

FOCAL ARTICLE

---

# Stubborn Reliance on Intuition and Subjectivity in Employee Selection

---

SCOTT HIGHHOUSE

*Bowling Green State University*

## Abstract

The focus of this article is on implicit beliefs that inhibit adoption of selection decision aids (e.g., paper-and-pencil tests, structured interviews, mechanical combination of predictors). Understanding these beliefs is just as important as understanding organizational constraints to the adoption of selection technologies and may be more useful for informing the design of successful interventions. One of these is the implicit belief that it is theoretically possible to achieve near-perfect precision in predicting performance on the job. That is, people have an inherent resistance to analytical approaches to selection because they fail to view selection as probabilistic and subject to error. Another is the implicit belief that prediction of human behavior is improved through experience. This myth of expertise results in an overreliance on intuition and a reluctance to undermine one's own credibility by using a selection decision aid.

Perhaps the greatest technological achievement in industrial and organizational (I–O) psychology over the past 100 years is the development of decision aids (e.g., paper-and-pencil tests, structured interviews, mechanical combination of predictors) that substantially reduce error in the prediction of employee performance (Schmidt & Hunter, 1998). Arguably, the greatest failure of I–O psychology has been the inability to convince employers to use them. A little over 10 years ago, Terpstra (1996) sampled 201 human resources (HR) executives about the perceived effectiveness of various selection methods. As the left side of Figure 1 shows, they considered the traditional unstructured interview more effective than any of the paper-and-pencil assessment procedures. Inspection of actual effectiveness of these procedures, however, shows that paper-and-pencil tests commonly outperform

unstructured interviews. For example, the right side of Figure 1 shows the results of a meta-analysis conducted on the actual effectiveness of these same procedures for predicting performance in sales (Vinchur Schippmann, Switzer, & Roth, 1998). Use of any one of the paper-and-pencil tests *alone* outperforms the unstructured interview—a procedure that is presumed to assess ability, personality, and aptitude concurrently.

Although one might argue that these data merely reflect a lack of knowledge about effective practice, there is considerable evidence that employers simply do not believe that the research is relevant to their own situation (Colbert, Rynes, & Brown, 2005; Johns, 1993; Muchinsky, 2004; Terpstra & Rozelle, 1997; Whyte & Latham, 1997). For example, Rynes, Colbert, and Brown (2002) found that HR professionals were well aware of the limitations of the unstructured interview. Similarly, one of my students conducted a yet-unpublished survey of HR professionals ( $n = 206$ ) about their views of selection practice. His data indicated that the HR professionals agreed, by

---

Correspondence concerning this article should be addressed to Scott Highhouse. E-mail: shighho@bgsu.edu  
Address: Bowling Green State University, Bowling Green, OH 43403.

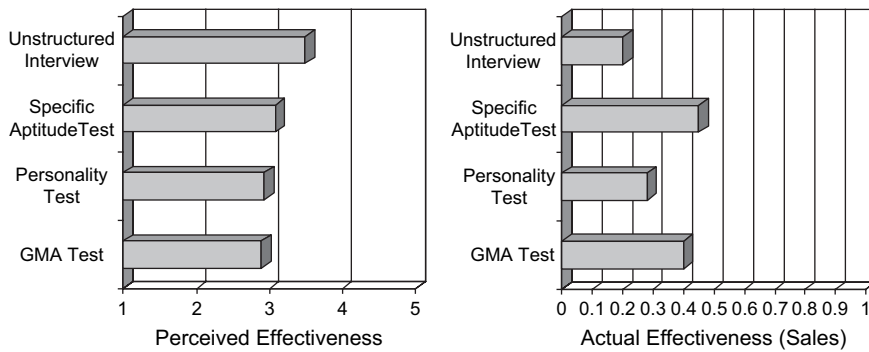


Figure 1. Perceived versus actual usefulness of various predictors.

