027)	.046 (.061)
1.25	60	24	72	48	.365 (.077)	.039 (.054)	.449 (.070)	.040 (.055)	.826 (.069)	.036 (.048)
1.50	72	24	72	48	.331 (.070)	.034 (.043)	.406 (.063)	.036 (.047)	.698 (.077)	.033 (.043)
1.75	84	24	72	48	.303 (.065)	.031 (.043)	.369 (.061)	.032 (.042)	.606 (.076)	.029 (.039)
2	96	24	72	48	.280 (.063)	.026 (.036)	.339 (.058)	.028 (.037)	.535 (.072)	.025 (.032)
3	144	24	72	48	.220 (.051)	.023 (.030)	.263 (.047)	.025 (.031)	.376 (.059)	.022 (.029)
4	192	24	72	48	.181 (.040)	.021 (.026)	.214 (.040)	.024 (.028)	.296 (.048)	.020 (.025)
5	240	24	72	48	.155 (.035)	.023 (.026)	.182 (.036)	.026 (.028)	.246 (.044)	.022 (.026)
6	288	24	72	48	.140 (.033)	.022 (.024)	.163 (.032)	.025 (.025)	.215 (.040)	.022 (.024)
10	480	24	72	48	.104 (.024)	.026 (.021)	.118 (.024)	.029 (.022)	.153 (.030)	.027 (.022)
25	**1,200**	**24**	**72**	**48**	**.069 (.016)**	**.033 (.018)**	**.074 (.016)**	**.035 (.019)**	**.097 (.020)**	**.038 (.021)**
50	2,400	24	72	48	.055 (.014)	.036 (.015)	.058 (.014)	.037 (.015)	.076 (.018)	.043 (.019)
100	4,800	24	72	48	.049 (.012)	.040 (.013)	.051 (.012)	.041 (.014)	.068 (.016)	.051 (.017)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.440 (.086)	.066 (.081)	.528 (.073)	.070 (.084)	.984 (.021)	.048 (.062)
1.25	60	24	72	48	.409 (.082)	.063 (.075)	.488 (.072)	.069 (.081)	.855 (.060)	.055 (.072)
1.50	72	24	72	48	.383 (.076)	.064 (.071)	.452 (.069)	.069 (.075)	.749 (.069)	.066 (.077)
1.75	84	24	72	48	.352 (.070)	.069 (.073)	.416 (.065)	.076 (.080)	.670 (.075)	.073 (.082)
2	**96**	**24**	**72**	**48**	**.339 (.069)**	**.061 (.063)**	**.397 (.065)**	**.067 (.064)**	**.618 (.077)**	**.073 (.075)**
3	144	24	72	48	.290 (.062)	.073 (.064)	.334 (.059)	.081 (.066)	.488 (.076)	.100 (.080)
4	192	24	72	48	.262 (.057)	.084 (.062)	.297 (.056)	.093 (.065)	.423 (.073)	.125 (.086)
5	240	24	72	48	.245 (.056)	.092 (.061)	.275 (.056)	.102 (.064)	.386 (.074)	.142 (.086)
6	288	24	72	48	.230 (.054)	.093 (.060)	.255 (.053)	.102 (.062)	.359 (.073)	.146 (.085)
10	480	24	72	48	.205 (.052)	.111 (.056)	.224 (.053)	.120 (.058)	.310 (.073)	.174 (.083)
25	1,200	24	72	48	.180 (.051)	.135 (.053)	.190 (.052)	.142 (.054)	.263 (.071)	.204 (.076)
50	2,400	24	72	48	.170 (.049)	.148 (.052)	.177 (.050)	.153 (.053)	.246 (.069)	.217 (.074)
100	4,800	24	72	48	.166 (.050)	.154 (.051)	.171 (.051)	.159 (.052)	.239 (.070)	.225 (.073)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.481 (.085)	.112 (.112)	.564 (.075)	.122 (.117)	.988 (.016)	.063 (.082)
1.25	60	24	72	48	.458 (.081)	.111 (.106)	.533 (.072)	.124 (.110)	.892 (.048)	.105 (.105)
1.50	**72**	**24**	**72**	**48**	**.432 (.079)**	**.121 (.106)**	**.503 (.073)**	**.137 (.113)**	**.808 (.063)**	**.144 (.124)**
1.75	84	24	72	48	.416 (.079)	.120 (.096)	.477 (.074)	.136 (.100)	.751 (.071)	.165 (.119)
2	96	24	72	48	.407 (.078)	.135 (.102)	.462 (.071)	.151 (.107)	.714 (.072)	.198 (.128)
3	144	24	72	48	.367 (.073)	.156 (.099)	.411 (.072)	.173 (.104)	.611 (.080)	.259 (.126)
4	192	24	72	48	.344 (.070)	.170 (.089)	.382 (.068)	.188 (.094)	.559 (.081)	.291 (.119)
5	240	24	72	48	.332 (.071)	.180 (.087)	.364 (.069)	.198 (.090)	.528 (.085)	.312 (.116)
6	288	24	72	48	.328 (.070)	.196 (.088)	.357 (.069)	.214 (.091)	.513 (.083)	.332 (.116)
10	480	24	72	48	.314 (.072)	.227 (.087)	.336 (.073)	.242 (.089)	.476 (.087)	.366 (.106)
25	1,200	24	72	48	.301 (.074)	.262 (.080)	.315 (.075)	.274 (.082)	.443 (.088)	.398 (.095)
50	2,400	24	72	48	.300 (.071)	.280 (.075)	.310 (.072)	.289 (.076)	.436 (.085)	.413 (.088)
100	4,800	24	72	48	.301 (.073)	.291 (.075)	.310 (.075)	.299 (.076)	.433 (.089)	.423 (.091)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.379 (.084)	.049 (.063)	.467 (.078)	.051 (.066)	.979 (.029)	.045 (.060)
1.25	60	24	72	48	.343 (.081)	.040 (.049)	.420 (.075)	.040 (.050)	.823 (.068)	.040 (.054)
1.50	72	24	72	48	.307 (.072)	.032 (.042)	.376 (.070)	.033 (.043)	.694 (.074)	.030 (.039)
1.75	84	24	72	48	.281 (.070)	.029 (.038)	.342 (.070)	.030 (.039)	.604 (.081)	.026 (.035)
2	96	24	72	48	.257 (.062)	.026 (.036)	.312 (.062)	.028 (.036)	.538 (.070)	.024 (.033)
3	144	24	72	48	.197 (.049)	.023 (.030)	.234 (.052)	.024 (.030)	.377 (.058)	.023 (.029)
4	192	24	72	48	.163 (.042)	.021 (.025)	.192 (.044)	.023 (.026)	.297 (.049)	.020 (.024)
5	240	24	72	48	.138 (.037)	.019 (.023)	.161 (.038)	.021 (.024)	.246 (.044)	.020 (.023)
6	**288**	**24**	**72**	**48**	**.121 (.031)**	**.020 (.022)**	**.141 (.033)**	**.022 (.023)**	**.216 (.038)**	**.022 (.024)**
10	480	24	72	48	.089 (.024)	.022 (.020)	.099 (.025)	.024 (.021)	.152 (.029)	.027 (.023)
25	1,200	24	72	48	.056 (.015)	.027 (.016)	.059 (.016)	.028 (.017)	.096 (.021)	.038 (.022)
50	2,400	24	72	48	.044 (.012)	.029 (.013)	.045 (.012)	.030 (.014)	.077 (.018)	.046 (.019)
100	4,800	24	72	48	.038 (.011)	.031 (.012)	.039 (.011)	.031 (.012)	.068 (.017)	.051 (.018)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.410 (.092)	.065 (.078)	.495 (.082)	.072 (.084)	.985 (.021)	.046 (.061)
1.25	60	24	72	48	.384 (.080)	.058 (.070)	.460 (.077)	.064 (.074)	.860 (.058)	.058 (.071)
1.50	**72**	**24**	**72**	**48**	**.353 (.079)**	**.060 (.070)**	**.420 (.075)**	**.067 (.075)**	**.752 (.070)**	**.067 (.076)**
1.75	84	24	72	48	.332 (.075)	.061 (.068)	.392 (.071)	.067 (.071)	.677 (.070)	.073 (.074)
2	96	24	72	48	.311 (.075)	.062 (.063)	.364 (.071)	.068 (.068)	.621 (.076)	.086 (.082)
3	144	24	72	48	.266 (.066)	.069 (.062)	.307 (.064)	.076 (.065)	.496 (.074)	.113 (.087)
4	192	24	72	48	.238 (.061)	.070 (.059)	.270 (.060)	.079 (.062)	.432 (.073)	.130 (.086)
5	240	24	72	48	.220 (.058)	.078 (.056)	.248 (.058)	.086 (.060)	.394 (.072)	.148 (.089)
6	288	24	72	48	.206 (.055)	.085 (.058)	.231 (.059)	.094 (.062)	.368 (.073)	.159 (.087)
10	480	24	72	48	.184 (.053)	.097 (.053)	.199 (.054)	.104 (.056)	.320 (.072)	.187 (.085)
25	1,200	24	72	48	.160 (.050)	.120 (.050)	.168 (.051)	.125 (.052)	.276 (.070)	.220 (.076)
50	2,400	24	72	48	.152 (.051)	.130 (.051)	.157 (.052)	.135 (.052)	.261 (.072)	.231 (.074)
100	4,800	24	72	48	.147 (.049)	.136 (.049)	.151 (.050)	.140 (.051)	.254 (.070)	.241 (.072)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	24	72	48	.446 (.086)	.095 (.100)	.529 (.081)	.103 (.106)	.988 (.018)	.058 (.069)
1.25	60	24	72	48	.433 (.088)	.101 (.099)	.502 (.079)	.112 (.106)	.890 (.049)	.098 (.100)
1.50	**72**	**24**	**72**	**48**	**.410 (.083)**	**.111 (.102)**	**.472 (.075)**	**.123 (.108)**	**.808 (.060)**	**.138 (.116)**
1.75	84	24	72	48	.390 (.081)	.114 (.093)	.448 (.077)	.127 (.099)	.749 (.069)	.166 (.119)
2	96	24	72	48	.372 (.078)	.121 (.102)	.425 (.076)	.134 (.105)	.702 (.071)	.194 (.134)
3	144	24	72	48	.340 (.073)	.136 (.089)	.381 (.074)	.152 (.094)	.609 (.076)	.259 (.123)
4	192	24	72	48	.321 (.075)	.151 (.088)	.355 (.074)	.166 (.092)	.560 (.080)	.292 (.119)
5	240	24	72	48	.312 (.074)	.169 (.087)	.341 (.075)	.185 (.090)	.537 (.085)	.320 (.118)
6	288	24	72	48	.302 (.074)	.180 (.087)	.330 (.075)	.195 (.090)	.514 (.081)	.334 (.111)
10	480	24	72	48	.281 (.074)	.196 (.083)	.300 (.076)	.209 (.085)	.470 (.086)	.356 (.106)
25	1,200	24	72	48	.270 (.072)	.233 (.077)	.282 (.074)	.242 (.079)	.441 (.082)	.396 (.091)
50	2,400	24	72	48	.267 (.077)	.247 (.079)	.276 (.078)	.255 (.081)	.428 (.086)	.404 (.091)
100	4,800	24	72	48	.264 (.078)	.254 (.079)	.272 (.079)	.262 (.081)	.425 (.086)	.413 (.088)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.444 (.056)	.040 (.044)	.535 (.047)	.045 (.047)	.993 (.009)	.021 (.027)
1.25	150	60	180	120	.412 (.050)	.040 (.042)	.494 (.044)	.047 (.044)	.846 (.039)	.027 (.033)
1.50	180	60	180	120	.385 (.048)	.044 (.041)	.459 (.044)	.049 (.045)	.739 (.044)	.034 (.037)
1.75	210	60	180	120	.362 (.046)	.046 (.043)	.428 (.041)	.052 (.045)	.659 (.047)	.044 (.040)
2	240	60	180	120	.341 (.044)	.048 (.039)	.402 (.040)	.055 (.041)	.599 (.046)	.048 (.041)
3	**360**	**60**	**180**	**120**	**.291 (.039)**	**.059 (.038)**	**.339 (.037)**	**.068 (.041)**	**.464 (.044)**	**.073 (.046)**
4	480	60	180	120	.266 (.037)	.068 (.036)	.306 (.036)	.078 (.039)	.400 (.043)	.089 (.045)
5	600	60	180	120	.250 (.036)	.077 (.037)	.284 (.035)	.087 (.039)	.360 (.042)	.103 (.045)
6	720	60	180	120	.236 (.033)	.084 (.036)	.267 (.033)	.095 (.038)	.333 (.040)	.113 (.045)
10	1,200	60	180	120	.209 (.031)	.104 (.034)	.231 (.031)	.115 (.035)	.277 (.038)	.138 (.042)
25	3,000	60	180	120	.182 (.030)	.133 (.033)	.195 (.031)	.142 (.034)	.226 (.037)	.168 (.040)
50	6,000	60	180	120	.179 (.031)	.151 (.032)	.187 (.032)	.159 (.033)	.216 (.037)	.185 (.038)
100	12,000	60	180	120	.174 (.030)	.160 (.031)	.181 (.031)	.166 (.032)	.207 (.036)	.191 (.037)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.502 (.054)	.111 (.081)	.584 (.047)	.127 (.084)	.996 (.006)	.027 (.036)
1.25	150	60	180	120	.480 (.053)	.123 (.075)	.553 (.046)	.142 (.082)	.899 (.029)	.108 (.078)
1.50	**180**	**60**	**180**	**120**	**.458 (.052)**	**.135 (.076)**	**.525 (.047)**	**.153 (.078)**	**.829 (.036)**	**.165 (.086)**
1.75	210	60	180	120	.446 (.050)	.146 (.072)	.509 (.046)	.167 (.075)	.782 (.040)	.209 (.089)
2	240	60	180	120	.436 (.048)	.157 (.070)	.493 (.044)	.178 (.074)	.740 (.043)	.243 (.089)
3	360	60	180	120	.408 (.050)	.190 (.071)	.454 (.047)	.213 (.072)	.655 (.050)	.319 (.088)
4	480	60	180	120	.392 (.048)	.215 (.067)	.431 (.047)	.238 (.070)	.611 (.052)	.360 (.083)
5	600	60	180	120	.386 (.050)	.232 (.067)	.420 (.049)	.254 (.068)	.587 (.056)	.385 (.079)
6	720	60	180	120	.384 (.050)	.251 (.063)	.415 (.049)	.273 (.065)	.573 (.054)	.404 (.075)
10	1,200	60	180	120	.371 (.047)	.283 (.060)	.395 (.047)	.301 (.061)	.537 (.053)	.435 (.067)
25	3,000	60	180	120	.362 (.049)	.323 (.054)	.378 (.049)	.337 (.055)	.504 (.056)	.463 (.063)
50	6,000	60	180	120	.362 (.052)	.342 (.056)	.375 (.053)	.354 (.057)	.496 (.059)	.475 (.062)
100	12,000	60	180	120	.366 (.051)	.355 (.052)	.377 (.051)	.366 (.053)	.494 (.056)	.483 (.058)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.533 (.060)	.165 (.101)	.608 (.049)	.190 (.103)	.997 (.005)	.032 (.040)
1.25	150	60	180	120	.517 (.055)	.187 (.094)	.584 (.047)	.212 (.097)	.920 (.024)	.169 (.095)
1.50	**180**	**60**	**180**	**120**	**.504 (.052)**	**.204 (.089)**	**.565 (.046)**	**.230 (.093)**	**.865 (.029)**	**.254 (.099)**
1.75	210	60	180	120	.491 (.052)	.217 (.088)	.548 (.046)	.245 (.090)	.824 (.034)	.306 (.098)
2	240	60	180	120	.485 (.053)	.234 (.085)	.537 (.048)	.264 (.086)	.795 (.038)	.353 (.096)
3	360	60	180	120	.467 (.050)	.275 (.079)	.509 (.046)	.303 (.079)	.726 (.042)	.440 (.085)
4	480	60	180	120	.460 (.052)	.307 (.074)	.496 (.049)	.333 (.074)	.692 (.044)	.483 (.074)
5	600	60	180	120	.453 (.049)	.326 (.073)	.486 (.048)	.350 (.074)	.672 (.045)	.504 (.071)
6	720	60	180	120	.450 (.051)	.339 (.069)	.480 (.050)	.362 (.069)	.657 (.046)	.517 (.067)
10	1,200	60	180	120	.445 (.051)	.371 (.066)	.468 (.051)	.391 (.066)	.630 (.048)	.547 (.062)
25	3,000	60	180	120	.443 (.052)	.412 (.056)	.458 (.052)	.427 (.056)	.605 (.050)	.572 (.054)
50	6,000	60	180	120	.445 (.051)	.429 (.054)	.458 (.052)	.442 (.054)	.600 (.048)	.583 (.050)
100	12,000	60	180	120	.444 (.054)	.436 (.056)	.455 (.054)	.447 (.056)	.594 (.050)	.586 (.052)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.394 (.059)	.032 (.038)	.469 (.060)	.036 (.041)	.993 (.010)	.018 (.025)
1.25	150	60	180	120	.349 (.058)	.034 (.037)	.414 (.057)	.037 (.039)	.833 (.042)	.021 (.027)
1.50	180	60	180	120	.319 (.053)	.034 (.035)	.374 (.055)	.036 (.036)	.716 (.047)	.024 (.029)
1.75	210	60	180	120	.293 (.049)	.036 (.035)	.343 (.052)	.038 (.036)	.636 (.047)	.027 (.030)
2	240	60	180	120	.274 (.048)	.038 (.033)	.316 (.051)	.041 (.034)	.573 (.046)	.032 (.031)
3	360	60	180	120	.217 (.040)	.041 (.030)	.245 (.042)	.046 (.032)	.433 (.041)	.043 (.032)
4	**480**	**60**	**180**	**120**	**.185 (.034)**	**.045 (.028)**	**.205 (.036)**	**.050 (.029)**	**.361 (.038)**	**.056 (.035)**
5	600	60	180	120	.165 (.034)	.048 (.026)	.180 (.034)	.052 (.027)	.316 (.038)	.064 (.035)
6	720	60	180	120	.149 (.029)	.050 (.026)	.161 (.030)	.053 (.027)	.288 (.034)	.070 (.033)
10	1,200	60	180	120	.117 (.023)	.056 (.023)	.124 (.023)	.059 (.023)	.230 (.032)	.089 (.034)
25	3,000	60	180	120	.088 (.017)	.062 (.019)	.090 (.016)	.064 (.019)	.179 (.027)	.117 (.031)
50	6,000	60	180	120	.077 (.014)	.064 (.016)	.079 (.014)	.066 (.016)	.162 (.027)	.130 (.029)
100	12,000	60	180	120	.072 (.013)	.066 (.014)	.073 (.013)	.067 (.014)	.154 (.026)	.137 (.027)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.460 (.061)	.096 (.073)	.532 (.059)	.109 (.079)	.995 (.006)	.026 (.033)
1.25	150	60	180	120	.434 (.061)	.104 (.072)	.495 (.059)	.116 (.076)	.893 (.030)	.094 (.072)
1.50	**180**	**60**	**180**	**120**	**.411 (.060)**	**.116 (.071)**	**.467 (.059)**	**.130 (.076)**	**.819 (.037)**	**.141 (.081)**
1.75	210	60	180	120	.393 (.058)	.121 (.068)	.443 (.059)	.136 (.072)	.766 (.042)	.185 (.084)
2	240	60	180	120	.380 (.056)	.128 (.064)	.426 (.057)	.142 (.068)	.727 (.045)	.208 (.090)
3	360	60	180	120	.341 (.056)	.149 (.062)	.376 (.059)	.163 (.065)	.633 (.049)	.281 (.085)
4	480	60	180	120	.319 (.054)	.167 (.059)	.346 (.055)	.180 (.062)	.589 (.050)	.324 (.079)
5	600	60	180	120	.307 (.052)	.179 (.058)	.329 (.054)	.191 (.060)	.560 (.052)	.350 (.074)
6	720	60	180	120	.296 (.054)	.183 (.058)	.316 (.056)	.195 (.060)	.539 (.053)	.358 (.073)
10	1,200	60	180	120	.275 (.054)	.203 (.058)	.288 (.055)	.212 (.059)	.504 (.055)	.397 (.069)
25	3,000	60	180	120	.254 (.052)	.223 (.053)	.262 (.052)	.230 (.054)	.468 (.057)	.424 (.062)
50	6,000	60	180	120	.253 (.052)	.238 (.052)	.259 (.053)	.243 (.053)	.461 (.058)	.440 (.060)
100	12,000	60	180	120	.251 (.049)	.243 (.051)	.255 (.050)	.248 (.051)	.459 (.055)	.448 (.057)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.505 (.061)	.151 (.093)	.576 (.055)	.171 (.096)	.996 (.005)	.032 (.043)
1.25	150	60	180	120	.487 (.060)	.172 (.089)	.550 (.056)	.194 (.090)	.917 (.023)	.154 (.091)
1.50	**180**	**60**	**180**	**120**	**.469 (.061)**	**.178 (.085)**	**.525 (.057)**	**.200 (.089)**	**.857 (.032)**	**.231 (.099)**
1.75	210	60	180	120	.455 (.061)	.198 (.083)	.504 (.058)	.221 (.084)	.816 (.036)	.291 (.094)
2	240	60	180	120	.446 (.057)	.206 (.083)	.491 (.056)	.227 (.085)	.786 (.037)	.329 (.094)
3	360	60	180	120	.425 (.057)	.243 (.077)	.460 (.057)	.264 (.078)	.717 (.044)	.422 (.084)
4	480	60	180	120	.410 (.057)	.266 (.072)	.439 (.057)	.285 (.074)	.680 (.044)	.460 (.075)
5	600	60	180	120	.403 (.059)	.280 (.071)	.428 (.059)	.299 (.073)	.658 (.046)	.486 (.069)
6	720	60	180	120	.396 (.058)	.290 (.070)	.419 (.060)	.308 (.072)	.644 (.048)	.498 (.067)
10	1,200	60	180	120	.382 (.059)	.315 (.067)	.399 (.059)	.329 (.068)	.614 (.048)	.528 (.062)
25	3,000	60	180	120	.375 (.059)	.347 (.064)	.387 (.060)	.358 (.065)	.592 (.046)	.559 (.053)
50	6,000	60	180	120	.369 (.060)	.355 (.062)	.378 (.061)	.364 (.064)	.579 (.050)	.562 (.053)
100	12,000	60	180	120	.369 (.059)	.362 (.060)	.377 (.060)	.369 (.061)	.575 (.049)	.567 (.050)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.341 (.079)	.030 (.038)	.407 (.088)	.032 (.039)	.993 (.010)	.018 (.024)
1.25	150	60	180	120	.302 (.076)	.027 (.033)	.355 (.087)	.030 (.033)	.836 (.039)	.020 (.027)
1.50	180	60	180	120	.273 (.071)	.029 (.034)	.319 (.082)	.031 (.034)	.721 (.044)	.026 (.030)
1.75	210	60	180	120	.250 (.069)	.028 (.030)	.291 (.080)	.031 (.032)	.641 (.046)	.029 (.032)
2	**240**	**60**	**180**	**120**	**.226 (.065)**	**.030 (.029)**	**.262 (.075)**	**.033 (.031)**	**.575 (.046)**	**.033 (.031)**
3	360	60	180	120	.176 (.055)	.032 (.027)	.197 (.062)	.035 (.028)	.438 (.045)	.049 (.035)
4	480	60	180	120	.148 (.046)	.035 (.026)	.163 (.052)	.038 (.028)	.368 (.040)	.061 (.037)
5	600	60	180	120	.126 (.043)	.036 (.024)	.136 (.045)	.038 (.024)	.322 (.037)	.069 (.035)
6	720	60	180	120	.112 (.037)	.037 (.022)	.119 (.039)	.040 (.023)	.296 (.036)	.077 (.036)
10	1,200	60	180	120	.086 (.029)	.041 (.021)	.089 (.028)	.042 (.021)	.240 (.034)	.100 (.036)
25	3,000	60	180	120	.058 (.016)	.042 (.017)	.059 (.015)	.043 (.017)	.187 (.032)	.126 (.035)
50	6,000	60	180	120	.050 (.012)	.042 (.014)	.050 (.012)	.042 (.014)	.170 (.031)	.137 (.032)
100	12,000	60	180	120	.046 (.011)	.043 (.012)	.047 (.011)	.043 (.012)	.164 (.031)	.148 (.032)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.407 (.084)	.080 (.068)	.472 (.090)	.089 (.073)	.995 (.006)	.025 (.033)
1.25	150	60	180	120	.380 (.083)	.082 (.065)	.432 (.090)	.092 (.069)	.891 (.031)	.087 (.067)
1.50	**180**	**60**	**180**	**120**	**.356 (.084)**	**.089 (.062)**	**.406 (.090)**	**.100 (.067)**	**.815 (.038)**	**.136 (.080)**
1.75	210	60	180	120	.339 (.080)	.095 (.065)	.383 (.086)	.106 (.068)	.763 (.043)	.177 (.081)
2	240	60	180	120	.322 (.077)	.100 (.062)	.362 (.085)	.112 (.066)	.722 (.044)	.206 (.087)
3	360	60	180	120	.285 (.076)	.117 (.060)	.313 (.081)	.128 (.063)	.628 (.050)	.276 (.082)
4	480	60	180	120	.268 (.076)	.133 (.065)	.289 (.080)	.143 (.068)	.582 (.050)	.313 (.076)
5	600	60	180	120	.247 (.072)	.138 (.064)	.264 (.075)	.146 (.067)	.554 (.053)	.342 (.077)
6	720	60	180	120	.239 (.069)	.143 (.060)	.253 (.073)	.151 (.063)	.530 (.051)	.352 (.068)
10	1,200	60	180	120	.219 (.068)	.160 (.064)	.228 (.070)	.166 (.066)	.500 (.052)	.393 (.064)
25	3,000	60	180	120	.195 (.061)	.170 (.061)	.199 (.061)	.173 (.061)	.463 (.054)	.420 (.060)
50	6,000	60	180	120	.186 (.056)	.173 (.056)	.190 (.057)	.176 (.057)	.454 (.053)	.433 (.056)
100	12,000	60	180	120	.184 (.055)	.178 (.056)	.187 (.056)	.181 (.057)	.451 (.056)	.441 (.057)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	60	180	120	.456 (.080)	.126 (.086)	.523 (.080)	.141 (.090)	.996 (.006)	.030 (.040)
1.25	150	60	180	120	.436 (.076)	.138 (.084)	.491 (.079)	.156 (.090)	.914 (.025)	.150 (.090)
1.50	**180**	**60**	**180**	**120**	**.419 (.078)**	**.147 (.081)**	**.469 (.082)**	**.164 (.085)**	**.854 (.032)**	**.222 (.095)**
1.75	210	60	180	120	.405 (.079)	.165 (.082)	.452 (.081)	.183 (.087)	.811 (.035)	.280 (.094)
2	240	60	180	120	.399 (.080)	.172 (.080)	.440 (.083)	.191 (.085)	.780 (.040)	.320 (.093)
3	360	60	180	120	.373 (.078)	.202 (.081)	.406 (.081)	.220 (.083)	.706 (.043)	.406 (.084)
4	480	60	180	120	.353 (.078)	.221 (.079)	.378 (.081)	.237 (.082)	.670 (.044)	.449 (.077)
5	600	60	180	120	.341 (.079)	.230 (.078)	.362 (.080)	.244 (.079)	.647 (.047)	.469 (.070)
6	720	60	180	120	.341 (.079)	.242 (.078)	.360 (.081)	.256 (.082)	.634 (.046)	.488 (.066)
10	1,200	60	180	120	.327 (.077)	.266 (.079)	.341 (.080)	.276 (.081)	.606 (.049)	.519 (.061)
25	3,000	60	180	120	.311 (.077)	.285 (.078)	.320 (.079)	.293 (.080)	.581 (.048)	.546 (.053)
50	6,000	60	180	120	.304 (.077)	.291 (.077)	.310 (.078)	.296 (.079)	.569 (.050)	.552 (.052)
100	12,000	60	180	120	.304 (.076)	.297 (.077)	.309 (.077)	.302 (.078)	.566 (.050)	.557 (.051)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.477 (.043)	.065 (.049)	.564 (.038)	.076 (.053)	.997 (.005)	.014 (.019)
1.25	255	102	306	204	.452 (.039)	.075 (.047)	.529 (.034)	.088 (.051)	.875 (.025)	.055 (.043)
1.50	306	102	306	204	.431 (.039)	.087 (.048)	.501 (.036)	.100 (.051)	.790 (.034)	.089 (.054)
1.75	**357**	**102**	**306**	**204**	**.411 (.038)**	**.092 (.048)**	**.475 (.034)**	**.105 (.051)**	**.728 (.037)**	**.117 (.060)**
2	408	102	306	204	.399 (.038)	.101 (.046)	.459 (.035)	.118 (.049)	.685 (.039)	.142 (.060)
3	612	102	306	204	.365 (.035)	.128 (.046)	.413 (.034)	.146 (.048)	.580 (.045)	.202 (.064)
4	816	102	306	204	.345 (.036)	.147 (.045)	.385 (.035)	.166 (.046)	.526 (.046)	.237 (.064)
5	1,020	102	306	204	.332 (.034)	.161 (.043)	.367 (.034)	.180 (.045)	.493 (.047)	.258 (.063)
6	1,224	102	306	204	.325 (.035)	.176 (.044)	.356 (.035)	.193 (.046)	.474 (.046)	.275 (.062)
10	2,040	102	306	204	.307 (.035)	.207 (.042)	.331 (.035)	.224 (.043)	.428 (.048)	.306 (.058)
25	5,100	102	306	204	.295 (.036)	.249 (.040)	.310 (.037)	.263 (.041)	.391 (.048)	.342 (.054)
50	10,200	102	306	204	.292 (.039)	.268 (.041)	.304 (.040)	.279 (.042)	.381 (.052)	.356 (.054)
100	20,400	102	306	204	.289 (.040)	.277 (.041)	.299 (.041)	.286 (.042)	.372 (.053)	.360 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.526 (.043)	.147 (.071)	.602 (.036)	.170 (.076)	.998 (.003)	.019 (.026)
1.25	255	102	306	204	.508 (.041)	.168 (.068)	.578 (.036)	.192 (.072)	.916 (.018)	.148 (.072)
1.50	**306**	**102**	**306**	**204**	**.496 (.039)**	**.188 (.068)**	**.558 (.035)**	**.215 (.071)**	**.859 (.024)**	**.236 (.077)**
1.75	357	102	306	204	.485 (.041)	.202 (.066)	.541 (.037)	.228 (.068)	.818 (.029)	.291 (.077)
2	408	102	306	204	.475 (.039)	.217 (.064)	.528 (.035)	.243 (.065)	.787 (.029)	.331 (.072)
3	612	102	306	204	.457 (.039)	.258 (.062)	.499 (.036)	.286 (.062)	.716 (.033)	.421 (.065)
4	816	102	306	204	.447 (.037)	.288 (.055)	.483 (.035)	.314 (.056)	.680 (.033)	.462 (.060)
5	1,020	102	306	204	.445 (.038)	.310 (.054)	.478 (.037)	.334 (.054)	.661 (.035)	.486 (.055)
6	1,224	102	306	204	.439 (.038)	.322 (.053)	.469 (.038)	.346 (.053)	.645 (.037)	.501 (.055)
10	2,040	102	306	204	.433 (.040)	.359 (.049)	.456 (.039)	.379 (.049)	.615 (.039)	.530 (.047)
25	5,100	102	306	204	.434 (.040)	.402 (.045)	.450 (.040)	.417 (.045)	.594 (.038)	.560 (.043)
50	10,200	102	306	204	.432 (.041)	.416 (.043)	.445 (.041)	.429 (.043)	.584 (.039)	.567 (.041)
100	20,400	102	306	204	.431 (.041)	.422 (.042)	.442 (.041)	.434 (.043)	.578 (.039)	.570 (.040)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.553 (.045)	.203 (.076)	.622 (.038)	.231 (.080)	.998 (.002)	.024 (.030)
1.25	255	102	306	204	.535 (.042)	.227 (.076)	.599 (.035)	.257 (.077)	.929 (.016)	.206 (.080)
1.50	**306**	**102**	**306**	**204**	**.528 (.041)**	**.249 (.072)**	**.584 (.036)**	**.279 (.075)**	**.880 (.021)**	**.305 (.080)**
1.75	357	102	306	204	.520 (.041)	.267 (.072)	.573 (.036)	.296 (.072)	.848 (.024)	.369 (.077)
2	408	102	306	204	.513 (.039)	.282 (.067)	.561 (.036)	.311 (.068)	.820 (.026)	.412 (.071)
3	612	102	306	204	.502 (.040)	.334 (.064)	.541 (.038)	.362 (.063)	.762 (.030)	.510 (.060)
4	816	102	306	204	.496 (.039)	.358 (.058)	.529 (.037)	.384 (.058)	.732 (.031)	.542 (.055)
5	1,020	102	306	204	.491 (.039)	.379 (.055)	.521 (.037)	.403 (.054)	.715 (.031)	.567 (.049)
6	1,224	102	306	204	.490 (.039)	.395 (.054)	.517 (.038)	.417 (.053)	.702 (.032)	.580 (.047)
10	2,040	102	306	204	.487 (.039)	.425 (.049)	.508 (.039)	.444 (.049)	.677 (.034)	.604 (.044)
25	5,100	102	306	204	.486 (.040)	.461 (.044)	.501 (.040)	.475 (.044)	.656 (.034)	.628 (.037)
50	10,200	102	306	204	.487 (.040)	.474 (.043)	.500 (.040)	.487 (.043)	.650 (.033)	.636 (.035)
100	20,400	102	306	204	.488 (.039)	.481 (.040)	.499 (.040)	.492 (.040)	.647 (.034)	.640 (.034)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.372 (.063)	.046 (.039)	.428 (.072)	.051 (.042)	.997 (.005)	.012 (.017)
1.25	255	102	306	204	.333 (.058)	.047 (.038)	.380 (.069)	.051 (.040)	.854 (.028)	.031 (.033)
1.50	306	102	306	204	.302 (.056)	.051 (.036)	.340 (.063)	.056 (.037)	.753 (.034)	.047 (.036)
1.75	**357**	**102**	**306**	**204**	**.277 (.056)**	**.053 (.033)**	**.308 (.060)**	**.058 (.034)**	**.682 (.036)**	**.058 (.038)**
2	408	102	306	204	.255 (.053)	.055 (.032)	.282 (.057)	.060 (.033)	.628 (.037)	.073 (.041)
3	612	102	306	204	.204 (.043)	.061 (.030)	.220 (.045)	.065 (.031)	.505 (.037)	.111 (.045)
4	816	102	306	204	.174 (.036)	.064 (.027)	.184 (.036)	.067 (.028)	.444 (.037)	.137 (.046)
5	1,020	102	306	204	.156 (.033)	.068 (.026)	.165 (.032)	.071 (.027)	.408 (.039)	.155 (.047)
6	1,224	102	306	204	.145 (.030)	.071 (.025)	.151 (.029)	.073 (.026)	.384 (.036)	.168 (.045)
10	2,040	102	306	204	.120 (.024)	.075 (.024)	.122 (.022)	.076 (.023)	.334 (.038)	.199 (.043)
25	5,100	102	306	204	.094 (.015)	.077 (.018)	.096 (.015)	.078 (.018)	.290 (.037)	.233 (.041)
50	10,200	102	306	204	.086 (.014)	.078 (.016)	.087 (.014)	.079 (.015)	.275 (.037)	.246 (.039)
100	20,400	102	306	204	.082 (.013)	.078 (.014)	.083 (.013)	.079 (.014)	.267 (.038)	.252 (.039)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.458 (.061)	.121 (.063)	.518 (.063)	.137 (.067)	.998 (.003)	.019 (.026)
1.25	255	102	306	204	.433 (.061)	.133 (.062)	.483 (.064)	.148 (.065)	.908 (.021)	.128 (.065)
1.50	**306**	**102**	**306**	**204**	**.410 (.061)**	**.145 (.062)**	**.455 (.065)**	**.161 (.065)**	**.846 (.026)**	**.205 (.073)**
1.75	357	102	306	204	.393 (.064)	.152 (.061)	.432 (.066)	.167 (.064)	.803 (.028)	.252 (.072)
2	408	102	306	204	.381 (.060)	.161 (.060)	.416 (.065)	.175 (.063)	.772 (.031)	.295 (.075)
3	612	102	306	204	.345 (.058)	.183 (.058)	.368 (.061)	.195 (.062)	.695 (.035)	.381 (.065)
4	816	102	306	204	.326 (.057)	.198 (.055)	.345 (.059)	.209 (.058)	.656 (.035)	.423 (.059)
5	1,020	102	306	204	.314 (.056)	.206 (.056)	.329 (.058)	.216 (.058)	.634 (.038)	.445 (.059)
6	1,224	102	306	204	.302 (.056)	.212 (.057)	.317 (.057)	.222 (.058)	.615 (.038)	.462 (.054)
10	2,040	102	306	204	.283 (.054)	.227 (.057)	.293 (.054)	.234 (.057)	.585 (.038)	.492 (.049)
25	5,100	102	306	204	.263 (.049)	.239 (.051)	.268 (.049)	.244 (.051)	.558 (.040)	.522 (.046)
50	10,200	102	306	204	.254 (.046)	.242 (.048)	.258 (.047)	.246 (.048)	.549 (.039)	.531 (.042)
100	20,400	102	306	204	.250 (.046)	.244 (.046)	.253 (.046)	.247 (.047)	.542 (.039)	.533 (.040)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.512 (.054)	.181 (.080)	.573 (.055)	.203 (.083)	.998 (.003)	.022 (.031)
1.25	255	102	306	204	.489 (.056)	.194 (.073)	.542 (.056)	.216 (.076)	.925 (.017)	.186 (.078)
1.50	**306**	**102**	**306**	**204**	**.477 (.057)**	**.215 (.072)**	**.523 (.056)**	**.238 (.074)**	**.876 (.023)**	**.287 (.080)**
1.75	357	102	306	204	.462 (.056)	.228 (.071)	.504 (.058)	.250 (.073)	.840 (.024)	.356 (.078)
2	408	102	306	204	.454 (.057)	.241 (.071)	.493 (.058)	.263 (.073)	.814 (.026)	.397 (.073)
3	612	102	306	204	.425 (.057)	.270 (.067)	.455 (.059)	.290 (.070)	.749 (.030)	.484 (.061)
4	816	102	306	204	.420 (.056)	.294 (.063)	.444 (.057)	.310 (.064)	.721 (.030)	.525 (.054)
5	1,020	102	306	204	.409 (.059)	.304 (.064)	.430 (.060)	.321 (.066)	.700 (.033)	.544 (.052)
6	1,224	102	306	204	.403 (.060)	.315 (.066)	.422 (.061)	.330 (.067)	.689 (.032)	.562 (.049)
10	2,040	102	306	204	.397 (.058)	.340 (.062)	.410 (.059)	.351 (.064)	.664 (.034)	.588 (.042)
25	5,100	102	306	204	.380 (.058)	.357 (.060)	.389 (.059)	.365 (.061)	.641 (.035)	.613 (.039)
50	10,200	102	306	204	.379 (.058)	.367 (.059)	.387 (.059)	.374 (.060)	.636 (.034)	.621 (.036)
100	20,400	102	306	204	.376 (.057)	.370 (.058)	.383 (.058)	.376 (.059)	.632 (.035)	.624 (.036)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.306 (.096)	.036 (.034)	.355 (.114)	.039 (.035)	.996 (.005)	.012 (.016)
1.25	255	102	306	204	.264 (.089)	.035 (.032)	.302 (.103)	.039 (.033)	.855 (.028)	.030 (.030)
1.50	**306**	**102**	**306**	**204**	**.240 (.085)**	**.035 (.030)**	**.269 (.097)**	**.039 (.033)**	**.756 (.034)**	**.049 (.037)**
1.75	357	102	306	204	.212 (.077)	.037 (.029)	.236 (.090)	.041 (.031)	.687 (.036)	.066 (.042)
2	408	102	306	204	.198 (.074)	.039 (.028)	.218 (.086)	.043 (.030)	.634 (.038)	.079 (.043)
3	612	102	306	204	.153 (.062)	.044 (.028)	.161 (.065)	.046 (.030)	.512 (.039)	.117 (.047)
4	816	102	306	204	.124 (.052)	.045 (.026)	.129 (.054)	.047 (.028)	.451 (.038)	.143 (.046)
5	1,020	102	306	204	.109 (.043)	.045 (.025)	.112 (.043)	.047 (.026)	.414 (.038)	.162 (.048)
6	1,224	102	306	204	.099 (.040)	.047 (.026)	.099 (.034)	.047 (.025)	.390 (.037)	.175 (.047)
10	2,040	102	306	204	.076 (.027)	.047 (.022)	.076 (.023)	.047 (.020)	.343 (.038)	.209 (.044)
25	5,100	102	306	204	.055 (.013)	.045 (.015)	.056 (.012)	.045 (.014)	.297 (.039)	.242 (.041)
50	10,200	102	306	204	.050 (.010)	.045 (.012)	.050 (.009)	.045 (.011)	.283 (.037)	.254 (.039)
100	20,400	102	306	204	.047 (.008)	.045 (.010)	.047 (.008)	.045 (.010)	.276 (.038)	.261 (.039)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.390 (.097)	.091 (.061)	.441 (.104)	.100 (.064)	.998 (.004)	.016 (.023)
1.25	**255**	**102**	**306**	**204**	**.361 (.096)**	**.100 (.061)**	**.404 (.107)**	**.111 (.063)**	**.905 (.021)**	**.124 (.066)**
1.50	306	102	306	204	.344 (.091)	.113 (.061)	.380 (.101)	.123 (.065)	.843 (.026)	.194 (.070)
1.75	357	102	306	204	.319 (.094)	.116 (.062)	.352 (.100)	.127 (.066)	.798 (.029)	.249 (.075)
2	408	102	306	204	.308 (.092)	.121 (.061)	.336 (.099)	.130 (.065)	.763 (.032)	.281 (.072)
3	612	102	306	204	.271 (.085)	.137 (.063)	.286 (.086)	.145 (.064)	.684 (.034)	.368 (.066)
4	816	102	306	204	.256 (.086)	.150 (.068)	.265 (.087)	.155 (.071)	.646 (.038)	.410 (.063)
5	1,020	102	306	204	.240 (.080)	.153 (.068)	.247 (.080)	.156 (.069)	.624 (.038)	.435 (.058)
6	1,224	102	306	204	.231 (.078)	.157 (.067)	.238 (.079)	.161 (.068)	.608 (.038)	.450 (.055)
10	2,040	102	306	204	.204 (.067)	.160 (.063)	.208 (.065)	.163 (.062)	.575 (.038)	.482 (.048)
25	5,100	102	306	204	.179 (.057)	.162 (.057)	.182 (.057)	.164 (.057)	.549 (.039)	.512 (.043)
50	10,200	102	306	204	.172 (.050)	.163 (.051)	.174 (.050)	.165 (.050)	.536 (.039)	.518 (.041)
100	20,400	102	306	204	.168 (.048)	.164 (.049)	.169 (.048)	.165 (.049)	.534 (.040)	.524 (.041)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table C27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	102	306	204	.450 (.088)	.142 (.073)	.505 (.093)	.159 (.078)	.998 (.003)	.023 (.029)
1.25	**255**	**102**	**306**	**204**	**.430 (.084)**	**.162 (.077)**	**.479 (.090)**	**.179 (.082)**	**.923 (.017)**	**.186 (.076)**
1.50	306	102	306	204	.411 (.088)	.170 (.076)	.449 (.094)	.185 (.080)	.873 (.023)	.275 (.076)
1.75	357	102	306	204	.401 (.089)	.183 (.077)	.438 (.097)	.200 (.084)	.837 (.025)	.343 (.076)
2	408	102	306	204	.390 (.088)	.196 (.077)	.422 (.092)	.212 (.081)	.808 (.027)	.384 (.074)
3	612	102	306	204	.367 (.090)	.220 (.080)	.388 (.093)	.233 (.084)	.744 (.031)	.473 (.063)
4	816	102	306	204	.353 (.087)	.236 (.081)	.374 (.091)	.250 (.084)	.712 (.032)	.511 (.058)
5	1,020	102	306	204	.338 (.088)	.243 (.079)	.354 (.089)	.254 (.082)	.693 (.033)	.533 (.051)
6	1,224	102	306	204	.335 (.088)	.253 (.083)	.347 (.091)	.262 (.085)	.681 (.033)	.550 (.049)
10	2,040	102	306	204	.316 (.084)	.265 (.081)	.325 (.086)	.273 (.083)	.655 (.034)	.578 (.043)
25	5,100	102	306	204	.300 (.080)	.280 (.080)	.305 (.082)	.284 (.081)	.632 (.034)	.601 (.038)
50	10,200	102	306	204	.294 (.078)	.283 (.078)	.298 (.079)	.287 (.079)	.625 (.035)	.611 (.036)
100	20,400	102	306	204	.286 (.074)	.281 (.074)	.290 (.075)	.284 (.075)	.621 (.034)	.614 (.035)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix D: Simulation Results for Items with Three Response Options - Item Condition Divided into Response Options

