

Chapter 19

New Concepts of Intelligence: Their Practical and Legal Implications for Employee Selection

Gerald V. Barrett, Alissa J. Kramen and Sarah B. Lueke

1. Introduction

In the 1920s and 1930s basic theories of intellectual ability were developed along with operational tests which proved effective in predicting job performance (Spearman 1927; Thorndike 1936). In a series of studies and meta-analyses throughout the 1970s and 1980s, Schmidt and Hunter showed that cognitive ability was the best overall predictor of job performance (Hunter & Hunter 1984; Hunter 1986; Schmidt & Hunter 1981). Partially in reaction to the meta-analytic findings, research to expand on the definitions of competencies continued. The development of competencies by McClelland (1973) was followed by a discussion of tacit knowledge (Wagner & Sternberg 1985), practical intelligence (Sternberg & Wagner 1986), and multiple intelligence (Gardner 1999). In the 1990s, emotional intelligence became the intelligence of interest (Feist & Barron 1996; Goleman 1995, 1998a, 1998b; Graves 1999; Mayer *et al.* 1990).

All these new theories and proposed measurement instruments pose a challenge to traditional cognitive ability tests since it is claimed that these tests are more valid and have lower adverse impact. It is our contention that many of these tests are nothing more than pop psychology. It is distressing to see such books (i.e. Goleman 1998b) quoted as if they had some merit. We will review the themes present throughout all of these “creative” concepts and examine whether they have practical implications and can withstand legal scrutiny in the public and private sector.

2. Legal Challenges and the Daubert Standards

It is our opinion that despite all these theorists’ claims of validity, if challenged in court, they would fail. The Daubert Standards for scientific tests are a set of guidelines for admissibility of scientific evidence into court (see Table 19.1).

Table 19.1: Daubert criteria.

-
- 1 The theory must have been tested, or is at least able to be tested.
 - 2 The theory (& expert) must have (been) published in peer reviewed publications.
 - 3 There must be a known or potential error rate.
 - 4 The theory must be generally accepted in the relevant scientific community.
 - 5 The methods for testing the theory must meet scientific standards.
-

The criteria were set forth in *Daubert v. Merrell Dow Pharmaceuticals* (1993) and clarified through subsequent supreme court cases (*General Electric Company v. Joiner* 1997; *Kumho Tire Company Ltd. v. Carmichael* 1999) and federal district and appeals court cases (e.g. *Black v. Rhone-Poulenc* 1998; *Butler v. Home Depot, Inc.* 1997; *Camp v. Lockheed Martin Corporation* 1998; *Clark v. Takata Corp.* 1999; *Gerlib v. R. R. Donnelley & Sons Co.* 2001; *Smith v. Ford Motor Co.* 2000). (See American College of Trial Lawyers 1994, and Barrett 2000, for a discussion). In the U.S., the Daubert standards pertain to any selection instrument used or advocated by a plaintiff. This is part of American case law, but the basic principles of scientific standards are relevant to everyone considering a particular measure or construct's use for personnel selection.

In a typical disparate impact discrimination lawsuit, a series of steps occur. First, the plaintiff demonstrates adverse impact. Next, the defendant demonstrates that despite the adverse impact, the test is still valid. Third, the plaintiff's expert shows that there are alternative selection procedures with equal validity that result in less adverse impact. Fourth, the defendant's expert attempts to demonstrate that the alternative selection procedure is not as valid and doesn't decrease disparate impact. This is where the problem with new theories of intelligence comes in. They either implicitly or explicitly imply that they have a better approach. This has been the situation in court cases in which we've been involved (e.g. *Adams v. City of Chicago* 1996; *Brown v. City of Chicago* 1996, 1998; *Firefighters Institute for Racial Equality v. City of St. Louis* 2000; *U.S. v. Garland* 2000). At the start of the process, the defendants attempt to have the plaintiff's expert's evidence stricken using the Daubert standards, which act as a gatekeeping function. For purposes of illustration, we will show why we believe these new concepts of intelligence would not be accepted under the Daubert Standards.

3. Key Studies Cited by Advocates of New Concepts of Intelligence

It is clear that the key studies cited repeatedly by advocates of these new concepts do not meet the Daubert criteria for scientific evidence. In this section, we will review these studies and challenge each one as they would be challenged in a court as a basis for use of a particular measure as a personnel selection instrument.

Table 19.2.: Prediction of Foreign Service Officers (FSO) job performance using competency measures.

Test	Reliability	Validity
U.S. Knowledge	0.70	0.32**
Empathy (PONS)	0.20	0.11
Relevant Experience	—	0.16

(Participation in student government in high school plus college academic record minus sum of participation in musical activities in high school & college).

Data from an unpublished report by McClelland, D. C., & Dailey, C. (1973), and from Table 1, p. 70 of Barrett (1994).

3.1. PONS

The PONS test is designed to measure ability to read emotions. An unpublished report by McClelland & Dailey (1973) examining the PONS as a predictor of Foreign Service Officers’ (FSO) job performance is often cited by advocates of emotional intelligence as supporting the validity of the PONS. However, as described in Barrett (1994), the U.S. Knowledge Test had higher reliability and validity compared to the PONS (see Table 19.2).

In fact, the validity coefficient for the Job Knowledge Test was significant, whereas the validity coefficient for the PONS did not reach statistical significance. By looking at Table 19.2, it is clear that the PONS, a measure of empathy, did not demonstrate a significant correlation with job performance. Despite this fact, Spencer & Spencer (1993) asserted that superior FSOs scored significantly higher on the PONS because they were better able to ‘tune into’ others’ feelings. In addition, Goleman (1998a) asserted that the exam that measured academic subject correlated negatively with job performance for Foreign Service Officer; conversely, ability to read emotions (PONS Test) did predict Foreign Service Officer job performance. “Their (academic) scores were a poor indicator of how adept these new diplomats were on their feet — in fact, their on-the-job performance ratings actually correlated negatively with how well they did on the very test used to select them” (Goleman 1998a: 18, 332, note 4). Again, this statement regarding the PONS did not receive support by the data. In our review of the literature, we found no peer-reviewed articles which demonstrated that the PONS predicted job performance.

3.2. Tacit Knowledge

Dr Sternberg often relies on a study by Scribner (1984) to support his concept of tacit knowledge (Wagner & Sternberg 1985) as a predictor of job performance. However, Scribner used a very specific sample in a milk processing plant (preloaders, inventory

workers, wholesale drivers, clerks) as well as 9th grade students. None of these would likely be considered representative samples by experts in the field. Four population groups (preloaders, inventory & drivers, clerks and students) were given 16 simulation problems and their strategies for solving the problem were examined (see Table 19.3).

From these results, Wagner and Sternberg concluded that tacit knowledge is a valid predictor of job performance. However, there was no measure of job performance in any of these studies, and none of the production workers were administered either an IQ test or a Tacit Knowledge inventory. The only relationship reported was that a standardized math test related to more optimum solutions for students. If a practitioner were to attempt to defend this study as evidence of the validity of tacit knowledge for predicting job performance, the study would certainly not meet the Daubert criteria. The theory was not tested and the methods used to test the theory would not meet scientific standards for test validation (APA 1999; EEOC 1978, 1980; SIOP 1987).

There have been studies that have attempted to empirically demonstrate a relationship between tacit knowledge and job performance. Wagner & Sternberg (1990) conducted a study at the Center for Creative Leadership using tacit knowledge to predict performance on two managerial simulations: Earth II and Energy International. *Tacit*

Table 19.3: Participants in Scribner (1984) descriptive case study of a dairy (from Table 1.1, p. 17).

Group	Number
Preloaders	5
Inventory	4
Wholesale Drivers	10
Clerks	11
Ninth Grade Students	30

Percent selected non-literal strategy when it is optimal (from Table 1.2, p. 23, Scriber 1984)

	(LPE & LME)
Preloaders	72%
Inventory & Drivers	65%
Clerks	47%
Students Scores on national math achievement test at or above grade level ($N=12$)	42% ¹
Student scores on national math achievement test below grade level ($N=18$)	15% ¹

Note: LPE = least physical effort; LME = least mental effort.