Note. Perceived effectiveness numbers are on a 1–5 scale (1 = *not good*; 3 = *average*; 5 = *extremely good*). Actual effectiveness numbers are correlations corrected for unreliability in the criterion and range restriction. Because Vinchur, Schippmann, Switzer, and Roth (1998) did not include interviews, the interview estimate is from Huffcutt and Arthur (1994) level 1 interview. GMA = general mental ability; personality = potency; specific aptitude = sales ability.

a factor of more than 3 to 1, that using tests was an effective way to evaluate a candidate's suitability and that tests that assess specific traits are effective for hiring employees. At the same time, however, these same professionals agreed, by more than 3 to 1, that you can learn more from an informal discussion with job candidates and that you can "read between the lines" to detect whether someone is suitable to hire. This apparent conflict between knowledge and belief seems loosely analogous to the common practice of preferring brand name cold remedies to store brand remedies containing the same ingredients. People know that the store brands are identical, but they do not trust them for their own colds.

Some might argue that the tide is turning. Much has been written on the merits of evidence-based management (Pfeffer & Sutton, 2006; Rousseau, 2006). This approach, much like evidence-based medicine, relies on the best available scientific evidence to make decisions. At the core of this movement is "analytics" or data-based decision making (e.g., Ayers, 2007). Discussions of number crunching in the arena of personnel selection, however, are almost always limited to anecdotes from professional sports (e.g., Davenport, 2006). Competing with the analytical point of view are books like

Malcolm Gladwell's (2005) *Blink: The Power of Thinking Without Thinking* and Gerd Gigerenzer's (2007) *Gut Feelings: The Intelligence of the Unconscious*, which extol the virtues of intuitive decision making. Although the assertions of these authors have little relevance for the prediction of human performance, the popularity of their work likely reinforces the common belief that good hiring is a matter of experience and intuition.

### Implicit Beliefs

My colleagues and I (Lievens, Highhouse, & De Corte, 2005) conducted a policy-capturing study of the decision processes of retail managers making hypothetical hiring decisions. We found that the managers placed more emphasis on competencies assessed by unstructured interviews than on competencies measured by tests, regardless of what those competencies were. They placed more emphasis, for instance, on Extraversion than on general mental ability when Extraversion was assessed using an unstructured interview (and general mental ability was assessed using a paper-and-pencil test). The *opposite* was found when Extraversion was assessed using a paper-and-pencil test and general mental ability was assessed using

an unstructured interview! Clearly, these managers believed that good old-fashioned “horse sense” was needed to accurately size up applicants (see Phelan & Smith, 1958).

The reluctance of employers to use analytical selection procedures is at least partially a reflection of broader misconceptions that the general public has about how to go about assessing and selecting people for jobs. Consider two high-profile policy opinions on testing and selection in the United States.

- In 1990, the National Commission on Testing and Public Policy (1990) issued eight recommendations for testing in schools and the workplace. Among those was the statement as follows: “Test scores are imperfect measures and should not be used alone to make important decisions about individuals” (National Commission on Testing and Public Policy, 1990, p. 30). The commission’s chairman, Bernard Gifford of Apple Computer, commented, “We just believe that under no circumstances should individuals be denied a job or college admission exclusively based on test scores” (“Panel Criticizes Standard Testing,” 1990).
- In the landmark Supreme Court decision on affirmative action at the University of Michigan, Justice Rehnquist concluded that consideration of race as a factor in student admission is acceptable—but it must be done at the *individual* level, with each applicant considered holistically. In concurrence, Justice O’Connor commented, “But the current [student selection] system, as I understand it, is a nonindividualized, mechanical one. As a result, I join the Court’s opinion . . .” (*Gratz v. Bollinger*, 2003, Concurrence 1).

Although these positions sound reasonable on the surface, they represent fundamentally flawed assumptions. No one disputes that test scores are imperfect measures, but the testing commission implies that combining them with something else will

correct the imperfections (rather than exacerbate them). The court’s majority opinion in *Gratz* suggests that individualized methods of selection are more fair and reliable than impersonal “mechanical” ones. Both of these examples illustrate two implicit beliefs about employee selection: (1) people believe that it is possible to achieve near-perfect precision in the prediction of employee success, and (2) people believe that there is such a thing as intuitive expertise in the prediction of human behavior. These implicit beliefs exert their influence on policy and practice, even though they may not be immediately accessible (Kahneman, 2003). I acknowledge that there are a number of contextual reasons for resistance to selection technologies, including organizational politics, habit, and culture, along with the existing legal climate (e.g., Johns, 1993; Muchinsky, 2004). However, whereas contextual issues are often situation specific, these are universal “truths” about people. As such, understanding and studying them provides hope for overcoming user resistance to selection decision aids.