Table D1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.436 (.060)	.037 (.047)	.529 (.052)	.040 (.048)	.992 (.011)	.024 (.035)
1.25	120	48	144	96	.403 (.056)	.036 (.043)	.486 (.049)	.040 (.047)	.838 (.044)	.026 (.032)
1.50	144	48	144	96	.369 (.049)	.035 (.038)	.444 (.045)	.039 (.041)	.717 (.048)	.028 (.034)
1.75	168	48	144	96	.348 (.050)	.036 (.037)	.417 (.046)	.041 (.039)	.641 (.053)	.033 (.036)
2	192	48	144	96	.328 (.049)	.038 (.037)	.391 (.045)	.042 (.040)	.581 (.053)	.034 (.036)
3	**288**	**48**	**144**	**96**	**.275 (.041)**	**.045 (.035)**	**.323 (.039)**	**.052 (.039)**	**.439 (.048)**	**.052 (.041)**
4	384	48	144	96	.245 (.036)	.051 (.035)	.284 (.035)	.058 (.037)	.368 (.042)	.063 (.042)
5	480	48	144	96	.224 (.035)	.056 (.033)	.258 (.035)	.064 (.035)	.325 (.041)	.070 (.038)
6	576	48	144	96	.210 (.034)	.062 (.033)	.240 (.033)	.070 (.035)	.295 (.040)	.079 (.040)
10	960	48	144	96	.181 (.003)	.076 (.032)	.202 (.030)	.085 (.034)	.239 (.035)	.099 (.038)
25	2,400	48	144	96	.155 (.028)	.103 (.029)	.166 (.028)	.111 (.030)	.190 (.033)	.128 (.035)
50	4,800	48	144	96	.145 (.027)	.116 (.029)	.153 (.028)	.122 (.029)	.172 (.032)	.139 (.034)
100	9,600	48	144	96	.142 (.027)	.127 (.028)	.148 (.028)	.132 (.028)	.166 (.031)	.149 (.032)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.488 (.061)	.100 (.084)	.573 (.051)	.115 (.089)	.994 (.009)	.030 (.040)
1.25	120	48	144	96	.461 (.059)	.100 (.075)	.538 (.052)	.116 (.083)	.890 (.034)	.088 (.076)
1.50	**144**	**48**	**144**	**96**	**.445 (.056)**	**.112 (.075)**	**.514 (.050)**	**.128 (.079)**	**.814 (.043)**	**.135 (.088)**
1.75	168	48	144	96	.433 (.054)	.121 (.073)	.495 (.050)	.140 (.080)	.760 (.049)	.177 (.094)
2	192	48	144	96	.415 (.057)	.127 (.074)	.474 (.052)	.144 (.077)	.715 (.053)	.196 (.094)
3	288	48	144	96	.386 (.053)	.162 (.072)	.432 (.050)	.181 (.076)	.623 (.058)	.270 (.095)
4	384	48	144	96	.369 (.053)	.186 (.072)	.408 (.050)	.207 (.074)	.575 (.059)	.313 (.092)
5	480	48	144	96	.359 (.052)	.201 (.067)	.394 (.052)	.220 (.069)	.549 (.062)	.335 (.086)
6	576	48	144	96	.352 (.053)	.210 (.067)	.384 (.053)	.230 (.070)	.530 (.064)	.347 (.089)
10	960	48	144	96	.345 (.055)	.251 (.066)	.369 (.056)	.269 (.067)	.498 (.066)	.388 (.080)
25	2,400	48	144	96	.333 (.056)	.291 (.061)	.349 (.057)	.305 (.063)	.463 (.067)	.418 (.074)
50	4,800	48	144	96	.327 (.055)	.305 (.058)	.339 (.056)	.316 (.060)	.446 (.066)	.423 (.069)
100	9,600	48	144	96	.326 (.054)	.315 (.056)	.337 (.056)	.325 (.057)	.443 (.065)	.432 (.067)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.526 (.064)	.148 (.103)	.604 (.055)	.169 (.108)	.995 (.006)	.039 (.051)
1.25	120	48	144	96	.507 (.059)	.166 (.095)	.575 (.052)	.189 (.100)	.913 (.028)	.145 (.096)
1.50	**144**	**48**	**144**	**96**	**.490 (.060)**	**.183 (.091)**	**.552 (.053)**	**.208 (.095)**	**.853 (.037)**	**.227 (.107)**
1.75	168	48	144	96	.480 (.058)	.202 (.095)	.537 (.053)	.225 (.099)	.811 (.043)	.282 (.111)
2	192	48	144	96	.474 (.056)	.211 (.088)	.527 (.051)	.237 (.091)	.781 (.044)	.317 (.103)
3	288	48	144	96	.448 (.058)	.249 (.083)	.492 (.053)	.276 (.083)	.706 (.049)	.405 (.093)
4	384	48	144	96	.438 (.054)	.277 (.080)	.474 (.053)	.301 (.081)	.668 (.052)	.445 (.086)
5	480	48	144	96	.436 (.057)	.301 (.078)	.469 (.056)	.324 (.079)	.648 (.055)	.473 (.084)
6	576	48	144	96	.431 (.055)	.317 (.075)	.460 (.054)	.339 (.075)	.634 (.053)	.489 (.078)
10	960	48	144	96	.423 (.056)	.349 (.070)	.446 (.055)	.367 (.070)	.603 (.056)	.515 (.072)
25	2,400	48	144	96	.420 (.060)	.387 (.065)	.435 (.060)	.402 (.065)	.576 (.059)	.542 (.063)
50	4,800	48	144	96	.422 (.059)	.406 (.062)	.435 (.059)	.418 (.063)	.571 (.058)	.553 (.061)
100	9,600	48	144	96	.423 (.058)	.415 (.061)	.435 (.059)	.426 (.061)	.569 (.056)	.560 (.058)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.395 (.063)	.033 (.041)	.479 (.059)	.035 (.043)	.991 (.012)	.023 (.030)
1.25	120	48	144	96	.356 (.057)	.032 (.038)	.428 (.056)	.033 (.039)	.829 (.046)	.020 (.026)
1.50	144	48	144	96	.322 (.055)	.029 (.034)	.382 (.054)	.032 (.037)	.707 (.052)	.023 (.028)
1.75	168	48	144	96	.297 (.049)	.033 (.037)	.351 (.049)	.036 (.039)	.624 (.052)	.024 (.030)
2	192	48	144	96	.275 (.048)	.032 (.033)	.321 (.049)	.034 (.034)	.557 (.052)	.025 (.027)
3	288	48	144	96	.219 (.039)	.033 (.030)	.251 (.042)	.036 (.032)	.410 (.045)	.031 (.030)
4	384	48	144	96	.183 (.034)	.036 (.027)	.207 (.034)	.039 (.028)	.334 (.039)	.038 (.030)
5	**480**	**48**	**144**	**96**	**.163 (.032)**	**.037 (.026)**	**.181 (.033)**	**.041 (.027)**	**.292 (.037)**	**.043 (.029)**
6	576	48	144	96	.148 (.028)	.041 (.025)	.164 (.029)	.044 (.025)	.262 (.034)	.050 (.029)
10	960	48	144	96	.114 (.022)	.047 (.022)	.123 (.022)	.050 (.023)	.203 (.030)	.065 (.029)
25	2,400	48	144	96	.084 (.017)	.056 (.019)	.087 (.017)	.058 (.020)	.151 (.026)	.088 (.027)
50	4,800	48	144	96	.072 (.015)	.059 (.017)	.075 (.015)	.061 (.017)	.133 (.025)	.100 (.027)
100	9,600	48	144	96	.066 (.014)	.059 (.015)	.068 (.014)	.060 (.015)	.123 (.024)	.106 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.451 (.066)	.084 (.076)	.529 (.062)	.096 (.082)	.994 (.009)	.028 (.038)
1.25	120	48	144	96	.430 (.062)	.090 (.072)	.499 (.058)	.102 (.076)	.885 (.036)	.074 (.070)
1.50	**144**	**48**	**144**	**96**	**.403 (.060)**	**.101 (.071)**	**.463 (.059)**	**.113 (.075)**	**.801 (.045)**	**.117 (.083)**
1.75	168	48	144	96	.388 (.058)	.104 (.070)	.443 (.057)	.119 (.075)	.749 (.046)	.150 (.088)
2	192	48	144	96	.373 (.060)	.112 (.069)	.421 (.058)	.125 (.071)	.705 (.053)	.175 (.089)
3	288	48	144	96	.334 (.054)	.130 (.063)	.370 (.055)	.144 (.066)	.603 (.057)	.241 (.091)
4	384	48	144	96	.309 (.054)	.150 (.065)	.339 (.055)	.164 (.067)	.550 (.060)	.276 (.088)
5	480	48	144	96	.297 (.055)	.158 (.062)	.323 (.056)	.172 (.064)	.523 (.061)	.300 (.085)
6	576	48	144	96	.289 (.054)	.171 (.060)	.310 (.056)	.183 (.062)	.505 (.064)	.320 (.084)
10	960	48	144	96	.264 (.052)	.187 (.057)	.280 (.053)	.198 (.060)	.458 (.063)	.345 (.079)
25	2,400	48	144	96	.245 (.051)	.212 (.054)	.253 (.052)	.219 (.056)	.427 (.063)	.382 (.070)
50	4,800	48	144	96	.240 (.053)	.222 (.055)	.247 (.054)	.228 (.056)	.417 (.065)	.393 (.067)
100	9,600	48	144	96	.238 (.054)	.230 (.055)	.244 (.055)	.235 (.056)	.413 (.065)	.401 (.067)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.502 (.066)	.141 (.095)	.575 (.058)	.160 (.100)	.995 (.007)	.038 (.049)
1.25	120	48	144	96	.481 (.064)	.155 (.094)	.546 (.058)	.175 (.097)	.911 (.030)	.139 (.097)
1.50	**144**	**48**	**144**	**96**	**.464 (.063)**	**.167 (.091)**	**.523 (.058)**	**.189 (.095)**	**.849 (.037)**	**.219 (.109)**
1.75	168	48	144	96	.451 (.061)	.178 (.087)	.505 (.057)	.201 (.090)	.804 (.040)	.258 (.105)
2	192	48	144	96	.437 (.063)	.189 (.086)	.486 (.061)	.211 (.087)	.770 (.046)	.297 (.103)
3	288	48	144	96	.413 (.058)	.222 (.085)	.450 (.057)	.242 (.086)	.693 (.048)	.384 (.094)
4	384	48	144	96	.397 (.061)	.241 (.080)	.429 (.060)	.261 (.081)	.655 (.053)	.423 (.088)
5	480	48	144	96	.390 (.059)	.261 (.077)	.418 (.058)	.280 (.077)	.636 (.053)	.452 (.084)
6	576	48	144	96	.382 (.060)	.270 (.075)	.406 (.061)	.288 (.077)	.616 (.057)	.462 (.081)
10	960	48	144	96	.369 (.059)	.296 (.069)	.387 (.060)	.311 (.070)	.585 (.054)	.493 (.069)
25	2,400	48	144	96	.364 (.062)	.333 (.066)	.376 (.062)	.344 (.067)	.563 (.057)	.527 (.063)
50	4,800	48	144	96	.359 (.063)	.343 (.064)	.368 (.064)	.352 (.065)	.550 (.060)	.532 (.062)
100	9,600	48	144	96	.360 (.061)	.351 (.063)	.368 (.063)	.360 (.064)	.548 (.057)	.538 (.059)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.354 (.075)	.029 (.036)	.424 (.084)	.030 (.038)	.991 (.013)	.022 (.031)
1.25	120	48	144	96	.315 (.074)	.028 (.034)	.373 (.080)	.029 (.036)	.830 (.047)	.022 (.030)
1.50	144	48	144	96	.282 (.069)	.025 (.030)	.333 (.076)	.027 (.032)	.711 (.053)	.022 (.028)
1.75	168	48	144	96	.257 (.062)	.028 (.032)	.303 (.072)	.030 (.034)	.626 (.052)	.024 (.031)
2	192	48	144	96	.235 (.063)	.027 (.028)	.275 (.068)	.029 (.030)	.562 (.051)	.026 (.030)
3	**288**	**48**	**144**	**96**	**.184 (.051)**	**.027 (.026)**	**.208 (.058)**	**.029 (.027)**	**.416 (.046)**	**.033 (.030)**
4	384	48	144	96	.152 (.044)	.029 (.024)	.170 (.049)	.032 (.025)	.341 (.040)	.042 (.031)
5	480	48	144	96	.133 (.040)	.031 (.023)	.146 (.044)	.033 (.024)	.295 (.037)	.047 (.031)
6	576	48	144	96	.117 (.035)	.033 (.023)	.128 (.038)	.035 (.024)	.266 (.036)	.052 (.030)
10	960	48	144	96	.088 (.027)	.036 (.020)	.093 (.029)	.037 (.020)	.209 (.030)	.070 (.031)
25	2,400	48	144	96	.059 (.017)	.040 (.018)	.061 (.016)	.041 (.017)	.157 (.027)	.095 (.030)
50	4,800	48	144	96	.049 (.012)	.040 (.013)	.049 (.011)	.040 (.013)	.137 (.026)	.105 (.028)
100	9,600	48	144	96	.044 (.010)	.039 (.012)	.044 (.010)	.040 (.012)	.128 (.025)	.111 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.408 (.083)	.072 (.070)	.477 (.084)	.080 (.074)	.994 (.009)	.030 (.040)
1.25	120	48	144	96	.385 (.080)	.079 (.067)	.445 (.083)	.088 (.069)	.883 (.036)	.077 (.068)
1.50	**144**	**48**	**144**	**96**	**.357 (.075)**	**.081 (.067)**	**.414 (.080)**	**.091 (.071)**	**.802 (.044)**	**.111 (.081)**
1.75	168	48	144	96	.341 (.072)	.084 (.063)	.392 (.077)	.095 (.067)	.743 (.048)	.142 (.084)
2	192	48	144	96	.331 (.074)	.090 (.065)	.374 (.079)	.100 (.069)	.701 (.053)	.172 (.088)
3	288	48	144	96	.282 (.070)	.106 (.063)	.311 (.075)	.116 (.067)	.598 (.058)	.237 (.091)
4	384	48	144	96	.262 (.069)	.119 (.062)	.286 (.074)	.128 (.065)	.548 (.059)	.271 (.085)
5	480	48	144	96	.250 (.068)	.131 (.062)	.270 (.072)	.140 (.065)	.517 (.060)	.294 (.082)
6	576	48	144	96	.239 (.067)	.135 (.061)	.258 (.070)	.146 (.065)	.499 (.059)	.311 (.078)
10	960	48	144	96	.214 (.062)	.150 (.060)	.225 (.065)	.158 (.063)	.457 (.060)	.344 (.072)
25	2,400	48	144	96	.196 (.061)	.167 (.059)	.201 (.061)	.171 (.060)	.425 (.060)	.378 (.068)
50	4,800	48	144	96	.186 (.054)	.171 (.054)	.190 (.055)	.174 (.055)	.415 (.060)	.391 (.063)
100	9,600	48	144	96	.183 (.057)	.175 (.057)	.186 (.057)	.178 (.058)	.406 (.061)	.394 (.063)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	98	48	144	96	.465 (.081)	.125 (.095)	.533 (.080)	.139 (.099)	.995 (.007)	.038 (.051)
1.25	120	48	144	96	.435 (.077)	.127 (.085)	.494 (.079)	.144 (.090)	.907 (.029)	.137 (.093)
1.50	**144**	**48**	**144**	**96**	**.420 (.076)**	**.142 (.083)**	**.476 (.076)**	**.158 (.088)**	**.845 (.037)**	**.211 (.102)**
1.75	168	48	144	96	.408 (.077)	.152 (.081)	.455 (.077)	.170 (.087)	.800 (.043)	.257 (.104)
2	192	48	144	96	.390 (.076)	.164 (.086)	.436 (.077)	.180 (.088)	.764 (.045)	.293 (.102)
3	288	48	144	96	.368 (.076)	.190 (.084)	.400 (.079)	.207 (.087)	.689 (.048)	.378 (.092)
4	384	48	144	96	.350 (.075)	.204 (.078)	.380 (.076)	.223 (.081)	.648 (.051)	.414 (.085)
5	480	48	144	96	.339 (.075)	.216 (.078)	.364 (.077)	.232 (.081)	.623 (.052)	.434 (.080)
6	576	48	144	96	.333 (.074)	.229 (.077)	.354 (.078)	.243 (.080)	.609 (.055)	.452 (.077)
10	960	48	144	96	.323 (.075)	.257 (.079)	.338 (.077)	.269 (.081)	.581 (.056)	.490 (.071)
25	2,400	48	144	96	.310 (.075)	.281 (.077)	.320 (.078)	.290 (.079)	.553 (.059)	.516 (.064)
50	4,800	48	144	96	.303 (.075)	.289 (.076)	.311 (.077)	.296 (.078)	.543 (.058)	.525 (.060)
100	9,600	48	144	96	.303 (.075)	.295 (.076)	.310 (.076)	.302 (.077)	.537 (.057)	.528 (.058)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.492 (.038)	.082 (.052)	.576 (.033)	.096 (.055)	.998 (.003)	.012 (.018)
1.25	300	120	360	240	.466 (.037)	.094 (.049)	.542 (.033)	.109 (.053)	.886 (.023)	.074 (.048)
1.50	**360**	**120**	**360**	**240**	**.446 (.037)**	**.101 (.048)**	**.515 (.033)**	**.118 (.051)**	**.811 (.029)**	**.120 (.056)**
1.75	420	120	360	240	.431 (.034)	.116 (.049)	.495 (.031)	.134 (.051)	.755 (.031)	.159 (.061)
2	480	120	360	240	.419 (.035)	.124 (.046)	.478 (.032)	.144 (.049)	.714 (.035)	.190 (.063)
3	720	120	360	240	.386 (.034)	.156 (.045)	.434 (.032)	.176 (.048)	.616 (.039)	.257 (.064)
4	960	120	360	240	.369 (.035)	.177 (.045)	.410 (.033)	.199 (.047)	.569 (.041)	.296 (.062)
5	1,200	120	360	240	.361 (.033)	.196 (.044)	.397 (.033)	.216 (.045)	.541 (.042)	.320 (.060)
6	1,440	120	360	240	.356 (.034)	.211 (.043)	.389 (.033)	.233 (.044)	.525 (.042)	.340 (.055)
10	2,400	120	360	240	.343 (.035)	.246 (.040)	.369 (.036)	.265 (.041)	.487 (.045)	.374 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.538 (.040)	.162 (.067)	.611 (.034)	.186 (.069)	.998 (.002)	.017 (.022)
1.25	300	120	360	240	.516 (.039)	.179 (.065)	.583 (.034)	.205 (.068)	.919 (.017)	.165 (.069)
1.50	**360**	**120**	**360**	**240**	**.503 (.038)**	**.201 (.065)**	**.563 (.034)**	**.228 (.068)**	**.864 (.021)**	**.255 (.072)**
1.75	420	120	360	240	.494 (.037)	.222 (.060)	.550 (.033)	.250 (.062)	.827 (.024)	.314 (.071)
2	480	120	360	240	.486 (.037)	.236 (.059)	.538 (.033)	.263 (.062)	.797 (.026)	.356 (.067)
3	720	120	360	240	.469 (.036)	.279 (.055)	.511 (.034)	.306 (.056)	.729 (.029)	.446 (.059)
4	960	120	360	240	.459 (.036)	.305 (.053)	.495 (.034)	.332 (.054)	.694 (.031)	.484 (.055)
5	1,200	120	360	240	.457 (.036)	.329 (.051)	.489 (.035)	.354 (.050)	.676 (.032)	.510 (.050)
6	1,440	120	360	240	.454 (.036)	.343 (.047)	.483 (.035)	.366 (.048)	.663 (.032)	.525 (.047)
10	2,400	120	360	240	.450 (.035)	.378 (.043)	.472 (.035)	.398 (.043)	.636 (.032)	.553 (.041)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.555 (.042)	.212 (.076)	.625 (.035)	.241 (.078)	.998 (.002)	.022 (.030)
1.25	300	120	360	240	.542 (.039)	.236 (.072)	.604 (.034)	.267 (.072)	.932 (.015)	.214 (.077)
1.50	**360**	**120**	**360**	**240**	**.533 (.038)**	**.260 (.071)**	**.589 (.035)**	**.290 (.072)**	**.885 (.019)**	**.321 (.075)**
1.75	420	120	360	240	.527 (.039)	.281 (.068)	.579 (.034)	.311 (.069)	.852 (.021)	.386 (.073)
2	480	120	360	240	.520 (.036)	.294 (.065)	.568 (.033)	.323 (.065)	.827 (.022)	.427 (.066)
3	720	120	360	240	.507 (.037)	.341 (.058)	.546 (.035)	.369 (.058)	.769 (.026)	.517 (.057)
4	960	120	360	240	.500 (.037)	.369 (.054)	.534 (.035)	.396 (.054)	.739 (.027)	.556 (.048)
5	1,200	120	360	240	.503 (.037)	.392 (.051)	.532 (.035)	.416 (.051)	.725 (.028)	.582 (.044)
6	1,440	120	360	240	.502 (.036)	.408 (.048)	.528 (.035)	.431 (.048)	.714 (.028)	.595 (.041)
10	2,400	120	360	240	.497 (.037)	.438 (.047)	.518 (.037)	.457 (.046)	.689 (.030)	.619 (.038)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.367 (.064)	.048 (.037)	.412 (.074)	.053 (.039)	.997 (.004)	.010 (.015)
1.25	300	120	360	240	.324 (.063)	.052 (.035)	.363 (.069)	.057 (.037)	.862 (.025)	.037 (.032)
1.50	**360**	**120**	**360**	**240**	**.294 (.060)**	**.055 (.035)**	**.329 (.067)**	**.060 (.037)**	**.768 (.031)**	**.060 (.039)**
1.75	420	120	360	240	.268 (.057)	.059 (.033)	.295 (.063)	.064 (.036)	.700 (.033)	.079 (.041)
2	480	120	360	240	.250 (.055)	.059 (.033)	.272 (.058)	.065 (.035)	.651 (.034)	.096 (.044)
3	720	120	360	240	.196 (.044)	.066 (.030)	.209 (.045)	.070 (.031)	.534 (.037)	.143 (.047)
4	960	120	360	240	.169 (.037)	.069 (.027)	.179 (.037)	.072 (.028)	.475 (.036)	.171 (.046)
5	1,200	120	360	240	.154 (.033)	.073 (.027)	.159 (.031)	.075 (.027)	.441 (.037)	.192 (.049)
6	1,440	120	360	240	.141 (.029)	.075 (.026)	.146 (.028)	.077 (.027)	.418 (.038)	.208 (.048)
10	2,400	120	360	240	.115 (.020)	.076 (.021)	.117 (.019)	.077 (.021)	.368 (.037)	.237 (.044)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.456 (.065)	.130 (.063)	.510 (.069)	.144 (.066)	.998 (.003)	.016 (.024)
1.25	300	120	360	240	.431 (.064)	.138 (.061)	.475 (.069)	.153 (.066)	.913 (.018)	.142 (.063)
1.50	**360**	**120**	**360**	**240**	**.406 (.065)**	**.153 (.059)**	**.446 (.068)**	**.167 (.063)**	**.853 (.023)**	**.221 (.071)**
1.75	420	120	360	240	.389 (.065)	.160 (.059)	.425 (.069)	.173 (.062)	.811 (.025)	.274 (.068)
2	480	120	360	240	.375 (.065)	.168 (.060)	.406 (.069)	.181 (.063)	.779 (.027)	.319 (.067)
3	720	120	360	240	.340 (.065)	.189 (.059)	.362 (.066)	.200 (.061)	.707 (.030)	.405 (.061)
4	960	120	360	240	.321 (.059)	.200 (.056)	.337 (.060)	.210 (.058)	.671 (.032)	.446 (.053)
5	1,200	120	360	240	.306 (.058)	.208 (.056)	.321 (.060)	.218 (.058)	.648 (.033)	.470 (.050)
6	1,440	120	360	240	.297 (.059)	.215 (.058)	.309 (.059)	.223 (.059)	.636 (.035)	.489 (.049)
10	2,400	120	360	240	.276 (.053)	.224 (.054)	.284 (.053)	.230 (.054)	.605 (.035)	.516 (.044)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.507 (.056)	.185 (.073)	.568 (.056)	.207 (.076)	.998 (.002)	.020 (.027)
1.25	300	120	360	240	.492 (.054)	.209 (.072)	.543 (.056)	.232 (.075)	.929 (.015)	.204 (.071)
1.50	**360**	**120**	**360**	**240**	**.476 (.056)**	**.223 (.071)**	**.522 (.056)**	**.245 (.074)**	**.881 (.019)**	**.306 (.075)**
1.75	420	120	360	240	.462 (.057)	.234 (.068)	.503 (.060)	.254 (.071)	.847 (.021)	.367 (.071)
2	480	120	360	240	.460 (.057)	.247 (.067)	.496 (.060)	.268 (.069)	.822 (.024)	.412 (.066)
3	720	120	360	240	.430 (.059)	.278 (.064)	.459 (.060)	.298 (.066)	.760 (.027)	.501 (.056)
4	960	120	360	240	.419 (.060)	.300 (.064)	.441 (.061)	.316 (.066)	.731 (.027)	.542 (.049)
5	1,200	120	360	240	.409 (.060)	.310 (.064)	.428 (.061)	.324 (.065)	.713 (.028)	.563 (.046)
6	1,440	120	360	240	.401 (.061)	.317 (.066)	.418 (.062)	.330 (.068)	.701 (.031)	.576 (.046)
10	2,400	120	360	240	.388 (.059)	.334 (.063)	.401 (.060)	.346 (.064)	.676 (.031)	.603 (.039)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.294 (.098)	.038 (.034)	.334 (.116)	.042 (.037)	.997 (.004)	.011 (.015)
1.25	300	120	360	240	.255 (.094)	.038 (.032)	.286 (.110)	.042 (.033)	.863 (.025)	.037 (.031)
1.50	**360**	**120**	**360**	**240**	**.225 (.086)**	**.039 (.029)**	**.248 (.096)**	**.042 (.031)**	**.768 (.030)**	**.059 (.037)**
1.75	420	120	360	240	.206 (.084)	.041 (.029)	.226 (.095)	.044 (.031)	.702 (.032)	.082 (.043)
2	480	120	360	240	.184 (.076)	.041 (.028)	.198 (.082)	.043 (.029)	.652 (.034)	.098 (.043)
3	720	120	360	240	.144 (.061)	.046 (.028)	.152 (.066)	.048 (.030)	.538 (.034)	.146 (.048)
4	960	120	360	240	.119 (.052)	.046 (.027)	.121 (.053)	.047 (.027)	.481 (.036)	.178 (.048)
5	1,200	120	360	240	.101 (.043)	.046 (.026)	.103 (.043)	.047 (.026)	.444 (.034)	.197 (.046)
6	1,440	120	360	240	.090 (.035)	.046 (.023)	.090 (.030)	.046 (.022)	.421 (.037)	.213 (.045)
10	2,400	120	360	240	.073 (.025)	.047 (.022)	.072 (.020)	.047 (.020)	.375 (.038)	.246 (.044)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.379 (.101)	.092 (.063)	.429 (.111)	.105 (.068)	.998 (.003)	.016 (.021)
1.25	**300**	**120**	**360**	**240**	**.355 (.103)**	**.104 (.060)**	**.389 (.109)**	**.114 (.063)**	**.912 (.018)**	**.136 (.063)**
1.50	360	120	360	240	.335 (.101)	.114 (.063)	.367 (.108)	.124 (.066)	.849 (.023)	.211 (.068)
1.75	420	120	360	240	.311 (.098)	.118 (.062)	.340 (.106)	.129 (.067)	.807 (.026)	.262 (.071)
2	480	120	360	240	.297 (.092)	.124 (.062)	.323 (.101)	.134 (.067)	.775 (.028)	.306 (.064)
3	720	120	360	240	.266 (.090)	.139 (.064)	.278 (.092)	.144 (.066)	.699 (.031)	.393 (.059)
4	960	120	360	240	.238 (.083)	.143 (.063)	.249 (.086)	.150 (.067)	.660 (.035)	.433 (.056)
5	1,200	120	360	240	.229 (.080)	.151 (.066)	.234 (.078)	.154 (.067)	.641 (.034)	.459 (.052)
6	1,440	120	360	240	.217 (.075)	.151 (.064)	.222 (.076)	.154 (.066)	.624 (.034)	.472 (.048)
10	2,400	120	360	240	.193 (.068)	.152 (.063)	.197 (.066)	.155 (.062)	.593 (.033)	.503 (.042)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	242	120	360	240	.443 (.092)	.146 (.073)	.497 (.098)	.164 (.078)	.998 (.002)	.021 (.025)
1.25	**300**	**120**	**360**	**240**	**.426 (.090)**	**.162 (.074)**	**.468 (.097)**	**.177 (.077)**	**.926 (.015)**	**.190 (.073)**
1.50	360	120	360	240	.409 (.090)	.174 (.076)	.450 (.098)	.194 (.081)	.877 (.020)	.293 (.074)
1.75	420	120	360	240	.394 (.095)	.188 (.079)	.429 (.102)	.205 (.085)	.841 (.022)	.355 (.072)
2	480	120	360	240	.384 (.093)	.193 (.078)	.413 (.097)	.208 (.083)	.815 (.024)	.399 (.066)
3	720	120	360	240	.353 (.090)	.215 (.079)	.376 (.096)	.230 (.085)	.754 (.026)	.490 (.055)
4	960	120	360	240	.346 (.091)	.236 (.082)	.362 (.095)	.246 (.085)	.724 (.028)	.529 (.050)
5	1,200	120	360	240	.334 (.093)	.245 (.085)	.348 (.094)	.255 (.087)	.706 (.031)	.554 (.047)
6	1,440	120	360	240	.319 (.089)	.244 (.084)	.331 (.091)	.253 (.087)	.691 (.031)	.563 (.045)
10	2,400	120	360	240	.314 (.085)	.268 (.083)	.320 (.085)	.272 (.084)	.669 (.030)	.595 (.039)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.526 (.030)	.132 (.046)	.604 (.025)	.154 (.048)	.999 (.002)	.009 (.012)
1.25	510	204	612	408	.505 (.029)	.153 (.048)	.576 (.025)	.177 (.050)	.911 (.014)	.136 (.049)
1.50	**612**	**204**	**612**	**408**	**.492 (.029)**	**.169 (.045)**	**.557 (.025)**	**.194 (.047)**	**.852 (.018)**	**.214 (.052)**
1.75	714	204	612	408	.479 (.027)	.185 (.046)	.539 (.024)	.211 (.047)	.808 (.020)	.271 (.052)
2	816	204	612	408	.471 (.025)	.199 (.043)	.528 (.023)	.225 (.044)	.777 (.020)	.309 (.053)
3	1,224	204	612	408	.451 (.027)	.243 (.042)	.497 (.025)	.270 (.042)	.703 (.024)	.398 (.046)
4	1,632	204	612	408	.442 (.025)	.272 (.038)	.481 (.024)	.298 (.038)	.666 (.024)	.438 (.040)
5	2,040	204	612	408	.436 (.026)	.290 (.037)	.471 (.025)	.316 (.037)	.644 (.025)	.462 (.039)
6	2,448	204	612	408	.434 (.026)	.310 (.036)	.466 (.025)	.335 (.036)	.630 (.025)	.481 (.037)
10	4,080	204	612	408	.428 (.027)	.347 (.033)	.453 (.027)	.368 (.034)	.599 (.028)	.509 (.035)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.554 (.031)	.206 (.056)	.624 (.026)	.234 (.058)	.999 (.001)	.013 (.017)
1.25	510	204	612	408	.541 (.030)	.229 (.056)	.603 (.026)	.259 (.058)	.930 (.011)	.209 (.060)
1.50	**612**	**204**	**612**	**408**	**.530 (.028)**	**.251 (.052)**	**.587 (.025)**	**.281 (.053)**	**.883 (.014)**	**.314 (.058)**
1.75	714	204	612	408	.523 (.030)	.270 (.051)	.575 (.027)	.300 (.052)	.849 (.017)	.377 (.056)
2	816	204	612	408	.517 (.028)	.288 (.049)	.566 (.025)	.319 (.049)	.824 (.017)	.422 (.051)
3	1,224	204	612	408	.504 (.028)	.332 (.044)	.543 (.026)	.360 (.044)	.764 (.021)	.508 (.043)
4	1,632	204	612	408	.500 (.027)	.364 (.041)	.533 (.026)	.389 (.041)	.737 (.021)	.551 (.036)
5	2,040	204	612	408	.497 (.027)	.382 (.039)	.527 (.026)	.407 (.039)	.719 (.021)	.571 (.034)
6	2,448	204	612	408	.495 (.028)	.398 (.038)	.523 (.027)	.422 (.037)	.707 (.022)	.586 (.033)
10	4,080	204	612	408	.494 (.028)	.432 (.035)	.516 (.027)	.452 (.034)	.685 (.023)	.613 (.030)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.572 (.031)	.244 (.060)	.637 (.026)	.276 (.059)	.999 (.001)	.014 (.018)
1.25	510	204	612	408	.559 (.032)	.273 (.059)	.618 (.027)	.305 (.058)	.938 (.010)	.244 (.060)
1.50	**612**	**204**	**612**	**408**	**.553 (.029)**	**.299 (.053)**	**.606 (.025)**	**.332 (.054)**	**.897 (.013)**	**.365 (.059)**
1.75	714	204	612	408	.548 (.029)	.320 (.053)	.596 (.026)	.352 (.053)	.867 (.015)	.432 (.054)
2	816	204	612	408	.542 (.028)	.333 (.050)	.588 (.025)	.364 (.049)	.845 (.016)	.476 (.051)
3	1,224	204	612	408	.535 (.030)	.385 (.045)	.570 (.027)	.413 (.045)	.794 (.019)	.566 (.038)
4	1,632	204	612	408	.532 (.028)	.412 (.042)	.562 (.026)	.438 (.041)	.768 (.019)	.601 (.034)
5	2,040	204	612	408	.530 (.028)	.432 (.039)	.557 (.027)	.455 (.038)	.752 (.020)	.621 (.032)
6	2,448	204	612	408	.527 (.029)	.444 (.039)	.553 (.028)	.467 (.038)	.741 (.020)	.635 (.030)
10	4,080	204	612	408	.528 (.029)	.477 (.035)	.548 (.028)	.495 (.035)	.722 (.021)	.659 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.314 (.075)	.063 (.035)	.342 (.082)	.067 (.037)	.999 (.002)	.007 (.010)
1.25	**510**	**204**	**612**	**408**	**.277 (.070)**	**.066 (.034)**	**.299 (.077)**	**.070 (.036)**	**.885 (.018)**	**.072 (.036)**
1.50	612	204	612	408	.246 (.065)	.069 (.032)	.260 (.066)	.073 (.034)	.809 (.021)	.119 (.042)
1.75	714	204	612	408	.226 (.060)	.072 (.032)	.236 (.058)	.074 (.033)	.752 (.024)	.153 (.046)
2	816	204	612	408	.206 (.052)	.071 (.030)	.215 (.049)	.074 (.031)	.712 (.025)	.186 (.047)
3	1,224	204	612	408	.163 (.039)	.075 (.029)	.167 (.034)	.076 (.028)	.617 (.027)	.255 (.045)
4	1,632	204	612	408	.142 (.029)	.075 (.024)	.144 (.026)	.076 (.024)	.569 (.027)	.295 (.043)
5	2,040	204	612	408	.126 (.024)	.076 (.023)	.128 (.020)	.077 (.022)	.539 (.029)	.316 (.042)
6	2,448	204	612	408	.119 (.022)	.077 (.022)	.121 (.018)	.077 (.021)	.521 (.030)	.335 (.043)
10	4,080	204	612	408	.101 (.014)	.077 (.018)	.103 (.013)	.078 (.018)	.482 (.030)	.370 (.037)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.436 (.074)	.146 (.059)	.476 (.080)	.158 (.063)	.999 (.001)	.011 (.015)
1.25	**510**	**204**	**612**	**408**	**.406 (.074)**	**.159 (.057)**	**.439 (.079)**	**.170 (.060)**	**.924 (.012)**	**.183 (.053)**
1.50	612	204	612	408	.378 (.071)	.167 (.054)	.408 (.075)	.178 (.058)	.873 (.016)	.281 (.056)
1.75	714	204	612	408	.358 (.071)	.168 (.057)	.382 (.073)	.179 (.060)	.837 (.018)	.346 (.054)
2	816	204	612	408	.345 (.070)	.173 (.057)	.366 (.073)	.183 (.060)	.808 (.019)	.381 (.051)
3	1,224	204	612	408	.311 (.066)	.188 (.057)	.325 (.066)	.196 (.059)	.746 (.021)	.472 (.042)
4	1,632	204	612	408	.291 (.061)	.199 (.056)	.301 (.061)	.205 (.057)	.714 (.023)	.515 (.040)
5	2,040	204	612	408	.278 (.058)	.203 (.054)	.286 (.058)	.208 (.054)	.696 (.023)	.539 (.036)
6	2,448	204	612	408	.267 (.052)	.201 (.050)	.274 (.053)	.206 (.051)	.681 (.022)	.550 (.034)
10	4,080	204	612	408	.247 (.047)	.208 (.047)	.252 (.046)	.212 (.047)	.657 (.025)	.580 (.031)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.494 (.065)	.199 (.063)	.544 (.066)	.219 (.066)	.999 (.001)	.012 (.016)
1.25	510	204	612	408	.480 (.062)	.223 (.066)	.525 (.066)	.244 (.068)	.936 (.010)	.232 (.059)
1.50	**612**	**204**	**612**	**408**	**.462 (.065)**	**.236 (.066)**	**.500 (.067)**	**.256 (.070)**	**.892 (.013)**	**.349 (.059)**
1.75	714	204	612	408	.448 (.065)	.245 (.064)	.482 (.067)	.265 (.066)	.861 (.015)	.411 (.055)
2	816	204	612	408	.445 (.062)	.259 (.063)	.474 (.065)	.275 (.066)	.839 (.017)	.460 (.051)
3	1,224	204	612	408	.414 (.065)	.280 (.064)	.436 (.067)	.294 (.066)	.784 (.019)	.546 (.041)
4	1,632	204	612	408	.401 (.065)	.297 (.065)	.421 (.066)	.311 (.067)	.757 (.020)	.584 (.037)
5	2,040	204	612	408	.388 (.063)	.303 (.064)	.403 (.064)	.314 (.066)	.740 (.020)	.605 (.032)
6	2,448	204	612	408	.383 (.063)	.310 (.064)	.395 (.064)	.319 (.066)	.729 (.020)	.617 (.031)
10	4,080	204	612	408	.368 (.061)	.324 (.061)	.378 (.061)	.333 (.062)	.710 (.022)	.644 (.028)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.241 (.113)	.045 (.033)	.260 (.125)	.047 (.034)	.999 (.002)	.007 (.010)
1.25	**510**	**204**	**612**	**408**	**.200 (.095)**	**.043 (.030)**	**.209 (.102)**	**.045 (.031)**	**.884 (.017)**	**.069 (.036)**
1.50	612	204	612	408	.179 (.092)	.045 (.030)	.187 (.096)	.047 (.031)	.807 (.021)	.114 (.041)
1.75	714	204	612	408	.159 (.085)	.045 (.030)	.166 (.091)	.046 (.032)	.750 (.024)	.150 (.044)
2	816	204	612	408	.138 (.069)	.045 (.030)	.142 (.072)	.046 (.031)	.709 (.026)	.178 (.046)
3	1,224	204	612	408	.103 (.050)	.045 (.026)	.105 (.045)	.046 (.025)	.612 (.026)	.251 (.044)
4	1,632	204	612	408	.088 (.040)	.046 (.026)	.087 (.034)	.045 (.024)	.564 (.028)	.289 (.043)
5	2,040	204	612	408	.075 (.027)	.045 (.021)	.075 (.024)	.045 (.020)	.535 (.028)	.312 (.039)
6	2,448	204	612	408	.069 (.025)	.044 (.021)	.069 (.019)	.043 (.019)	.515 (.029)	.330 (.039)
10	4,080	204	612	408	.056 (.012)	.042 (.014)	.056 (.011)	.042 (.014)	.477 (.030)	.364 (.036)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into three binary items; "Dummy-coded Options" are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.344 (.116)	.103 (.059)	.380 (.129)	.112 (.064)	.999 (.001)	.011 (.015)
1.25	**510**	**204**	**612**	**408**	**.313 (.112)**	**.109 (.061)**	**.337 (.121)**	**.117 (.064)**	**.922 (.012)**	**.171 (.053)**
1.50	612	204	612	408	.297 (.108)	.118 (.063)	.318 (.117)	.126 (.069)	.868 (.016)	.267 (.058)
1.75	714	204	612	408	.275 (.105)	.119 (.063)	.291 (.111)	.125 (.068)	.831 (.018)	.329 (.056)
2	816	204	612	408	.265 (.103)	.124 (.064)	.275 (.104)	.128 (.066)	.804 (.019)	.370 (.051)
3	1,224	204	612	408	.223 (.092)	.130 (.066)	.228 (.090)	.133 (.066)	.738 (.022)	.462 (.043)
4	1,632	204	612	408	.208 (.085)	.135 (.066)	.211 (.082)	.136 (.065)	.706 (.023)	.502 (.039)
5	2,040	204	612	408	.187 (.072)	.130 (.060)	.191 (.070)	.132 (.059)	.687 (.024)	.525 (.036)
6	2,448	204	612	408	.176 (.064)	.130 (.056)	.179 (.061)	.132 (.055)	.673 (.024)	.539 (.035)
10	4,080	204	612	408	.158 (.052)	.131 (.050)	.160 (.049)	.132 (.048)	.647 (.024)	.568 (.030)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table D27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Three Response Options where the Number of Items for the Simulation Condition are Divided into Three Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	410	204	612	408	.427 (.103)	.159 (.073)	.471 (.115)	.175 (.079)	.999 (.001)	.012 (.016)
1.25	**510**	**204**	**612**	**408**	**.403 (.104)**	**.169 (.074)**	**.438 (.114)**	**.184 (.082)**	**.934 (.011)**	**.224 (.059)**
1.50	612	204	612	408	.385 (.102)	.181 (.074)	.416 (.110)	.196 (.081)	.889 (.014)	.336 (.060)
1.75	714	204	612	408	.375 (.104)	.194 (.080)	.401 (.110)	.206 (.085)	.858 (.016)	.402 (.053)
2	816	204	612	408	.351 (.102)	.193 (.079)	.378 (.107)	.206 (.084)	.833 (.018)	.447 (.050)
3	1,224	204	612	408	.328 (.102)	.216 (.086)	.342 (.102)	.223 (.086)	.778 (.019)	.537 (.041)
4	1,632	204	612	408	.311 (.098)	.223 (.086)	.323 (.098)	.231 (.088)	.750 (.021)	.572 (.036)
5	2,040	204	612	408	.299 (.093)	.226 (.084)	.305 (.092)	.230 (.084)	.734 (.020)	.594 (.033)
6	2,448	204	612	408	.294 (.089)	.231 (.083)	.301 (.089)	.236 (.084)	.722 (.022)	.607 (.032)
10	4,080	204	612	408	.272 (.077)	.235 (.074)	.277 (.077)	.238 (.075)	.700 (.022)	.631 (.028)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into three binary items; “Dummy-coded Options” are the number of base items split into three binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix E: Simulation Results for Items with Four Response Options – Dummy-coded Response Options Summed Equals Item Condition