¹ From Note C.

Table 19.4: Center for Creative Leadership Studies/tacit knowledge (Wagner & Sternberg 1990).

Behavioral Assessment Data Rating (BAD)¹

TK	-0.61
IQ	0.38

¹ Sum of 8 dimensions for both simulations.

Separate correlations with the 8 dimensions were not reported.

Knowledge (or street smarts) was defined as the “work-related practical know-how that is learned informally on the job” and was measured by the *Tacit Knowledge Inventory for Managers* (TKIM; Wagner & Sternberg 1991), which consists of a series of work related situations. Wagner and Sternberg concluded that “scores on research measures of street smarts are quite predictive of ability to learn to solve practical problems in the office” (pp. 494–495) (see Table 19.4).

However, these conclusions would not be accepted under the Daubert standards. Several characteristics of the methodology used to demonstrate the predictability of tacit knowledge are problematic. First, tacit knowledge scores were calculated based on deviations of responses from an expert prototype of 15 additional participants in the Center for Creative Leadership’s Leadership Development Program (LDP). They obtain scores for this “expert” group, obtain means on various dimensions, and subtract the individual score in the validation sample from the “prototype” score. You, in effect, have a difference score, which are known to have low reliability (Edwards 1994). This is not a typical scoring method used in personnel selection, and the authors do not provide evidence that this was acceptable. The criterion measures consisted of behavioral assessment data ratings on eight dimensions based on performance in the managerial simulations (activity level, led the discussion, influenced others, problem analysis, task orientation, motivated others, verbal effectiveness and interpersonal skills). In addition to the unusual method of using a deviation score, there are other problems with the scoring used in this study. When there were problems with finding correlations with the difference scores, they standardize scores on the test so that everyone has the same standard deviation (Wagner 1994). This adjustment of individuals’ test scores would receive scrutiny under Civil Service rules and regulations. It is difficult to tell individuals that he/she has a lower test score than his/her observed score because it was changed to reflect the group standard deviation.

There are several issues with the dependent variables when this study is presented as a validation study. These are often used as predictors in assessment centers. There were no data presented relating managerial job performance and performance on the managerial simulations, neither correlational evidence nor evidence of content validity. Even if the simulations were related to job performance, this does not mean that tacit knowledge would also be related to job performance simply based on the evidence of a correlation between tacit knowledge and performance in the simulation (McCornack

1956). No data were presented relating predictors to each dimension. Participants were 45 participants in the Leadership Development Program from various jobs and organizational levels. This sample size is particularly small when one considers how many predictors were used in the regression equations. It is not evident exactly how the subtest scores were used, but even if one considers that subtests were combined to be one predictor score, there were eight predictors. This sample, both in terms of size and nature, will have problems with generalizability to managerial jobs. Wagner and Sternberg do not describe the nature of participants' jobs. Finally, there is a problem with the definition of tacit knowledge itself, which is defined as something learned on the job. The Uniform Guidelines (EEOC 1978) explicitly prohibit employers from selecting applicants based on knowledge learned on the job.

In another study on tacit knowledge, Colonia-Willner (1998) examined the relationship between the TKIM and job performance ratings of non-expert Brazilian bank managers.

Based on the results presented in Table 19.5, it is clear that although this study meets the Daubert standard of a tested theory, the test of this theory resulted in no relationship between tacit knowledge and job performance ratings. The questionable scoring method of the TKIM discussed with regard to the Wagner & Sternberg (1990) study (i.e. difference scores based on comparison to an expert prototype) was also used in this

Table 19.5: Expert versus non-expert bank managers in Brazil. In Raven's, DAT and TKIM Time Data from Table 5 of Colonia-Willner (1998).

Brazilian Bank Managers

Test	Expert (N = 43) M	Non-Expert (N = 157) M	P
Raven's score	17.67	14.83	<0.01
DAT score	27.60	22.41	<0.01
TKIM Time (min)	29.67	33.01	<0.01

(Time in minutes taken by the participants to grade the 91 strategies presented by TKIM)

TKIM Prediction of job performance for 157 Non-Expert Bank Managers (data from Table 4 of Colonia-Willner 1998).

TKIM	Job Performance
Overall	0.00
Self	0.01
Others	-0.06
Task	0.03

study. Finally, the results alluded to in the abstract of this article are somewhat misleading. Colonia-Willner stated in the abstract that the “TKIM predicted managerial skill, while the DAT and Raven’s did not” (p. 45). However, when the results are examined more closely, it is evident that the TKIM did not predict job performance ratings. The statement made in the abstract refers to a relationship of $r = -0.12$ between overall TKIM score and an index of managerial success, which was a composite of management span (number of personnel supervised directly and indirectly), salary and job performance ratings. This is not a typical measure of job performance for a validation study. Based on the studies reviewed above, tacit knowledge would not survive legal scrutiny under the Daubert standards.

3.3. Competencies

Daniel (1992) proposed to identify critical leadership competencies for manufacturing supervisors and to test whether they can differentiate between top supervisors and a control group of supervisors in an electronics company. This study has never been cited by any of the major proponents of new concepts of intelligence. Two studies were conducted. The first study involved critical behavior interviews with 9 high-performing and 8 control supervisors and identified 13 competencies. In the second study, these 13 competencies were rated by the managers of 15 top supervisors and a control group of 23 supervisors ($N = 38$). Only the competency “image and reputation” differentiated the top supervisors from the control group ($t(32) = 3.11, p < 0.02$). However, “image and reputation” measures work behavior on the job. This study is an example of confusion between predictors and criteria in this study, and it does not meet the Daubert standards.

3.4. Interpersonal Accuracy

Davis & Kraus (1997) conducted a meta-analysis of relationships between individual differences and interpersonal accuracy (see Table 19.6).

Interpersonal accuracy has been defined in many ways, including accuracy in assessing another person’s personality, affective or non-affective state, the role, identity or status of a target individual, or accuracy in predicting the actual behavior of a target person. Based on the results of this meta-analysis, Davis and Kraus concluded that a good judge (i.e. someone with a high level of interpersonal accuracy) is intelligent, more likely to view the world in a cognitively complex manner, and has good psychological adjustment. However, social intelligence was not significantly related to interpersonal accuracy, as would be expected. The findings of this meta-analysis have been cited by leading proponents of emotional intelligence. Goleman (1998b) asserted that “Those who are trusting — tend to be more highly attuned to feelings” (pp. 142, 350, note 20). However, it is evident from Table 19.6 that this relationship is rather small. No evidence regarding job performance was presented in this meta-analysis.

Table 19.6: (From Davis & Kraus (1997), Table 5.2, p. 157).

Individual Differences	Mean Effect	Fail Safe
Intellectual Functioning	0.23	465
Cognitive Complexity	0.27	30
Positive Adjustment	0.24	34
Social Intelligence	0.08	1
Interpersonal Trust	0.20	21

Note: Fail Safe = “. . . whenever the combined significance level was reliable, a ‘fail-safe’ number was calculated using the formula recommended by Cooper (1979), which provides an estimate of how many findings of zero association, not included in these analyses, would be necessary in order to make the combined significance level unreliable” (pg. 156).

3.5. Emotional Intelligence

Graves (1999) tested Mayer *et al.*'s (1999) Emotional Knowledge Test (EKT) as a predictor of performance on four assessment center exercises (see Table 19.7).

Graves found that emotional intelligence as scored by experts was significantly related to performance on the assessment center exercises. However, the magnitude of correlations was similar to the magnitude of the correlations between IQ and performance. The problems with using this study as validation evidence are similar to those discussed with regard to the Wagner & Sternberg (1990) study. The criterion measures were simulated jobs, which are assessment center exercises.

Table 19.8 reviews the available validation evidence.