### Irreducible Unpredictability

I recently came across an article in a popular trade magazine for executives, purportedly summarizing the state of the science on executive assessment (Sindelar, 2002). I was struck by a statement made by the author: “For many top-level positions, technical competence accounts for only 20 percent of a successful alignment. Psychological factors account for the rest” (pp. 13–14).<sup>1</sup> Whether intentional or not, the author was clearly implying what is shown on the top of Figure 2—that 80% of the variance in executive success can be explained by psychological factors (presumably temperament or personality). Reality, however, is much more like the chart on the bottom of Figure 2—showing that most of the variance in executive success is simply not

---

1. The author identified his affiliation as the “Institute for Advanced Business Psychology.”

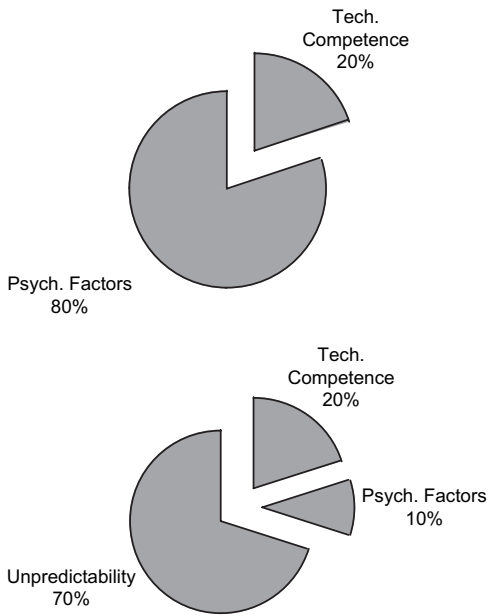


Figure 2. Variance in success accounted for by technical competence and psychological factors.

predictable prior to employment. The business of assessment and selection involves considerable irreducible unpredictability; yet, many seem to believe that all failures in prediction are because of mistakes in the assessment process. Put another way, people seem to believe that, as long as the applicant is the right person for the job and the applicant is accurately assessed, success is certain. The “validity ceiling” has been a continually vexing problem for I–O psychology (see Campbell, 1990; Rundquist, 1969). Enormous resources and effort are focused on the quixotic quest for new and better predictors that will explain more and more variance in performance. This represents a refusal, by knowledgeable people, to recognize that many determinants of performance are not knowable at the time of hire. The notion that it is still possible to achieve large gains in the prediction of employee success reflects a failure to accept that there is no such thing as perfect prediction in this domain. Campbell noted that our poor professional self-esteem is based on an unrealistic notion of what can be achieved in the prediction of employee

success. Campbell wrote: “No external source imposed this [validity ceiling] standard on the discipline or even argued that there should be a standard at all” (p. 689).

Recall the earlier comment by the national testing commission, cautioning that tests are “imperfect” and must be supplemented with other things. It is remarkably similar to Viteles’ (1925) observation that “objective scores of vocational tests are at best uncertain diagnostic criteria” (p. 132). This early pioneer of I–O was arguing that standardized methods of assessment could only fill the proverbial glass halfway. Intuitive judgment was needed to fill it the rest of the way. Viteles wrote: “It is the opinion of the writer that in the cause of greater scientific accuracy in vocational selection in industry the statistical point of view must be supplemented by a clinical point of view” (p. 134). Countering this position was Freyd (1926), who cautioned against allowing intuition to creep into hiring decisions. Freyd, who represented the analytical viewpoint of selection, argued “allowing selection to be influenced by personal interpretations with their unavoidable prejudices instead of relying upon objective measures gives even less consideration to the well-being and interest of the individual worker” (p. 354). History proved Freyd prescient.

Table 1 shows the results of the earliest study investigating the relative effectiveness of standardized procedures alone versus supplementing those procedures with intuitive judgment (Sarbin, 1943). As you can see, academic achievement was better predicted by the standardized scores alone than by the scores plus clinical judgment. The notion that analysis outperforms intuition in the prediction of human behavior is among the most well-established findings in the behavioral sciences (Grove & Meehl, 1994; Grove, Zald, Lebow, Snitz, & Nelson, 2000).<sup>2</sup> Why, therefore, does the intuitive

2. Although few studies in I–O have explicitly made this comparison, there are a number of examples where tests alone outperformed tests + intuition (e.g., Borneman, Cooper, Klieger, & Kuncel, 2007; Huse, 1962; Meyer, 1956).