Table E1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.410 (.084)	.045 (.062)	.516 (.071)	.046 (.062)	.979 (.029)	.043 (.055)
1.25	60	16	64	48	.380 (.078)	.039 (.054)	.473 (.071)	.038 (.052)	.825 (.069)	.038 (.050)
1.50	72	16	64	48	.340 (.070)	.032 (.043)	.426 (.066)	.033 (.045)	.693 (.075)	.033 (.045)
1.75	84	16	64	48	.307 (.067)	.029 (.039)	.384 (.062)	.029 (.039)	.596 (.074)	.028 (.036)
2	96	16	64	48	.292 (.060)	.028 (.037)	.362 (.054)	.029 (.037)	.535 (.069)	.026 (.034)
3	144	16	64	48	.227 (.049)	.021 (.028)	.279 (.047)	.022 (.030)	.374 (.059)	.020 (.026)
4	192	16	64	48	.190 (.042)	.020 (.025)	.233 (.042)	.022 (.027)	.295 (.051)	.021 (.026)
5	**240**	**16**	**64**	**48**	**.164 (.036)**	**.019 (.023)**	**.201 (.036)**	**.021 (.024)**	**.245 (.043)**	**.021 (.025)**
6	288	16	64	48	.148 (.034)	.020 (.023)	.180 (.034)	.023 (.025)	.215 (.039)	.023 (.025)
10	480	16	64	48	.111 (.026)	.021 (.020)	.132 (.027)	.024 (.021)	.150 (.031)	.025 (.022)
25	1,200	16	64	48	.075 (.021)	.029 (.020)	.086 (.021)	.033 (.021)	.094 (.023)	.035 (.023)
50	2,400	16	64	48	.064 (.019)	.037 (.019)	.070 (.020)	.041 (.020)	.076 (.022)	.044 (.022)
100	4,800	16	64	48	.058 (.018)	.044 (.019)	.062 (.019)	.047 (.020)	.067 (.020)	.051 (.021)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.441 (.085)	.056 (.069)	.541 (.071)	.060 (.075)	.983 (.023)	.047 (.061)
1.25	60	16	64	48	.409 (.080)	.059 (.072)	.499 (.070)	.062 (.076)	.843 (.064)	.046 (.057)
1.50	72	16	64	48	.380 (.074)	.051 (.063)	.461 (.069)	.056 (.066)	.733 (.071)	.047 (.057)
1.75	84	16	64	48	.357 (.071)	.056 (.068)	.429 (.065)	.061 (.070)	.655 (.075)	.054 (.064)
2	96	16	64	48	.336 (.067)	.053 (.059)	.404 (.064)	.059 (.063)	.592 (.074)	.058 (.066)
3	**144**	**16**	**64**	**48**	**.279 (.062)**	**.058 (.059)**	**.332 (.061)**	**.065 (.063)**	**.451 (.076)**	**.073 (.072)**
4	192	16	64	48	.250 (.056)	.058 (.052)	.293 (.055)	.066 (.056)	.382 (.068)	.081 (.067)
5	240	16	64	48	.232 (.055)	.068 (.054)	.270 (.054)	.077 (.059)	.343 (.069)	.097 (.073)
6	288	16	64	48	.218 (.052)	.072 (.054)	.252 (.052)	.082 (.057)	.316 (.067)	.104 (.073)
10	480	16	64	48	.187 (.052)	.085 (.056)	.211 (.055)	.096 (.059)	.257 (.069)	.122 (.075)
25	1,200	16	64	48	.162 (.050)	.112 (.053)	.176 (.053)	.123 (.056)	.210 (.066)	.151 (.070)
50	2,400	16	64	48	.151 (.052)	.124 (.054)	.162 (.055)	.133 (.056)	.191 (.067)	.161 (.070)
100	4,800	16	64	48	.149 (.054)	.134 (.055)	.158 (.056)	.142 (.057)	.186 (.069)	.170 (.070)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.485 (.088)	.102 (.105)	.577 (.074)	.115 (.111)	.987 (.020)	.062 (.083)
1.25	60	16	64	48	.458 (.084)	.099 (.101)	.544 (.073)	.110 (.108)	.890 (.051)	.103 (.112)
1.50	**72**	**16**	**64**	**48**	**.431 (.083)**	**.096 (.092)**	**.509 (.075)**	**.109 (.099)**	**.806 (.070)**	**.142 (.130)**
1.75	84	16	64	48	.420 (.081)	.113 (.100)	.489 (.076)	.130 (.107)	.750 (.078)	.179 (.141)
2	96	16	64	48	.402 (.082)	.114 (.094)	.467 (.075)	.129 (.102)	.701 (.086)	.193 (.142)
3	144	16	64	48	.361 (.076)	.139 (.095)	.415 (.072)	.158 (.103)	.606 (.098)	.264 (.151)
4	192	16	64	48	.342 (.080)	.150 (.093)	.387 (.081)	.169 (.100)	.555 (.113)	.287 (.158)
5	240	16	64	48	.322 (.083)	.160 (.097)	.362 (.084)	.180 (.104)	.515 (.115)	.295 (.154)
6	288	16	64	48	.322 (.081)	.178 (.097)	.359 (.082)	.200 (.104)	.510 (.116)	.328 (.155)
10	480	16	64	48	.307 (.086)	.212 (.099)	.337 (.089)	.232 (.105)	.475 (.124)	.365 (.151)
25	1,200	16	64	48	.283 (.091)	.238 (.097)	.303 (.096)	.255 (.101)	.424 (.132)	.378 (.143)
50	2,400	16	64	48	.286 (.092)	.264 (.097)	.304 (.097)	.281 (.101)	.422 (.129)	.400 (.134)
100	4,800	16	64	48	.279 (.095)	.268 (.096)	.295 (.099)	.283 (.101)	.411 (.134)	.401 (.136)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.405 (.082)	.044 (.058)	.506 (.069)	.047 (.062)	.981 (.028)	.043 (.057)
1.25	60	16	64	48	.369 (.074)	.036 (.047)	.458 (.068)	.037 (.048)	.821 (.070)	.034 (.047)
1.50	72	16	64	48	.332 (.072)	.032 (.043)	.414 (.065)	.033 (.042)	.694 (.078)	.029 (.040)
1.75	84	16	64	48	.301 (.066)	.030 (.040)	.375 (.063)	.032 (.042)	.594 (.077)	.027 (.037)
2	96	16	64	48	.282 (.059)	.025 (.033)	.347 (.057)	.027 (.035)	.528 (.071)	.024 (.031)
3	144	16	64	48	.216 (.049)	.020 (.027)	.263 (.046)	.021 (.028)	.368 (.059)	.018 (.024)
4	192	16	64	48	.175 (.039)	.018 (.024)	.213 (.038)	.018 (.024)	.282 (.048)	.016 (.023)
5	240	16	64	48	.153 (.034)	.019 (.021)	.184 (.033)	.020 (.023)	.238 (.041)	.018 (.021)
6	288	16	64	48	.136 (.030)	.018 (.021)	.163 (.030)	.021 (.024)	.206 (.038)	.018 (.021)
10	480	16	64	48	.099 (.023)	.019 (.018)	.116 (.023)	.021 (.019)	.143 (.028)	.019 (.018)
25	**1,200**	**16**	**64**	**48**	**.061 (.015)**	**.022 (.014)**	**.069 (.015)**	**.025 (.015)**	**.083 (.019)**	**.025 (.016)**
50	2,400	16	64	48	.049 (.013)	.028 (.013)	.053 (.013)	.030 (.014)	.065 (.016)	.033 (.016)
100	4,800	16	64	48	.042 (.012)	.031 (.013)	.044 (.012)	.033 (.013)	.055 (.015)	.038 (.016)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.427 (.083)	.058 (.074)	.526 (.069)	.062 (.078)	.982 (.025)	.047 (.060)
1.25	60	16	64	48	.400 (.080)	.050 (.062)	.489 (.071)	.054 (.066)	.844 (.063)	.045 (.059)
1.50	72	16	64	48	.371 (.076)	.051 (.061)	.447 (.067)	.055 (.065)	.733 (.073)	.047 (.061)
1.75	84	16	64	48	.344 (.071)	.050 (.057)	.415 (.065)	.054 (.060)	.649 (.076)	.050 (.060)
2	**96**	**16**	**64**	**48**	**.325 (.064)**	**.047 (.054)**	**.389 (.059)**	**.051 (.057)**	**.589 (.073)**	**.055 (.061)**
3	144	16	64	48	.268 (.060)	.055 (.054)	.318 (.057)	.061 (.059)	.449 (.075)	.072 (.071)
4	192	16	64	48	.235 (.052)	.059 (.051)	.275 (.051)	.066 (.056)	.377 (.069)	.083 (.070)
5	240	16	64	48	.217 (.052)	.061 (.052)	.251 (.051)	.070 (.056)	.337 (.069)	.093 (.074)
6	288	16	64	48	.203 (.052)	.068 (.052)	.233 (.052)	.077 (.055)	.310 (.069)	.098 (.070)
10	480	16	64	48	.173 (.046)	.078 (.050)	.194 (.048)	.086 (.053)	.254 (.066)	.117 (.073)
25	1,200	16	64	48	.144 (.047)	.098 (.050)	.157 (.049)	.107 (.053)	.203 (.067)	.145 (.071)
50	2,400	16	64	48	.135 (.045)	.110 (.049)	.144 (.048)	.117 (.051)	.188 (.065)	.156 (.069)
100	4,800	16	64	48	.132 (.047)	.119 (.049)	.139 (.049)	.125 (.051)	.181 (.066)	.165 (.068)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.471 (.090)	.091 (.100)	.565 (.076)	.103 (.109)	.988 (.019)	.063 (.080)
1.25	60	16	64	48	.452 (.084)	.102 (.098)	.535 (.074)	.115 (.103)	.888 (.054)	.105 (.112)
1.50	**72**	**16**	**64**	**48**	**.421 (.082)**	**.102 (.097)**	**.497 (.074)**	**.116 (.107)**	**.805 (.069)**	**.139 (.125)**
1.75	84	16	64	48	.408 (.081)	.108 (.097)	.477 (.076)	.123 (.104)	.749 (.080)	.179 (.140)
2	96	16	64	48	.387 (.079)	.106 (.090)	.449 (.075)	.121 (.101)	.701 (.086)	.184 (.143)
3	144	16	64	48	.352 (.078)	.132 (.095)	.402 (.076)	.149 (.102)	.604 (.101)	.256 (.147)
4	192	16	64	48	.324 (.073)	.145 (.092)	.368 (.074)	.163 (.098)	.551 (.104)	.289 (.149)
5	240	16	64	48	.311 (.076)	.155 (.092)	.349 (.077)	.174 (.098)	.519 (.109)	.304 (.145)
6	288	16	64	48	.313 (.080)	.170 (.092)	.348 (.082)	.190 (.098)	.515 (.111)	.332 (.145)
10	480	16	64	48	.286 (.082)	.195 (.093)	.312 (.085)	.213 (.098)	.464 (.121)	.355 (.144)
25	1,200	16	64	48	.267 (.083)	.224 (.088)	.286 (.086)	.241 (.092)	.429 (.123)	.384 (.134)
50	2,400	16	64	48	.272 (.091)	.248 (.092)	.288 (.095)	.263 (.097)	.431 (.129)	.407 (.135)
100	4,800	16	64	48	.266 (.088)	.254 (.091)	.280 (.093)	.267 (.095)	.417 (.127)	.406 (.131)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.398 (.083)	.041 (.055)	.493 (.078)	.042 (.057)	.981 (.025)	.042 (.057)
1.25	60	16	64	48	.354 (.078)	.035 (.047)	.439 (.071)	.035 (.046)	.822 (.070)	.035 (.046)
1.50	72	16	64	48	.318 (.072)	.031 (.041)	.393 (.068)	.031 (.042)	.693 (.079)	.030 (.038)
1.75	84	16	64	48	.285 (.066)	.026 (.035)	.354 (.064)	.028 (.037)	.594 (.076)	.026 (.035)
2	96	16	64	48	.260 (.062)	.025 (.033)	.321 (.059)	.026 (.033)	.523 (.072)	.023 (.031)
3	144	16	64	48	.200 (.047)	.020 (.026)	.244 (.046)	.021 (.027)	.366 (.057)	.020 (.029)
4	192	16	64	48	.164 (.040)	.018 (.023)	.199 (.040)	.020 (.025)	.285 (.049)	.018 (.024)
5	240	16	64	48	.140 (.033)	.017 (.021)	.167 (.034)	.019 (.022)	.238 (.042)	.018 (.022)
6	288	16	64	48	.122 (.029)	.016 (.020)	.146 (.030)	.019 (.021)	.204 (.036)	.018 (.022)
10	**480**	**16**	**64**	**48**	**.088 (.024)**	**.016 (.016)**	**.103 (.024)**	**.018 (.018)**	**.142 (.029)**	**.018 (.018)**
25	1,200	16	64	48	.052 (.015)	.020 (.014)	.058 (.015)	.022 (.015)	.083 (.020)	.027 (.018)
50	2,400	16	64	48	.040 (.012)	.023 (.013)	.043 (.013)	.025 (.013)	.065 (.018)	.033 (.018)
100	4,800	16	64	48	.034 (.011)	.025 (.012)	.036 (.011)	.027 (.012)	.055 (.016)	.038 (.017)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.415 (.086)	.060 (.073)	.511 (.076)	.062 (.076)	.985 (.022)	.043 (.061)
1.25	60	16	64	48	.385 (.079)	.054 (.067)	.468 (.073)	.056 (.071)	.847 (.060)	.047 (.062)
1.50	**72**	**16**	**64**	**48**	**.355 (.076)**	**.051 (.063)**	**.429 (.071)**	**.057 (.067)**	**.737 (.074)**	**.057 (.067)**
1.75	84	16	64	48	.330 (.070)	.049 (.060)	.397 (.066)	.053 (.064)	.657 (.077)	.058 (.070)
2	96	16	64	48	.309 (.068)	.047 (.057)	.367 (.065)	.051 (.060)	.594 (.077)	.060 (.070)
3	144	16	64	48	.257 (.058)	.051 (.051)	.303 (.057)	.058 (.055)	.465 (.077)	.084 (.076)
4	192	16	64	48	.225 (.056)	.053 (.052)	.264 (.056)	.061 (.056)	.394 (.079)	.098 (.086)
5	240	16	64	48	.205 (.052)	.059 (.051)	.238 (.053)	.066 (.054)	.353 (.073)	.105 (.079)
6	288	16	64	48	.192 (.051)	.063 (.049)	.219 (.053)	.070 (.053)	.325 (.076)	.115 (.081)
10	480	16	64	48	.163 (.049)	.073 (.047)	.183 (.051)	.081 (.050)	.274 (.078)	.138 (.086)
25	1,200	16	64	48	.137 (.049)	.096 (.049)	.149 (.052)	.103 (.052)	.224 (.076)	.168 (.082)
50	2,400	16	64	48	.128 (.049)	.104 (.048)	.136 (.052)	.111 (.051)	.213 (.077)	.182 (.080)
100	4,800	16	64	48	.119 (.046)	.107 (.046)	.126 (.048)	.112 (.048)	.198 (.076)	.183 (.078)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	16	64	48	.459 (.089)	.088 (.097)	.549 (.075)	.098 (.103)	.989 (.015)	.061 (.078)
1.25	**60**	**16**	**64**	**48**	**.429 (.086)**	**.087 (.091)**	**.512 (.079)**	**.098 (.098)**	**.889 (.054)**	**.106 (.114)**
1.50	72	16	64	48	.406 (.081)	.093 (.094)	.480 (.077)	.105 (.101)	.809 (.065)	.143 (.128)
1.75	84	16	64	48	.384 (.078)	.092 (.087)	.449 (.074)	.104 (.095)	.749 (.076)	.178 (.141)
2	96	16	64	48	.370 (.078)	.101 (.090)	.433 (.075)	.116 (.098)	.709 (.079)	.202 (.136)
3	144	16	64	48	.328 (.077)	.115 (.089)	.377 (.076)	.130 (.094)	.606 (.093)	.254 (.138)
4	192	16	64	48	.309 (.075)	.131 (.090)	.349 (.077)	.145 (.094)	.561 (.101)	.291 (.145)
5	240	16	64	48	.291 (.076)	.139 (.083)	.328 (.078)	.156 (.090)	.528 (.107)	.311 (.146)
6	288	16	64	48	.281 (.076)	.147 (.085)	.313 (.079)	.164 (.091)	.507 (.107)	.330 (.142)
10	480	16	64	48	.265 (.079)	.176 (.086)	.290 (.083)	.193 (.091)	.469 (.116)	.358 (.138)
25	1,200	16	64	48	.254 (.080)	.211 (.085)	.271 (.085)	.225 (.089)	.445 (.115)	.399 (.126)
50	2,400	16	64	48	.249 (.087)	.228 (.090)	.264 (.092)	.242 (.095)	.438 (.124)	.418 (.129)
100	4,800	16	64	48	.241 (.087)	.231 (.088)	.254 (.092)	.242 (.093)	.421 (.125)	.411 (.127)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.432 (.054)	.032 (.040)	.527 (.047)	.035 (.041)	.993 (.010)	.019 (.027)
1.25	150	40	160	120	.393 (.051)	.029 (.033)	.480 (.043)	.032 (.036)	.832 (.042)	.020 (.026)
1.50	180	40	160	120	.365 (.047)	.030 (.035)	.445 (.043)	.035 (.038)	.720 (.047)	.025 (.030)
1.75	210	40	160	120	.343 (.044)	.031 (.032)	.415 (.042)	.036 (.035)	.638 (.049)	.027 (.030)
2	240	40	160	120	.320 (.042)	.033 (.032)	.386 (.039)	.039 (.036)	.575 (.047)	.032 (.034)
3	**360**	**40**	**160**	**120**	**.268 (.037)**	**.038 (.030)**	**.319 (.035)**	**.043 (.033)**	**.433 (.043)**	**.044 (.034)**
4	480	40	160	120	.238 (.034)	.044 (.030)	.281 (.033)	.051 (.032)	.363 (.041)	.056 (.035)
5	600	40	160	120	.217 (.033)	.050 (.029)	.253 (.033)	.058 (.032)	.318 (.039)	.065 (.036)
6	720	40	160	120	.204 (.030)	.056 (.029)	.237 (.030)	.064 (.031)	.291 (.037)	.073 (.035)
10	1,200	40	160	120	.174 (.030)	.070 (.031)	.198 (.031)	.080 (.034)	.234 (.037)	.094 (.040)
25	3,000	40	160	120	.144 (.027)	.094 (.030)	.158 (.028)	.103 (.031)	.181 (.033)	.120 (.036)
50	6,000	40	160	120	.137 (.027)	.110 (.029)	.147 (.028)	.118 (.030)	.167 (.032)	.136 (.035)
100	12,000	40	160	120	.129 (.026)	.115 (.027)	.138 (.028)	.123 (.029)	.156 (.032)	.139 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.494 (.057)	.085 (.072)	.582 (.047)	.099 (.079)	.996 (.006)	.028 (.038)
1.25	150	40	160	120	.465 (.056)	.102 (.072)	.546 (.049)	.122 (.078)	.899 (.031)	.113 (.087)
1.50	**180**	**40**	**160**	**120**	**.446 (.054)**	**.109 (.069)**	**.520 (.049)**	**.128 (.076)**	**.830 (.042)**	**.172 (.099)**
1.75	210	40	160	120	.432 (.053)	.122 (.070)	.500 (.049)	.142 (.075)	.780 (.050)	.216 (.111)
2	240	40	160	120	.417 (.052)	.130 (.070)	.481 (.049)	.151 (.074)	.744 (.054)	.249 (.109)
3	360	40	160	120	.388 (.053)	.161 (.071)	.439 (.052)	.184 (.076)	.660 (.067)	.327 (.115)
4	480	40	160	120	.369 (.055)	.182 (.067)	.415 (.055)	.206 (.071)	.616 (.074)	.368 (.109)
5	600	40	160	120	.356 (.056)	.198 (.071)	.397 (.057)	.221 (.074)	.590 (.076)	.390 (.110)
6	720	40	160	120	.350 (.055)	.209 (.067)	.387 (.056)	.233 (.071)	.570 (.076)	.401 (.103)
10	1,200	40	160	120	.343 (.063)	.248 (.072)	.372 (.065)	.271 (.075)	.541 (.086)	.443 (.104)
25	3,000	40	160	120	.327 (.064)	.286 (.069)	.350 (.067)	.306 (.072)	.507 (.089)	.467 (.097)
50	6,000	40	160	120	.326 (.065)	.305 (.067)	.346 (.068)	.323 (.071)	.499 (.089)	.479 (.093)
100	12,000	40	160	120	.323 (.063)	.312 (.065)	.340 (.066)	.328 (.068)	.490 (.087)	.479 (.090)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.533 (.057)	.157 (.093)	.615 (.049)	.184 (.100)	.997 (.004)	.039 (.050)
1.25	**150**	**40**	**160**	**120**	**.515 (.056)**	**.175 (.094)**	**.591 (.049)**	**.202 (.098)**	**.931 (.023)**	**.218 (.115)**
1.50	180	40	160	120	.499 (.058)	.191 (.090)	.569 (.051)	.219 (.096)	.882 (.032)	.314 (.117)
1.75	210	40	160	120	.492 (.058)	.206 (.092)	.555 (.053)	.236 (.097)	.848 (.038)	.379 (.119)
2	240	40	160	120	.480 (.057)	.219 (.091)	.539 (.053)	.249 (.094)	.819 (.042)	.418 (.109)
3	360	40	160	120	.464 (.058)	.263 (.085)	.512 (.056)	.292 (.087)	.763 (.052)	.507 (.102)
4	480	40	160	120	.449 (.059)	.286 (.080)	.493 (.058)	.317 (.082)	.732 (.056)	.540 (.091)
5	600	40	160	120	.444 (.061)	.307 (.080)	.483 (.061)	.337 (.082)	.713 (.058)	.565 (.086)
6	720	40	160	120	.443 (.061)	.330 (.078)	.480 (.061)	.359 (.080)	.705 (.059)	.585 (.083)
10	1,200	40	160	120	.433 (.068)	.358 (.079)	.463 (.068)	.384 (.081)	.676 (.067)	.604 (.082)
25	3,000	40	160	120	.435 (.067)	.403 (.074)	.458 (.069)	.425 (.076)	.659 (.066)	.631 (.073)
50	6,000	40	160	120	.434 (.066)	.417 (.069)	.455 (.068)	.438 (.071)	.654 (.064)	.640 (.067)
100	12,000	40	160	120	.438 (.067)	.429 (.069)	.458 (.070)	.448 (.072)	.653 (.067)	.646 (.069)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.403 (.057)	.027 (.033)	.486 (.053)	.029 (.034)	.992 (.011)	.018 (.025)
1.25	150	40	160	120	.358 (.052)	.027 (.032)	.432 (.050)	.030 (.033)	.825 (.041)	.017 (.021)
1.50	180	40	160	120	.326 (.046)	.025 (.030)	.394 (.046)	.029 (.033)	.707 (.045)	.019 (.026)
1.75	210	40	160	120	.298 (.044)	.027 (.029)	.356 (.045)	.031 (.031)	.622 (.046)	.020 (.024)
2	240	40	160	120	.278 (.044)	.026 (.027)	.330 (.044)	.029 (.029)	.559 (.048)	.021 (.024)
3	360	40	160	120	.220 (.035)	.029 (.024)	.255 (.037)	.032 (.026)	.409 (.040)	.028 (.025)
4	480	40	160	120	.186 (.030)	.033 (.024)	.212 (.032)	.037 (.025)	.334 (.037)	.036 (.026)
5	**600**	**40**	**160**	**120**	**.161 (.028)**	**.035 (.023)**	**.182 (.029)**	**.039 (.024)**	**.287 (.035)**	**.042 (.028)**
6	720	40	160	120	.147 (.025)	.037 (.022)	.165 (.027)	.041 (.023)	.259 (.032)	.046 (.027)
10	1,200	40	160	120	.115 (.021)	.045 (.021)	.126 (.021)	.049 (.022)	.202 (.029)	.063 (.028)
25	3,000	40	160	120	.082 (.016)	.052 (.018)	.086 (.016)	.055 (.019)	.147 (.026)	.085 (.027)
50	6,000	40	160	120	.070 (.014)	.055 (.015)	.073 (.014)	.057 (.016)	.128 (.024)	.095 (.025)
100	12,000	40	160	120	.065 (.014)	.057 (.014)	.067 (.014)	.059 (.015)	.120 (.024)	.103 (.025)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.464 (.057)	.083 (.070)	.543 (.054)	.093 (.075)	.995 (.007)	.026 (.034)
1.25	150	40	160	120	.437 (.055)	.086 (.064)	.509 (.053)	.099 (.070)	.894 (.032)	.094 (.076)
1.50	**180**	**40**	**160**	**120**	**.409 (.054)**	**.096 (.065)**	**.476 (.053)**	**.110 (.069)**	**.820 (.042)**	**.153 (.093)**
1.75	210	40	160	120	.390 (.052)	.102 (.064)	.449 (.053)	.117 (.068)	.767 (.049)	.190 (.100)
2	240	40	160	120	.379 (.055)	.109 (.065)	.433 (.054)	.125 (.069)	.731 (.054)	.223 (.106)
3	360	40	160	120	.337 (.054)	.130 (.060)	.378 (.055)	.147 (.065)	.639 (.063)	.294 (.105)
4	480	40	160	120	.315 (.054)	.146 (.061)	.349 (.055)	.163 (.063)	.590 (.071)	.329 (.103)
5	600	40	160	120	.301 (.054)	.158 (.060)	.330 (.057)	.174 (.064)	.566 (.073)	.359 (.102)
6	720	40	160	120	.290 (.055)	.167 (.060)	.317 (.058)	.183 (.063)	.547 (.075)	.371 (.097)
10	1,200	40	160	120	.273 (.056)	.192 (.062)	.292 (.060)	.205 (.064)	.517 (.077)	.412 (.094)
25	3,000	40	160	120	.249 (.056)	.215 (.058)	.263 (.059)	.226 (.061)	.479 (.082)	.438 (.090)
50	6,000	40	160	120	.245 (.059)	.228 (.060)	.256 (.062)	.238 (.063)	.473 (.084)	.452 (.088)
100	12,000	40	160	120	.237 (.058)	.228 (.059)	.247 (.061)	.237 (.062)	.462 (.084)	.451 (.086)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.512 (.059)	.144 (.092)	.589 (.052)	.166 (.098)	.997 (.005)	.038 (.048)
1.25	**150**	**40**	**160**	**120**	**.491 (.060)**	**.165 (.089)**	**.561 (.055)**	**.188 (.093)**	**.926 (.024)**	**.202 (.111)**
1.50	180	40	160	120	.473 (.058)	.174 (.085)	.537 (.056)	.199 (.090)	.877 (.033)	.294 (.119)
1.75	210	40	160	120	.461 (.060)	.189 (.085)	.520 (.057)	.216 (.089)	.841 (.038)	.358 (.113)
2	240	40	160	120	.456 (.058)	.206 (.086)	.510 (.058)	.233 (.090)	.816 (.041)	.405 (.111)
3	360	40	160	120	.425 (.059)	.234 (.081)	.469 (.059)	.261 (.083)	.753 (.051)	.490 (.096)
4	480	40	160	120	.408 (.062)	.253 (.078)	.446 (.063)	.278 (.083)	.721 (.057)	.522 (.097)
5	600	40	160	120	.408 (.063)	.279 (.077)	.443 (.064)	.304 (.080)	.705 (.057)	.552 (.087)
6	720	40	160	120	.398 (.063)	.288 (.076)	.429 (.065)	.312 (.078)	.691 (.058)	.564 (.084)
10	1,200	40	160	120	.391 (.067)	.318 (.076)	.415 (.069)	.339 (.079)	.668 (.063)	.594 (.077)
25	3,000	40	160	120	.377 (.066)	.346 (.072)	.395 (.069)	.363 (.074)	.644 (.064)	.614 (.071)
50	6,000	40	160	120	.376 (.068)	.361 (.071)	.392 (.070)	.376 (.073)	.638 (.063)	.624 (.066)
100	12,000	40	160	120	.375 (.070)	.367 (.071)	.390 (.073)	.382 (.074)	.633 (.068)	.625 (.069)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.362 (.072)	.025 (.034)	.440 (.075)	.027 (.036)	.993 (.009)	.019 (.026)
1.25	150	40	160	120	.320 (.064)	.024 (.030)	.385 (.071)	.026 (.031)	.829 (.040)	.019 (.025)
1.50	180	40	160	120	.290 (.061)	.022 (.027)	.346 (.070)	.024 (.028)	.709 (.048)	.019 (.023)
1.75	210	40	160	120	.261 (.058)	.024 (.027)	.310 (.067)	.026 (.030)	.624 (.048)	.023 (.028)
2	240	40	160	120	.246 (.053)	.024 (.025)	.289 (.062)	.026 (.027)	.563 (.047)	.023 (.026)
3	**360**	**40**	**160**	**120**	**.183 (.044)**	**.025 (.022)**	**.212 (.052)**	**.027 (.024)**	**.411 (.039)**	**.032 (.026)**
4	480	40	160	120	.153 (.039)	.026 (.022)	.171 (.044)	.029 (.023)	.337 (.038)	.039 (.028)
5	600	40	160	120	.133 (.037)	.028 (.020)	.147 (.041)	.031 (.021)	.296 (.036)	.046 (.029)
6	720	40	160	120	.117 (.031)	.031 (.021)	.129 (.034)	.033 (.022)	.264 (.033)	.052 (.031)
10	1,200	40	160	120	.087 (.024)	.034 (.017)	.093 (.025)	.035 (.018)	.205 (.031)	.067 (.030)
25	3,000	40	160	120	.059 (.016)	.038 (.015)	.061 (.015)	.039 (.015)	.154 (.027)	.091 (.029)
50	6,000	40	160	120	.049 (.013)	.039 (.013)	.050 (.013)	.040 (.014)	.136 (.027)	.104 (.028)
100	12,000	40	160	120	.044 (.012)	.039 (.012)	.045 (.012)	.040 (.012)	.128 (.028)	.111 (.029)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.418 (.070)	.069 (.063)	.494 (.075)	.077 (.067)	.995 (.007)	.029 (.037)
1.25	**150**	**40**	**160**	**120**	**.391 (.071)**	**.071 (.058)**	**.458 (.076)**	**.082 (.064)**	**.895 (.030)**	**.100 (.080)**
1.50	180	40	160	120	.368 (.069)	.075 (.059)	.425 (.074)	.086 (.062)	.822 (.040)	.153 (.091)
1.75	210	40	160	120	.346 (.070)	.085 (.059)	.399 (.076)	.097 (.065)	.768 (.049)	.195 (.102)
2	240	40	160	120	.333 (.070)	.091 (.056)	.380 (.075)	.104 (.061)	.732 (.053)	.227 (.105)
3	360	40	160	120	.293 (.065)	.109 (.058)	.330 (.073)	.122 (.061)	.643 (.060)	.303 (.101)
4	480	40	160	120	.268 (.068)	.118 (.058)	.297 (.073)	.131 (.063)	.600 (.064)	.342 (.097)
5	600	40	160	120	.259 (.067)	.129 (.058)	.285 (.073)	.142 (.062)	.573 (.069)	.366 (.096)
6	720	40	160	120	.246 (.065)	.136 (.058)	.267 (.071)	.149 (.063)	.554 (.069)	.383 (.092)
10	1,200	40	160	120	.221 (.064)	.149 (.059)	.236 (.068)	.159 (.064)	.513 (.074)	.407 (.093)
25	3,000	40	160	120	.200 (.063)	.168 (.061)	.208 (.066)	.175 (.065)	.484 (.076)	.440 (.083)
50	6,000	40	160	120	.192 (.061)	.177 (.060)	.200 (.065)	.184 (.064)	.477 (.081)	.455 (.084)
100	12,000	40	160	120	.186 (.061)	.178 (.061)	.192 (.064)	.184 (.064)	.466 (.081)	.456 (.083)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	40	160	120	.478 (.070)	.123 (.086)	.549 (.072)	.141 (.091)	.997 (.005)	.038 (.048)
1.25	**150**	**40**	**160**	**120**	**.452 (.069)**	**.135 (.082)**	**.518 (.071)**	**.154 (.088)**	**.923 (.024)**	**.190 (.106)**
1.50	180	40	160	120	.433 (.073)	.145 (.083)	.491 (.074)	.165 (.088)	.875 (.033)	.289 (.111)
1.75	210	40	160	120	.423 (.072)	.165 (.087)	.476 (.074)	.186 (.092)	.839 (.036)	.354 (.112)
2	240	40	160	120	.411 (.073)	.169 (.082)	.458 (.076)	.190 (.086)	.811 (.043)	.393 (.108)
3	360	40	160	120	.375 (.075)	.194 (.081)	.414 (.079)	.215 (.086)	.746 (.051)	.474 (.102)
4	480	40	160	120	.367 (.078)	.220 (.083)	.400 (.082)	.240 (.088)	.716 (.054)	.518 (.092)
5	600	40	160	120	.354 (.078)	.233 (.080)	.384 (.083)	.253 (.085)	.698 (.056)	.542 (.085)
6	720	40	160	120	.355 (.078)	.246 (.084)	.383 (.083)	.266 (.089)	.688 (.057)	.559 (.081)
10	1,200	40	160	120	.337 (.082)	.267 (.084)	.356 (.087)	.283 (.089)	.661 (.063)	.584 (.077)
25	3,000	40	160	120	.323 (.083)	.294 (.083)	.339 (.087)	.308 (.088)	.640 (.064)	.609 (.070)
50	6,000	40	160	120	.318 (.082)	.303 (.083)	.331 (.086)	.316 (.087)	.631 (.063)	.616 (.066)
100	12,000	40	160	120	.320 (.084)	.311 (.084)	.331 (.089)	.322 (.089)	.628 (.068)	.620 (.070)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.459 (.042)	.044 (.037)	.551 (.036)	.051 (.041)	.996 (.005)	.012 (.015)
1.25	255	68	272	204	.428 (.040)	.048 (.039)	.510 (.034)	.057 (.043)	.857 (.028)	.031 (.032)
1.50	306	68	272	204	.399 (.038)	.055 (.038)	.476 (.034)	.064 (.040)	.757 (.035)	.050 (.039)
1.75	357	68	272	204	.380 (.039)	.061 (.038)	.450 (.035)	.073 (.041)	.688 (.041)	.069 (.046)
2	**408**	**68**	**272**	**204**	**.362 (.036)**	**.062 (.037)**	**.426 (.034)**	**.073 (.040)**	**.634 (.040)**	**.078 (.045)**
3	612	68	272	204	.319 (.034)	.082 (.037)	.371 (.034)	.095 (.040)	.512 (.042)	.119 (.051)
4	816	68	272	204	.295 (.033)	.096 (.037)	.340 (.033)	.111 (.039)	.454 (.044)	.147 (.054)
5	1,020	68	272	204	.279 (.033)	.107 (.036)	.318 (.033)	.123 (.039)	.416 (.043)	.165 (.055)
6	1,224	68	272	204	.269 (.034)	.118 (.038)	.304 (.034)	.135 (.040)	.393 (.045)	.182 (.054)
10	2,040	68	272	204	.246 (.033)	.143 (.037)	.273 (.034)	.159 (.039)	.343 (.045)	.211 (.053)
25	5,100	68	272	204	.225 (.032)	.177 (.035)	.243 (.034)	.191 (.036)	.298 (.043)	.242 (.047)
50	10,200	68	272	204	.221 (.034)	.195 (.036)	.236 (.036)	.208 (.038)	.286 (.045)	.258 (.048)
100	20,400	68	272	204	.214 (.036)	.201 (.036)	.226 (.037)	.213 (.038)	.274 (.047)	.260 (.048)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.526 (.042)	.136 (.071)	.608 (.037)	.160 (.073)	.998 (.003)	.023 (.029)
1.25	**255**	**68**	**272**	**204**	**.504 (.042)**	**.152 (.067)**	**.579 (.037)**	**.179 (.072)**	**.925 (.019)**	**.191 (.085)**
1.50	306	68	272	204	.489 (.043)	.170 (.065)	.557 (.038)	.198 (.069)	.874 (.025)	.290 (.088)
1.75	357	68	272	204	.478 (.041)	.187 (.068)	.541 (.038)	.215 (.073)	.838 (.030)	.347 (.089)
2	408	68	272	204	.469 (.042)	.201 (.067)	.529 (.039)	.230 (.070)	.811 (.033)	.392 (.087)
3	612	68	272	204	.448 (.046)	.239 (.067)	.498 (.044)	.269 (.069)	.748 (.043)	.479 (.083)
4	816	68	272	204	.438 (.045)	.267 (.063)	.481 (.044)	.297 (.065)	.719 (.044)	.520 (.077)
5	1,020	68	272	204	.432 (.046)	.289 (.061)	.471 (.045)	.319 (.063)	.700 (.046)	.545 (.071)
6	1,224	68	272	204	.427 (.048)	.304 (.060)	.464 (.048)	.333 (.063)	.685 (.050)	.554 (.070)
10	2,040	68	272	204	.419 (.051)	.340 (.061)	.450 (.051)	.367 (.063)	.661 (.052)	.585 (.065)
25	5,100	68	272	204	.416 (.051)	.381 (.056)	.440 (.053)	.404 (.058)	.640 (.054)	.610 (.059)
50	10,200	68	272	204	.419 (.051)	.401 (.054)	.440 (.053)	.422 (.056)	.634 (.053)	.620 (.055)
100	20,400	68	272	204	.415 (.051)	.406 (.053)	.435 (.053)	.425 (.055)	.626 (.054)	.619 (.055)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.559 (.044)	.203 (.082)	.635 (.037)	.235 (.087)	.998 (.002)	.031 (.038)
1.25	**255**	**68**	**272**	**204**	**.548 (.044)**	**.223 (.079)**	**.616 (.038)**	**.256 (.082)**	**.944 (.014)**	**.279 (.095)**
1.50	306	68	272	204	.532 (.044)	.249 (.077)	.595 (.038)	.283 (.080)	.904 (.020)	.396 (.091)
1.75	357	68	272	204	.526 (.043)	.270 (.073)	.585 (.038)	.304 (.075)	.878 (.022)	.469 (.081)
2	408	68	272	204	.521 (.044)	.285 (.074)	.575 (.041)	.320 (.074)	.856 (.026)	.514 (.080)
3	612	68	272	204	.506 (.046)	.331 (.069)	.552 (.044)	.365 (.071)	.809 (.032)	.596 (.067)
4	816	68	272	204	.500 (.048)	.360 (.068)	.540 (.047)	.394 (.069)	.786 (.035)	.631 (.061)
5	1,020	68	272	204	.497 (.045)	.380 (.063)	.534 (.045)	.411 (.063)	.772 (.035)	.651 (.055)
6	1,224	68	272	204	.495 (.048)	.398 (.061)	.530 (.047)	.428 (.062)	.763 (.036)	.665 (.053)
10	2,040	68	272	204	.492 (.049)	.429 (.059)	.521 (.049)	.456 (.059)	.745 (.037)	.687 (.047)
25	5,100	68	272	204	.492 (.051)	.465 (.055)	.515 (.052)	.488 (.056)	.727 (.039)	.704 (.043)
50	10,200	68	272	204	.493 (.053)	.480 (.055)	.513 (.054)	.500 (.056)	.722 (.041)	.711 (.043)
100	20,400	68	272	204	.492 (.051)	.485 (.052)	.512 (.052)	.505 (.053)	.718 (.040)	.713 (.041)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.390 (.053)	.035 (.034)	.461 (.057)	.039 (.036)	.996 (.006)	.011 (.016)
1.25	255	68	272	204	.350 (.050)	.036 (.033)	.410 (.054)	.040 (.034)	.843 (.031)	.022 (.025)
1.50	306	68	272	204	.313 (.049)	.038 (.030)	.362 (.054)	.042 (.031)	.734 (.036)	.029 (.028)
1.75	357	68	272	204	.291 (.046)	.040 (.030)	.333 (.052)	.045 (.031)	.658 (.037)	.038 (.032)
2	**408**	**68**	**272**	**204**	**.270 (.045)**	**.040 (.028)**	**.309 (.049)**	**.045 (.029)**	**.600 (.037)**	**.045 (.032)**
3	612	68	272	204	.213 (.036)	.050 (.027)	.237 (.039)	.055 (.029)	.467 (.037)	.073 (.038)
4	816	68	272	204	.182 (.034)	.053 (.025)	.199 (.036)	.058 (.026)	.399 (.038)	.087 (.039)
5	1,020	68	272	204	.162 (.029)	.056 (.024)	.175 (.031)	.061 (.025)	.360 (.036)	.103 (.040)
6	1,224	68	272	204	.149 (.027)	.059 (.024)	.159 (.028)	.063 (.025)	.333 (.036)	.113 (.040)
10	2,040	68	272	204	.117 (.021)	.064 (.020)	.123 (.021)	.067 (.021)	.278 (.036)	.138 (.039)
25	5,100	68	272	204	.091 (.016)	.069 (.018)	.094 (.016)	.071 (.018)	.231 (.035)	.170 (.038)
50	10,200	68	272	204	.082 (.015)	.072 (.017)	.084 (.016)	.073 (.017)	.216 (.038)	.185 (.040)
100	20,400	68	272	204	.077 (.014)	.071 (.014)	.078 (.014)	.073 (.014)	.205 (.036)	.190 (.036)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.472 (.056)	.114 (.064)	.540 (.056)	.130 (.069)	.998 (.003)	.020 (.027)
1.25	**255**	**68**	**272**	**204**	**.446 (.054)**	**.125 (.061)**	**.505 (.057)**	**.143 (.065)**	**.919 (.019)**	**.166 (.080)**
1.50	306	68	272	204	.424 (.055)	.136 (.059)	.478 (.059)	.154 (.063)	.863 (.026)	.251 (.085)
1.75	357	68	272	204	.407 (.055)	.147 (.061)	.454 (.056)	.164 (.064)	.821 (.033)	.305 (.090)
2	408	68	272	204	.395 (.055)	.156 (.060)	.439 (.060)	.174 (.065)	.794 (.036)	.350 (.091)
3	612	68	272	204	.357 (.056)	.179 (.058)	.389 (.059)	.195 (.061)	.722 (.044)	.431 (.084)
4	816	68	272	204	.342 (.056)	.199 (.057)	.368 (.059)	.215 (.060)	.694 (.045)	.482 (.072)
5	1,020	68	272	204	.327 (.058)	.207 (.059)	.350 (.061)	.223 (.061)	.669 (.048)	.500 (.071)
6	1,224	68	272	204	.316 (.056)	.214 (.058)	.337 (.059)	.228 (.061)	.656 (.052)	.514 (.071)
10	2,040	68	272	204	.296 (.056)	.231 (.057)	.310 (.059)	.243 (.061)	.627 (.053)	.545 (.066)
25	5,100	68	272	204	.278 (.054)	.250 (.055)	.288 (.057)	.259 (.058)	.603 (.055)	.571 (.060)
50	10,200	68	272	204	.272 (.058)	.258 (.058)	.281 (.060)	.267 (.061)	.598 (.056)	.581 (.059)
100	20,400	68	272	204	.265 (.055)	.258 (.054)	.273 (.057)	.266 (.056)	.592 (.055)	.584 (.057)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.524 (.052)	.181 (.078)	.593 (.050)	.206 (.082)	.998 (.002)	.028 (.036)
1.25	**255**	**68**	**272**	**204**	**.506 (.050)**	**.202 (.075)**	**.567 (.050)**	**.229 (.079)**	**.940 (.015)**	**.260 (.096)**
1.50	306	68	272	204	.488 (.056)	.216 (.079)	.544 (.055)	.243 (.082)	.899 (.021)	.375 (.092)
1.75	357	68	272	204	.480 (.056)	.231 (.074)	.530 (.057)	.258 (.078)	.870 (.026)	.440 (.087)
2	408	68	272	204	.471 (.054)	.243 (.074)	.517 (.055)	.269 (.076)	.850 (.028)	.489 (.084)
3	612	68	272	204	.449 (.056)	.282 (.071)	.486 (.059)	.308 (.075)	.799 (.032)	.575 (.068)
4	816	68	272	204	.436 (.059)	.304 (.070)	.469 (.061)	.327 (.073)	.775 (.036)	.613 (.063)
5	1,020	68	272	204	.426 (.060)	.317 (.069)	.454 (.062)	.339 (.072)	.758 (.038)	.631 (.057)
6	1,224	68	272	204	.423 (.059)	.330 (.068)	.449 (.061)	.351 (.071)	.748 (.038)	.643 (.057)
10	2,040	68	272	204	.413 (.065)	.353 (.068)	.433 (.067)	.372 (.071)	.729 (.041)	.667 (.051)
25	5,100	68	272	204	.404 (.065)	.378 (.067)	.419 (.068)	.393 (.070)	.713 (.042)	.688 (.046)
50	10,200	68	272	204	.401 (.064)	.389 (.065)	.415 (.067)	.402 (.069)	.706 (.042)	.694 (.045)
100	20,400	68	272	204	.403 (.068)	.397 (.069)	.417 (.071)	.410 (.072)	.705 (.042)	.699 (.043)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.334 (.077)	.027 (.029)	.399 (.094)	.030 (.031)	.996 (.005)	.011 (.017)
1.25	255	68	272	204	.290 (.076)	.030 (.029)	.334 (.091)	.033 (.031)	.845 (.032)	.024 (.027)
1.50	**306**	**68**	**272**	**204**	**.260 (.073)**	**.030 (.027)**	**.301 (.085)**	**.033 (.028)**	**.740 (.036)**	**.035 (.032)**
1.75	357	68	272	204	.237 (.068)	.031 (.026)	.269 (.079)	.035 (.028)	.668 (.039)	.047 (.036)
2	408	68	272	204	.217 (.065)	.033 (.025)	.248 (.078)	.036 (.027)	.610 (.039)	.055 (.038)
3	612	68	272	204	.168 (.055)	.037 (.024)	.182 (.060)	.039 (.025)	.481 (.041)	.086 (.042)
4	816	68	272	204	.139 (.047)	.037 (.021)	.151 (.052)	.040 (.022)	.416 (.039)	.104 (.044)
5	1,020	68	272	204	.119 (.040)	.040 (.021)	.125 (.040)	.042 (.022)	.377 (.041)	.122 (.049)
6	1,224	68	272	204	.107 (.037)	.041 (.021)	.113 (.039)	.043 (.022)	.352 (.043)	.133 (.049)
10	2,040	68	272	204	.080 (.024)	.043 (.019)	.082 (.022)	.044 (.019)	.298 (.043)	.160 (.047)
25	5,100	68	272	204	.058 (.015)	.044 (.015)	.059 (.014)	.044 (.015)	.253 (.043)	.193 (.048)
50	10,200	68	272	204	.051 (.012)	.044 (.013)	.052 (.011)	.045 (.012)	.238 (.043)	.207 (.044)
100	20,400	68	272	204	.047 (.010)	.044 (.011)	.048 (.010)	.044 (.011)	.230 (.043)	.215 (.044)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.413 (.086)	.090 (.061)	.477 (.096)	.103 (.065)	.998 (.003)	.021 (.027)
1.25	**255**	**68**	**272**	**204**	**.384 (.082)**	**.099 (.060)**	**.435 (.092)**	**.113 (.064)**	**.916 (.020)**	**.157 (.078)**
1.50	306	68	272	204	.365 (.084)	.104 (.057)	.410 (.092)	.117 (.060)	.860 (.027)	.245 (.088)
1.75	357	68	272	204	.347 (.081)	.113 (.059)	.387 (.091)	.126 (.064)	.819 (.033)	.300 (.085)
2	408	68	272	204	.330 (.081)	.116 (.059)	.363 (.091)	.128 (.064)	.790 (.035)	.338 (.083)
3	612	68	272	204	.291 (.081)	.137 (.062)	.315 (.089)	.148 (.066)	.721 (.040)	.431 (.077)
4	816	68	272	204	.270 (.079)	.148 (.063)	.288 (.082)	.158 (.067)	.684 (.044)	.472 (.072)
5	1,020	68	272	204	.254 (.075)	.154 (.062)	.271 (.081)	.165 (.067)	.665 (.046)	.497 (.069)
6	1,224	68	272	204	.248 (.076)	.161 (.065)	.263 (.080)	.171 (.069)	.649 (.049)	.507 (.068)
10	2,040	68	272	204	.222 (.068)	.168 (.063)	.230 (.070)	.175 (.065)	.621 (.051)	.536 (.063)
25	5,100	68	272	204	.204 (.062)	.182 (.062)	.209 (.064)	.187 (.063)	.599 (.053)	.566 (.058)
50	10,200	68	272	204	.193 (.058)	.182 (.057)	.198 (.059)	.186 (.059)	.592 (.054)	.576 (.056)
100	20,400	68	272	204	.188 (.056)	.182 (.056)	.192 (.058)	.187 (.058)	.586 (.054)	.578 (.056)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table E27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	68	272	204	.477 (.075)	.151 (.077)	.540 (.080)	.170 (.082)	.998 (.002)	.025 (.033)
1.25	**255**	**68**	**272**	**204**	**.454 (.079)**	**.168 (.076)**	**.508 (.082)**	**.189 (.081)**	**.938 (.015)**	**.254 (.089)**
1.50	306	68	272	204	.439 (.079)	.180 (.079)	.487 (.086)	.201 (.084)	.897 (.020)	.367 (.092)
1.75	357	68	272	204	.422 (.079)	.190 (.075)	.467 (.087)	.212 (.081)	.866 (.025)	.433 (.085)
2	408	68	272	204	.411 (.082)	.197 (.077)	.451 (.087)	.216 (.081)	.845 (.028)	.477 (.080)
3	612	68	272	204	.388 (.082)	.230 (.080)	.420 (.089)	.250 (.086)	.794 (.032)	.564 (.067)
4	816	68	272	204	.374 (.084)	.251 (.082)	.398 (.090)	.267 (.086)	.770 (.035)	.605 (.060)
5	1,020	68	272	204	.366 (.082)	.261 (.081)	.390 (.089)	.279 (.087)	.750 (.038)	.619 (.060)
6	1,224	68	272	204	.353 (.086)	.265 (.082)	.375 (.090)	.282 (.086)	.740 (.038)	.630 (.055)
10	2,040	68	272	204	.340 (.086)	.287 (.082)	.357 (.090)	.302 (.087)	.721 (.039)	.656 (.050)
25	5,100	68	272	204	.326 (.083)	.302 (.083)	.338 (.087)	.314 (.087)	.703 (.042)	.678 (.046)
50	10,200	68	272	204	.320 (.081)	.309 (.081)	.330 (.085)	.318 (.084)	.696 (.043)	.684 (.045)
100	20,400	68	272	204	.315 (.082)	.309 (.083)	.324 (.086)	.318 (.087)	.693 (.043)	.687 (.044)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix F: Simulation Results for Items with Four Response Options - Item Condition Divided into Response Options