Despite many claims that measures of emotional intelligence have been demonstrated as valid predictors of job performance (e.g. Bachman *et al.* 2000; Bar-On 1997, 2000; Goleman 1998a, 1998b), there is very little empirical evidence to support this statement.

Table 19.7: Correlation between EKT, IQ & performance Composites ($N=149$). (Data from Graves 1999, Table 24, p. 171).

Composite	Emotional Intelligence		IQ
	Expert Scored	Consensus Scored	
Peer	0.24**	0.12	0.25**
Assessor	0.27**	0.10	0.24**
Combined	0.31**	0.13	0.29**

Note: Criterion was peer and assessor ratings of performance on four job simulation exercises.

Table 19.8: Evidence of the relation between emotional intelligence and job performance.

Measure	Study	Sample	Results
Bar-On EQ-I 133 self-report items	Bachman <i>et al.</i> (2000)	Study 1 N = 36 debt collection officers Study 2 N = 34	No validity coefficients were reported. Study 1 No mean difference between Bar-on scores for most and least successful employees Study 2 Significant difference for empathy in the opposite direction
	Handley (1997)	Air force recruiters	No validity coefficients were reported. Handley reported that there were significant differences between successful and unsuccessful recruiters on 11 of 16 scales.
Emotional Competence Inventory (ECI) 110 self-report items asking participants to describe how representative each item is of their typical work behavior	Murensky (2000)	N = 90 executives in key leadership roles in an international oil corporation (13 female).	The criteria for "leader performance" was obtained using the Balanced Scorecard. • The correlation between the ECI Clusters and the Balanced scorecard was not significant.
EI 34 word pair items (e.g. insecure/secure) rated by interviewers.	Carrothers <i>et al.</i> (2000)	N = 147 medical school applicants	No validity coefficients were reported. No significant difference in EI scores between accepted and rejected applicants.

Table 19.8: Continued.

Measure	Study	Sample	Results
Short Version of the Multifactor Emotional Intelligence Scale (MEIS) Scenario based questions for 3 "branches" of emotional intelligence	Rice (1999)	<i>N</i> = 26 teams 9 of 26 teams were composed of 2 people, one of whom was the team leader.	Correlations of Total EIQ Scores with Overall Team Performance or Team Leader (pp. 65–73) <i>Predictor</i> <i>r</i> <i>Criterion</i> Team Mean EIQ Scores 0.25/0.30/0.08 Manager ranking/rating/Average team member ratings of team performance. Team leader EIQ scores –0.01/0.03/0.05 Manager team ranking/rating/Average team member team performance ratings. Team Average EIQ scores 0.34/0.03 Manager team leader ranking./ performance ratings. Team EIQ 0.11 Average team member ratings of team leader. Team Leader EIQ 0.51/0.25/–0.18 Manager team leader ranking/ ratings/ Team member average performance ratings of team leaders. • Using a two-tailed test, none of the correlations shown in the table above were found to be significant (Rice shows two of the correlations to be significant).
Emotional Intelligence Scale by Schutte et al. (1998) 33 self-report items on a Likert-type scale.	Malouff & Schutte (1998)	<i>N</i> = 26 college seniors	$r = 0.38, p < 0.05$, one-tailed. After four months, supervisors rated the students' para-professional performance.

Table 19.8: Continued.

Measure	Study	Sample	Results
Emotional Knowledge Test (EKT) 5 dimensions scored by expert and consensus ratings, 1 dimension was also scored by the target.	Graves (1999)	N = 149 participants who were recruited through a newspaper advertisement and paid \$130-\$150	Criterion measures were assessor and peer performance ratings on assessment center exercises, not job performance. Expert vs. consensus scores give very different results. Coefficient alpha ranged from 0.15 to 0.82. Emotional Intelligence composite with performance composite: Table 24 (Graves 1999, p. 171). Correlations between EIQ, IQ, and Performance Composites. (N = 149). Composite Emotional Intelligence IQ Expert Scored Consensus Scored Peer 0.24** 0.12 0.25** Assessor 0.27** 0.10 0.24** Combined 0.31** 0.13 0.29** Adapted from Graves (1999) ** $p < .01$, two-tailed. Note: Peer = Peer composite based on average factor scores; Assessor = Assessor composite based on the average of the adjusted observed ratings. Combined = Combined peer and assessor composites.

Table 19.8: Continued.

Measure	Study	Sample	Results
PONS (Measure of empathy)	McClelland & Dailey (1972)	N = 115 Foreign Service Officers	Test
			Reliability Validity
			U.S. Knowledge Test
			0.70
			0.32**
			Empathy (PONS)
			0.20
			0.11
			<i>Note:</i> Table adapted from Table 1 of Barrett (1994).
			** $p < 0.01$, two-tailed.

It is important to note that none of the tests in this table are objective tests with correct answers. They are either self-report or rated by others. As can be seen in the table, none of the studies would pass all of the Daubert standards.

3.6. Group Intelligence

Williams & Sternberg (1988) defined group intelligence as “the full potential of a group of people working as a unit” (p. 356). They conducted a study to demonstrate the relationship between IQ and group performance (group product quality). Williams and Sternberg concluded that IQ was an essential component of group intelligence and as a predictor of group performance (see Table 19.9).

They also concluded that talkativeness and dominance were part of group intelligence. However, Goleman came to very different conclusions when he described this study in his 1998 book. Goleman (1998b) asserted that “In a classic study of group IQ by Wendy Williams and Robert Sternberg . . . the interpersonal skills and compatibility of the group members emerged as key to their performance (a result found time and time again). Williams and Sternberg found out that those who were socially inept . . . were a drag on the whole effort . . . All in all, the social effectiveness of the group predicted how well it would do, more than did the individual IQ of its members” (pp. 205, 358, notes 15, 16).

This is quite a different interpretation than that offered by the authors of the study: Williams & Sternberg (1988) stated that “. . . IQ was an essential component of intelligence; not only is a lot of IQ on average desirable, but also, one group member particularly high in IQ” (p. 375) and that “. . . having a fellow group member with a high desire to participate in such situations (demanding and uncomfortable social situations)

Table 19.9: Williams & Sternberg (1988).
Cognitive correlations of IQ with group product quality (from Table 7, p. 369).

IQ	Group Product Quality
Maximum (group member who scored highest)	0.65***
Mean (average score of 3 group members)	0.65***
Minimum (group member who scored lowest)	0.43*
Personality characteristics (from Table 7, p. 369)	
	Group Product Quality
Empathy	NS
Extraversion	NS
Maximum Desire to Participate	-0.64***
Mean Private Self-Consciousness	0.41*
Multiple regressions of all written predictors on group product quality (from Table 8, p. 373)	
R ² = 0.64	Beta
Maximum IQ	0.50**
Maximum Desire to Participate	-0.48**

has a negative impact on the group's performance" (p. 370). Neither empathy nor extraversion predicted group product quality.

4. Conclusions Regarding the Daubert Standards

The first of the Daubert criteria states that the theory must have been tested, or is at least able to be tested. Gardner (1999) admits that he is not going to develop tests and attempt to empirically prove his theory. This admission would automatically rule out his expert testimony and preclude his concept of multiple intelligences from having any value for personnel selection in a real context. The concept of Emotional Intelligence also would fail to meet the Daubert standards, since there are no instruments which have demonstrated validity in predicting job performance. Goleman's (1995, 1998a, 1998b) theory is so diffuse it can never be tested, while Mayer *et al.* (1990) have a theory but negative results. Competencies as developed by McClelland (1973), Boyatzis (1994), and Spencer & Spencer (1993) is not actually a theory that can be tested, but a process of obtaining scores based on expert responses. Many competencies rely on past

performance, but there is no definition of the construct of past performance. This has not been done with the measures of specific intelligence described here. Tacit knowledge (Wagner & Sternberg 1985), has been criticized due to its lack of a coherent definition (Gottfredson, in press). Without a definable construct/latent trait, there is no way a coherent measure can be developed. Constructs need to be defined, measures need to be developed, tested and then cross-validated.