**Table 1.** Sarbin's (1943) Investigation of Two Methods for Predicting Success of University of Minnesota Undergraduates Admitted in 1939

| Predictor composite   | Correlation with criterion ( <i>r</i> ) |
|---|---|
| High school rank + college aptitude test                                    | .45                                     |
| High school rank + college aptitude test + intuitive judgment of counselors | .35                                     |

perspective remain so appealing? Einhorn (1986) observed that a crucial distinction between the intuitive and the analytical approaches to human prediction is the worldview of the people making the judgments. According to Einhorn, the intuitive approach reflects a deterministic worldview, one that rejects the idea that the future is inherently probabilistic. This is contrasted with the analytical worldview, which accepts uncertainty as inevitable. Consider the San Diego Chargers professional football team who, despite having a regular season record of 14-2 in 2006, fired its head coach following a play-off loss. The fired coach had a reputation for leading teams to successful regular season records, only to lose the big games. The Chargers organization evidently failed to consider that the contribution of uncertainty to a play-off outcome is much greater than to a 16-game season record. Abelson (1985) found that knowledgeable baseball fans overestimated by a factor of 75 the contribution of skill (vs. chance) to the likelihood of a major league baseball player getting a hit in a given turn at bat.

Intuitive approaches to employee selection make the errors in selection ambiguous. Analytical approaches make them part of the process—hence, visible. Considerable research suggests that ambiguity about the likelihood of an outcome (e.g., the operation has an unknown chance of success) encourages more optimism than a low known probability (e.g., the operation has a 20% chance of success; see Kuhn, 1997). There is little room for optimism when a composite of predictors is known to leave 75% of the variance unexplained. This may explain why selection procedures that are difficult to evaluate (e.g., feelings about “fit”) are so attractive. Einhorn

(1986) noted, however, that one must be willing to accept error to make less error.

### Myth of Expertise

I have argued that one of the reasons that people have an inherent resistance to analytical approaches to hiring is that they fail to view selection in probabilistic terms. A related but different reason for employer reticence to use selection decision aids is that most people believe in the myth of selection expertise. By this I mean the belief that one can become skilled in making intuitive judgments about a candidate's likelihood of success. This is reflected in the survey responses of the HR professionals who believed in “reading between the lines” to size up job candidates. It is also evidenced in the phenomenal growth of the professional recruiter or “headhunter” profession (Finlay & Coverdill, 1999) and the perseverance of the holistic approach to managerial assessment (Highhouse, 2002).

Despite this widespread belief in intuitive expertise, the data suggest that it is a myth. For example, the considerable research on predicting human behavior per se shows that experience does not improve predictions made by clinicians, social workers, parole boards, judges, auditors, admission committees, marketers, and business planners (Camerer & Johnson, 1991; Dawes, Faust, & Meehl, 1989; Grove et al., 2000; Sherden, 1998). Although it is commonly accepted that some (employment) interviewers are better than others, research on variance in interviewer validity suggests that differences are due entirely to sampling error (Pulakos, Schmitt, Whitney, & Smith, 1996). Existing evidence suggests that the interrater reliability of the traditional (unstructured)

interview is so low that, even with a perfectly reliable and valid criterion, interview-based judgments could never account for more than 10% of the variance in job performance (Conway, Jako, & Goodman, 1995).<sup>3</sup> This empirical evidence is troubling for a procedure that is supposed to simultaneously take into account ability, motivation, and person–organization fit. Keep in mind also that these findings are based on interviews that had ratings associated with the interviewers' judgments. Thus, the unstructured interviews subjected to meta-analyses are almost certainly unusual and on the high end of rigor. The data do not paint a sanguine picture of intuitive judgment in the hiring process.

There are commonly two scholarly rebuttals to the arguments against prediction expertise. I will consider these in turn. One response to the limitations of intuitive approaches to selection is to focus on the ability of experts to spot idiosyncrasies in a candidate's profile (Jeanneret & Silzer, 1998). Meehl (1954) noted that one limitation of analytical formulas was their inability to incorporate "broken-leg" cues. The term comes from an anecdotal example in which one is trying to predict whether or not a person will go to the movie on a particular day. A mechanical formula might take into account things like the nature of the movie (e.g., less likely to go to romantic comedy) or the weather (e.g., more likely to go on a rainy day). The mechanical procedure would not take into account, however, an event that is extremely rare (e.g., the person has a broken leg), and thus, the mechanical prediction will not be as accurate as a prediction based on a simple intuitive observation. A mechanical approach to selection would not, the logic goes, consider idiosyncratic characteristics of any particular job candidate—a seasoned expert would.