Table F1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.438 (.050)	.032 (.037)	.532 (.043)	.037 (.041)	.994 (.008)	.016 (.023)
1.25	180	48	192	144	.406 (.046)	.034 (.034)	.492 (.041)	.039 (.039)	.840 (.036)	.022 (.029)
1.50	216	48	192	144	.373 (.043)	.034 (.033)	.450 (.040)	.041 (.037)	.726 (.043)	.029 (.032)
1.75	252	48	192	144	.353 (.041)	.038 (.034)	.423 (.038)	.044 (.037)	.650 (.043)	.034 (.033)
2	288	48	192	144	.336 (.041)	.039 (.033)	.400 (.037)	.045 (.037)	.594 (.044)	.041 (.035)
3	**432**	**48**	**192**	**144**	**.282 (.036)**	**.048 (.030)**	**.334 (.035)**	**.056 (.033)**	**.455 (.044)**	**.061 (.039)**
4	576	48	192	144	.250 (.034)	.057 (.033)	.293 (.033)	.067 (.035)	.382 (.041)	.076 (.040)
5	720	48	192	144	.234 (.032)	.065 (.032)	.272 (.032)	.075 (.035)	.344 (.040)	.088 (.041)
6	864	48	192	144	.222 (.031)	.069 (.031)	.256 (.031)	.080 (.033)	.319 (.038)	.096 (.040)
10	1,440	48	192	144	.193 (.030)	.088 (.032)	.218 (.031)	.100 (.034)	.263 (.038)	.124 (.043)
25	3,600	48	192	144	.168 (.029)	.117 (.031)	.184 (.003)	.128 (.033)	.214 (.035)	.153 (.039)
50	7,200	48	192	144	.160 (.030)	.132 (.032)	.171 (.031)	.142 (.033)	.198 (.037)	.167 (.040)
100	14,400	48	192	144	.155 (.029)	.141 (.031)	.165 (.031)	.150 (.032)	.190 (.037)	.174 (.038)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.506 (.052)	.104 (.071)	.590 (.044)	.122 (.078)	.997 (.004)	.027 (.035)
1.25	**180**	**48**	**192**	**144**	**.479 (.049)**	**.114 (.070)**	**.559 (.043)**	**.134 (.077)**	**.911 (.026)**	**.141 (.086)**
1.50	216	48	192	144	.460 (.050)	.129 (.069)	.532 (.045)	.151 (.076)	.848 (.037)	.212 (.102)
1.75	252	48	192	144	.444 (.052)	.139 (.072)	.511 (.048)	.162 (.077)	.802 (.044)	.261 (.106)
2	288	48	192	144	.440 (.051)	.155 (.071)	.501 (.047)	.178 (.074)	.773 (.048)	.304 (.109)
3	432	48	192	144	.408 (.051)	.187 (.070)	.460 (.050)	.213 (.075)	.695 (.058)	.389 (.106)
4	576	48	192	144	.391 (.053)	.210 (.067)	.437 (.053)	.236 (.071)	.656 (.065)	.427 (.100)
5	720	48	192	144	.383 (.052)	.226 (.067)	.424 (.053)	.252 (.070)	.633 (.066)	.448 (.095)
6	864	48	192	144	.378 (.053)	.241 (.068)	.416 (.054)	.268 (.070)	.617 (.068)	.466 (.095)
10	1,440	48	192	144	.368 (.058)	.277 (.067)	.399 (.059)	.302 (.070)	.588 (.073)	.497 (.089)
25	3,600	48	192	144	.361 (.060)	.321 (.065)	.383 (.062)	.342 (.068)	.560 (.076)	.524 (.083)
50	7,200	48	192	144	.360 (.063)	.339 (.066)	.380 (.065)	.358 (.068)	.552 (.077)	.534 (.080)
100	14,400	48	192	144	.361 (.061)	.350 (.062)	.379 (.063)	.368 (.065)	.552 (.074)	.543 (.076)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.543 (.051)	.175 (.095)	.623 (.044)	.203 (.099)	.998 (.003)	.035 (.046)
1.25	**180**	**48**	**192**	**144**	**.527 (.050)**	**.197 (.087)**	**.599 (.044)**	**.228 (.092)**	**.935 (.020)**	**.241 (.107)**
1.50	216	48	192	144	.515 (.051)	.213 (.089)	.581 (.045)	.244 (.094)	.892 (.026)	.350 (.109)
1.75	252	48	192	144	.502 (.054)	.227 (.086)	.565 (.051)	.260 (.091)	.859 (.033)	.413 (.106)
2	288	48	192	144	.496 (.051)	.245 (.081)	.554 (.048)	.277 (.085)	.836 (.034)	.456 (.096)
3	432	48	192	144	.478 (.053)	.283 (.080)	.526 (.052)	.316 (.083)	.781 (.043)	.539 (.086)
4	576	48	192	144	.471 (.055)	.317 (.075)	.514 (.053)	.350 (.078)	.756 (.046)	.584 (.079)
5	720	48	192	144	.466 (.057)	.336 (.076)	.504 (.057)	.367 (.077)	.740 (.050)	.602 (.075)
6	864	48	192	144	.463 (.058)	.353 (.074)	.499 (.058)	.382 (.077)	.728 (.051)	.614 (.073)
10	1,440	48	192	144	.458 (.057)	.387 (.070)	.488 (.058)	.414 (.072)	.705 (.051)	.639 (.066)
25	3,600	48	192	144	.458 (.059)	.428 (.064)	.482 (.060)	.450 (.066)	.689 (.052)	.663 (.057)
50	7,200	48	192	144	.454 (.065)	.438 (.067)	.475 (.066)	.459 (.069)	.679 (.058)	.667 (.061)
100	14,400	48	192	144	.456 (.059)	.449 (.060)	.476 (.061)	.468 (.062)	.676 (.053)	.669 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.397 (.052)	.029 (.034)	.480 (.053)	.032 (.036)	.994 (.009)	.015 (.020)
1.25	180	48	192	144	.356 (.048)	.029 (.032)	.426 (.050)	.031 (.035)	.831 (.037)	.018 (.024)
1.50	216	48	192	144	.323 (.047)	.028 (.030)	.383 (.049)	.033 (.032)	.711 (.044)	.020 (.024)
1.75	252	48	192	144	.296 (.043)	.029 (.028)	.349 (.046)	.033 (.030)	.631 (.043)	.022 (.025)
2	288	48	192	144	.274 (.041)	.033 (.030)	.320 (.045)	.036 (.031)	.566 (.042)	.027 (.027)
3	432	48	192	144	.217 (.035)	.034 (.025)	.249 (.038)	.039 (.026)	.422 (.040)	.037 (.030)
4	**576**	**48**	**192**	**144**	**.185 (.031)**	**.039 (.024)**	**.208 (.032)**	**.043 (.025)**	**.351 (.033)**	**.047 (.029)**
5	720	48	192	144	.163 (.028)	.040 (.024)	.181 (.031)	.045 (.025)	.305 (.034)	.053 (.030)
6	864	48	192	144	.147 (.026)	.043 (.023)	.162 (.027)	.047 (.024)	.277 (.033)	.059 (.030)
10	1,440	48	192	144	.116 (.021)	.050 (.020)	.125 (.021)	.054 (.021)	.221 (.031)	.081 (.031)
25	3,600	48	192	144	.085 (.017)	.058 (.018)	.089 (.017)	.060 (.018)	.169 (.029)	.107 (.031)
50	7,200	48	192	144	.074 (.015)	.061 (.015)	.076 (.015)	.063 (.016)	.150 (.027)	.118 (.028)
100	14,400	48	192	144	.070 (.014)	.063 (.015)	.072 (.015)	.064 (.015)	.143 (.027)	.126 (.028)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.466 (.056)	.091 (.064)	.544 (.052)	.106 (.069)	.996 (.005)	.025 (.033)
1.25	**180**	**48**	**192**	**144**	**.440 (.054)**	**.102 (.066)**	**.509 (.052)**	**.118 (.071)**	**.903 (.028)**	**.119 (.080)**
1.50	216	48	192	144	.414 (.053)	.108 (.063)	.477 (.054)	.123 (.068)	.835 (.037)	.185 (.093)
1.75	252	48	192	144	.398 (.052)	.118 (.063)	.454 (.054)	.136 (.067)	.791 (.043)	.235 (.099)
2	288	48	192	144	.386 (.056)	.125 (.064)	.436 (.057)	.141 (.068)	.753 (.050)	.269 (.101)
3	432	48	192	144	.349 (.054)	.152 (.060)	.390 (.055)	.171 (.063)	.674 (.057)	.354 (.098)
4	576	48	192	144	.325 (.054)	.165 (.060)	.357 (.057)	.181 (.063)	.630 (.060)	.386 (.092)
5	720	48	192	144	.313 (.053)	.178 (.060)	.341 (.056)	.195 (.063)	.606 (.064)	.411 (.091)
6	864	48	192	144	.302 (.055)	.186 (.060)	.326 (.059)	.202 (.063)	.588 (.066)	.428 (.091)
10	1,440	48	192	144	.280 (.054)	.206 (.058)	.298 (.056)	.220 (.060)	.552 (.071)	.455 (.087)
25	3,600	48	192	144	.265 (.056)	.231 (.056)	.277 (.059)	.242 (.059)	.531 (.071)	.493 (.077)
50	7,200	48	192	144	.252 (.054)	.235 (.055)	.262 (.057)	.244 (.057)	.514 (.074)	.493 (.077)
100	14,400	48	192	144	.250 (.056)	.242 (.056)	.260 (.058)	.251 (.058)	.511 (.068)	.501 (.069)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.516 (.057)	.156 (.087)	.590 (.052)	.180 (.090)	.998 (.003)	.035 (.043)
1.25	**180**	**48**	**192**	**144**	**.496 (.056)**	**.174 (.084)**	**.562 (.051)**	**.197 (.086)**	**.932 (.020)**	**.217 (.103)**
1.50	216	48	192	144	.482 (.057)	.191 (.085)	.543 (.053)	.218 (.089)	.886 (.028)	.331 (.105)
1.75	252	48	192	144	.469 (.054)	.200 (.081)	.524 (.053)	.225 (.084)	.853 (.034)	.389 (.104)
2	288	48	192	144	.459 (.055)	.215 (.078)	.511 (.053)	.242 (.081)	.828 (.035)	.434 (.098)
3	432	48	192	144	.434 (.057)	.254 (.078)	.475 (.058)	.280 (.082)	.771 (.042)	.524 (.088)
4	576	48	192	144	.425 (.059)	.280 (.074)	.460 (.061)	.305 (.076)	.743 (.047)	.564 (.079)
5	720	48	192	144	.417 (.060)	.296 (.075)	.450 (.061)	.321 (.078)	.729 (.049)	.587 (.074)
6	864	48	192	144	.412 (.062)	.305 (.073)	.440 (.063)	.328 (.075)	.713 (.051)	.594 (.073)
10	1,440	48	192	144	.403 (.065)	.335 (.071)	.427 (.067)	.356 (.073)	.693 (.053)	.623 (.065)
25	3,600	48	192	144	.386 (.070)	.358 (.072)	.404 (.073)	.375 (.075)	.667 (.056)	.640 (.062)
50	7,200	48	192	144	.388 (.069)	.374 (.070)	.404 (.072)	.389 (.074)	.665 (.057)	.651 (.060)
100	14,400	48	192	144	.386 (.068)	.378 (.069)	.400 (.071)	.392 (.071)	.661 (.056)	.654 (.058)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.350 (.072)	.026 (.031)	.422 (.081)	.028 (.033)	.994 (.008)	.015 (.021)
1.25	180	48	192	144	.314 (.068)	.025 (.030)	.373 (.079)	.027 (.030)	.832 (.036)	.017 (.022)
1.50	216	48	192	144	.282 (.064)	.023 (.026)	.335 (.073)	.026 (.028)	.716 (.043)	.021 (.025)
1.75	252	48	192	144	.254 (.059)	.024 (.025)	.298 (.068)	.027 (.027)	.632 (.043)	.026 (.027)
2	**288**	**48**	**192**	**144**	**.237 (.058)**	**.024 (.024)**	**.276 (.068)**	**.027 (.026)**	**.575 (.043)**	**.029 (.027)**
3	432	48	192	144	.180 (.048)	.028 (.023)	.206 (.054)	.031 (.025)	.430 (.039)	.043 (.033)
4	576	48	192	144	.148 (.043)	.030 (.022)	.165 (.049)	.032 (.023)	.358 (.039)	.052 (.033)
5	720	48	192	144	.129 (.037)	.031 (.021)	.142 (.041)	.034 (.022)	.315 (.036)	.061 (.033)
6	864	48	192	144	.116 (.034)	.034 (.019)	.125 (.037)	.036 (.020)	.287 (.035)	.069 (.033)
10	1,440	48	192	144	.086 (.026)	.037 (.018)	.090 (.026)	.038 (.019)	.231 (.033)	.089 (.034)
25	3,600	48	192	144	.058 (.015)	.040 (.015)	.060 (.015)	.041 (.015)	.178 (.031)	.116 (.033)
50	7,200	48	192	144	.050 (.013)	.041 (.014)	.051 (.013)	.041 (.014)	.161 (.032)	.130 (.034)
100	14,400	48	192	144	.046 (.012)	.041 (.012)	.046 (.011)	.042 (.012)	.154 (.032)	.137 (.032)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.422 (.074)	.074 (.060)	.492 (.080)	.084 (.066)	.997 (.005)	.023 (.031)
1.25	**180**	**48**	**192**	**144**	**.389 (.074)**	**.080 (.058)**	**.455 (.080)**	**.092 (.063)**	**.904 (.027)**	**.118 (.079)**
1.50	216	48	192	144	.368 (.075)	.089 (.059)	.421 (.080)	.100 (.064)	.836 (.036)	.184 (.091)
1.75	252	48	192	144	.348 (.070)	.094 (.058)	.397 (.077)	.107 (.062)	.788 (.043)	.232 (.093)
2	288	48	192	144	.342 (.073)	.103 (.059)	.386 (.081)	.116 (.063)	.754 (.047)	.268 (.098)
3	432	48	192	144	.290 (.070)	.116 (.056)	.322 (.078)	.129 (.062)	.669 (.054)	.344 (.095)
4	576	48	192	144	.273 (.069)	.132 (.059)	.298 (.075)	.144 (.063)	.632 (.059)	.391 (.092)
5	720	48	192	144	.258 (.069)	.139 (.059)	.282 (.075)	.152 (.064)	.606 (.061)	.411 (.087)
6	864	48	192	144	.253 (.068)	.150 (.061)	.272 (.074)	.162 (.066)	.593 (.063)	.429 (.089)
10	1,440	48	192	144	.226 (.066)	.159 (.059)	.238 (.070)	.168 (.062)	.557 (.063)	.458 (.080)
25	3,600	48	192	144	.203 (.065)	.176 (.065)	.210 (.068)	.182 (.067)	.528 (.071)	.489 (.078)
50	7,200	48	192	144	.199 (.064)	.185 (.064)	.205 (.066)	.190 (.066)	.519 (.067)	.500 (.071)
100	14,400	48	192	144	.192 (.063)	.185 (.062)	.199 (.066)	.191 (.065)	.513 (.070)	.503 (.072)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	146	48	192	144	.481 (.072)	.136 (.086)	.554 (.074)	.157 (.091)	.998 (.003)	.032 (.041)
1.25	**180**	**48**	**192**	**144**	**.457 (.073)**	**.145 (.078)**	**.520 (.075)**	**.166 (.084)**	**.930 (.020)**	**.216 (.102)**
1.50	216	48	192	144	.441 (.072)	.158 (.079)	.498 (.076)	.180 (.083)	.884 (.028)	.317 (.106)
1.75	252	48	192	144	.422 (.073)	.169 (.081)	.471 (.078)	.191 (.087)	.850 (.031)	.375 (.106)
2	288	48	192	144	.413 (.072)	.183 (.082)	.459 (.079)	.204 (.087)	.824 (.035)	.427 (.098)
3	432	48	192	144	.383 (.075)	.213 (.081)	.418 (.084)	.234 (.087)	.766 (.043)	.516 (.086)
4	576	48	192	144	.371 (.075)	.233 (.079)	.402 (.079)	.254 (.083)	.737 (.047)	.552 (.079)
5	720	48	192	144	.357 (.079)	.240 (.080)	.384 (.084)	.258 (.085)	.717 (.048)	.570 (.075)
6	864	48	192	144	.353 (.079)	.253 (.080)	.376 (.084)	.271 (.084)	.706 (.052)	.586 (.073)
10	1,440	48	192	144	.338 (.083)	.275 (.084)	.356 (.088)	.291 (.088)	.682 (.053)	.612 (.065)
25	3,600	48	192	144	.327 (.083)	.299 (.084)	.341 (.087)	.312 (.088)	.663 (.057)	.634 (.062)
50	7,200	48	192	144	.322 (.082)	.309 (.083)	.333 (.086)	.320 (.087)	.654 (.057)	.640 (.059)
100	14,400	48	192	144	.322 (.082)	.315 (.082)	.333 (.086)	.326 (.086)	.652 (.056)	.645 (.057)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.505 (.033)	.095 (.046)	.591 (.028)	.113 (.051)	.999 (.002)	.010 (.014)
1.25	450	120	480	360	.482 (.031)	.106 (.045)	.560 (.028)	.127 (.049)	.907 (.018)	.125 (.058)
1.50	**540**	**120**	**480**	**360**	**.461 (.032)**	**.123 (.045)**	**.533 (.029)**	**.146 (.048)**	**.841 (.026)**	**.194 (.069)**
1.75	630	120	480	360	.450 (.031)	.137 (.045)	.516 (.029)	.161 (.048)	.798 (.030)	.250 (.072)
2	720	120	480	360	.436 (.032)	.146 (.044)	.500 (.030)	.171 (.047)	.764 (.033)	.285 (.075)
3	1,080	120	480	360	.408 (.033)	.180 (.043)	.460 (.032)	.207 (.046)	.683 (.042)	.365 (.074)
4	1,440	120	480	360	.395 (.033)	.204 (.042)	.441 (.033)	.232 (.044)	.643 (.045)	.403 (.070)
5	1,800	120	480	360	.386 (.035)	.223 (.044)	.428 (.035)	.251 (.047)	.621 (.049)	.430 (.072)
6	2,160	120	480	360	.381 (.035)	.241 (.043)	.420 (.036)	.268 (.045)	.607 (.050)	.449 (.067)
10	3,600	120	480	360	.374 (.037)	.281 (.043)	.405 (.038)	.306 (.044)	.579 (.052)	.485 (.063)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.556 (.033)	.189 (.059)	.632 (.028)	.222 (.064)	.999 (.001)	.016 (.020)
1.25	**450**	**120**	**480**	**360**	**.541 (.034)**	**.213 (.060)**	**.610 (.029)**	**.244 (.061)**	**.942 (.011)**	**.272 (.072)**
1.50	540	120	480	360	.528 (.035)	.235 (.059)	.591 (.031)	.267 (.060)	.902 (.016)	.386 (.070)
1.75	630	120	480	360	.520 (.032)	.254 (.056)	.579 (.030)	.289 (.057)	.875 (.018)	.458 (.065)
2	720	120	480	360	.514 (.034)	.272 (.058)	.570 (.032)	.306 (.061)	.854 (.020)	.501 (.063)
3	1,080	120	480	360	.500 (.034)	.317 (.052)	.546 (.033)	.352 (.054)	.805 (.025)	.587 (.052)
4	1,440	120	480	360	.494 (.034)	.348 (.050)	.535 (.034)	.382 (.052)	.780 (.026)	.622 (.046)
5	1,800	120	480	360	.490 (.035)	.370 (.048)	.528 (.034)	.402 (.049)	.766 (.027)	.641 (.041)
6	2,160	120	480	360	.487 (.036)	.385 (.047)	.522 (.036)	.416 (.047)	.755 (.028)	.653 (.041)
10	3,600	120	480	360	.486 (.036)	.422 (.044)	.515 (.036)	.449 (.044)	.738 (.028)	.678 (.035)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.581 (.035)	.249 (.069)	.652 (.029)	.287 (.071)	.999 (.001)	.019 (.023)
1.25	**450**	**120**	**480**	**360**	**.572 (.034)**	**.278 (.064)**	**.635 (.030)**	**.316 (.065)**	**.953 (.009)**	**.337 (.073)**
1.50	540	120	480	360	.561 (.035)	.302 (.063)	.619 (.031)	.339 (.064)	.920 (.013)	.469 (.068)
1.75	630	120	480	360	.554 (.033)	.321 (.060)	.608 (.030)	.358 (.061)	.897 (.015)	.534 (.059)
2	720	120	480	360	.551 (.034)	.341 (.058)	.602 (.031)	.377 (.059)	.881 (.016)	.581 (.055)
3	1,080	120	480	360	.541 (.036)	.388 (.052)	.584 (.033)	.422 (.053)	.840 (.020)	.660 (.042)
4	1,440	120	480	360	.534 (.036)	.411 (.049)	.572 (.035)	.444 (.049)	.819 (.022)	.687 (.038)
5	1,800	120	480	360	.537 (.035)	.436 (.049)	.572 (.034)	.468 (.049)	.809 (.023)	.708 (.036)
6	2,160	120	480	360	.535 (.036)	.450 (.047)	.567 (.036)	.479 (.047)	.800 (.023)	.716 (.033)
10	3,600	120	480	360	.537 (.036)	.483 (.044)	.564 (.036)	.509 (.044)	.786 (.024)	.738 (.031)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.363 (.064)	.053 (.034)	.411 (.074)	.059 (.036)	.998 (.002)	.007 (.011)
1.25	450	120	480	360	.322 (.060)	.057 (.032)	.359 (.068)	.062 (.033)	.877 (.021)	.058 (.038)
1.50	**540**	**120**	**480**	**360**	**.289 (.057)**	**.059 (.029)**	**.317 (.063)**	**.064 (.031)**	**.793 (.026)**	**.092 (.043)**
1.75	630	120	480	360	.259 (.054)	.062 (.029)	.283 (.058)	.066 (.031)	.736 (.030)	.127 (.048)
2	720	120	480	360	.241 (.049)	.064 (.028)	.261 (.050)	.069 (.030)	.690 (.033)	.150 (.052)
3	1,080	120	480	360	.193 (.042)	.072 (.027)	.204 (.040)	.075 (.028)	.587 (.039)	.214 (.059)
4	1,440	120	480	360	.166 (.034)	.074 (.026)	.175 (.035)	.077 (.027)	.536 (.042)	.250 (.060)
5	1,800	120	480	360	.149 (.029)	.075 (.024)	.154 (.027)	.077 (.025)	.503 (.042)	.269 (.056)
6	2,160	120	480	360	.137 (.025)	.075 (.023)	.142 (.023)	.077 (.023)	.484 (.043)	.289 (.057)
10	3,600	120	480	360	.114 (.019)	.078 (.021)	.117 (.017)	.079 (.020)	.442 (.045)	.322 (.055)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.469 (.061)	.146 (.058)	.522 (.068)	.163 (.062)	.999 (.001)	.016 (.020)
1.25	**450**	**120**	**480**	**360**	**.438 (.063)**	**.157 (.056)**	**.484 (.069)**	**.174 (.060)**	**.935 (.012)**	**.235 (.071)**
1.50	540	120	480	360	.418 (.062)	.172 (.057)	.458 (.068)	.188 (.061)	.892 (.017)	.349 (.070)
1.75	630	120	480	360	.397 (.063)	.179 (.057)	.433 (.069)	.195 (.062)	.860 (.019)	.414 (.064)
2	720	120	480	360	.383 (.064)	.184 (.055)	.414 (.069)	.199 (.059)	.838 (.021)	.457 (.063)
3	1,080	120	480	360	.350 (.063)	.204 (.054)	.373 (.067)	.217 (.058)	.784 (.026)	.546 (.053)
4	1,440	120	480	360	.330 (.062)	.216 (.056)	.347 (.063)	.227 (.059)	.757 (.028)	.581 (.048)
5	1,800	120	480	360	.315 (.060)	.220 (.057)	.330 (.062)	.231 (.059)	.739 (.029)	.601 (.046)
6	2,160	120	480	360	.307 (.057)	.227 (.055)	.320 (.059)	.236 (.057)	.730 (.031)	.617 (.045)
10	3,600	120	480	360	.286 (.056)	.237 (.054)	.295 (.057)	.245 (.056)	.708 (.031)	.641 (.038)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.525 (.054)	.215 (.068)	.583 (.056)	.241 (.071)	.999 (.001)	.020 (.024)
1.25	**450**	**120**	**480**	**360**	**.506 (.054)**	**.233 (.066)**	**.558 (.058)**	**.259 (.069)**	**.949 (.010)**	**.318 (.073)**
1.50	540	120	480	360	.494 (.055)	.255 (.065)	.540 (.059)	.280 (.070)	.917 (.013)	.448 (.067)
1.75	630	120	480	360	.482 (.056)	.264 (.063)	.523 (.059)	.289 (.066)	.892 (.016)	.517 (.061)
2	720	120	480	360	.469 (.059)	.273 (.065)	.507 (.061)	.297 (.069)	.873 (.017)	.558 (.054)
3	1,080	120	480	360	.449 (.061)	.305 (.063)	.478 (.064)	.327 (.068)	.831 (.021)	.638 (.045)
4	1,440	120	480	360	.435 (.063)	.322 (.067)	.459 (.066)	.341 (.070)	.810 (.023)	.671 (.041)
5	1,800	120	480	360	.426 (.064)	.333 (.066)	.447 (.066)	.351 (.069)	.797 (.024)	.689 (.038)
6	2,160	120	480	360	.422 (.064)	.341 (.066)	.440 (.067)	.357 (.068)	.790 (.024)	.700 (.036)
10	3,600	120	480	360	.409 (.064)	.359 (.066)	.423 (.067)	.372 (.069)	.773 (.026)	.721 (.034)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.293 (.098)	.038 (.029)	.327 (.117)	.042 (.032)	.998 (.002)	.008 (.011)
1.25	**450**	**120**	**480**	**360**	**.250 (.088)**	**.039 (.028)**	**.272 (.100)**	**.042 (.029)**	**.879 (.021)**	**.061 (.037)**
1.50	540	120	480	360	.214 (.083)	.040 (.027)	.232 (.091)	.043 (.029)	.798 (.027)	.102 (.046)
1.75	630	120	480	360	.194 (.077)	.042 (.027)	.210 (.085)	.045 (.029)	.741 (.030)	.133 (.052)
2	720	120	480	360	.177 (.072)	.042 (.026)	.188 (.077)	.045 (.028)	.697 (.033)	.158 (.052)
3	1,080	120	480	360	.137 (.058)	.047 (.026)	.139 (.054)	.047 (.025)	.598 (.037)	.227 (.059)
4	1,440	120	480	360	.111 (.044)	.048 (.024)	.111 (.039)	.048 (.023)	.547 (.039)	.267 (.057)
5	1,800	120	480	360	.098 (.037)	.047 (.022)	.099 (.036)	.048 (.023)	.518 (.040)	.290 (.055)
6	2,160	120	480	360	.086 (.030)	.048 (.023)	.087 (.027)	.048 (.022)	.495 (.045)	.305 (.057)
10	3,600	120	480	360	.067 (.017)	.045 (.016)	.068 (.015)	.046 (.016)	.458 (.044)	.341 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.388 (.101)	.109 (.058)	.433 (.112)	.122 (.063)	.999 (.001)	.014 (.020)
1.25	**450**	**120**	**480**	**360**	**.356 (.098)**	**.116 (.059)**	**.391 (.110)**	**.128 (.065)**	**.934 (.012)**	**.226 (.068)**
1.50	540	120	480	360	.340 (.096)	.127 (.059)	.369 (.106)	.138 (.064)	.888 (.016)	.335 (.064)
1.75	630	120	480	360	.316 (.099)	.127 (.061)	.339 (.104)	.137 (.066)	.856 (.020)	.402 (.064)
2	720	120	480	360	.302 (.094)	.135 (.061)	.322 (.100)	.144 (.066)	.831 (.022)	.444 (.063)
3	1,080	120	480	360	.266 (.088)	.147 (.064)	.277 (.091)	.152 (.068)	.778 (.027)	.534 (.052)
4	1,440	120	480	360	.243 (.083)	.151 (.066)	.253 (.086)	.157 (.069)	.749 (.029)	.571 (.049)
5	1,800	120	480	360	.228 (.077)	.154 (.064)	.235 (.077)	.159 (.065)	.732 (.029)	.592 (.046)
6	2,160	120	480	360	.218 (.077)	.155 (.065)	.224 (.077)	.159 (.066)	.722 (.031)	.606 (.045)
10	3,600	120	480	360	.197 (.068)	.160 (.062)	.202 (.069)	.164 (.063)	.699 (.033)	.633 (.040)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	362	120	480	360	.463 (.088)	.167 (.070)	.512 (.098)	.187 (.078)	.999 (.001)	.019 (.024)
1.25	**450**	**120**	**480**	**360**	**.444 (.094)**	**.188 (.076)**	**.489 (.101)**	**.208 (.081)**	**.948 (.010)**	**.310 (.074)**
1.50	540	120	480	360	.422 (.094)	.196 (.074)	.459 (.102)	.215 (.081)	.912 (.014)	.432 (.069)
1.75	630	120	480	360	.414 (.092)	.212 (.075)	.447 (.100)	.230 (.082)	.888 (.015)	.505 (.059)
2	720	120	480	360	.401 (.093)	.219 (.079)	.431 (.101)	.236 (.086)	.870 (.018)	.547 (.057)
3	1,080	120	480	360	.372 (.091)	.240 (.081)	.392 (.096)	.253 (.085)	.825 (.022)	.626 (.045)
4	1,440	120	480	360	.352 (.094)	.250 (.087)	.370 (.099)	.263 (.091)	.806 (.023)	.664 (.041)
5	1,800	120	480	360	.349 (.091)	.265 (.083)	.364 (.095)	.276 (.087)	.791 (.024)	.680 (.038)
6	2,160	120	480	360	.336 (.090)	.265 (.085)	.349 (.093)	.276 (.087)	.783 (.025)	.690 (.036)
10	3,600	120	480	360	.318 (.087)	.275 (.083)	.329 (.089)	.285 (.086)	.765 (.025)	.711 (.031)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.546 (.026)	.157 (.043)	.626 (.022)	.186 (.045)	.999 (.001)	.008 (.011)
1.25	**765**	**204**	**816**	**612**	**.525 (.025)**	**.177 (.043)**	**.599 (.022)**	**.207 (.045)**	**.934 (.010)**	**.226 (.051)**
1.50	918	204	816	612	.513 (.025)	.200 (.041)	.581 (.022)	.232 (.045)	.889 (.013)	.337 (.054)
1.75	1,071	204	816	612	.504 (.024)	.217 (.040)	.568 (.021)	.250 (.042)	.858 (.015)	.404 (.050)
2	1,224	204	816	612	.497 (.025)	.231 (.038)	.557 (.023)	.265 (.040)	.835 (.017)	.449 (.048)
3	1,836	204	816	612	.480 (.025)	.279 (.040)	.531 (.024)	.314 (.041)	.780 (.020)	.538 (.043)
4	2,448	204	816	612	.472 (.024)	.308 (.035)	.518 (.024)	.342 (.036)	.752 (.022)	.574 (.038)
5	3,060	204	816	612	.469 (.026)	.332 (.034)	.510 (.025)	.365 (.034)	.736 (.022)	.597 (.033)
6	3,672	204	816	612	.467 (.027)	.349 (.035)	.505 (.027)	.382 (.036)	.725 (.024)	.610 (.035)
10	6,120	204	816	612	.464 (.027)	.387 (.033)	.496 (.028)	.417 (.034)	.702 (.026)	.634 (.032)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.579 (.026)	.242 (.052)	.650 (.021)	.278 (.053)	1.00 (.001)	.013 (.015)
1.25	**765**	**204**	**816**	**612**	**.565 (.025)**	**.270 (.049)**	**.630 (.022)**	**.308 (.050)**	**.951 (.007)**	**.331 (.058)**
1.50	918	204	816	612	.555 (.026)	.290 (.048)	.615 (.023)	.328 (.048)	.918 (.010)	.457 (.053)
1.75	1,071	204	816	612	.551 (.026)	.311 (.046)	.607 (.023)	.348 (.047)	.895 (.012)	.527 (.046)
2	1,224	204	816	612	.545 (.026)	.327 (.045)	.597 (.024)	.364 (.046)	.877 (.013)	.569 (.043)
3	1,836	204	816	612	.536 (.026)	.376 (.039)	.580 (.025)	.412 (.040)	.837 (.016)	.649 (.033)
4	2,448	204	816	612	.531 (.027)	.403 (.039)	.569 (.025)	.437 (.039)	.815 (.016)	.679 (.029)
5	3,060	204	816	612	.532 (.027)	.427 (.038)	.567 (.026)	.458 (.038)	.804 (.017)	.699 (.028)
6	3,672	204	816	612	.529 (.028)	.440 (.036)	.563 (.027)	.471 (.036)	.795 (.018)	.709 (.027)
10	6,120	204	816	612	.532 (.028)	.475 (.034)	.559 (.028)	.502 (.034)	.780 (.019)	.729 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.598 (.025)	.289 (.054)	.665 (.021)	.328 (.055)	1.00 (0.00)	.014 (.017)
1.25	**765**	**204**	**816**	**612**	**.588 (.026)**	**.318 (.051)**	**.648 (.023)**	**.357 (.051)**	**.958 (.006)**	**.380 (.057)**
1.50	918	204	816	612	.582 (.027)	.342 (.051)	.636 (.024)	.380 (.051)	.930 (.008)	.511 (.052)
1.75	1,071	204	816	612	.578 (.027)	.364 (.047)	.628 (.025)	.401 (.048)	.910 (.010)	.582 (.044)
2	1,224	204	816	612	.573 (.027)	.381 (.045)	.621 (.025)	.418 (.045)	.895 (.011)	.623 (.039)
3	1,836	204	816	612	.567 (.027)	.427 (.042)	.606 (.025)	.462 (.042)	.859 (.014)	.696 (.031)
4	2,448	204	816	612	.565 (.027)	.456 (.039)	.600 (.026)	.487 (.038)	.842 (.015)	.725 (.027)
5	3,060	204	816	612	.563 (.028)	.473 (.037)	.595 (.027)	.503 (.037)	.831 (.016)	.739 (.024)
6	3,672	204	816	612	.561 (.028)	.485 (.037)	.591 (.028)	.513 (.037)	.823 (.016)	.748 (.023)
10	6,120	204	816	612	.562 (.028)	.514 (.034)	.588 (.028)	.539 (.034)	.810 (.016)	.766 (.021)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.303 (.073)	.065 (.031)	.329 (.079)	.070 (.033)	.999 (.001)	.006 (.009)
1.25	**765**	**204**	**816**	**612**	**.261 (.061)**	**.069 (.030)**	**.277 (.062)**	**.072 (.030)**	**.907 (.013)**	**.121 (.042)**
1.50	918	204	816	612	.236 (.057)	.071 (.029)	.248 (.056)	.073 (.030)	.844 (.017)	.195 (.046)
1.75	1,071	204	816	612	.213 (.049)	.072 (.027)	.223 (.049)	.075 (.027)	.800 (.021)	.249 (.050)
2	1,224	204	816	612	.196 (.045)	.073 (.027)	.205 (.045)	.075 (.027)	.765 (.022)	.283 (.050)
3	1,836	204	816	612	.155 (.031)	.076 (.023)	.160 (.028)	.077 (.023)	.688 (.027)	.372 (.049)
4	2,448	204	816	612	.135 (.024)	.076 (.022)	.138 (.021)	.077 (.022)	.649 (.030)	.414 (.046)
5	3,060	204	816	612	.123 (.019)	.077 (.018)	.125 (.017)	.077 (.018)	.626 (.030)	.438 (.044)
6	3,672	204	816	612	.116 (.017)	.077 (.018)	.118 (.015)	.078 (.018)	.611 (.031)	.455 (.043)
10	6,120	204	816	612	.101 (.013)	.078 (.016)	.102 (.012)	.079 (.015)	.579 (.033)	.484 (.041)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.434 (.072)	.157 (.053)	.476 (.080)	.172 (.058)	1.00 (.001)	.012 (.013)
1.25	**765**	**204**	**816**	**612**	**.399 (.074)**	**.164 (.055)**	**.432 (.078)**	**.177 (.059)**	**.946 (.008)**	**.293 (.056)**
1.50	918	204	816	612	.382 (.072)	.176 (.055)	.407 (.077)	.188 (.060)	.910 (.011)	.417 (.051)
1.75	1,071	204	816	612	.369 (.069)	.187 (.055)	.390 (.071)	.197 (.058)	.885 (.013)	.489 (.047)
2	1,224	204	816	612	.351 (.068)	.188 (.055)	.372 (.070)	.199 (.057)	.864 (.015)	.529 (.046)
3	1,836	204	816	612	.311 (.061)	.198 (.052)	.324 (.062)	.205 (.054)	.820 (.017)	.615 (.038)
4	2,448	204	816	612	.291 (.055)	.203 (.050)	.301 (.054)	.209 (.050)	.796 (.019)	.648 (.033)
5	3,060	204	816	612	.278 (.055)	.208 (.052)	.288 (.056)	.214 (.052)	.783 (.019)	.667 (.030)
6	3,672	204	816	612	.270 (.049)	.210 (.048)	.276 (.049)	.215 (.048)	.773 (.019)	.678 (.028)
10	6,120	204	816	612	.249 (.046)	.214 (.046)	.255 (.046)	.218 (.046)	.756 (.021)	.700 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.513 (.061)	.226 (.057)	.562 (.066)	.250 (.062)	1.00 (.001)	.013 (.017)
1.25	**765**	**204**	**816**	**612**	**.494 (.062)**	**.247 (.062)**	**.535 (.067)**	**.268 (.066)**	**.956 (.007)**	**.364 (.057)**
1.50	918	204	816	612	.476 (.065)	.262 (.063)	.513 (.067)	.283 (.066)	.926 (.009)	.497 (.052)
1.75	1,071	204	816	612	.465 (.063)	.272 (.062)	.497 (.066)	.292 (.065)	.905 (.011)	.563 (.044)
2	1,224	204	816	612	.454 (.065)	.280 (.061)	.484 (.068)	.300 (.065)	.888 (.012)	.604 (.040)
3	1,836	204	816	612	.427 (.067)	.303 (.065)	.448 (.070)	.319 (.069)	.852 (.014)	.680 (.033)
4	2,448	204	816	612	.411 (.066)	.313 (.064)	.430 (.068)	.329 (.066)	.833 (.016)	.708 (.028)
5	3,060	204	816	612	.402 (.066)	.324 (.065)	.419 (.068)	.338 (.067)	.822 (.016)	.726 (.026)
6	3,672	204	816	612	.399 (.065)	.332 (.063)	.415 (.066)	.345 (.065)	.815 (.016)	.736 (.024)
10	6,120	204	816	612	.387 (.062)	.345 (.062)	.397 (.064)	.354 (.064)	.801 (.017)	.755 (.022)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.227 (.103)	.044 (.028)	.241 (.114)	.046 (.029)	.999 (.001)	.006 (.008)
1.25	**765**	**204**	**816**	**612**	**.188 (.091)**	**.044 (.029)**	**.198 (.097)**	**.046 (.030)**	**.906 (.013)**	**.120 (.039)**
1.50	918	204	816	612	.159 (.071)	.044 (.026)	.166 (.074)	.045 (.027)	.844 (.017)	.195 (.048)
1.75	1,071	204	816	612	.144 (.068)	.045 (.027)	.148 (.069)	.046 (.028)	.799 (.020)	.248 (.047)
2	1,224	204	816	612	.128 (.059)	.045 (.025)	.129 (.057)	.045 (.025)	.767 (.022)	.287 (.049)
3	1,836	204	816	612	.096 (.041)	.045 (.024)	.097 (.037)	.045 (.024)	.687 (.026)	.371 (.047)
4	2,448	204	816	612	.079 (.027)	.043 (.019)	.080 (.026)	.043 (.020)	.648 (.028)	.412 (.045)
5	3,060	204	816	612	.072 (.024)	.044 (.019)	.071 (.021)	.044 (.018)	.625 (.029)	.435 (.043)
6	3,672	204	816	612	.065 (.018)	.043 (.017)	.065 (.014)	.043 (.016)	.608 (.031)	.450 (.043)
10	6,120	204	816	612	.054 (.010)	.041 (.012)	.054 (.008)	.041 (.011)	.577 (.032)	.483 (.040)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into four binary items; "Dummy-coded Options" are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.344 (.115)	.110 (.058)	.373 (.125)	.120 (.064)	.999 (.001)	.010 (.013)
1.25	**765**	**204**	**816**	**612**	**.318 (.110)**	**.122 (.065)**	**.334 (.117)**	**.128 (.069)**	**.944 (.008)**	**.284 (.053)**
1.50	918	204	816	612	.286 (.106)	.120 (.061)	.299 (.109)	.125 (.065)	.906 (.011)	.408 (.055)
1.75	1,071	204	816	612	.267 (.099)	.124 (.062)	.281 (.103)	.130 (.065)	.879 (.013)	.471 (.048)
2	1,224	204	816	612	.255 (.096)	.126 (.061)	.265 (.099)	.130 (.065)	.859 (.015)	.518 (.046)
3	1,836	204	816	612	.217 (.082)	.131 (.060)	.222 (.080)	.132 (.059)	.813 (.017)	.603 (.037)
4	2,448	204	816	612	.197 (.073)	.131 (.058)	.201 (.070)	.133 (.057)	.789 (.018)	.637 (.032)
5	3,060	204	816	612	.187 (.068)	.135 (.057)	.190 (.064)	.136 (.055)	.775 (.020)	.657 (.030)
6	3,672	204	816	612	.176 (.057)	.133 (.050)	.177 (.055)	.134 (.049)	.766 (.020)	.668 (.029)
10	6,120	204	816	612	.158 (.049)	.132 (.046)	.160 (.047)	.134 (.045)	.746 (.021)	.688 (.027)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table F27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Four Response Options where the Number of Items for the Simulation Condition are Divided into Four Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	614	204	816	612	.433 (.102)	.176 (.072)	.477 (.115)	.195 (.080)	1.00 (.001)	.013 (.015)
1.25	**765**	**204**	**816**	**612**	**.410 (.104)**	**.189 (.075)**	**.444 (.112)**	**.205 (.082)**	**.954 (.007)**	**.355 (.056)**
1.50	918	204	816	612	.392 (.104)	.201 (.078)	.421 (.113)	.215 (.085)	.924 (.009)	.484 (.049)
1.75	1,071	204	816	612	.376 (.101)	.207 (.078)	.401 (.108)	.220 (.083)	.902 (.011)	.551 (.047)
2	1,224	204	816	612	.365 (.103)	.213 (.079)	.387 (.109)	.226 (.083)	.886 (.012)	.594 (.038)
3	1,836	204	816	612	.337 (.100)	.228 (.085)	.351 (.101)	.237 (.086)	.847 (.014)	.672 (.032)
4	2,448	204	816	612	.318 (.096)	.234 (.084)	.328 (.097)	.241 (.086)	.827 (.016)	.701 (.027)
5	3,060	204	816	612	.305 (.093)	.235 (.085)	.312 (.094)	.241 (.087)	.816 (.016)	.718 (.026)
6	3,672	204	816	612	.299 (.089)	.242 (.083)	.307 (.091)	.247 (.085)	.809 (.016)	.728 (.025)
10	6,120	204	816	612	.274 (.078)	.240 (.075)	.280 (.078)	.245 (.076)	.794 (.017)	.747 (.022)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into four binary items; “Dummy-coded Options” are the number of base items split into four binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix G: Simulation Results for Items with Five Response Options – Dummy-coded Response Options Summed Equals Item Condition