The second of the Daubert criteria, that the theory (& expert) must have (been) published in peer reviewed publications, has also not been met by these new concepts. One of the outstanding features of all new concepts of intelligence is that they are primarily published in trade books and book chapters without adequate peer review. The number of actual publications in peer reviewed journals by these mentioned authors is very few, despite the fact that they have been around for 30 years.

The third is that there must be a known or potential error rate. Within selection contexts, either insufficient data exists or the data show that the instruments are not valid. It is impossible to determine the number of true positives, true negatives, false positives and false negatives to be expected in using the technique.

The fourth of the criteria says that the theory must be generally accepted in the relevant scientific community. All of the conceptualizations have been critiqued by practitioners and professionals in the area of Industrial/Organizational Psychology. It is doubtful that there is any acceptance of the concepts in the relevant field.

The fifth of the Daubert criteria asserts that the methods for testing the theory must meet scientific standards. All of the attempts to validate the instruments discussed here use non-standard procedures. They confuse predictors and criteria, make unwarranted statistical adjustments to the data, try schemes of using experts to develop prototypes, and manipulate data to get desired results when all else fails. Finally, none of the studies have ever used a predictive validation design.

5. Common Characteristics of Advocates of “New” Intelligences

All of the new concepts, including emotional intelligence, tacit knowledge, practical intelligence, competencies and multiple intelligence have common themes throughout their theories and research.

5.1. All of the “New and Innovative” Theorists Use a Strawman Approach

These theorists start by claiming that cognitive ability is given too much weight and that other attributes are important in predicting job performance or life outcomes. They say that the correlation between cognitive ability and job performance is 0.20, which accounts for 4% of the variance, so 96% of the variance is unaccounted for (Ghiselli 1966; Goleman 1998b). Of course, not even the early theorists of intellectual ability (e.g. Spearman 1904) ever said that it was the sole determinant of success in life. Certainly, it is possible that personality traits, attitudes and values might add incremental validity to cognitive ability in predicting job performance. The problem still

is one of obtaining relevant non-cognitive measures that do add incremental validity in predicting job performance (Barrett *et al.* 2001). Research concerning these new concepts has continually failed to show incremental validity of any of their measures over cognitive ability in predicting job performance.

Goleman (1998b) implies that emotional intelligence should have incremental validity over cognitive ability with statements such as “Paradoxically, IQ has the least power in predicting success among that pool of people smart enough to handle the most cognitively demanding fields, and the value of emotional intelligence for success grows more powerful the higher the intelligence barriers for entry into a field. In MBA programs or in careers like engineering, law, or medicine, where professional selection focuses almost exclusively on intellectual abilities, emotional intelligence carries much more weight than IQ in determining who emerges as a leader”. Goleman also relies on anecdotes (e.g. p. 22) in which two people have high ability, but what determines success is their emotional intelligence. Despite Goleman’s claims and anecdotes, he provides absolutely no data to show the incremental validity of EQ tests.

Even personality measures, which Goleman cites as a scientific basis for emotional intelligence have failed to demonstrate incremental validity. It is doubtful that newly developed personality-based measures (i.e. emotional intelligence) will be able to find incremental validity where decades of research have failed. Goleman stated that emotional intelligence is based on 5 elements: self-awareness, motivation, self-regulation, empathy and adeptness in relationships. Each of these are non-cognitive constructs that have been researched elsewhere, and he relies on personality based research to support his propositions. For example, he cites research on early career self-confidence in predicting promotions and success in later career (Howard & Bray 1988), and longitudinal research on high-IQ individuals that found those most self-confident in their early career were most successful in their later career. This is consistent with Goleman’s description of the essence of emotional intelligence, that “the new measure takes for granted having enough intellectual ability and technical know-how to do our jobs; it focuses instead on personal qualities, such as initiative and empathy, adaptability and persuasiveness” (p. 3). Goleman also describes Spencer and Spencer’s (1993) research by saying that the need to achieve was found to be the strongest competence that distinguished star from average executives. Again, he is using this to say that among top level executives, achievement drive is what distinguishes among level of performance. No evidence is presented to support the incremental validity of emotional intelligence over cognitive ability in predicting performance. The majority of Goleman’s evidence for the importance of empathy for superior job performance comes in anecdotal form (pp. 133–162). He also relies on Spencer & Spencer’s (1993) work to say that developing others (sensing others’ development needs and bolstering their abilities) was paramount to sales’ manager performance because it was the competence most frequently found by top performers in the field. Again, this evidence is in no way a validation study following professional and legal guidelines, and none of the 286 studies reviewed by Spencer and Spencer were published in peer review journals. The ability to regulate oneself and effectively interact with others is part of Goleman’s definition of emotional intelligence. A study conducted by Stewart & Carson (1995) was used by Goleman as an example of the importance of this trait. This study found that

extraversion was negatively related to job performance. Extraversion is one of the Big Five factors of personality, which have been studied extensively as predictors of job performance.

Table 19.10 illustrates the lack of evidence that personality provides incremental validity over cognitive ability.

Many people cite Schmidt & Hunter (1998) in order to demonstrate an alternative selection procedure exists that is both valid and results in incremental validity over cognitive ability: personality (specifically conscientiousness and integrity). However, Schmidt & Hunter (1998) rely on simulations, not actual validity studies, so this cannot be used as evidence of incremental validity. This article is also often misinterpreted as a meta-analysis, which is untrue.

One of the few studies to empirically demonstrate incremental validity was Day & Silverman (1989). Day and Silverman found that interpersonal orientation had incremental validity over cognitive ability in predicting cooperation and a global job performance measure, and that ascendancy had incremental validity over cognitive ability in predicting potential for success. However, the manner of computing these personality variables was unusual. An individual's score on interpersonal orientation was calculated by subtracting the sum of two subscale scores from the sum of four other subscale scores. No definitions of the constructs themselves were given, nor have Day and Silverman's results ever been cross-validated. Finally, most of the studies presented in Table 19.10 tend "stack the deck" when looking for a personality trait that will result in incremental validity. These studies, as is often the case in practice, correlate all personality traits in a measure with job performance and then use only those that correlate in the regression equation. This is contrary to a rational model of test development where constructs are defined, specific hypotheses are made, and those hypotheses are tested. Despite claims and anecdotal evidence that new measures will have incremental validity, there is simply no empirical evidence.

5.2. Research of Dubious Relevance is Cited to Add Legitimacy to Their "New" Concepts

Often these researchers interpret research results directly opposite to actual findings (Barrett 1994), leave out of their discussion any positive results for cognitive ability, cite unpublished research they claim supports their viewpoint, which is often unattainable, and ignore early research on the topic.

Cherniss (2000) used a longitudinal study completed by Snarey & Valliant (1985) to assert that IQ has little relationship with how well people do at work. However, the authors of the study said that "... whenever intelligence is included among the variables, it emerges as a more significant factor than social or personality measures". Cherniss (2000) was simply wrong in his interpretation of the Snarey & Valliant (1985) article.

The most egregious case of citing inaccurate evidence is Goleman, who cites hundreds of articles in professional literature to support his propositions. In examining these studies, we found that he was often factually incorrect in his reporting. For

Table 19.10: Incremental validity of personality measures over cognitive ability tests for predicting performance in published studies.