Another common response to criticisms of intuitive selection is to focus on the

expert's ability to interpret configurations of traits (Prien, Schippmann, & Prien, 2003). The notion behind this argument is that each candidate is unique, and one must consider each piece of information about the candidate in light of all the other pieces of information. In other words, assessing patterns of traits is more accurate than assessing traits individually. For example, Prien et al. noted that executive assessment requires a "dynamic interpretation" of applicant data, one that takes into account interactions between test scores and other observations (p. 125). This view is reinforced by leadership theorists who assert that leader characteristics exhibit complex configural relations with leadership outcomes (e.g., Zaccaro, 2007).

Even if we do accept that decision makers incorporate broken-leg cues and configurations of traits, existing evidence suggests that these things account for negligible variance in the predicted outcome. For example, Dawes (1971) modeled admission decisions of a four-person graduate admissions committee using a bootstrapping procedure. This is shown in Figure 3. Dawes found that the model (i.e., paramorphic representation) of the admission committee's judgments outperformed the committee itself. More relevant to this discussion, however, was the fact that, whereas a linear combination of the expert cues correlated significantly ( $r = .25$ ) with the criterion, the residual—which included configural judgments, broken-leg cues, and error—was inconsequential ( $r = .01$ ). Camerer and Johnson (1991) noted that, despite accounting for a large portion of the error term, broken-leg cues and configural judgments consistently provide little incremental gain in prediction—even for so-called experts. The problem with broken-leg cues is that people rely too much on them because they present compelling stories. The tendency to be seduced by detailed stories causes people to ignore relevant information and to violate simple rules of logic (see Highhouse, 1997, 2001). Also, as one reviewer noted, broken legs are themselves constructs that can and should be measured reliably. The problem with trait configurations, on the other hand, is that

3. Meta-analysis suggests that it accounts for negligible incremental validity over simple paper-and-pencil tests of cognitive ability and conscientiousness (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000).

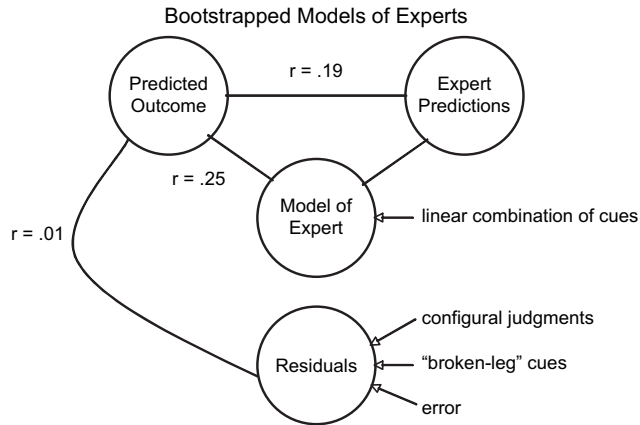


Figure 3. Results from Dawes' (1971) examination of graduate admissions decisions.

they require feats of information integration that contradict current understanding of human cognitive limitations (Ruscio, 2003). And true real-world examples of predictive interactions between job applicant characteristics are difficult to find (e.g., Sackett, Gruys, & Ellingson, 1998). Hastie and Dawes (2001) distilled from the vast literature on prediction "experts" the following stylized facts:

- They rely on few pieces of information.
- They lack insight into how they arrive at predictions.
- They exhibit poor interjudge agreement.
- They become more confident in their accuracy when irrelevant information is presented.

The obvious remedy to the limitations of expertise is to structure expert intuition and mechanically combine it with other decision aids, such as paper-and-pencil inventories. However, there would likely be considerable resistance to structuring or mechanizing the judgment process (e.g., Lievens et al., 2005; van der Zee, Bakker, & Bakker, 2002). Most people believe that aspects of an applicant's character are far too complex to be assessed by scores, ratings, and formulas.