Table G1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.416 (.083)	.045 (.057)	.524 (.072)	.042 (.056)	.981 (.024)	.044 (.056)
1.25	60	12	60	48	.377 (.073)	.039 (.050)	.475 (.067)	.040 (.052)	.824 (.066)	.037 (.050)
1.50	72	12	60	48	.339 (.070)	.031 (.042)	.426 (.063)	.031 (.044)	.689 (.075)	.031 (.041)
1.75	84	12	60	48	.311 (.065)	.027 (.036)	.389 (.061)	.027 (.035)	.593 (.075)	.028 (.039)
2	96	12	60	48	.285 (.064)	.025 (.032)	.358 (.059)	.025 (.033)	.527 (.073)	.026 (.032)
3	144	12	60	48	.221 (.048)	.019 (.026)	.277 (.047)	.021 (.027)	.366 (.058)	.020 (.026)
4	192	12	60	48	.185 (.041)	.016 (.021)	.229 (.040)	.018 (.023)	.286 (.048)	.017 (.022)
5	**240**	**12**	**60**	**48**	**.158 (.037)**	**.015 (.020)**	**.195 (.037)**	**.017 (.021)**	**.236 (.043)**	**.017 (.021)**
6	288	12	60	48	.139 (.032)	.015 (.018)	.171 (.032)	.017 (.020)	.202 (.037)	.017 (.020)
10	480	12	60	48	.103 (.025)	.015 (.016)	.124 (.026)	.018 (.018)	.140 (.029)	.018 (.018)
25	1,200	12	60	48	.066 (.020)	.022 (.018)	.077 (.020)	.025 (.019)	.083 (.022)	.027 (.020)
50	2,400	12	60	48	.053 (.017)	.026 (.016)	.059 (.017)	.030 (.017)	.063 (.018)	.031 (.018)
100	4,800	12	60	48	.047 (.016)	.033 (.016)	.052 (.017)	.036 (.017)	.055 (.018)	.038 (.018)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.433 (.085)	.055 (.070)	.538 (.072)	.058 (.074)	.982 (.026)	.046 (.061)
1.25	60	12	60	48	.406 (.076)	.046 (.062)	.497 (.067)	.047 (.064)	.841 (.059)	.041 (.054)
1.50	72	12	60	48	.367 (.075)	.044 (.056)	.454 (.067)	.049 (.060)	.716 (.072)	.044 (.056)
1.75	84	12	60	48	.341 (.069)	.045 (.056)	.419 (.065)	.050 (.061)	.633 (.075)	.043 (.057)
2	96	12	60	48	.324 (.065)	.042 (.052)	.394 (.061)	.046 (.056)	.573 (.073)	.044 (.053)
3	**144**	**12**	**60**	**48**	**.262 (.060)**	**.042 (.046)**	**.316 (.059)**	**.048 (.051)**	**.421 (.073)**	**.053 (.058)**
4	192	12	60	48	.230 (.055)	.044 (.045)	.275 (.055)	.051 (.048)	.351 (.069)	.057 (.056)
5	240	12	60	48	.210 (.053)	.048 (.043)	.248 (.053)	.055 (.047)	.307 (.064)	.065 (.057)
6	288	12	60	48	.197 (.052)	.053 (.046)	.231 (.053)	.061 (.051)	.282 (.065)	.073 (.061)
10	480	12	60	48	.164 (.045)	.063 (.047)	.188 (.047)	.071 (.051)	.222 (.057)	.086 (.061)
25	1,200	12	60	48	.139 (.050)	.088 (.050)	.153 (.052)	.097 (.053)	.175 (.063)	.115 (.064)
50	2,400	12	60	48	.126 (.049)	.099 (.049)	.136 (.052)	.107 (.052)	.155 (.061)	.124 (.062)
100	4,800	12	60	48	.119 (.047)	.104 (.048)	.127 (.049)	.111 (.051)	.143 (.056)	.127 (.059)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.475 (.085)	.077 (.087)	.574 (.072)	.086 (.094)	.987 (.019)	.059 (.077)
1.25	60	12	60	48	.442 (.083)	.082 (.093)	.533 (.072)	.092 (.099)	.874 (.057)	.082 (.104)
1.50	**72**	**12**	**60**	**48**	**.422 (.081)**	**.085 (.087)**	**.504 (.073)**	**.095 (.093)**	**.785 (.078)**	**.118 (.129)**
1.75	84	12	60	48	.399 (.085)	.088 (.091)	.474 (.080)	.102 (.099)	.718 (.088)	.134 (.139)
2	96	12	60	48	.378 (.081)	.094 (.090)	.448 (.078)	.107 (.099)	.667 (.098)	.150 (.148)
3	144	12	60	48	.334 (.078)	.100 (.082)	.391 (.078)	.115 (.090)	.558 (.113)	.192 (.153)
4	192	12	60	48	.313 (.081)	.120 (.092)	.359 (.082)	.138 (.100)	.501 (.122)	.224 (.162)
5	240	12	60	48	.297 (.081)	.128 (.091)	.339 (.083)	.146 (.098)	.467 (.125)	.237 (.158)
6	288	12	60	48	.292 (.084)	.142 (.092)	.329 (.088)	.161 (.101)	.451 (.131)	.261 (.164)
10	480	12	60	48	.267 (.084)	.165 (.093)	.297 (.089)	.184 (.100)	.407 (.133)	.287 (.156)
25	1,200	12	60	48	.257 (.098)	.209 (.103)	.277 (.103)	.226 (.108)	.379 (.151)	.328 (.161)
50	2,400	12	60	48	.247 (.100)	.222 (.102)	.263 (.106)	.237 (.108)	.354 (.151)	.330 (.156)
100	4,800	12	60	48	.244 (.100)	.231 (.102)	.259 (.106)	.245 (.108)	.348 (.148)	.336 (.152)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.410 (.085)	.042 (.055)	.516 (.073)	.042 (.056)	.981 (.025)	.037 (.048)
1.25	60	12	60	48	.369 (.073)	.037 (.049)	.465 (.066)	.037 (.048)	.819 (.066)	.036 (.051)
1.50	72	12	60	48	.333 (.070)	.031 (.041)	.419 (.064)	.031 (.039)	.688 (.075)	.029 (.040)
1.75	84	12	60	48	.308 (.066)	.028 (.039)	.385 (.061)	.028 (.038)	.595 (.075)	.026 (.035)
2	96	12	60	48	.280 (.059)	.025 (.032)	.351 (.056)	.026 (.034)	.525 (.073)	.023 (.032)
3	144	12	60	48	.214 (.047)	.020 (.027)	.266 (.045)	.020 (.027)	.360 (.057)	.019 (.026)
4	192	12	60	48	.176 (.037)	.016 (.021)	.217 (.037)	.017 (.022)	.279 (.045)	.015 (.020)
5	240	12	60	48	.152 (.035)	.016 (.021)	.186 (.035)	.017 (.022)	.232 (.042)	.015 (.019)
6	288	12	60	48	.132 (.030)	.014 (.018)	.161 (.030)	.016 (.020)	.197 (.036)	.014 (.019)
10	480	12	60	48	.094 (.023)	.014 (.015)	.113 (.023)	.016 (.016)	.133 (.028)	.015 (.016)
25	1,200	12	60	48	.058 (.015)	.019 (.014)	.066 (.015)	.021 (.015)	.076 (.018)	.020 (.015)
50	**2,400**	**12**	**60**	**48**	**.044 (.013)**	**.022 (.013)**	**.048 (.013)**	**.024 (.014)**	**.056 (.016)**	**.025 (.015)**
100	4,800	12	60	48	.038 (.013)	.026 (.014)	.040 (.013)	.028 (.014)	.047 (.015)	.031 (.016)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.431 (.085)	.051 (.067)	.535 (.071)	.053 (.072)	.983 (.022)	.044 (.058)
1.25	60	12	60	48	.395 (.080)	.050 (.064)	.490 (.071)	.052 (.067)	.840 (.062)	.040 (.057)
1.50	72	12	60	48	.368 (.074)	.047 (.058)	.450 (.068)	.049 (.060)	.720 (.076)	.043 (.054)
1.75	84	12	60	48	.339 (.071)	.044 (.056)	.415 (.066)	.047 (.059)	.636 (.077)	.042 (.053)
2	96	12	60	48	.315 (.069)	.044 (.051)	.383 (.063)	.047 (.056)	.569 (.075)	.042 (.049)
3	**144**	**12**	**60**	**48**	**.259 (.057)**	**.043 (.047)**	**.312 (.056)**	**.049 (.052)**	**.427 (.071)**	**.052 (.056)**
4	192	12	60	48	.224 (.053)	.044 (.044)	.267 (.053)	.051 (.049)	.350 (.067)	.060 (.058)
5	240	12	60	48	.205 (.052)	.048 (.045)	.241 (.052)	.055 (.049)	.310 (.068)	.065 (.059)
6	288	12	60	48	.188 (.045)	.047 (.042)	.220 (.046)	.054 (.046)	.278 (.060)	.067 (.058)
10	480	12	60	48	.159 (.045)	.062 (.044)	.182 (.046)	.070 (.046)	.226 (.060)	.089 (.059)
25	1,200	12	60	48	.126 (.042)	.078 (.044)	.138 (.044)	.086 (.047)	.169 (.056)	.109 (.061)
50	2,400	12	60	48	.116 (.043)	.090 (.044)	.125 (.045)	.097 (.046)	.153 (.057)	.122 (.059)
100	4,800	12	60	48	.109 (.043)	.095 (.045)	.116 (.045)	.101 (.047)	.142 (.058)	.126 (.060)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.468 (.087)	.077 (.088)	.566 (.075)	.086 (.093)	.987 (.017)	.061 (.082)
1.25	**60**	**12**	**60**	**48**	**.438 (.082)**	**.078 (.088)**	**.528 (.072)**	**.089 (.095)**	**.873 (.060)**	**.093 (.112)**
1.50	72	12	60	48	.413 (.080)	.078 (.087)	.496 (.073)	.090 (.094)	.791 (.076)	.114 (.126)
1.75	84	12	60	48	.390 (.078)	.080 (.080)	.464 (.073)	.091 (.088)	.724 (.086)	.137 (.135)
2	96	12	60	48	.372 (.076)	.087 (.084)	.438 (.072)	.099 (.090)	.674 (.094)	.159 (.145)
3	144	12	60	48	.330 (.079)	.104 (.087)	.382 (.079)	.121 (.094)	.566 (.114)	.215 (.161)
4	192	12	60	48	.307 (.078)	.117 (.089)	.351 (.078)	.133 (.095)	.516 (.121)	.239 (.162)
5	240	12	60	48	.293 (.081)	.128 (.087)	.333 (.083)	.146 (.094)	.486 (.134)	.267 (.170)
6	288	12	60	48	.277 (.082)	.131 (.089)	.313 (.085)	.148 (.096)	.454 (.134)	.265 (.165)
10	480	12	60	48	.254 (.082)	.156 (.089)	.282 (.087)	.174 (.095)	.411 (.139)	.293 (.161)
25	1,200	12	60	48	.232 (.087)	.185 (.089)	.250 (.091)	.200 (.094)	.366 (.142)	.317 (.153)
50	2,400	12	60	48	.232 (.094)	.208 (.097)	.246 (.099)	.222 (.103)	.360 (.149)	.337 (.155)
100	4,800	12	60	48	.231 (.096)	.219 (.097)	.244 (.101)	.231 (.102)	.354 (.149)	.343 (.153)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.401 (.080)	.045 (.061)	.503 (.072)	.046 (.061)	.981 (.025)	.043 (.056)
1.25	60	12	60	48	.361 (.077)	.038 (.050)	.452 (.069)	.037 (.050)	.818 (.071)	.037 (.050)
1.50	72	12	60	48	.326 (.070)	.033 (.042)	.407 (.065)	.033 (.043)	.688 (.074)	.031 (.041)
1.75	84	12	60	48	.291 (.066)	.027 (.033)	.367 (.061)	.027 (.034)	.593 (.074)	.025 (.033)
2	96	12	60	48	.268 (.060)	.023 (.032)	.334 (.057)	.024 (.035)	.521 (.070)	.023 (.033)
3	144	12	60	48	.205 (.047)	.019 (.025)	.253 (.045)	.019 (.025)	.361 (.057)	.017 (.024)
4	192	12	60	48	.166 (.039)	.016 (.022)	.204 (.039)	.017 (.023)	.278 (.048)	.016 (.022)
5	240	12	60	48	.142 (.033)	.015 (.019)	.173 (.034)	.016 (.020)	.231 (.042)	.015 (.020)
6	288	12	60	48	.126 (.029)	.014 (.018)	.152 (.029)	.015 (.018)	.200 (.035)	.014 (.017)
10	**480**	**12**	**60**	**48**	**.086 (.021)**	**.013 (.015)**	**.102 (.021)**	**.014 (.016)**	**.132 (.027)**	**.014 (.015)**
25	1,200	12	60	48	.051 (.014)	.017 (.013)	.057 (.015)	.019 (.013)	.076 (.019)	.021 (.015)
50	2,400	12	60	48	.038 (.012)	.020 (.012)	.042 (.012)	.021 (.013)	.057 (.016)	.026 (.016)
100	4,800	12	60	48	.032 (.011)	.023 (.011)	.034 (.011)	.024 (.012)	.048 (.015)	.031 (.016)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.424 (.083)	.055 (.071)	.523 (.074)	.055 (.071)	.981 (.028)	.047 (.061)
1.25	60	12	60	48	.385 (.080)	.046 (.059)	.476 (.071)	.048 (.064)	.840 (.063)	.042 (.056)
1.50	72	12	60	48	.355 (.077)	.047 (.061)	.435 (.068)	.050 (.063)	.725 (.073)	.048 (.064)
1.75	84	12	60	48	.332 (.072)	.044 (.054)	.402 (.067)	.049 (.057)	.642 (.080)	.048 (.061)
2	**96**	**12**	**60**	**48**	**.306 (.067)**	**.042 (.051)**	**.371 (.062)**	**.046 (.055)**	**.577 (.079)**	**.047 (.058)**
3	144	12	60	48	.250 (.058)	.041 (.046)	.299 (.057)	.046 (.050)	.436 (.075)	.060 (.067)
4	192	12	60	48	.215 (.054)	.042 (.043)	.257 (.055)	.048 (.047)	.363 (.075)	.068 (.066)
5	240	12	60	48	.195 (.049)	.045 (.042)	.230 (.050)	.051 (.045)	.319 (.068)	.075 (.066)
6	288	12	60	48	.179 (.047)	.046 (.042)	.209 (.048)	.051 (.045)	.290 (.067)	.079 (.065)
10	480	12	60	48	.148 (.046)	.058 (.043)	.169 (.049)	.065 (.047)	.234 (.069)	.100 (.070)
25	1,200	12	60	48	.117 (.042)	.072 (.041)	.129 (.045)	.079 (.044)	.179 (.065)	.118 (.067)
50	2,400	12	60	48	.106 (.043)	.081 (.043)	.113 (.046)	.087 (.045)	.162 (.066)	.130 (.068)
100	4,800	12	60	48	.105 (.046)	.092 (.046)	.111 (.049)	.097 (.049)	.156 (.066)	.140 (.067)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	50	12	60	48	.454 (.089)	.072 (.085)	.551 (.076)	.079 (.091)	.986 (.020)	.058 (.072)
1.25	**60**	**12**	**60**	**48**	**.423 (.079)**	**.078 (.085)**	**.510 (.072)**	**.086 (.090)**	**.878 (.058)**	**.094 (.109)**
1.50	72	12	60	48	.398 (.083)	.080 (.085)	.478 (.076)	.090 (.091)	.796 (.074)	.129 (.129)
1.75	84	12	60	48	.380 (.080)	.075 (.077)	.452 (.076)	.087 (.086)	.735 (.086)	.159 (.143)
2	96	12	60	48	.361 (.074)	.082 (.081)	.428 (.072)	.093 (.090)	.688 (.093)	.177 (.152)
3	144	12	60	48	.316 (.076)	.095 (.081)	.369 (.077)	.109 (.089)	.583 (.114)	.232 (.164)
4	192	12	60	48	.293 (.077)	.110 (.085)	.337 (.079)	.125 (.094)	.536 (.127)	.268 (.172)
5	240	12	60	48	.281 (.074)	.119 (.080)	.319 (.076)	.135 (.085)	.508 (.123)	.288 (.161)
6	288	12	60	48	.264 (.077)	.121 (.083)	.297 (.078)	.136 (.088)	.475 (.129)	.288 (.165)
10	480	12	60	48	.244 (.082)	.149 (.087)	.270 (.087)	.165 (.094)	.439 (.141)	.325 (.168)
25	1,200	12	60	48	.227 (.091)	.184 (.095)	.245 (.096)	.198 (.101)	.406 (.150)	.360 (.161)
50	2,400	12	60	48	.217 (.087)	.192 (.089)	.231 (.092)	.205 (.094)	.387 (.148)	.363 (.154)
100	4,800	12	60	48	.217 (.092)	.205 (.092)	.230 (.097)	.217 (.098)	.381 (.150)	.370 (.153)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.422 (.053)	.025 (.034)	.523 (.046)	.027 (.034)	.993 (.011)	.017 (.024)
1.25	150	30	150	120	.388 (.049)	.024 (.031)	.478 (.043)	.027 (.033)	.828 (.042)	.019 (.024)
1.50	180	30	150	120	.358 (.044)	.024 (.029)	.440 (.041)	.027 (.031)	.709 (.046)	.019 (.025)
1.75	210	30	150	120	.331 (.044)	.024 (.026)	.405 (.040)	.027 (.029)	.623 (.048)	.021 (.026)
2	240	30	150	120	.310 (.041)	.024 (.026)	.379 (.039)	.028 (.029)	.561 (.046)	.023 (.027)
3	**360**	**30**	**150**	**120**	**.256 (.036)**	**.026 (.023)**	**.308 (.035)**	**.030 (.025)**	**.413 (.043)**	**.030 (.026)**
4	480	30	150	120	.219 (.032)	.031 (.024)	.263 (.032)	.037 (.026)	.336 (.039)	.038 (.028)
5	600	30	150	120	.198 (.030)	.035 (.024)	.236 (.030)	.041 (.026)	.291 (.036)	.045 (.029)
6	720	30	150	120	.183 (.031)	.038 (.024)	.217 (.031)	.045 (.027)	.262 (.036)	.049 (.030)
10	1,200	30	150	120	.153 (.026)	.050 (.025)	.177 (.027)	.058 (.027)	.205 (.032)	.066 (.031)
25	3,000	30	150	120	.122 (.025)	.070 (.025)	.135 (.026)	.078 (.027)	.152 (.029)	.089 (.030)
50	6,000	30	150	120	.111 (.024)	.082 (.025)	.120 (.025)	.089 (.026)	.133 (.028)	.100 (.029)
100	12,000	30	150	120	.107 (.024)	.092 (.025)	.114 (.026)	.098 (.027)	.126 (.028)	.109 (.029)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.481 (.059)	.071 (.063)	.574 (.050)	.084 (.071)	.995 (.007)	.025 (.035)
1.25	**150**	**30**	**150**	**120**	**.454 (.056)**	**.077 (.066)**	**.538 (.050)**	**.092 (.071)**	**.889 (.037)**	**.092 (.083)**
1.50	180	30	150	120	.429 (.055)	.086 (.065)	.507 (.051)	.101 (.071)	.813 (.053)	.145 (.108)
1.75	210	30	150	120	.411 (.055)	.092 (.063)	.483 (.051)	.109 (.068)	.757 (.062)	.182 (.122)
2	240	30	150	120	.397 (.054)	.098 (.061)	.463 (.051)	.117 (.067)	.713 (.067)	.199 (.116)
3	360	30	150	120	.361 (.057)	.123 (.066)	.415 (.057)	.143 (.071)	.622 (.087)	.274 (.134)
4	480	30	150	120	.341 (.059)	.144 (.069)	.387 (.059)	.165 (.074)	.575 (.092)	.311 (.133)
5	600	30	150	120	.327 (.060)	.158 (.070)	.369 (.062)	.180 (.075)	.544 (.100)	.331 (.134)
6	720	30	150	120	.319 (.060)	.169 (.068)	.358 (.062)	.192 (.074)	.532 (.101)	.353 (.132)
10	1,200	30	150	120	.306 (.066)	.204 (.073)	.337 (.069)	.226 (.077)	.493 (.111)	.383 (.133)
25	3,000	30	150	120	.288 (.070)	.242 (.075)	.310 (.073)	.261 (.079)	.456 (.116)	.411 (.127)
50	6,000	30	150	120	.286 (.074)	.262 (.077)	.305 (.078)	.279 (.081)	.444 (.117)	.421 (.122)
100	12,000	30	150	120	.286 (.071)	.273 (.072)	.303 (.075)	.289 (.076)	.440 (.115)	.428 (.117)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.530 (.061)	.132 (.092)	.616 (.052)	.157 (.099)	.997 (.004)	.043 (.052)
1.25	**150**	**30**	**150**	**120**	**.512 (.061)**	**.156 (.088)**	**.589 (.054)**	**.181 (.096)**	**.934 (.025)**	**.242 (.127)**
1.50	180	30	150	120	.492 (.059)	.172 (.090)	.565 (.054)	.199 (.094)	.888 (.037)	.350 (.137)
1.75	210	30	150	120	.484 (.061)	.187 (.092)	.551 (.056)	.217 (.098)	.857 (.044)	.410 (.137)
2	240	30	150	120	.474 (.061)	.204 (.093)	.538 (.058)	.235 (.099)	.833 (.052)	.454 (.134)
3	360	30	150	120	.446 (.062)	.233 (.091)	.499 (.061)	.265 (.096)	.771 (.065)	.526 (.126)
4	480	30	150	120	.440 (.068)	.268 (.088)	.487 (.068)	.301 (.092)	.749 (.072)	.575 (.115)
5	600	30	150	120	.434 (.070)	.291 (.091)	.476 (.070)	.322 (.095)	.731 (.074)	.591 (.109)
6	720	30	150	120	.434 (.070)	.310 (.088)	.474 (.071)	.341 (.091)	.725 (.074)	.610 (.103)
10	1,200	30	150	120	.423 (.074)	.340 (.087)	.455 (.075)	.368 (.090)	.700 (.080)	.631 (.099)
25	3,000	30	150	120	.420 (.079)	.385 (.084)	.446 (.082)	.409 (.087)	.681 (.086)	.654 (.093)
50	6,000	30	150	120	.421 (.082)	.403 (.084)	.444 (.086)	.425 (.088)	.677 (.086)	.664 (.090)
100	12,000	30	150	120	.420 (.083)	.411 (.085)	.441 (.086)	.432 (.088)	.670 (.088)	.664 (.089)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.405 (.052)	.023 (.030)	.498 (.046)	.025 (.033)	.993 (.010)	.017 (.023)
1.25	150	30	150	120	.364 (.047)	.021 (.027)	.445 (.045)	.022 (.028)	.822 (.042)	.015 (.021)
1.50	180	30	150	120	.330 (.046)	.023 (.028)	.402 (.044)	.025 (.029)	.701 (.048)	.017 (.022)
1.75	210	30	150	120	.304 (.043)	.021 (.024)	.368 (.042)	.024 (.026)	.613 (.049)	.017 (.023)
2	240	30	150	120	.282 (.041)	.021 (.025)	.340 (.040)	.023 (.026)	.550 (.047)	.017 (.021)
3	360	30	150	120	.218 (.032)	.023 (.023)	.259 (.033)	.026 (.024)	.391 (.038)	.021 (.022)
4	480	30	150	120	.184 (.028)	.024 (.020)	.214 (.029)	.027 (.022)	.315 (.035)	.024 (.021)
5	600	30	150	120	.163 (.026)	.029 (.021)	.189 (.028)	.032 (.022)	.272 (.033)	.030 (.022)
6	720	30	150	120	.146 (.023)	.028 (.020)	.167 (.025)	.032 (.021)	.240 (.031)	.031 (.022)
10	**1,200**	**30**	**150**	**120**	**.111 (.019)**	**.035 (.019)**	**.123 (.020)**	**.039 (.019)**	**.179 (.026)**	**.044 (.023)**
25	3,000	30	150	120	.078 (.016)	.043 (.016)	.083 (.016)	.046 (.017)	.125 (.023)	.063 (.024)
50	6,000	30	150	120	.065 (.015)	.048 (.016)	.069 (.015)	.051 (.016)	.106 (.023)	.073 (.024)
100	12,000	30	150	120	.059 (.014)	.051 (.015)	.061 (.014)	.053 (.015)	.097 (.022)	.080 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.458 (.057)	.067 (.060)	.546 (.051)	.077 (.065)	.995 (.008)	.027 (.035)
1.25	**150**	**30**	**150**	**120**	**.432 (.054)**	**.072 (.058)**	**.509 (.051)**	**.084 (.063)**	**.887 (.037)**	**.087 (.083)**
1.50	180	30	150	120	.401 (.054)	.075 (.058)	.473 (.051)	.088 (.064)	.806 (.050)	.128 (.093)
1.75	210	30	150	120	.385 (.054)	.083 (.059)	.449 (.052)	.095 (.064)	.752 (.057)	.166 (.107)
2	240	30	150	120	.366 (.053)	.087 (.058)	.425 (.053)	.100 (.062)	.707 (.066)	.187 (.109)
3	360	30	150	120	.324 (.052)	.105 (.056)	.370 (.054)	.121 (.061)	.611 (.081)	.257 (.122)
4	480	30	150	120	.301 (.052)	.120 (.057)	.341 (.054)	.137 (.061)	.569 (.086)	.302 (.123)
5	600	30	150	120	.286 (.055)	.132 (.059)	.319 (.057)	.148 (.062)	.534 (.095)	.316 (.127)
6	720	30	150	120	.278 (.055)	.142 (.060)	.307 (.058)	.157 (.064)	.520 (.097)	.337 (.127)
10	1,200	30	150	120	.251 (.054)	.161 (.059)	.273 (.057)	.176 (.063)	.479 (.100)	.368 (.120)
25	3,000	30	150	120	.231 (.061)	.192 (.064)	.246 (.065)	.204 (.068)	.442 (.109)	.397 (.119)
50	6,000	30	150	120	.223 (.060)	.202 (.060)	.234 (.064)	.212 (.064)	.430 (.110)	.408 (.115)
100	12,000	30	150	120	.221 (.061)	.211 (.062)	.231 (.065)	.220 (.065)	.424 (.109)	.413 (.111)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.513 (.061)	.125 (.086)	.597 (.053)	.147 (.093)	.997 (.004)	.041 (.054)
1.25	**150**	**30**	**150**	**120**	**.492 (.060)**	**.141 (.084)**	**.566 (.054)**	**.165 (.090)**	**.932 (.025)**	**.224 (.123)**
1.50	180	30	150	120	.473 (.061)	.159 (.085)	.541 (.058)	.185 (.091)	.885 (.036)	.327 (.132)
1.75	210	30	150	120	.456 (.060)	.169 (.087)	.520 (.058)	.197 (.094)	.850 (.046)	.390 (.131)
2	240	30	150	120	.446 (.060)	.179 (.082)	.504 (.058)	.207 (.087)	.824 (.050)	.431 (.132)
3	360	30	150	120	.424 (.065)	.221 (.087)	.472 (.065)	.249 (.092)	.769 (.062)	.523 (.116)
4	480	30	150	120	.406 (.065)	.244 (.086)	.448 (.066)	.270 (.090)	.738 (.069)	.556 (.117)
5	600	30	150	120	.395 (.066)	.258 (.078)	.433 (.068)	.284 (.083)	.720 (.071)	.578 (.105)
6	720	30	150	120	.389 (.070)	.266 (.081)	.424 (.072)	.292 (.085)	.711 (.073)	.592 (.102)
10	1,200	30	150	120	.383 (.072)	.303 (.082)	.410 (.075)	.327 (.087)	.689 (.078)	.619 (.098)
25	3,000	30	150	120	.368 (.079)	.334 (.083)	.389 (.083)	.353 (.087)	.660 (.086)	.633 (.094)
50	6,000	30	150	120	.368 (.083)	.350 (.085)	.385 (.087)	.367 (.089)	.659 (.088)	.645 (.092)
100	12,000	30	150	120	.370 (.077)	.361 (.079)	.386 (.081)	.377 (.082)	.659 (.086)	.652 (.088)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.376 (.062)	.024 (.033)	.461 (.064)	.024 (.033)	.992 (.010)	.018 (.022)
1.25	150	30	150	120	.332 (.057)	.022 (.029)	.409 (.062)	.024 (.029)	.820 (.045)	.017 (.023)
1.50	180	30	150	120	.304 (.053)	.021 (.028)	.369 (.061)	.023 (.030)	.701 (.047)	.016 (.022)
1.75	210	30	150	120	.278 (.051)	.019 (.023)	.335 (.057)	.022 (.025)	.617 (.048)	.018 (.024)
2	240	30	150	120	.253 (.049)	.019 (.022)	.305 (.056)	.021 (.024)	.548 (.046)	.018 (.022)
3	**360**	**30**	**150**	**120**	**.194 (.040)**	**.018 (.018)**	**.227 (.045)**	**.021 (.020)**	**.395 (.040)**	**.021 (.021)**
4	480	30	150	120	.159 (.033)	.022 (.019)	.185 (.038)	.025 (.020)	.320 (.036)	.028 (.023)
5	600	30	150	120	.139 (.032)	.023 (.018)	.158 (.037)	.025 (.019)	.274 (.035)	.032 (.024)
6	720	30	150	120	.121 (.028)	.023 (.017)	.137 (.031)	.026 (.018)	.244 (.031)	.036 (.024)
10	1,200	30	150	120	.090 (.022)	.028 (.016)	.098 (.024)	.031 (.017)	.184 (.028)	.048 (.025)
25	3,000	30	150	120	.058 (.015)	.033 (.014)	.061 (.016)	.035 (.015)	.129 (.025)	.068 (.027)
50	6,000	30	150	120	.047 (.013)	.035 (.013)	.049 (.013)	.036 (.013)	.112 (.025)	.079 (.026)
100	12,000	30	150	120	.043 (.012)	.037 (.012)	.044 (.012)	.038 (.012)	.103 (.025)	.086 (.026)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.431 (.067)	.060 (.061)	.514 (.063)	.070 (.067)	.996 (.006)	.026 (.034)
1.25	**150**	**30**	**150**	**120**	**.401 (.061)**	**.064 (.056)**	**.470 (.063)**	**.073 (.060)**	**.890 (.034)**	**.095 (.083)**
1.50	180	30	150	120	.373 (.062)	.067 (.055)	.437 (.066)	.079 (.060)	.818 (.049)	.149 (.097)
1.75	210	30	150	120	.354 (.060)	.070 (.053)	.412 (.064)	.080 (.056)	.764 (.054)	.183 (.107)
2	240	30	150	120	.338 (.061)	.075 (.053)	.392 (.066)	.086 (.057)	.723 (.060)	.215 (.119)
3	360	30	150	120	.292 (.060)	.091 (.053)	.334 (.065)	.104 (.057)	.630 (.074)	.285 (.121)
4	480	30	150	120	.266 (.061)	.100 (.057)	.299 (.065)	.114 (.061)	.580 (.086)	.318 (.126)
5	600	30	150	120	.249 (.060)	.112 (.056)	.276 (.065)	.124 (.060)	.552 (.090)	.340 (.123)
6	720	30	150	120	.237 (.061)	.117 (.056)	.262 (.065)	.130 (.060)	.536 (.092)	.360 (.120)
10	1,200	30	150	120	.213 (.062)	.133 (.058)	.232 (.068)	.145 (.063)	.496 (.095)	.388 (.116)
25	3,000	30	150	120	.196 (.065)	.162 (.063)	.206 (.068)	.170 (.066)	.471 (.102)	.428 (.111)
50	6,000	30	150	120	.185 (.063)	.166 (.062)	.194 (.067)	.175 (.066)	.454 (.105)	.432 (.110)
100	12,000	30	150	120	.184 (.065)	.175 (.065)	.191 (.069)	.182 (.068)	.452 (.110)	.441 (.113)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	122	30	150	120	.483 (.068)	.117 (.086)	.562 (.065)	.136 (.091)	.997 (.004)	.039 (.050)
1.25	**150**	**30**	**150**	**120**	**.460 (.069)**	**.129 (.084)**	**.533 (.070)**	**.149 (.092)**	**.931 (.026)**	**.224 (.123)**
1.50	180	30	150	120	.436 (.068)	.137 (.081)	.501 (.070)	.157 (.086)	.882 (.037)	.320 (.128)
1.75	210	30	150	120	.422 (.067)	.150 (.082)	.483 (.071)	.173 (.087)	.851 (.043)	.388 (.127)
2	240	30	150	120	.414 (.070)	.157 (.081)	.468 (.074)	.179 (.087)	.825 (.049)	.426 (.130)
3	360	30	150	120	.386 (.071)	.183 (.077)	.428 (.075)	.206 (.083)	.768 (.059)	.517 (.113)
4	480	30	150	120	.368 (.072)	.209 (.080)	.405 (.077)	.231 (.085)	.734 (.067)	.551 (.108)
5	600	30	150	120	.358 (.075)	.224 (.081)	.391 (.080)	.246 (.086)	.718 (.071)	.574 (.105)
6	720	30	150	120	.357 (.081)	.241 (.085)	.387 (.086)	.262 (.091)	.711 (.070)	.594 (.098)
10	1,200	30	150	120	.334 (.080)	.258 (.084)	.358 (.086)	.278 (.090)	.682 (.080)	.611 (.098)
25	3,000	30	150	120	.321 (.086)	.286 (.087)	.338 (.091)	.302 (.092)	.663 (.080)	.634 (.088)
50	6,000	30	150	120	.322 (.088)	.305 (.089)	.336 (.093)	.319 (.094)	.659 (.082)	.646 (.085)
100	12,000	30	150	120	.320 (.092)	.311 (.092)	.333 (.097)	.325 (.097)	.650 (.085)	.643 (.086)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.444 (.045)	.031 (.032)	.542 (.037)	.037 (.036)	.996 (.005)	.011 (.015)
1.25	255	51	255	204	.412 (.039)	.034 (.031)	.499 (.034)	.041 (.035)	.843 (.030)	.022 (.024)
1.50	306	51	255	204	.384 (.038)	.039 (.031)	.461 (.034)	.046 (.034)	.738 (.035)	.032 (.028)
1.75	357	51	255	204	.360 (.036)	.042 (.031)	.432 (.033)	.049 (.035)	.660 (.038)	.041 (.032)
2	408	51	255	204	.342 (.035)	.043 (.030)	.409 (.033)	.052 (.034)	.604 (.039)	.048 (.033)
3	**612**	**51**	**255**	**204**	**.293 (.032)**	**.055 (.032)**	**.347 (.031)**	**.066 (.035)**	**.472 (.039)**	**.076 (.040)**
4	816	51	255	204	.267 (.031)	.067 (.032)	.312 (.031)	.079 (.035)	.406 (.040)	.096 (.042)
5	1,020	51	255	204	.250 (.031)	.077 (.031)	.289 (.031)	.089 (.034)	.367 (.040)	.111 (.044)
6	1,224	51	255	204	.236 (.031)	.083 (.033)	.271 (.032)	.096 (.036)	.339 (.040)	.120 (.047)
10	2,040	51	255	204	.212 (.030)	.104 (.034)	.238 (.031)	.118 (.036)	.289 (.040)	.149 (.045)
25	5,100	51	255	204	.185 (.032)	.135 (.034)	.202 (.034)	.148 (.036)	.239 (.042)	.180 (.045)
50	10,200	51	255	204	.177 (.031)	.150 (.032)	.190 (.033)	.161 (.034)	.222 (.040)	.192 (.041)
100	20,400	51	255	204	.176 (.031)	.162 (.032)	.187 (.033)	.172 (.034)	.218 (.040)	.203 (.042)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.518 (.043)	.113 (.067)	.604 (.037)	.135 (.073)	.998 (.003)	.023 (.031)
1.25	**255**	**51**	**255**	**204**	**.494 (.046)**	**.130 (.067)**	**.573 (.040)**	**.154 (.072)**	**.926 (.022)**	**.199 (.098)**
1.50	306	51	255	204	.481 (.045)	.150 (.065)	.554 (.042)	.177 (.071)	.879 (.032)	.309 (.108)
1.75	357	51	255	204	.468 (.046)	.159 (.070)	.535 (.043)	.188 (.075)	.842 (.039)	.361 (.111)
2	408	51	255	204	.456 (.046)	.172 (.064)	.520 (.043)	.200 (.069)	.816 (.043)	.407 (.107)
3	612	51	255	204	.433 (.048)	.212 (.064)	.486 (.047)	.243 (.068)	.757 (.052)	.500 (.096)
4	816	51	255	204	.421 (.052)	.243 (.069)	.468 (.053)	.274 (.073)	.727 (.062)	.535 (.100)
5	1,020	51	255	204	.414 (.053)	.261 (.067)	.457 (.054)	.292 (.071)	.709 (.064)	.558 (.094)
6	1,224	51	255	204	.407 (.053)	.273 (.068)	.446 (.054)	.303 (.071)	.695 (.064)	.569 (.091)
10	2,040	51	255	204	.401 (.059)	.314 (.068)	.434 (.061)	.342 (.071)	.673 (.071)	.599 (.086)
25	5,100	51	255	204	.391 (.062)	.352 (.068)	.416 (.065)	.376 (.071)	.644 (.075)	.614 (.082)
50	10,200	51	255	204	.395 (.060)	.375 (.062)	.418 (.063)	.397 (.064)	.645 (.070)	.630 (.073)
100	20,400	51	255	204	.395 (.063)	.384 (.064)	.415 (.065)	.405 (.067)	.640 (.071)	.633 (.073)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.558 (.046)	.189 (.083)	.639 (.040)	.221 (.089)	.999 (.002)	.035 (.042)
1.25	**255**	**51**	**255**	**204**	**.542 (.047)**	**.212 (.080)**	**.615 (.042)**	**.247 (.084)**	**.950 (.015)**	**.333 (.104)**
1.50	306	51	255	204	.534 (.047)	.241 (.076)	.600 (.042)	.278 (.080)	.918 (.020)	.463 (.096)
1.75	357	51	255	204	.521 (.048)	.251 (.078)	.583 (.043)	.286 (.081)	.892 (.025)	.519 (.096)
2	408	51	255	204	.520 (.049)	.273 (.078)	.579 (.046)	.310 (.080)	.878 (.028)	.572 (.086)
3	612	51	255	204	.501 (.052)	.318 (.077)	.551 (.050)	.356 (.079)	.835 (.036)	.647 (.074)
4	816	51	255	204	.495 (.050)	.345 (.072)	.539 (.050)	.381 (.074)	.814 (.037)	.679 (.065)
5	1,020	51	255	204	.491 (.055)	.368 (.074)	.531 (.055)	.402 (.076)	.801 (.040)	.694 (.062)
6	1,224	51	255	204	.493 (.055)	.387 (.071)	.531 (.055)	.420 (.073)	.795 (.042)	.707 (.062)
10	2,040	51	255	204	.488 (.056)	.420 (.067)	.519 (.057)	.449 (.070)	.777 (.043)	.727 (.054)
25	5,100	51	255	204	.482 (.062)	.454 (.068)	.507 (.064)	.478 (.069)	.759 (.048)	.739 (.052)
50	10,200	51	255	204	.488 (.061)	.474 (.063)	.511 (.063)	.497 (.065)	.758 (.048)	.748 (.050)
100	20,400	51	255	204	.491 (.063)	.484 (.065)	.512 (.065)	.505 (.067)	.757 (.048)	.753 (.049)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.397 (.048)	.028 (.029)	.478 (.049)	.032 (.031)	.996 (.006)	.011 (.015)
1.25	255	51	255	204	.358 (.044)	.028 (.028)	.425 (.048)	.032 (.030)	.831 (.031)	.017 (.021)
1.50	306	51	255	204	.324 (.042)	.030 (.027)	.381 (.046)	.034 (.029)	.717 (.036)	.020 (.024)
1.75	357	51	255	204	.298 (.039)	.032 (.026)	.348 (.043)	.037 (.028)	.638 (.035)	.026 (.024)
2	408	51	255	204	.276 (.038)	.031 (.025)	.320 (.042)	.036 (.026)	.577 (.037)	.029 (.026)
3	**612**	**51**	**255**	**204**	**.220 (.033)**	**.038 (.023)**	**.250 (.036)**	**.043 (.024)**	**.438 (.036)**	**.046 (.028)**
4	816	51	255	204	.186 (.029)	.044 (.023)	.208 (.031)	.048 (.024)	.367 (.034)	.059 (.031)
5	1,020	51	255	204	.164 (.026)	.045 (.021)	.181 (.028)	.050 (.022)	.322 (.033)	.067 (.031)
6	1,224	51	255	204	.149 (.025)	.048 (.021)	.163 (.026)	.052 (.022)	.295 (.033)	.076 (.032)
10	2,040	51	255	204	.118 (.020)	.055 (.019)	.126 (.020)	.058 (.020)	.240 (.032)	.098 (.034)
25	5,100	51	255	204	.088 (.016)	.062 (.017)	.091 (.016)	.064 (.017)	.188 (.030)	.127 (.031)
50	10,200	51	255	204	.078 (.015)	.065 (.015)	.080 (.015)	.067 (.015)	.172 (.031)	.140 (.032)
100	20,400	51	255	204	.072 (.015)	.066 (.015)	.074 (.015)	.067 (.016)	.162 (.031)	.146 (.031)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.480 (.052)	.100 (.061)	.552 (.051)	.116 (.065)	.998 (.003)	.022 (.028)
1.25	**255**	**51**	**255**	**204**	**.447 (.051)**	**.111 (.062)**	**.516 (.053)**	**.128 (.066)**	**.920 (.022)**	**.176 (.090)**
1.50	306	51	255	204	.430 (.051)	.123 (.061)	.491 (.052)	.142 (.065)	.867 (.031)	.270 (.100)
1.75	357	51	255	204	.410 (.051)	.131 (.059)	.465 (.053)	.151 (.063)	.828 (.038)	.326 (.104)
2	408	51	255	204	.394 (.052)	.139 (.058)	.445 (.055)	.158 (.063)	.799 (.043)	.363 (.101)
3	612	51	255	204	.360 (.054)	.164 (.057)	.398 (.057)	.183 (.061)	.734 (.053)	.454 (.100)
4	816	51	255	204	.338 (.053)	.179 (.058)	.371 (.057)	.198 (.062)	.694 (.059)	.484 (.095)
5	1,020	51	255	204	.328 (.056)	.194 (.059)	.355 (.060)	.211 (.063)	.681 (.062)	.516 (.093)
6	1,224	51	255	204	.316 (.054)	.203 (.057)	.341 (.059)	.220 (.062)	.666 (.064)	.531 (.088)
10	2,040	51	255	204	.296 (.058)	.223 (.058)	.315 (.061)	.238 (.061)	.638 (.071)	.557 (.086)
25	5,100	51	255	204	.277 (.060)	.245 (.061)	.289 (.062)	.256 (.064)	.613 (.072)	.581 (.079)
50	10,200	51	255	204	.272 (.059)	.255 (.060)	.283 (.062)	.266 (.063)	.611 (.074)	.595 (.077)
100	20,400	51	255	204	.267 (.061)	.259 (.060)	.277 (.064)	.269 (.064)	.604 (.073)	.596 (.075)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.532 (.051)	.171 (.077)	.606 (.048)	.199 (.081)	.999 (.002)	.033 (.042)
1.25	**255**	**51**	**255**	**204**	**.513 (.054)**	**.191 (.078)**	**.578 (.050)**	**.219 (.082)**	**.948 (.015)**	**.313 (.104)**
1.50	306	51	255	204	.499 (.053)	.212 (.077)	.560 (.053)	.242 (.082)	.913 (.021)	.437 (.100)
1.75	357	51	255	204	.484 (.053)	.224 (.078)	.539 (.054)	.254 (.081)	.888 (.026)	.505 (.095)
2	408	51	255	204	.482 (.054)	.244 (.078)	.532 (.055)	.273 (.082)	.870 (.029)	.548 (.086)
3	612	51	255	204	.453 (.059)	.275 (.073)	.496 (.061)	.304 (.077)	.824 (.038)	.626 (.076)
4	816	51	255	204	.440 (.060)	.298 (.074)	.475 (.063)	.325 (.076)	.802 (.040)	.659 (.069)
5	1,020	51	255	204	.431 (.062)	.311 (.073)	.464 (.066)	.337 (.077)	.787 (.043)	.673 (.067)
6	1,224	51	255	204	.431 (.063)	.327 (.072)	.460 (.067)	.352 (.076)	.782 (.043)	.691 (.062)
10	2,040	51	255	204	.417 (.067)	.352 (.072)	.441 (.071)	.374 (.075)	.763 (.048)	.709 (.059)
25	5,100	51	255	204	.411 (.068)	.383 (.071)	.430 (.071)	.401 (.075)	.751 (.047)	.729 (.053)
50	10,200	51	255	204	.404 (.073)	.390 (.074)	.421 (.076)	.406 (.078)	.740 (.050)	.729 (.052)
100	20,400	51	255	204	.403 (.073)	.396 (.073)	.419 (.076)	.412 (.076)	.740 (.050)	.734 (.051)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.355 (.067)	.023 (.026)	.423 (.079)	.025 (.028)	.996 (.006)	.011 (.016)
1.25	255	51	255	204	.311 (.063)	.025 (.026)	.365 (.075)	.027 (.027)	.835 (.032)	.018 (.021)
1.50	306	51	255	204	.280 (.061)	.025 (.025)	.328 (.072)	.028 (.027)	.723 (.037)	.022 (.024)
1.75	**357**	**51**	**255**	**204**	**.258 (.057)**	**.025 (.023)**	**.299 (.069)**	**.028 (.025)**	**.644 (.036)**	**.029 (.027)**
2	408	51	255	204	.235 (.057)	.027 (.023)	.271 (.067)	.031 (.024)	.585 (.039)	.038 (.030)
3	612	51	255	204	.175 (.047)	.031 (.022)	.196 (.053)	.033 (.022)	.444 (.037)	.054 (.034)
4	816	51	255	204	.144 (.039)	.031 (.019)	.158 (.044)	.034 (.020)	.379 (.037)	.069 (.034)
5	1,020	51	255	204	.126 (.037)	.034 (.019)	.137 (.039)	.037 (.020)	.337 (.037)	.081 (.039)
6	1,224	51	255	204	.113 (.033)	.035 (.019)	.120 (.035)	.038 (.020)	.307 (.038)	.088 (.040)
10	2,040	51	255	204	.085 (.023)	.039 (.016)	.088 (.023)	.040 (.017)	.254 (.037)	.114 (.041)
25	5,100	51	255	204	.059 (.015)	.041 (.013)	.060 (.014)	.042 (.013)	.206 (.036)	.145 (.039)
50	10,200	51	255	204	.050 (.012)	.042 (.012)	.051 (.012)	.042 (.012)	.188 (.037)	.157 (.038)
100	20,400	51	255	204	.046 (.011)	.042 (.012)	.047 (.011)	.043 (.011)	.178 (.037)	.162 (.038)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.426 (.071)	.076 (.052)	.495 (.080)	.089 (.058)	.998 (.003)	.022 (.027)
1.25	**255**	**51**	**255**	**204**	**.402 (.074)**	**.089 (.055)**	**.459 (.084)**	**.102 (.060)**	**.920 (.022)**	**.175 (.087)**
1.50	306	51	255	204	.374 (.072)	.094 (.054)	.426 (.081)	.108 (.059)	.866 (.031)	.263 (.101)
1.75	357	51	255	204	.360 (.072)	.105 (.054)	.406 (.083)	.120 (.060)	.829 (.037)	.325 (.102)
2	408	51	255	204	.343 (.074)	.109 (.056)	.385 (.082)	.123 (.059)	.799 (.042)	.366 (.105)
3	612	51	255	204	.303 (.073)	.129 (.058)	.332 (.080)	.142 (.063)	.735 (.052)	.458 (.097)
4	816	51	255	204	.285 (.072)	.146 (.059)	.310 (.080)	.160 (.065)	.700 (.056)	.495 (.091)
5	1,020	51	255	204	.266 (.075)	.153 (.063)	.287 (.081)	.164 (.067)	.680 (.060)	.518 (.086)
6	1,224	51	255	204	.257 (.070)	.157 (.059)	.275 (.076)	.169 (.065)	.667 (.060)	.530 (.084)
10	2,040	51	255	204	.233 (.069)	.170 (.064)	.245 (.073)	.180 (.067)	.645 (.064)	.566 (.079)
25	5,100	51	255	204	.211 (.066)	.185 (.063)	.218 (.068)	.191 (.065)	.619 (.066)	.587 (.072)
50	10,200	51	255	204	.204 (.067)	.191 (.066)	.211 (.070)	.197 (.069)	.610 (.070)	.595 (.073)
100	20,400	51	255	204	.198 (.063)	.191 (.062)	.204 (.066)	.197 (.066)	.608 (.071)	.600 (.072)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table G27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Sum of the Dummy-coded Response Options Equals the Item Condition