Study	Sample	Predictor(s)	Criterion	R ²	ΔR ²
Allworth & Hesketh (2000)	245 Hotel employees	Cognitive ability • Raven's Progressive Matrices • Ball Clerical Speed and Accuracy Test • Numerical Reasoning Test	Supervisor ratings of job performance	0.0729	—
		Goldberg Agreeableness, Conscientiousness and Extroversion		0.1225	0.0496
Armeson <i>et al.</i> (1993)	50 insurance claims examiners	Claims Examiner Inventory and Basic Skills Test	Overall performance	0.35	—
		Hogan Personality Inventory and PROFILE		0.41	0.06
Black (2000)	284 New Zealand police recruits	Cognitive ability	Overall performance on practical and academic tests	0.11	—
		NEO-PI-R Conscientiousness		0.17	0.06
Cortina <i>et al.</i> (1992) ^a	314 State police recruits	Civil service exam	Supervisor final training ratings of recruits	0.16	—
		Inwald Personality Inventory		0.20	0.04
Crant (1995)	146 real estate agents	General mental ability	Job performance (houses sold, listings generated and commission income)	0.129	—
		NEO-FFI Conscientiousness		0.130	0.001
Day & Silverman (1989)	43 accountants	Wesman	Global composite of Potential for Success, Technical Ability, Timeliness of Work, Client Relations and Cooperation	0.019	—
		GPA		0.082	0.063
		Interpersonal Orientation (Positive weight in regression equation) Interpersonal Orientation was composed of: (affiliation + nurturance + degree of exhibition + social recognition) — (aggression + defence)		0.231	0.034
		Ascendancy (Negative weight in regression equation) Ascendancy was composed of: (dominance — abasement)		0.259	0.028
		Work Orientation (Positive Weight in regression equation) Work Orientation was composed of: (achievement + endurance) — play			

Table 19.10: Continued.

Study	Sample	Predictor(s)	Criterion	R ²	ΔR ²
Fetzer <i>et al.</i> (2001)	152 customer service managers	Reasoning Ability	Performance Appraisal composite	0.027	—
		Personal Characteristics Inventory — Agreeableness		0.051	0.024
		Reasoning Ability	Interpersonal performance	0.030	—
		Personal Characteristics Inventory — Agreeableness		0.049	0.019
		Numeric Ability	Interpersonal performance	0.020	—
		Personal Characteristics Inventory — Agreeableness		0.039	0.019
Numeric Ability	Integrity-type performance	0.046	—		
Personal Characteristics Inventory — Agreeableness		0.062	0.016		
Gellatly <i>et al.</i> (1991)	114 food- service organization unit managers	Personnel Assessment Form (PAF)	Overall supervisor ratings	0.004	—
		Personality Research Form-E (PRF- E) Self Reliant		0.107	0.103
Hattrup <i>et al.</i> (1998)	103 entry-level customer service and sales representatives	Cognitive ability	Sales performance	0.0961	—
		Conscientiousness (O'Connell 1994)		0.0970	0.0009
Jackson <i>et al.</i> (2000)	187 security officers	Cognitive ability tests	Standardized incident report score	0.57	—
		Jackson Personality Inventory- Revised Dependability		0.58	0.01
McHenry <i>et al.</i> (1990)	4,039 soldiers in nine army jobs	General Cognitive Ability (ASVAB)	Core Technical Proficiency	0.63	—
		Temperament/Personality composite computed from ABLE (Achievement Orientation, Dependability, Adjustment & Physical Condition)		0.63	0.00

Table 19.10: Continued.

Study	Sample	Predictor(s)	Criterion	R ²	ΔR ²	
Mount <i>et al.</i> (1999)	146 civilian U.S. Army Managers	Wonderlic Personality Test	Supervisor ratings	0.029	—	
		Personal Characteristics Inventory Conscientiousness		0.084	0.055	
	103 sales representatives	Wonderlic Personality Test		0.04	—	
		Personal Characteristics Inventory Conscientiousness		0.105	0.065	
	121 district managers	Wonderlic Personality Test		0.058	—	
		Personal Characteristics Inventory Conscientiousness		0.122	0.064	
Mount <i>et al.</i> (2000)	376 clerical employees	Wonderlic Personnel Test	Quantity/Quality	0.02	—	
		Personal Characteristics Inventory (Conscientiousness, Extraversion, Agreeableness, Openness, Emotional Stability)		0.136	0.116	
	Wonderlic Personnel Test	Problem Solving	0.047	—		
	Personal Characteristics Inventory (Conscientiousness, Extraversion, Agreeableness, Openness, Emotional Stability)		0.168	0.121		
	Wonderlic Personnel Test	Interpersonal Relationships	0.001	—		
	Personal Characteristics Inventory (Conscientiousness, Extraversion, Agreeableness, Openness, Emotional Stability)		0.051	0.049		
	Wonderlic Personnel Test	Retention	0.002	—		
	Personal Characteristics Inventory (Conscientiousness, Extraversion, Agreeableness, Openness, Emotional Stability)		0.080	0.078		
	Neuman & Wright (1999) (individual level data)	316 HR representatives	Skills (Checking & Forms Completion) and Cognitive ability	Peer ratings of task performance	0.12	—
			NEO-PI-R Conscientiousness and Agreeableness		0.20	0.08
Neuman & Wright (1999) (group level data)	79 four-person work teams	Skills (Checking & Forms Completion) and Cognitive ability	Supervisor ratings of task performance	0.28	—	
		NEO-PI-R Conscientiousness and Agreeableness		0.36	0.08	

Table 19.10: Continued.

Study	Sample	Predictor(s)	Criterion	R ²	ΔR ²
Oakes <i>et al.</i> (2001)	9,793 Air Traffic Controller trainees	Cognitive ability	Skill acquisition	0.039	—
		16PF (Q2, N, M, Q1, A, E, F, B, I, G, O, Q3 Factors)		0.041	0.002
Siem (1992)	325 Air Force Pilot Trainees	Battery without Automated Aircrew Personality Profiler (AAPP) (Air Force Officer Qualifying Test and Basic Attributes Tests only)	Training success	0.29	—
		Full model (AAPP included)		0.33	0.04

^a Cortina *et al.* (1992) stated that “the incremental validity of one inventory over the other is not assessed . . . Analyses with the IPI entered before and after the MMPI were conducted. Because these analyses took up valuable space and added little information to those that are now presented, they were removed”.

example, he cites McClelland’s research to show that EIQ is more important than cognitive ability tests (Goleman 1998a). However, the results of that study showed that cognitive ability correlated 0.32 for Foreign Service Officers and empathy correlated 0.11. In addition, Goleman (1998b) referred to a study completed by Boyatzis *et al.* (1990) when he said that team leaders who were trained in team leadership competencies later had higher morale and had cut product development time by 30%. However, Boyatzis *et al.* stated that while the training in team leadership competencies clarified leadership it did not necessarily help with lack of management skills.

Goleman (1998a) also exhibited some selective reporting by not citing relevant articles that don’t support his claims, including results that show cognitive ability as the best predictor (Crant 1995; Daniel 1992; Davis & Kraus 1997; Holahan *et al.* 1995). These articles consistently showed that intellectual functioning is a better predictor than concepts such as interpersonal trust, machiavellianism, personality and self-confidence. Sternberg has also ignored evidence contradicting his theories. Colonia-Willner (1998) found that the TKIM did not predict job performance. However, this study has been completely ignored by Sternberg (2001, 2002), as it does not support his theory that tacit knowledge should predict job performance.

5.3. Reinventing the Wheel and Ignoring Contradictory Evidence

Much of the literature on these creative concepts of intelligence uses old concepts without acknowledging their historical basis. As early as 1904, Spearman had measurement tools for academic intelligence and common sense. Certainly, Sternberg and Wagner’s practical intelligence is not a new concept. While Spearman believed that common sense was highly related to general cognitive ability, this did not preclude

Sternberg and Wagner from asserting, with little or no evidence, that this is a separate concept.

In a similar fashion, competencies were used in industrial psychology tests as early as the 1930s (Bingham 1937). The term competencies has no agreed upon definition yet they seem to be no more than what has been called Knowledge, Skills, Abilities and Other characteristics (KSAOs) (Barrett 2000). There is no way to distinguish what have been referred to as abilities and what are now referred to as competencies. In 1995, abilities were defined as “an underlying characteristic of the person that leads to or causes superior or effective performance” (Boyatzis *et al.* 1995). Competencies were then defined in exactly the same way in a 1996 article (Boyatzis 1996). Such researchers did not acknowledge Daniel’s (1992) published study examining leadership competencies of supervisors at an electronics company. In fact, the term competencies has been around for 60 years, first used by Walter Van Dyke Bingham in 1937. The new theorists have just recycled an old term without acknowledging the original researchers.