An example of the irrationality of this bias against decision aids is the contempt with which most college football fans and com-

mentators hold the Bowl Championship Series, which is a mechanical formula that incorporates expert ratings (e.g., coaches poll) and computer rankings (e.g., wins and losses of opponents) into an overall ranking of football teams. The nature of the complaints ("unplug the computers") suggests that people do not want mechanical formulas making their expert decisions about who attends bowl games. A University of Oregon coach infamously declared: "I liken the BCS to a bad disease, like cancer" (Vondersmith, 2001). Another example of this bias against decision aids is the considerable patient resistance to diagnostic decision aids (Arkes, Shaffer, & Medow, 2007). Arkes and his colleagues found that physicians who made computer-based diagnoses of ankle injuries were perceived less competent, professional, and thorough than physicians who made diagnoses without any aids. Indeed, the idea that (with the appropriate data) a physician might not even need to meet or interact with a patient to understand his or her personal health issues would be a hard sell to most people. Physicians, aware of this lay bias against "cookbook medicine," grossly underutilize these valuable technologies in practice (Kaplan, 2001).<sup>4</sup> Hastie and Dawes (2001) noted that relying on

4. This underutilization also results from overconfidence on the part of physicians in their own diagnostic expertise.

expertise is more socially acceptable than relying on test scores or formulas. Research on medical decision making supports this contention. It is no wonder, therefore, that HR practitioners would be reluctant to undermine their status by administering a paper-and-pencil test, structuring an employment interview, or plugging ratings into a mechanical formula.

### Concluding Remarks

We know quite a bit about *applicant* reactions to hiring methods (Hausknecht, Day, & Thomas, 2004), but very little attention has been given to user resistance to selection decision aids. Campbell (1990) noted: "We still do not know much about how to best communicate selection results to people outside the [I-O] profession" (p. 704). Fifteen years later, Anderson (2005) lamented: "In fact, the whole area of practitioner beliefs about selection methods and processes is a gargantuan one which research has made little or no inroads into" (p. 19). I have inferred from the general psychological literature, and the specific selection literature, two implicit beliefs that likely inhibit the widespread acceptance of selection technologies. These include the belief that it is possible to achieve near-perfect precision in predicting performance on the job and the belief that intuitive prediction can be improved by experience. People trust that the complex characteristics of applicants can be best assessed by a sensitive, equally complex human being. This does not stand up to scientific scrutiny, and I-O psychologists need to begin focusing their efforts on understanding how to navigate these waters. We can begin by drawing from the judgment and decision making and human factors literatures on how to better communicate uncertainty and error. We also need to learn how to better calibrate user expectations. Consider Muchinsky's (2004) experience in communicating a .50 validity coefficient for a mechanical comprehension test:

my pleasure regarding the findings was highly apparent to the client organi-

zation. It was at this point a senior company official said to me, "I fail to see the basis for your enthusiasm." (p. 194)

Research on probability neglect (Sunstein, 2002) suggests that people make little distinction between probabilities that they consider small. In addition, research on evaluability (Hsee, 1996) has shown that most attributes cannot be evaluated without appropriate context. Perhaps if Muchinsky (2004) had compared his .50 to flipping a coin (.00) or to an unstructured interview (.20), management would have been more impressed. Perhaps management would have been more impressed by a common-language effect size indicator or by an expectancy chart. We simply do not have the research to guide these communication decisions.

The traditional unstructured interview has remained the most popular and widely used selection procedure for over 100 years (Buckley, Norris, & Wiese, 2000). This is despite the fact that, during this same period, there have been significant advancements in the development of selection decision aids. Guion (1965) argued that the waste of human resources caused by poor selection procedures should pain the professional conscience of I-O psychologists. It is true that people are not very predictable, but selection decision aids help.

### References

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 128–132.
- Anderson, N. (2005). Relationship between practice and research in personnel selection: Does the left hand know what the right is doing? In A. Evers, N. Anderson, & O. Voskuil (Eds.), *The Blackwell handbook of personnel selection* (pp. 1–24). Malden, MA: Blackwell.
- Arkes, H., Shaffer, V. A., & Medow, M. A. (2007). Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making*, *27*, 189–202.
- Ayers, I. (2007). *Super crunchers: Why thinking-by-numbers is the new way to be smart*. New York: Random House.
- Borneman, M. J., Cooper, S. R., Klieger, D. M., & Kuncel, N. R. (2007, April). The efficacy of the admissions interview: A meta-analysis. In N. R. Kuncel (Chair), *Alternative predictors of academic performance:*