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	206	51	255	204	.492 (.069)	.147 (.076)	.564 (.074)	.170 (.080)	.999 (.002)	.032 (.040)
1.25	**255**	**51**	**255**	**204**	**.464 (.073)**	**.159 (.077)**	**.527 (.078)**	**.183 (.083)**	**.946 (.015)**	**.303 (.107)**
1.50	306	51	255	204	.447 (.071)	.176 (.076)	.504 (.076)	.200 (.083)	.909 (.023)	.423 (.101)
1.75	357	51	255	204	.436 (.073)	.184 (.074)	.485 (.079)	.208 (.081)	.883 (.027)	.490 (.096)
2	408	51	255	204	.429 (.074)	.200 (.076)	.477 (.080)	.224 (.081)	.866 (.028)	.539 (.084)
3	612	51	255	204	.397 (.081)	.229 (.081)	.433 (.088)	.251 (.087)	.821 (.036)	.621 (.076)
4	816	51	255	204	.384 (.080)	.249 (.081)	.414 (.086)	.272 (.088)	.798 (.039)	.653 (.067)
5	1,020	51	255	204	.374 (.083)	.262 (.081)	.402 (.089)	.283 (.088)	.784 (.042)	.669 (.065)
6	1,224	51	255	204	.366 (.086)	.269 (.085)	.390 (.091)	.288 (.089)	.774 (.043)	.681 (.062)
10	2,040	51	255	204	.357 (.087)	.295 (.086)	.375 (.092)	.311 (.092)	.760 (.045)	.705 (.056)
25	5,100	51	255	204	.349 (.087)	.324 (.086)	.363 (.092)	.338 (.091)	.744 (.048)	.723 (.052)
50	10,200	51	255	204	.339 (.090)	.326 (.090)	.351 (.094)	.338 (.094)	.739 (.047)	.728 (.049)
100	20,400	51	255	204	.340 (.091)	.333 (.091)	.351 (.096)	.344 (.095)	.738 (.048)	.733 (.049)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Appendix H: Simulation Results for Items with Five Response Options - Item Condition Divided into Response Options