Social intelligence or emotional intelligence has been a subject of research since the 1920s. This line has continually faced the challenge of developing an operational definition and establishing validity of its use in predicting job performance. In fact, the George Washington Social Intelligence Test had validity levels higher than emotional intelligence tests that have been developed in recent times (Hunt 1928).

There continues to be confusion in the definition of emotional intelligence. There are two ways to measure emotional intelligence: objective and self-report tests. Both forms of EIQ tests should be considered to be extensions of objective and self-report empathy tests. There is no reason to believe that the new tests of EIQ are measuring a new concept and they share the same problems with past personality tests. These problems include low reliability, low or no criterion-related validity, limited construct validity, easily faked and differential validity for broad versus narrow trait assessment (Barrett *et al.* 2001). While the use of ability based EIQ measures helps with the problem of faking, the reliability and validity are too low (Barrett *et al.* 2001). There may not even be a single emotional intelligence construct. Barrett *et al.* (2001) found that the concept of emotional intelligence may not be one single construct. Two of the subscales of EQ, Emotion Perception (Music) and Emotion Perception (Sound) were negatively correlated (-0.38) (Davies *et al.* 1998). Even the label originated earlier than is typically noted. Mayer *et al.* (1990) are often credited for coining the term emotional intelligence, when in fact Payne (1985, 1986) used the term in his dissertation five years earlier.

Gardner’s (1983) Multiple Intelligences theory presented human intelligence as a set of intellectual potentials including linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, intrapersonal and interpersonal. These are proposed by Gardner to be distinct components of human intelligence. However, he does not cite Spearman (1904), who also included a musical component, pitch discrimination, in his studies on intelligence. Contrary to Gardner, Spearman found that pitch discrimination was not distinct from general mental ability and found evidence for one general factor of intelligence. We do not wish to take a stand on the matter of whether musical ability is distinct from general mental ability, but simply wish to point out that Gardner completely ignored evidence contrary to his theory that had been available for almost 80 years.

5.4. Research Has Used Unorthodox Validation Methods

Many of the studies cited by these innovative researchers involve measures which confuse the independent variable (i.e. the predictor) and the dependent variable of job performance. The most graphic example of this was the Wagner & Sternberg (1990) study performed at the Center for Creative Leadership. They used as a measure of job performance two simulations involving groups of subjects interacting on a business problem. The performance of managers in the simulation were rated on nine dimensions. These simulations are routinely used in assessment centers to predict present or subsequent job performance. By a slight of hand, Wagner and Sternberg have turned performance on a predictor into job performance.

5.5. They Use Complex, Unorthodox, Usually Changing Methods to Score Tests

The evidence for competency testing, practical intelligence, and tacit knowledge all use unusual scoring procedures for their predictors. There are obvious problems with their expert prototype scoring and simulation approaches. One is that the “Psychological Corporation recommends that each organization conduct local validation studies to determine the job relatedness of TKIM” (Wagner & Sternberg 1991). Few organizations have the resources to complete a criterion related, concurrent validation study every time they want to use a valid selection test. The point is that there are no universal right answers with the Tacit Knowledge Inventory and the correct answers depend upon expert groups used in any one organization, in contrast to cognitive ability tests. In *Fink v. Finegan* (1936), the basic principles of a competitive examination procedure were outlined. These principles included having an objective standard or measure that is capable of being challenged and reviewed by other experts in the area, and there must be definite standards. It is clear from our review that the TKIM cannot be used in civil service testing because it does not meet the requirements of a competitive examination. This is true in every jurisdiction. Competitive examinations require that there be an effective competition among candidates. It also means that the test must be as objective as possible. This is impossible with the TKIM. Wagner & Sternberg (1991) pointed out that “there are no right or wrong answers for the response alternatives; the scoring is based on the amount of agreement between experts and applicants” (p. 23). Therefore, the measure will not satisfy the requirements of a competitive examination.

In addition, the distinction between concurrent and predictive validation designs is not well understood. Barrett *et al.* (1981) are often cited as evidence that validity coefficients obtained under concurrent and predictive designs tend to be similar. With non-cognitive tests, however, validity coefficients obtained using a predictive design are generally lower compared to validities obtained using a concurrent design (Hough 1998; Ones *et al.* 1993). The studies cited to support the use of specific intelligences are not based on predictive designs.

An additional problem with the TKIM is that there has been no cross-validation of the items. In the manual, Wagner & Sternberg (1991) state that there are 12 scenarios, each with 9 to 20 items. This means that there were over 120 items in the original TKIM,

however, only 39 are reported to be significant. The alpha of these items is only 0.68, which is too low for use in individual selection. The process used to validate this test is similar to that used for BIBs (Biographical Information Blanks), which capitalizes on chance.

5.6. Ignore Professional and Legal Standards for Initial Selection and Promotion

The literature on these “creative” concepts of intelligence never refer to the U.S. professional and legal guidelines because they violate them. In our review of validation studies of actual incumbents or candidates, we did not find one study which met professional standards from any of these novel researchers. For example, Bar-On (1997) inappropriately used two studies (Handley 1997; Wagner & Morse 1975) to attempt to show the validity of his Emotional Quotient Inventory (EQ-I). One of the studies (Handley 1997) was an unpublished study in an Air Force Newsletter that only reported mean differences and did not report validity coefficients. In addition, the U.S. GAO stated that it is too early to evaluate its effectiveness (U.S. GAO 1998). Wagner & Sternberg’s (1991) concept of tacit knowledge has little evidence to support its validity. In fact, the Users Manual reports no evidence of validity and no studies that use job performance (Barrett *et al.* 2001). There isn’t one study where there was a direct comparison of these theories with cognitive ability to determine relative or incremental validity.

The Uniform Guidelines (Equal Employment Opportunity Commission. (1979, p. 2253-24; question #74) state that “content validity is not appropriate where the selection procedure involves knowledges, skills, or abilities which the employee will be expected to learn ‘on the job’ However, if such an ability . . . takes a substantial amount of time to learn, is required for successful job performance, and is not taught to those initial hires who possess it in advance, a test for that ability may be supported on a content validity basis”. That is, selection tests cannot test for information that could be learned on the job in a short amount of time. The TKIM manual states “TKIM is also an excellent tool for training and development . . . Training directors can lead group discussions that will elaborate on the practical know-how and “rule of thumb” that underlie the expert managers’ ratings” (Wagner & Sternberg 1991). This is an admission that whatever the TKIM measures can be trained in a short time and is therefore inappropriate to be used as a selection device. The test contains various work-related situations and items relevant to handling those situations. For example, one of the situations is a role play in which the vice-president of an electronics company needs to decide what to do about the company losing market share. These items do not display adequate coverage of performance on the job. The TKIM contains no content valid items.

A problem with competencies is that there are no empirical studies that show competencies are valid. Spencer & Spencer (1993) stated that criterion validity was the most important aspect of assessing a selection instrument. They also state that predictive validation studies are superior to other forms of validation. However, they provide no predictive validity evidence to support the use of competencies. While Spencer and

Spencer do claim that they have shown that competencies have predictive validity, they have confused validation studies with pilot studies. The use of very small sample sized precludes the use of any type of meaningful analyses (Schmidt *et al.* 1976).

5.7. Claim Their Approach Will Have Superior Validity to CATs With No Adverse Impact

All of the “novel” researchers claim that their new measures of competencies, emotional intelligence, practical intelligence and tacit knowledge have lower adverse impact when predicting job performance. There are three problems with this contention. First, there are very few examples, if any, demonstrating that any of these constructs actually predict job performance in real organizations. Second, we could find no evidence based on studies in any organization that these measures reduce adverse impact. Third, even when one considers personality tests, the alternative selection tests most researched as a means to reduce adverse impact, there is no conclusive evidence that a consistent reduction in adverse impact occurs. In fact, there is evidence that the use of a personality test in addition to a cognitive ability test may produce larger mean group differences than the use of a cognitive ability test alone (Kriska 2001). This result is a function of the correlation between predictors in the composite and the mean differences between races on the alternative test (Schmitt *et al.* 1997). In other words, a general statement that the introduction of a personality test to a cognitive ability test will result in a decrease in adverse impact cannot be made. A more accurate statement would be to say that in *some* cases, the use of personality tests in conjunction with ability tests *may* reduce adverse impact, but that in other cases it may in fact increase adverse impact. Therefore, even if there were empirical examples of these newer intelligence constructs’ ability to predict job performance or produce less adverse impact than a traditional ability test, this would still not provide evidence of their ability to do so in a composite selection battery with a cognitive ability test.

References

- Adams v. City of Chicago 1996 U.S. Dist. LEXIS 3567 (N.D. Ill. 1996)
- Allworth, E., & Hesketh, B. (2000). Job requirements biodata as a predictor of performance in customer service roles. *International Journal of Selection and Assessment*, 8, 137–147.
- American College of Trial Lawyers (1994, April 15). *Standards and procedures for determining the admissibility of expert evidence after Daubert*.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Arneson, S., Millikin-Davies, M., & Hogan, J. (1993). Validation of personality and cognitive measures for insurance claims examiners. *Journal of Business and Psychology*, 7, 459–473.
- Bachman, J., Stein, S., Campbell, K., & Sitarenios, G. (2000). Emotional intelligence in the collection of debt. *International Journal of Selection and Assessment*, 8, 176–182.
- Bar-On, N. (1997). *Bar-On emotional quotient inventory: Technical manual*. Toronto, ON: Multi-Health Systems, Inc.

- Bar-On, R. (2000). Emotional and social intelligence: Insights from the Emotional Quotient Inventory. In: R. Bar-On, & J. D. A. Parker (Eds), *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 363–388). San Francisco, CA: Jossey-Bass.
- Barrett, G. V. (1994). Empirical data say it all. *American Psychologist*, 49 (1), 69–71.
- Barrett, G. V. (2000, April). Recommended educational experience to prepare an I/O psychologist to be an expert witness. In: D. Zink (Chair), *Implications of legal developments*. Symposium conducted at the 15th annual meeting of the Society for Industrial and Organizational Psychologists, New Orleans, LA.
- Barrett, G. V., Miguel, R. F., Tan, J. A., & Hurd, J. M. (2001, April). *Emotional Intelligence: The Madison Avenue approach to science and professional practice*. Presented at the 16th annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology*, 66, 1–6.
- Bingham, W. V. D. (1937). *Aptitudes and aptitudes testing*. New York: Harper & Brothers.
- Black, J. (2000). Personality testing and police selection: Utility of the “big five”. *New Zealand Journal of Psychology*, 29, 2–9.
- Black v. Rhone-Poulenc, Inc. 19 F. Supp. 2d 592 (S.D.W. Va. 1998).
- Boyatzis, R. E. (1994). Rendering unto competence the things that are competent. *American Psychologist*, 49 (1), 64–66.
- Boyatzis, R. E. (1996). Consequences and rejuvenation of competency-based human resource and organization development. *Research in Organizational Change and Development*, 9, 101–122.
- Boyatzis, R. E., Cowen, S. S., & Kolb, D. A. (Eds) (1995). *Innovation in professional education*. San Francisco: Jossey-Bass.
- Boyatzis, R. E., Esteves, M. R., & Spencer, L. M. (1990). Entrepreneurial innovation in pharmaceutical research and development. *Human Resource Planning*, 15 (4), 15–29.
- Brown v. City of Chicago, 917 F. Supp. 577 (N.D. Ill. 1996).
- Brown v. City of Chicago, 8 F. Supp. 2d 1095 (N.D. Ill. 1998).
- Butler v. Home Depot, Inc., 984 F.Supp. 1257 (N.D. Cal. 1997).
- Camp v. Lockheed Martin Corporation, U.S. Dist. LEXIS 20750 (S.D. Tex. 1998).
- Carrothers, R. M., Gregory, S. W., & Gallagher, T. J. (2000). Measuring emotional intelligence of medical school applicants. *Academic Medicine*, 751, 456–463.
- Cherniss, C. (2000, April). Emotional intelligence: What it is and why it matters. In: R. C. Page (Chair), *Competency models and emotional intelligence: Are they useful constructs?* Symposium conducted at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Clark v. Takata Corp. 192 F. 3d 750 (7th Cir. 1999).
- Colonia-Willner, R. (1998). Practical Intelligence at work: Relationship between aging and cognitive efficiency among managers in a bank environment. *Psychology and Aging*, 13 (1), 45–57.
- Cooper, H. (1979). Statistically combining independent studies. A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology*, 37(1), 131–176.
- Cortina, J. M., Doherty, M. L., Schmitt, N., Kaufman, G., & Smith, R. G. (1992). The “big five” personality factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology*, 45, 119–140.
- Crant, J. M. (1995). The proactive personality scale and objective job performance among real estate agents. *Journal of Applied Psychology*, 80, 532–537.

- Holahan, C. K., Sears, R. R., & Cronbach, L. J. (1995). *The gifted group in later maturity*. Stanford, CA: Stanford University Press.
- Daniel, T. L. (1992). Identifying critical leadership competencies of manufacturing supervisors in a major electronics corporation. *Group and Organization Management*, 17 (1), 57–71.
- Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).
- Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology*, 75 (4), 989–1015.
- Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In: W. Ickes (Ed.). *Empathic Accuracy* (pp. 144–168). New York: Guilford Press.
- Day, D. V., & Silverman, S. B. (1989). Personality and job performances: Evidence of incremental validity. *Personnel Psychology*, 42 (1), 25–36.
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior & Human Decision Processes*, 58 (1), 51–100.
- Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43 (166), 38290–38315.
- Equal Employment Opportunity Commission. (1979). Adoption of questions and answers to clarify and provide a common interpretation of the uniform guidelines on employee selection procedure. *Federal Register*, 44 (43), 11996–12009.
- Equal Employment Opportunity Commission. (1980). Adoption of additional questions and answers to clarify and provide a common interpretation of the uniform guidelines on employee selection procedure. *Federal Register*, 45 (87), 29530–29531.
- Feist, G. J., & Barron, F. (1996, June). *Emotional intelligence and academic intelligence in career and life success*. Paper presented at the annual convention of the American Psychological Society, San Francisco, CA.
- Fetzer, M. S., Fortunato, V. J., Kudisch, J. D., & Eidson, C. E. Jr. (2001). *Predictive and incremental validity in a customer service setting*. Poster presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Fink v. Finegan (1936) 270 N. Y. 356.
- Firefighters Institute for Racial Equality v. City of St. Louis 2000 U.S. App. LEXIS 19191 (8th Cir. 2000).
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books, Inc.
- Gardner, H. (1999). *Intelligence Framed: Multiple intelligences for the 21st Century*. New York: Basic Books.
- Gellatly, I. R., Paunonen, S. V., Meyer, J. P., Jackson, D. N., & Goffin, R. D. (1991). Personality, vocational interest, and cognitive predictors of managerial job performance and satisfaction. *Personality and Individual Differences*, 12, 221–231.
- General Electric Company v. Joiner, 522 U.S. 136 (1997).
- Gerlib v. R. R. Donnelley & Sons Co. 2001 U.S. Dist. LEXIS 17879 (N.D. Ill. 2001).
- Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: John Wiley & Sons, Inc.
- Goleman, D. (1995). *Emotional Intelligence*. New York: Bantam Books.
- Goleman, D. (1998a). What makes a leader? *Harvard Business Review*, 76 (6), 92–102.
- Goleman, D. (1998b). *Working with emotional intelligence*. New York: Bantam Books.
- Gottfredson, L. S. (in press). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence*.