Table H1

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.444 (.043)	.030 (.032)	.540 (.036)	.034 (.035)	.996 (.006)	.011 (.017)
1.25	240	48	240	192	.408 (.040)	.033 (.032)	.494 (.036)	.040 (.036)	.841 (.032)	.020 (.023)
1.50	288	48	240	192	.380 (.039)	.036 (.031)	.458 (.036)	.041 (.033)	.733 (.038)	.030 (.030)
1.75	336	48	240	192	.357 (.037)	.037 (.031)	.429 (.034)	.044 (.033)	.656 (.038)	.037 (.032)
2	384	48	240	192	.336 (.036)	.041 (.030)	.404 (.033)	.048 (.033)	.597 (.040)	.046 (.035)
3	**576**	**48**	**240**	**192**	**.288 (.032)**	**.051 (.031)**	**.341 (.031)**	**.061 (.034)**	**.462 (.039)**	**.067 (.038)**
4	768	48	240	192	.260 (.033)	.059 (.031)	.304 (.033)	.070 (.034)	.395 (.041)	.084 (.041)
5	960	48	240	192	.243 (.031)	.069 (.031)	.282 (.031)	.081 (.034)	.356 (.040)	.100 (.042)
6	1,152	48	240	192	.230 (.029)	.076 (.031)	.265 (.030)	.088 (.034)	.330 (.038)	.109 (.043)
10	1,920	48	240	192	.202 (.030)	.096 (.033)	.228 (.031)	.109 (.035)	.275 (.039)	.136 (.044)
25	4,800	48	240	192	.178 (.031)	.127 (.033)	.195 (.032)	.139 (.035)	.228 (.040)	.167 (.043)
50	9,600	48	240	192	.170 (.031)	.142 (.032)	.182 (.032)	.152 (.034)	.211 (.039)	.180 (.041)
100	19,200	48	240	192	.165 (.031)	.151 (.032)	.176 (.033)	.161 (.034)	.203 (.039)	.188 (.040)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H2

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.515 (.048)	.113 (.068)	.603 (.039)	.135 (.073)	.998 (.003)	.025 (.031)
1.25	**240**	**48**	**240**	**192**	**.493 (.046)**	**.125 (.067)**	**.572 (.042)**	**.149 (.073)**	**.924 (.023)**	**.192 (.099)**
1.50	288	48	240	192	.473 (.046)	.139 (.072)	.546 (.042)	.163 (.077)	.871 (.034)	.281 (.109)
1.75	336	48	240	192	.461 (.047)	.154 (.069)	.530 (.045)	.181 (.074)	.835 (.041)	.346 (.113)
2	384	48	240	192	.450 (.048)	.164 (.066)	.513 (.045)	.191 (.072)	.807 (.046)	.386 (.114)
3	576	48	240	192	.425 (.049)	.203 (.066)	.478 (.049)	.233 (.070)	.743 (.059)	.476 (.106)
4	768	48	240	192	.409 (.053)	.226 (.069)	.457 (.053)	.256 (.073)	.713 (.065)	.513 (.106)
5	960	48	240	192	.402 (.054)	.247 (.069)	.445 (.055)	.277 (.073)	.692 (.069)	.535 (.101)
6	1,152	48	240	192	.397 (.055)	.261 (.067)	.437 (.056)	.291 (.071)	.680 (.070)	.551 (.097)
10	1,920	48	240	192	.389 (.057)	.301 (.066)	.422 (.059)	.328 (.069)	.657 (.070)	.581 (.085)
25	4,800	48	240	192	.383 (.062)	.343 (.068)	.408 (.065)	.366 (.071)	.632 (.077)	.601 (.084)
50	9,600	48	240	192	.384 (.064)	.364 (.066)	.407 (.067)	.385 (.069)	.627 (.080)	.612 (.083)
100	19,200	48	240	192	.384 (.065)	.374 (.066)	.404 (.068)	.394 (.069)	.622 (.079)	.614 (.080)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H3

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.554 (.048)	.184 (.083)	.635 (.040)	.215 (.089)	.999 (.002)	.037 (.045)
1.25	**240**	**48**	**240**	**192**	**.540 (.049)**	**.202 (.081)**	**.613 (.043)**	**.236 (.087)**	**.949 (.016)**	**.320 (.106)**
1.50	288	48	240	192	.529 (.050)	.231 (.083)	.596 (.045)	.265 (.088)	.915 (.023)	.447 (.104)
1.75	336	48	240	192	.521 (.050)	.251 (.077)	.583 (.046)	.288 (.081)	.891 (.028)	.523 (.096)
2	384	48	240	192	.511 (.051)	.263 (.083)	.570 (.047)	.299 (.086)	.870 (.031)	.552 (.095)
3	576	48	240	192	.498 (.051)	.312 (.076)	.548 (.050)	.347 (.078)	.830 (.037)	.637 (.075)
4	768	48	240	192	.492 (.053)	.341 (.074)	.536 (.052)	.376 (.076)	.810 (.041)	.672 (.072)
5	960	48	240	192	.490 (.055)	.362 (.075)	.530 (.055)	.397 (.077)	.797 (.043)	.687 (.067)
6	1,152	48	240	192	.485 (.058)	.377 (.075)	.523 (.058)	.410 (.077)	.788 (.046)	.700 (.067)
10	1,920	48	240	192	.482 (.061)	.413 (.072)	.513 (.062)	.442 (.074)	.772 (.048)	.719 (.061)
25	4,800	48	240	192	.483 (.064)	.455 (.069)	.508 (.066)	.479 (.070)	.758 (.050)	.738 (.054)
50	9,600	48	240	192	.478 (.066)	.463 (.068)	.501 (.068)	.486 (.070)	.748 (.054)	.738 (.057)
100	19,200	48	240	192	.484 (.066)	.476 (.068)	.506 (.069)	.498 (.070)	.750 (.052)	.745 (.053)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H4

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.399 (.046)	.025 (.028)	.480 (.047)	.029 (.030)	.996 (.006)	.011 (.014)
1.25	240	48	240	192	.358 (.045)	.028 (.027)	.427 (.047)	.032 (.029)	.831 (.032)	.015 (.020)
1.50	288	48	240	192	.327 (.041)	.029 (.028)	.386 (.046)	.033 (.030)	.718 (.038)	.019 (.022)
1.75	336	48	240	192	.298 (.039)	.029 (.026)	.351 (.044)	.033 (.026)	.633 (.039)	.023 (.024)
2	384	48	240	192	.277 (.039)	.031 (.026)	.325 (.043)	.035 (.027)	.574 (.040)	.027 (.025)
3	**576**	**48**	**240**	**192**	**.219 (.032)**	**.036 (.023)**	**.251 (.037)**	**.041 (.025)**	**.430 (.036)**	**.041 (.026)**
4	768	48	240	192	.187 (.029)	.040 (.023)	.210 (.031)	.045 (.023)	.358 (.034)	.052 (.028)
5	960	48	240	192	.164 (.026)	.043 (.021)	.182 (.028)	.047 (.022)	.315 (.034)	.061 (.029)
6	1,152	48	240	192	.150 (.025)	.045 (.021)	.165 (.026)	.050 (.021)	.288 (.031)	.068 (.030)
10	1,920	48	240	192	.117 (.021)	.053 (.019)	.125 (.021)	.057 (.019)	.230 (.031)	.089 (.031)
25	4,800	48	240	192	.087 (.017)	.060 (.017)	.091 (.017)	.062 (.017)	.180 (.031)	.117 (.032)
50	9,600	48	240	192	.077 (.015)	.063 (.015)	.079 (.015)	.065 (.016)	.161 (.028)	.129 (.030)
100	19,200	48	240	192	.071 (.014)	.064 (.015)	.073 (.015)	.066 (.015)	.154 (.030)	.137 (.031)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H5

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.475 (.050)	.092 (.061)	.553 (.050)	.108 (.066)	.998 (.003)	.021 (.028)
1.25	**240**	**48**	**240**	**192**	**.449 (.052)**	**.106 (.061)**	**.518 (.053)**	**.124 (.066)**	**.917 (.024)**	**.166 (.089)**
1.50	288	48	240	192	.425 (.050)	.115 (.059)	.488 (.051)	.133 (.065)	.861 (.033)	.253 (.104)
1.75	336	48	240	192	.408 (.050)	.123 (.058)	.464 (.053)	.142 (.061)	.821 (.040)	.303 (.106)
2	384	48	240	192	.395 (.052)	.133 (.057)	.445 (.055)	.152 (.062)	.791 (.046)	.347 (.107)
3	576	48	240	192	.359 (.052)	.158 (.056)	.399 (.056)	.177 (.060)	.721 (.056)	.431 (.101)
4	768	48	240	192	.333 (.054)	.173 (.060)	.366 (.058)	.192 (.064)	.685 (.063)	.471 (.102)
5	960	48	240	192	.321 (.054)	.186 (.059)	.348 (.057)	.203 (.063)	.664 (.065)	.493 (.097)
6	1,152	48	240	192	.314 (.057)	.197 (.061)	.339 (.061)	.214 (.065)	.653 (.069)	.513 (.095)
10	1,920	48	240	192	.292 (.056)	.219 (.059)	.311 (.060)	.234 (.063)	.624 (.071)	.541 (.086)
25	4,800	48	240	192	.275 (.058)	.242 (.059)	.288 (.062)	.254 (.063)	.601 (.077)	.568 (.084)
50	9,600	48	240	192	.266 (.058)	.249 (.059)	.277 (.061)	.259 (.062)	.590 (.077)	.574 (.080)
100	19,200	48	240	192	.266 (.063)	.257 (.063)	.275 (.066)	.266 (.066)	.585 (.082)	.576 (.084)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H6

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.531 (.052)	.166 (.082)	.606 (.049)	.193 (.088)	.999 (.002)	.034 (.043)
1.25	**240**	**48**	**240**	**192**	**.512 (.053)**	**.188 (.075)**	**.580 (.050)**	**.218 (.080)**	**.946 (.017)**	**.301 (.105)**
1.50	288	48	240	192	.496 (.050)	.203 (.076)	.556 (.050)	.231 (.080)	.909 (.023)	.422 (.107)
1.75	336	48	240	192	.484 (.053)	.217 (.076)	.538 (.054)	.246 (.079)	.884 (.028)	.492 (.094)
2	384	48	240	192	.471 (.055)	.232 (.075)	.524 (.055)	.261 (.078)	.865 (.030)	.536 (.093)
3	576	48	240	192	.449 (.057)	.269 (.074)	.492 (.058)	.296 (.076)	.818 (.040)	.613 (.081)
4	768	48	240	192	.436 (.062)	.291 (.076)	.472 (.063)	.318 (.079)	.795 (.044)	.648 (.077)
5	960	48	240	192	.431 (.061)	.311 (.072)	.463 (.063)	.336 (.075)	.783 (.045)	.669 (.069)
6	1,152	48	240	192	.428 (.066)	.324 (.077)	.456 (.068)	.348 (.080)	.774 (.047)	.680 (.069)
10	1,920	48	240	192	.416 (.065)	.348 (.071)	.440 (.068)	.370 (.074)	.757 (.050)	.700 (.062)
25	4,800	48	240	192	.412 (.071)	.384 (.073)	.431 (.074)	.402 (.077)	.742 (.054)	.721 (.059)
50	9,600	48	240	192	.407 (.072)	.393 (.073)	.423 (.076)	.409 (.077)	.735 (.053)	.724 (.055)
100	19,200	48	240	192	.405 (.071)	.398 (.071)	.421 (.074)	.413 (.074)	.733 (.053)	.728 (.054)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H7