- Graves, J. G. (1999). Emotional intelligence and cognitive ability: Predicting performance in job-simulated activities. (Doctoral dissertation, California School of Professional Psychology, San Diego 1999). *Dissertation Abstracts International*, 60 (5-B), 2398.
- Handley, R. (1997). *Leveraging corporate performance through human capital profiling*. Unpublished manuscript.
- Hattrup, K., O'Connell, M. J., & Wingate, P. H. (1998). Prediction of multidimensional criteria distinguishing task & contextual performance. *Human Performance*, 11, 305–319.
- Holahan, C. K., & Sears, R. R., & Cronbach, L. J. (1995). *Self-confidence in high-IQ people and career success. The gifted group in later maturity*. Stanford: Stanford University Press.
- Hough, L. M. (1998). Personality at work. In: M. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives and traditional testing for selection* (pp. 131–159). Hillsdale, NJ: Erlbaum.
- Howard, A., & Bray, D. W. (1988). Self-confidence early in career predicts success later at AT&T. *Managerial lives in transition*. New York: Guilford Press.
- Hunt, T. (1928). The measurement of social intelligence. *Journal of Applied Psychology*, 12, 317–334.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340–362.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96 (1), 72–98.
- Jackson, D. N., Harris, W. G., Ashton, M. C., McCarthy, J. M., & Tremblay, P. F. (2000). How useful are work samples in validation studies? *International Journal of Selection and Assessment*, 8, 29–33.
- Kriska, S. D. (2001, April). The validity-adverse impact trade-off: Real data and mathematical model estimates. In: J. P. Wiesen (Chair), *Reducing adverse impact while maintaining validity: The public sector experience*. Symposium conducted at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Kumho Tire Company, Ltd. v. Carmichael, 526 U.S. 137 (1999).
- Malouff, J., & Schutte, N. (1998). *Emotional intelligence scale scores predict counselor performance*. Paper presented at the Annual Convention of the American Psychological Society, Washington, D.C.
- Mayer, J. D., DiPaolo, M., & Salovey, P. (1990). Perceiving affective content in ambiguous visual stimuli: A component of emotional intelligence. *Journal of Personality Assessment*, 54 (3&4), 772–781.
- Mayer, J. D., Salovey, P., & Caruso, D. (1999). *MSCEIT V. 2: The Mayer, Salovey, & Caruso Emotional Intelligence Test technical manual, Version 2: Mockup draft of 1st edition, July 28 2000*. Toronto, ON: Multi-Health Systems.
- McClelland, D. C. (1973). Testing for competence rather than for “intelligence”. *American Psychologist*, 28 (1), 1–14.
- McClelland, D. C., & Dailey, C. (1972). *Improving officer selection for the Foreign Service*. Boston, MA: Hay/McBer.
- McClelland, D. C., & Dailey, C. (1973). *Evaluating new methods of measuring the qualities needed in superior Foreign Service Officers*. Boston: McBer.
- McCormack, R. L. (1956). A criticism of studies comparing item-weighting methods. *Journal of Applied Psychology*, 40 (5), 343–344.
- McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335–354.

- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1999). The joint relationship of conscientiousness and ability with performance: Test of the interaction hypothesis. *Journal of Management*, 25, 707–721.
- 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.
- Murensky, C. L. (2000). *The relationships between emotional intelligence, personality, critical thinking ability, and organizational leadership performance at upper levels of management*. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.
- Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology*, 84, 376–389.
- Oakes, D. W., Ferris, G. R., Martocchio, J. J., Buckley, M. R., & Broach, D. (2001). Cognitive ability and personality predictors of training program skill acquisition and job performance. *Journal of Business and Psychology*, 15, 523–548.
- O'Connell, M. S. (1994). *Development and validation in Mexico of a conscientiousness scale*. Dallas, TX: Select International.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679–703.
- Payne, W. L. (1985/1986). A study of emotion: Developing emotional intelligence; self-integration; relating to fear, pain and desire (theory, structure of reality, problem-solving, contraction/expansion, tuning in/coming out/letting go). *Dissertation Abstracts International*, 47 (1), 203A. (UMI No. 8605928).
- Rice, C. L. (1999). *A quantitative study of emotional intelligence and its impact on team performance*. Unpublished master's thesis, Pepperdine University, Malibu, CA.
- Rudder v. District of Columbia, 890 F. Supp. 23 (D.D.C. 1995).
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128–1137.
- 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., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, 61 (4), 473–485.
- 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 (5), 719–730.
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. S., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25(2), 167–177.
- Scribner, S. (1984). Studying working intelligence. In: B. Rogoff, & J. Lave (Eds), *Everyday cognition: Its development in social context* (pp. 9–40). Cambridge, MA: Harvard University Press.
- Siem, F. M. (1992). Predictive validity of an automated personality inventory for Air Force pilot selection. *International Journal of Aviation Psychology*, 2, 261–270.
- Smith v. Ford Motor Co., 215 F.3d 713 (7th Cir. 2000).
- Snarey, J. R., & Vaillant, G. E. (1985). How lower- and working-class youth become middle-class adults: The association between ego defense mechanisms and upward social mobility. *Child Development*, 56 (4), 899–910.

- Society for Industrial and Organizational Psychology (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.
- Spearman, C. (1904). "General Intelligence" objectively determined and measured. *The American Journal of Psychology*, 15, 259–293.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan Co.
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. New York: John Wiley & Sons.
- Sternberg, R. J. (2001). Successful intelligence: Understanding what Spearman had rather than what he studied. In: J. M. Collis, & S. Messick (Eds), *Intelligence and Personality: Bridging the Gap in Theory and Measurement* (pp. 347–373). Mahwah, NJ: Lawrence Erlbaum.
- Sternberg, R. J. (2002). Successful intelligence: A new approach to leadership. In: R. E. Riggio, S. E. Murphy, & F. J. Pirozzolo (Eds). *Multiple intelligences and leadership*. Mahway, NJ: Lawrence Erlbaum Associates.
- Sternberg, R. J., & Wagner, R. K. (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge University Press.
- Stewart, G. L., & Carson, K. P. (1995). Personality dimensions and domains of service performance: A field investigation. *Journal of Business and Psychology*, 9 (4), 365–378.
- Thorndike, R. L. (1936). Factor analysis of social and abstract intelligence. *Journal of Educational Psychology*, 27 (3), 231–233.
- United States General Accounting Office. (1998, January). *Military recruiting: DOD could improve recruiter selection and incentive systems (GAO/NSIAD-98-58)*. Washington, D.C.: U.S.GAO.
- United States of America v. City of Garland, Texas, No. 3-98-CV-0307-L (N.D. Tex.). (2000).
- Wagner, R. K. (1994). Context counts: the case of cognitive ability testing for job selection. In: R. J. Sternberg, & R. K. Wagner (Eds), *Mind in context: Interactionist perspectives on human intelligence* (pp. 133–151). New York: Cambridge University Press.
- Wagner, F. R., & Morse, J. J. (1975). A measure of individual sense of competence. *Psychological Reports*, 36, 451–459.
- Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: the role of tacit knowledge. *Journal of Personality and Social Psychology*, 49 (2), 436–458.
- Wagner, R. K., & Sternberg, R. J. (1990). Street smarts. In: K. E. Clark, & M. B. Clark (Eds), *Measures of leadership* (pp. 493–504). West Orange, NJ: Leadership Library of America, Inc.
- Wagner, R. K., & Sternberg, R. J. (1991). *TKIM. The common sense manager. Tacit Knowledge Inventory for Managers: User Manual*. The Psychological Corporation.
- Williams, W. M., & Sternberg, R. J. (1988). Group intelligence: Why some groups are better than others. *Intelligence*, 12, 351–377.