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.358 (.064)	.024 (.029)	.429 (.078)	.026 (.030)	.996 (.006)	.011 (.014)
1.25	240	48	240	192	.319 (.062)	.022 (.024)	.377 (.073)	.025 (.026)	.833 (.033)	.016 (.021)
1.50	288	48	240	192	.283 (.060)	.024 (.025)	.330 (.071)	.027 (.026)	.720 (.038)	.021 (.024)
1.75	336	48	240	192	.260 (.055)	.023 (.022)	.303 (.067)	.026 (.024)	.639 (.039)	.025 (.024)
2	**384**	**48**	**240**	**192**	**.239 (.054)**	**.026 (.022)**	**.276 (.063)**	**.028 (.024)**	**.579 (.037)**	**.031 (.026)**
3	576	48	240	192	.180 (.046)	.029 (.021)	.203 (.053)	.032 (.022)	.439 (.037)	.047 (.031)
4	768	48	240	192	.149 (.040)	.031 (.020)	.164 (.046)	.034 (.021)	.370 (.036)	.061 (.034)
5	960	48	240	192	.127 (.035)	.033 (.019)	.137 (.037)	.035 (.020)	.325 (.036)	.071 (.035)
6	1,152	48	240	192	.114 (.032)	.034 (.018)	.123 (.035)	.037 (.019)	.298 (.035)	.080 (.036)
10	1,920	48	240	192	.084 (.023)	.037 (.017)	.088 (.023)	.039 (.017)	.242 (.035)	.101 (.037)
25	4,800	48	240	192	.058 (.014)	.040 (.015)	.060 (.014)	.041 (.015)	.193 (.035)	.132 (.038)
50	9,600	48	240	192	.051 (.013)	.042 (.013)	.052 (.013)	.043 (.013)	.178 (.037)	.145 (.038)
100	19,200	48	240	192	.045 (.011)	.041 (.011)	.046 (.011)	.042 (.011)	.166 (.036)	.150 (.037)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H8

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.432 (.070)	.078 (.055)	.501 (.077)	.090 (.060)	.998 (.003)	.025 (.032)
1.25	**240**	**48**	**240**	**192**	**.401 (.072)**	**.088 (.057)**	**.462 (.079)**	**.101 (.060)**	**.919 (.022)**	**.167 (.085)**
1.50	288	48	240	192	.377 (.069)	.096 (.055)	.430 (.077)	.110 (.059)	.861 (.033)	.255 (.100)
1.75	336	48	240	192	.360 (.070)	.102 (.054)	.408 (.079)	.116 (.059)	.824 (.039)	.316 (.102)
2	384	48	240	192	.346 (.069)	.108 (.054)	.388 (.078)	.123 (.059)	.794 (.043)	.354 (.104)
3	576	48	240	192	.302 (.071)	.125 (.057)	.333 (.079)	.138 (.061)	.725 (.054)	.436 (.102)
4	768	48	240	192	.279 (.070)	.137 (.057)	.305 (.079)	.150 (.063)	.687 (.059)	.474 (.094)
5	960	48	240	192	.270 (.072)	.148 (.059)	.292 (.077)	.162 (.064)	.671 (.062)	.502 (.091)
6	1,152	48	240	192	.252 (.069)	.151 (.059)	.268 (.075)	.161 (.063)	.651 (.066)	.510 (.091)
10	1,920	48	240	192	.231 (.069)	.167 (.061)	.244 (.073)	.177 (.065)	.627 (.068)	.545 (.082)
25	4,800	48	240	192	.212 (.066)	.185 (.064)	.221 (.069)	.192 (.067)	.604 (.069)	.572 (.076)
50	9,600	48	240	192	.203 (.068)	.189 (.067)	.209 (.071)	.195 (.069)	.597 (.072)	.580 (.076)
100	19,200	48	240	192	.195 (.064)	.188 (.064)	.200 (.067)	.193 (.067)	.587 (.074)	.579 (.075)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H9

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 48, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	194	48	240	192	.494 (.068)	.143 (.081)	.565 (.070)	.166 (.086)	.999 (.002)	.031 (.037)
1.25	**240**	**48**	**240**	**192**	**.465 (.073)**	**.159 (.076)**	**.527 (.075)**	**.182 (.082)**	**.944 (.016)**	**.292 (.103)**
1.50	288	48	240	192	.451 (.072)	.169 (.076)	.508 (.076)	.192 (.082)	.907 (.023)	.410 (.105)
1.75	336	48	240	192	.434 (.073)	.183 (.079)	.486 (.078)	.208 (.087)	.881 (.028)	.486 (.099)
2	384	48	240	192	.422 (.074)	.193 (.077)	.467 (.082)	.215 (.083)	.862 (.030)	.529 (.096)
3	576	48	240	192	.402 (.076)	.227 (.077)	.439 (.081)	.250 (.083)	.816 (.039)	.610 (.079)
4	768	48	240	192	.386 (.078)	.246 (.080)	.416 (.086)	.268 (.086)	.792 (.042)	.643 (.070)
5	960	48	240	192	.375 (.082)	.261 (.081)	.403 (.087)	.283 (.085)	.781 (.044)	.667 (.067)
6	1,152	48	240	192	.367 (.082)	.269 (.082)	.393 (.089)	.289 (.088)	.769 (.048)	.672 (.068)
10	1,920	48	240	192	.351 (.085)	.289 (.085)	.372 (.091)	.307 (.090)	.752 (.050)	.695 (.062)
25	4,800	48	240	192	.337 (.086)	.310 (.086)	.350 (.091)	.323 (.091)	.736 (.051)	.714 (.055)
50	9,600	48	240	192	.340 (.090)	.326 (.089)	.353 (.095)	.339 (.094)	.730 (.054)	.719 (.057)
100	19,200	48	240	192	.331 (.093)	.324 (.093)	.341 (.097)	.334 (.097)	.724 (.056)	.718 (.057)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H10

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.515 (.030)	.102 (.044)	.604 (.026)	.124 (.048)	.999 (.001)	.010 (.013)
1.25	**600**	**120**	**600**	**480**	**.493 (.030)**	**.120 (.043)**	**.573 (.027)**	**.143 (.047)**	**.919 (.017)**	**.168 (.067)**
1.50	720	120	600	480	.475 (.029)	.132 (.040)	.549 (.027)	.157 (.043)	.866 (.024)	.258 (.077)
1.75	840	120	600	480	.463 (.030)	.148 (.043)	.532 (.028)	.175 (.046)	.829 (.029)	.324 (.081)
2	960	120	600	480	.451 (.030)	.160 (.044)	.517 (.028)	.188 (.046)	.799 (.034)	.362 (.082)
3	1,440	120	600	480	.425 (.031)	.197 (.044)	.479 (.031)	.227 (.047)	.729 (.044)	.446 (.083)
4	1,920	120	600	480	.413 (.034)	.226 (.045)	.462 (.034)	.257 (.048)	.699 (.050)	.491 (.080)
5	2,400	120	600	480	.406 (.035)	.247 (.044)	.450 (.036)	.277 (.047)	.678 (.052)	.513 (.077)
6	2,880	120	600	480	.400 (.036)	.261 (.044)	.442 (.038)	.291 (.047)	.664 (.056)	.527 (.077)
10	4,800	120	600	480	.393 (.038)	.299 (.045)	.427 (.040)	.329 (.048)	.637 (.058)	.555 (.072)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H11

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.570 (.030)	.205 (.056)	.646 (.025)	.241 (.058)	1.00 (.001)	.017 (.020)
1.25	**600**	**120**	**600**	**480**	**.554 (.031)**	**.234 (.055)**	**.624 (.027)**	**.270 (.058)**	**.955 (.008)**	**.364 (.069)**
1.50	720	120	600	480	.546 (.029)	.258 (.054)	.611 (.027)	.297 (.055)	.926 (.012)	.495 (.063)
1.75	840	120	600	480	.538 (.032)	.279 (.052)	.598 (.028)	.317 (.054)	.904 (.014)	.561 (.057)
2	960	120	600	480	.533 (.032)	.294 (.052)	.591 (.030)	.334 (.055)	.888 (.016)	.604 (.052)
3	1,440	120	600	480	.520 (.034)	.342 (.051)	.568 (.033)	.380 (.053)	.852 (.020)	.681 (.043)
4	1,920	120	600	480	.513 (.033)	.372 (.049)	.556 (.033)	.408 (.050)	.831 (.022)	.706 (.040)
5	2,400	120	600	480	.513 (.035)	.395 (.047)	.552 (.034)	.430 (.047)	.822 (.024)	.725 (.037)
6	2,880	120	600	480	.510 (.036)	.409 (.047)	.547 (.036)	.442 (.048)	.814 (.024)	.735 (.035)
10	4,800	120	600	480	.509 (.038)	.445 (.046)	.540 (.039)	.475 (.047)	.798 (.026)	.751 (.033)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H12

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.598 (.031)	.271 (.062)	.669 (.026)	.312 (.063)	1.00 (.001)	.022 (.026)
1.25	**600**	**120**	**600**	**480**	**.586 (.031)**	**.300 (.060)**	**.650 (.027)**	**.340 (.062)**	**.964 (.007)**	**.441 (.069)**
1.50	720	120	600	480	.580 (.032)	.326 (.059)	.639 (.029)	.367 (.060)	.941 (.010)	.574 (.056)
1.75	840	120	600	480	.576 (.033)	.350 (.058)	.631 (.030)	.390 (.058)	.924 (.011)	.640 (.050)
2	960	120	600	480	.569 (.033)	.362 (.056)	.621 (.030)	.402 (.057)	.911 (.013)	.676 (.045)
3	1,440	120	600	480	.562 (.033)	.411 (.050)	.605 (.031)	.449 (.050)	.881 (.016)	.741 (.035)
4	1,920	120	600	480	.561 (.036)	.444 (.049)	.599 (.035)	.479 (.049)	.867 (.017)	.768 (.030)
5	2,400	120	600	480	.559 (.036)	.462 (.047)	.594 (.035)	.495 (.047)	.858 (.018)	.781 (.029)
6	2,880	120	600	480	.557 (.037)	.475 (.048)	.591 (.037)	.507 (.048)	.852 (.020)	.788 (.028)
10	4,800	120	600	480	.557 (.038)	.506 (.045)	.586 (.038)	.535 (.045)	.840 (.020)	.804 (.024)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H13

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.358 (.059)	.053 (.030)	.402 (.067)	.059 (.031)	.999 (.001)	.006 (.009)
1.25	**600**	**120**	**600**	**480**	**.318 (.056)**	**.058 (.030)**	**.351 (.063)**	**.063 (.032)**	**.889 (.018)**	**.078 (.041)**
1.50	720	120	600	480	.286 (.055)	.063 (.029)	.313 (.059)	.068 (.030)	.814 (.026)	.133 (.053)
1.75	840	120	600	480	.259 (.049)	.063 (.026)	.282 (.053)	.068 (.028)	.761 (.031)	.173 (.059)
2	960	120	600	480	.238 (.045)	.065 (.025)	.257 (.047)	.070 (.027)	.721 (.033)	.199 (.063)
3	1,440	120	600	480	.188 (.035)	.071 (.024)	.198 (.036)	.074 (.025)	.628 (.042)	.273 (.069)
4	1,920	120	600	480	.165 (.030)	.075 (.022)	.172 (.029)	.078 (.023)	.583 (.046)	.315 (.068)
5	2,400	120	600	480	.150 (.026)	.077 (.023)	.154 (.024)	.079 (.022)	.556 (.048)	.341 (.068)
6	2,880	120	600	480	.138 (.024)	.078 (.022)	.142 (.023)	.080 (.022)	.535 (.050)	.354 (.067)
10	4,800	120	600	480	.114 (.018)	.079 (.018)	.117 (.017)	.081 (.018)	.496 (.052)	.385 (.062)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H14

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.470 (.060)	.153 (.052)	.523 (.066)	.171 (.056)	1.00 (.001)	.015 (.019)
1.25	**600**	**120**	**600**	**480**	**.445 (.060)**	**.168 (.052)**	**.490 (.066)**	**.186 (.055)**	**.950 (.009)**	**.322 (.069)**
1.50	720	120	600	480	.419 (.065)	.177 (.055)	.458 (.070)	.194 (.059)	.915 (.014)	.446 (.068)
1.75	840	120	600	480	.405 (.063)	.184 (.055)	.438 (.068)	.200 (.059)	.892 (.016)	.517 (.064)
2	960	120	600	480	.393 (.063)	.197 (.055)	.424 (.069)	.213 (.058)	.874 (.018)	.560 (.058)
3	1,440	120	600	480	.358 (.061)	.216 (.055)	.378 (.065)	.228 (.059)	.832 (.022)	.641 (.046)
4	1,920	120	600	480	.332 (.061)	.221 (.056)	.350 (.064)	.233 (.059)	.811 (.026)	.672 (.047)
5	2,400	120	600	480	.323 (.060)	.231 (.056)	.337 (.062)	.242 (.058)	.799 (.026)	.691 (.040)
6	2,880	120	600	480	.309 (.058)	.232 (.056)	.321 (.059)	.242 (.057)	.790 (.028)	.701 (.040)
10	4,800	120	600	480	.293 (.056)	.246 (.053)	.301 (.057)	.253 (.054)	.775 (.028)	.723 (.035)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H15

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.538 (.049)	.228 (.062)	.597 (.053)	.256 (.066)	1.00 (.001)	.020 (.023)
1.25	**600**	**120**	**600**	**480**	**.517 (.055)**	**.248 (.064)**	**.568 (.059)**	**.275 (.068)**	**.962 (.007)**	**.418 (.072)**
1.50	720	120	600	480	.503 (.057)	.267 (.062)	.548 (.060)	.293 (.065)	.937 (.010)	.556 (.058)
1.75	840	120	600	480	.494 (.056)	.281 (.060)	.534 (.060)	.306 (.065)	.918 (.012)	.616 (.050)
2	960	120	600	480	.484 (.059)	.291 (.063)	.523 (.062)	.317 (.067)	.905 (.014)	.658 (.048)
3	1,440	120	600	480	.462 (.059)	.323 (.062)	.491 (.062)	.345 (.065)	.874 (.017)	.728 (.037)
4	1,920	120	600	480	.451 (.065)	.341 (.067)	.476 (.069)	.362 (.072)	.859 (.018)	.754 (.033)
5	2,400	120	600	480	.439 (.063)	.350 (.066)	.462 (.066)	.369 (.070)	.849 (.020)	.767 (.031)
6	2,880	120	600	480	.437 (.063)	.360 (.064)	.458 (.066)	.378 (.068)	.843 (.020)	.775 (.029)
10	4,800	120	600	480	.419 (.068)	.371 (.068)	.435 (.071)	.386 (.071)	.827 (.023)	.788 (.028)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H16

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.286 (.093)	.039 (.027)	.315 (.111)	.042 (.029)	.999 (.002)	.007 (.010)
1.25	**600**	**120**	**600**	**480**	**.244 (.085)**	**.040 (.024)**	**.267 (.096)**	**.043 (.026)**	**.894 (.017)**	**.087 (.043)**
1.50	720	120	600	480	.213 (.079)	.041 (.025)	.227 (.088)	.044 (.027)	.822 (.024)	.147 (.055)
1.75	840	120	600	480	.189 (.072)	.043 (.025)	.200 (.074)	.046 (.026)	.770 (.028)	.188 (.058)
2	960	120	600	480	.165 (.063)	.042 (.023)	.177 (.067)	.044 (.024)	.733 (.031)	.220 (.059)
3	1,440	120	600	480	.126 (.046)	.045 (.022)	.129 (.043)	.046 (.022)	.641 (.040)	.296 (.065)
4	1,920	120	600	480	.107 (.038)	.047 (.022)	.108 (.035)	.047 (.022)	.601 (.044)	.341 (.065)
5	2,400	120	600	480	.091 (.028)	.045 (.019)	.092 (.025)	.046 (.019)	.571 (.045)	.362 (.064)
6	2,880	120	600	480	.085 (.026)	.046 (.019)	.085 (.022)	.046 (.018)	.552 (.047)	.377 (.063)
10	4,800	120	600	480	.068 (.017)	.046 (.016)	.068 (.015)	.046 (.015)	.519 (.050)	.413 (.060)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H17

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.391 (.099)	.109 (.055)	.435 (.111)	.123 (.062)	.999 (.001)	.014 (.017)
1.25	**600**	**120**	**600**	**480**	**.367 (.098)**	**.124 (.056)**	**.401 (.111)**	**.136 (.064)**	**.948 (.010)**	**.310 (.069)**
1.50	720	120	600	480	.339 (.098)	.131 (.060)	.365 (.105)	.141 (.065)	.913 (.013)	.438 (.063)
1.75	840	120	600	480	.319 (.094)	.134 (.060)	.339 (.101)	.143 (.065)	.889 (.016)	.509 (.061)
2	960	120	600	480	.303 (.096)	.136 (.063)	.324 (.101)	.146 (.068)	.869 (.019)	.546 (.059)
3	1,440	120	600	480	.264 (.084)	.148 (.061)	.276 (.086)	.155 (.064)	.826 (.024)	.630 (.050)
4	1,920	120	600	480	.248 (.081)	.157 (.063)	.254 (.082)	.160 (.065)	.806 (.023)	.665 (.041)
5	2,400	120	600	480	.226 (.075)	.154 (.064)	.234 (.079)	.160 (.067)	.794 (.025)	.682 (.040)
6	2,880	120	600	480	.222 (.075)	.161 (.065)	.228 (.075)	.166 (.066)	.786 (.026)	.696 (.038)
10	4,800	120	600	480	.196 (.060)	.161 (.055)	.201 (.061)	.164 (.056)	.767 (.029)	.715 (.037)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H18

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 120, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	482	120	600	480	.473 (.089)	.180 (.072)	.521 (.100)	.199 (.079)	1.00 (.001)	.019 (.023)
1.25	**600**	**120**	**600**	**480**	**.452 (.092)**	**.201 (.071)**	**.496 (.101)**	**.223 (.079)**	**.961 (.007)**	**.409 (.069)**
1.50	720	120	600	480	.428 (.091)	.208 (.075)	.466 (.099)	.228 (.082)	.934 (.011)	.542 (.061)
1.75	840	120	600	480	.415 (.090)	.215 (.076)	.445 (.098)	.232 (.083)	.916 (.012)	.606 (.054)
2	960	120	600	480	.407 (.094)	.229 (.079)	.437 (.101)	.247 (.086)	.902 (.014)	.647 (.048)
3	1,440	120	600	480	.373 (.094)	.244 (.082)	.394 (.101)	.259 (.089)	.870 (.018)	.717 (.038)
4	1,920	120	600	480	.366 (.095)	.264 (.085)	.383 (.099)	.278 (.089)	.853 (.020)	.744 (.034)
5	2,400	120	600	480	.349 (.094)	.266 (.087)	.367 (.099)	.280 (.093)	.844 (.020)	.759 (.031)
6	2,880	120	600	480	.341 (.092)	.272 (.087)	.354 (.096)	.283 (.091)	.838 (.019)	.769 (.029)
10	4,800	120	600	480	.320 (.087)	.277 (.084)	.329 (.090)	.285 (.087)	.824 (.022)	.783 (.027)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H19

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.557 (.022)	.167 (.039)	.639 (.018)	.201 (.042)	1.00 (0.00)	.008 (.010)
1.25	**1,020**	**204**	**1020**	**816**	**.541 (.022)**	**.196 (.038)**	**.616 (.019)**	**.231 (.041)**	**.948 (.007)**	**.308 (.053)**
1.50	1,224	204	1020	816	.529 (.022)	.218 (.039)	.600 (.020)	.254 (.042)	.914 (.011)	.441 (.050)
1.75	1,428	204	1020	816	.522 (.023)	.235 (.041)	.587 (.021)	.273 (.042)	.890 (.013)	.509 (.048)
2	1,632	204	1020	816	.514 (.024)	.251 (.039)	.576 (.022)	.289 (.040)	.870 (.015)	.549 (.045)
3	2,448	204	1020	816	.500 (.024)	.300 (.037)	.552 (.023)	.339 (.038)	.827 (.018)	.631 (.039)
4	3,264	204	1020	816	.494 (.025)	.334 (.036)	.541 (.024)	.372 (.037)	.806 (.020)	.665 (.035)
5	4,080	204	1020	816	.491 (.026)	.357 (.036)	.535 (.026)	.394 (.037)	.794 (.021)	.683 (.033)
6	4,896	204	1020	816	.491 (.026)	.376 (.035)	.532 (.026)	.412 (.036)	.786 (.022)	.695 (.032)
10	8,160	204	1020	816	.488 (.028)	.414 (.034)	.523 (.029)	.447 (.035)	.767 (.024)	.713 (.030)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H20

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.594 (.024)	.260 (.047)	.665 (.020)	.300 (.049)	1.00 (0.00)	.012 (.015)
1.25	**1,020**	**204**	**1020**	**816**	**.583 (.023)**	**.292 (.044)**	**.648 (.020)**	**.332 (.046)**	**.964 (.005)**	**.430 (.051)**
1.50	1,224	204	1020	816	.575 (.024)	.313 (.045)	.635 (.021)	.354 (.045)	.939 (.008)	.564 (.044)
1.75	1,428	204	1020	816	.570 (.025)	.337 (.043)	.626 (.023)	.378 (.043)	.922 (.009)	.630 (.038)
2	1,632	204	1020	816	.566 (.025)	.354 (.044)	.619 (.023)	.395 (.045)	.909 (.011)	.668 (.036)
3	2,448	204	1020	816	.559 (.026)	.402 (.039)	.603 (.025)	.440 (.040)	.878 (.012)	.734 (.027)
4	3,264	204	1020	816	.557 (.027)	.436 (.038)	.597 (.026)	.472 (.038)	.864 (.014)	.762 (.024)
5	4,080	204	1020	816	.553 (.027)	.452 (.036)	.589 (.027)	.486 (.037)	.854 (.015)	.774 (.023)
6	4,896	204	1020	816	.553 (.029)	.468 (.037)	.587 (.028)	.500 (.037)	.847 (.015)	.782 (.022)
10	8,160	204	1020	816	.554 (.029)	.500 (.035)	.583 (.029)	.529 (.035)	.835 (.015)	.797 (.019)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H21

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .00, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.615 (.024)	.313 (.050)	.682 (.021)	.356 (.051)	1.00 (0.00)	.014 (.016)
1.25	**1,020**	**204**	**1020**	**816**	**.607 (.025)**	**.344 (.048)**	**.667 (.022)**	**.386 (.049)**	**.969 (.004)**	**.489 (.052)**
1.50	1,224	204	1020	816	.601 (.025)	.369 (.045)	.656 (.023)	.411 (.046)	.949 (.006)	.618 (.041)
1.75	1,428	204	1020	816	.598 (.025)	.390 (.045)	.649 (.023)	.431 (.044)	.934 (.007)	.680 (.034)
2	1,632	204	1020	816	.594 (.025)	.408 (.043)	.643 (.023)	.448 (.043)	.923 (.008)	.715 (.030)
3	2,448	204	1020	816	.589 (.028)	.454 (.040)	.629 (.026)	.491 (.040)	.897 (.011)	.775 (.024)
4	3,264	204	1020	816	.588 (.027)	.483 (.039)	.624 (.026)	.518 (.039)	.884 (.012)	.798 (.021)
5	4,080	204	1020	816	.585 (.027)	.499 (.036)	.618 (.027)	.531 (.036)	.876 (.012)	.808 (.019)
6	4,896	204	1020	816	.586 (.028)	.515 (.036)	.618 (.028)	.545 (.036)	.872 (.013)	.816 (.018)
10	8,160	204	1020	816	.587 (.029)	.542 (.034)	.613 (.029)	.569 (.034)	.861 (.013)	.829 (.017)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H22

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.300 (.067)	.067 (.029)	.322 (.071)	.071 (.029)	1.00 (.001)	.006 (.008)
1.25	**1,020**	**204**	**1020**	**816**	**.259 (.058)**	**.070 (.027)**	**.275 (.060)**	**.074 (.029)**	**.922 (.010)**	**.175 (.047)**
1.50	1,224	204	1020	816	.230 (.050)	.072 (.026)	.240 (.049)	.075 (.027)	.870 (.015)	.272 (.053)
1.75	1,428	204	1020	816	.210 (.045)	.075 (.024)	.218 (.044)	.077 (.024)	.834 (.019)	.334 (.054)
2	1,632	204	1020	816	.190 (.038)	.073 (.025)	.197 (.037)	.076 (.025)	.806 (.021)	.379 (.054)
3	2,448	204	1020	816	.153 (.028)	.077 (.021)	.157 (.024)	.078 (.021)	.739 (.027)	.463 (.050)
4	3,264	204	1020	816	.135 (.020)	.078 (.019)	.137 (.018)	.079 (.019)	.708 (.028)	.505 (.048)
5	4,080	204	1020	816	.123 (.019)	.078 (.018)	.125 (.016)	.079 (.018)	.688 (.031)	.526 (.045)
6	4,896	204	1020	816	.115 (.016)	.078 (.017)	.117 (.014)	.080 (.016)	.675 (.031)	.542 (.044)
10	8,160	204	1020	816	.101 (.012)	.078 (.014)	.101 (.011)	.079 (.014)	.648 (.035)	.569 (.043)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H23

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.436 (.069)	.165 (.053)	.474 (.076)	.180 (.058)	1.00 (0.00)	.011 (.013)
1.25	**1,020**	**204**	**1020**	**816**	**.411 (.072)**	**.179 (.053)**	**.440 (.077)**	**.190 (.056)**	**.959 (.006)**	**.391 (.054)**
1.50	1,224	204	1020	816	.386 (.070)	.185 (.054)	.413 (.073)	.198 (.058)	.931 (.008)	.524 (.048)
1.75	1,428	204	1020	816	.363 (.066)	.187 (.052)	.386 (.069)	.199 (.055)	.911 (.011)	.592 (.043)
2	1,632	204	1020	816	.352 (.063)	.193 (.051)	.370 (.066)	.202 (.053)	.898 (.011)	.632 (.038)
3	2,448	204	1020	816	.316 (.060)	.204 (.051)	.329 (.060)	.212 (.052)	.863 (.014)	.704 (.031)
4	3,264	204	1020	816	.297 (.059)	.212 (.052)	.307 (.059)	.219 (.053)	.846 (.015)	.733 (.027)
5	4,080	204	1020	816	.281 (.051)	.212 (.047)	.288 (.051)	.217 (.048)	.836 (.016)	.747 (.026)
6	4,896	204	1020	816	.273 (.049)	.215 (.046)	.281 (.049)	.221 (.047)	.829 (.016)	.755 (.025)
10	8,160	204	1020	816	.257 (.047)	.221 (.045)	.261 (.046)	.225 (.045)	.815 (.019)	.772 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H24

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .10, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.522 (.061)	.240 (.060)	.570 (.067)	.264 (.064)	1.00 (0.00)	.012 (.014)
1.25	**1,020**	**204**	**1020**	**816**	**.505 (.060)**	**.264 (.059)**	**.545 (.065)**	**.286 (.063)**	**.967 (.005)**	**.470 (.052)**
1.50	1,224	204	1020	816	.485 (.064)	.274 (.061)	.522 (.068)	.296 (.065)	.945 (.007)	.600 (.041)
1.75	1,428	204	1020	816	.474 (.065)	.287 (.063)	.505 (.069)	.306 (.068)	.930 (.008)	.662 (.036)
2	1,632	204	1020	816	.463 (.065)	.294 (.061)	.492 (.069)	.314 (.066)	.918 (.009)	.699 (.032)
3	2,448	204	1020	816	.434 (.066)	.315 (.062)	.455 (.069)	.331 (.065)	.890 (.012)	.761 (.025)
4	3,264	204	1020	816	.422 (.066)	.327 (.064)	.440 (.068)	.341 (.067)	.876 (.013)	.784 (.023)
5	4,080	204	1020	816	.414 (.065)	.337 (.063)	.430 (.068)	.351 (.066)	.869 (.013)	.798 (.021)
6	4,896	204	1020	816	.408 (.066)	.343 (.065)	.422 (.069)	.355 (.068)	.864 (.013)	.805 (.019)
10	8,160	204	1020	816	.393 (.067)	.353 (.067)	.405 (.070)	.364 (.070)	.853 (.014)	.819 (.018)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. "Base Items" are the number of continuous items simulated; "Full Response Options" are the number of base items each split into five binary items; "Dummy-coded Options" are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key's performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H25

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .05 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.218 (.094)	.044 (.027)	.228 (.098)	.045 (.027)	1.00 (.001)	.005 (.007)
1.25	**1,020**	**204**	**1020**	**816**	**.182 (.085)**	**.044 (.026)**	**.189 (.088)**	**.045 (.028)**	**.922 (.011)**	**.174 (.044)**
1.50	1,224	204	1020	816	.156 (.073)	.044 (.025)	.158 (.068)	.045 (.025)	.870 (.015)	.273 (.053)
1.75	1,428	204	1020	816	.139 (.060)	.044 (.024)	.141 (.055)	.045 (.024)	.832 (.019)	.332 (.053)
2	1,632	204	1020	816	.124 (.052)	.044 (.022)	.125 (.048)	.044 (.021)	.805 (.020)	.375 (.052)
3	2,448	204	1020	816	.091 (.028)	.043 (.018)	.091 (.022)	.043 (.017)	.741 (.025)	.467 (.046)
4	3,264	204	1020	816	.078 (.025)	.043 (.018)	.078 (.021)	.043 (.017)	.709 (.027)	.506 (.044)
5	4,080	204	1020	816	.069 (.017)	.042 (.015)	.069 (.014)	.042 (.014)	.689 (.029)	.529 (.043)
6	4,896	204	1020	816	.065 (.017)	.043 (.015)	.064 (.013)	.043 (.013)	.677 (.030)	.544 (.043)
10	8,160	204	1020	816	.054 (.008)	.041 (.010)	.054 (.007)	.041 (.010)	.650 (.031)	.570 (.039)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H26

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .10 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.345 (.111)	.116 (.058)	.369 (.122)	.124 (.065)	1.00 (0.00)	.010 (.013)
1.25	**1,020**	**204**	**1020**	**816**	**.312 (.103)**	**.121 (.056)**	**.334 (.112)**	**.129 (.062)**	**.957 (.006)**	**.378 (.052)**
1.50	1,224	204	1020	816	.291 (.102)	.129 (.061)	.305 (.108)	.134 (.065)	.929 (.008)	.508 (.048)
1.75	1,428	204	1020	816	.269 (.096)	.128 (.061)	.276 (.096)	.130 (.061)	.909 (.010)	.578 (.043)
2	1,632	204	1020	816	.252 (.090)	.128 (.060)	.260 (.090)	.132 (.061)	.894 (.011)	.622 (.038)
3	2,448	204	1020	816	.212 (.073)	.129 (.053)	.217 (.069)	.131 (.051)	.858 (.015)	.694 (.032)
4	3,264	204	1020	816	.195 (.066)	.131 (.054)	.198 (.064)	.133 (.053)	.840 (.016)	.722 (.028)
5	4,080	204	1020	816	.180 (.059)	.129 (.050)	.182 (.056)	.131 (.048)	.829 (.017)	.736 (.026)
6	4,896	204	1020	816	.175 (.054)	.133 (.048)	.177 (.053)	.134 (.047)	.822 (.017)	.746 (.025)
10	8,160	204	1020	816	.156 (.043)	.131 (.040)	.157 (.041)	.132 (.039)	.809 (.019)	.765 (.023)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.

Table H27

Variance Explained for the Vertical Percent Method, Point-Biserial Correlation Method, and Option-Keyed Multiple Regression in the Simulation Conditions where Items = 204, Mean Intercorrelation of Items = .20, and the Standard Deviation of Validities and Item Intercorrelations = .15 for Items with Five Response Options where the Number of Items for the Simulation Condition are Divided into Five Response Options

Ratio of Sample Size to Items	Actual Sample Size	Base Items	Full Response Options	Dummy-coded Options	VPM WD	VPM CV	PBC WD	PBC CV	MR WD	MR CV
1 + 2	818	204	1020	816	.443 (.100)	.187 (.071)	.481 (.109)	.203 (.077)	1.00 (0.00)	.014 (.018)
1.25	**1,020**	**204**	**1020**	**816**	**.416 (.107)**	**.198 (.076)**	**.445 (.115)**	**.213 (.083)**	**.966 (.005)**	**.460 (.050)**
1.50	1,224	204	1020	816	.395 (.107)	.203 (.079)	.422 (.116)	.217 (.086)	.943 (.007)	.589 (.041)
1.75	1,428	204	1020	816	.388 (.104)	.218 (.081)	.412 (.110)	.232 (.088)	.927 (.008)	.653 (.036)
2	1,632	204	1020	816	.368 (.104)	.217 (.083)	.383 (.106)	.225 (.086)	.915 (.010)	.687 (.033)
3	2,448	204	1020	816	.341 (.099)	.235 (.085)	.353 (.100)	.243 (.087)	.888 (.012)	.756 (.025)
4	3,264	204	1020	816	.319 (.093)	.238 (.083)	.329 (.096)	.245 (.086)	.872 (.012)	.778 (.022)
5	4,080	204	1020	816	.306 (.088)	.240 (.081)	.315 (.090)	.247 (.083)	.864 (.013)	.790 (.021)
6	4,896	204	1020	816	.302 (.088)	.245 (.081)	.310 (.089)	.252 (.082)	.859 (.013)	.798 (.019)
10	8,160	204	1020	816	.282 (.077)	.248 (.074)	.288 (.079)	.254 (.076)	.848 (.015)	.812 (.018)

Note. VPM WD = vertical percent method in the weight-derivation sample; VPM CV = vertical percent method in the cross-validation sample; PBC WD = point-biserial correlation method in the weight-derivation sample; PBC CV = point-biserial correlation method in the cross-validation sample; MR WD = option-keyed multiple regression in the weight-derivation sample; MR CV = option-keyed multiple regression in the cross-validation sample. “Base Items” are the number of continuous items simulated; “Full Response Options” are the number of base items each split into five binary items; “Dummy-coded Options” are the number of base items split into five binary items, one of which is dropped per base item. Each table entry for an empirical key’s performance is the mean variance explained across 1,000 simulation trials with the standard deviation of those trials in parentheses. All conditions assume a mean continuous validity of .04. The bolded text is the sample size to items ratio at which option-keyed multiple regression ties or exceeds the other two methods in the cross-validation sample.