Emotional Intelligence as an Aspect of General Intelligence: What Would David Wechsler Say?

Alan S. Kaufman
Yale University School of Medicine

James C. Kaufman
Educational Testing Service

R. D. Roberts, M. Zeidner, and G. Matthews (2001) have carefully examined the controversial issue of whether emotional intelligence (EI) should be classified as an intelligence and whether EI’s constructs meet the same psychometric standards as general intelligence’s constructs. This article casts their efforts into the framework of both historical and modern IQ-testing theory and research. It details David Wechsler’s attempts to integrate EI into his tests and how his conception of a good clinician would be that of an emotionally intelligent clinician. Current theories and research on IQ also have a role in EI beyond what Roberts et al. described, including J. L. Horn’s (1989) expanded model and A. R. Luria’s (1966) neuropsychological research, and better criteria than the Armed Services Vocational Aptitude Battery should be used in future EI studies. The authors look forward to more research being conducted on EI, particularly in future performance-based assessments.

In Roberts, Zeidner, and Matthews’s (2001) article, the authors asked pertinent questions and applied appropriate statistics to attempt to deal objectively with the controversial issues surrounding emotional intelligence (EI), most notably whether EI is legitimately classified as an intelligence and whether its constructs meet state-of-the-art psychometric standards. The cautious conclusions that they reached are logical within their chosen framework. However, they have inadvertently slighted the field of intelligence testing, particularly the clinical assessment of IQ. The authors failed to give adequate consideration to the pioneers of clinical IQ testing or to the theories and ideas that underlie the more recent intelligence tests.

The names Binet and Wechsler are mentioned in passing, but their implicit beliefs about EI are not explored, despite the unprecedented historical roles that these men played in defining psychometric tests of intellectual abilities. For example, the Army Alpha test that Mayer, Caruso, and Salovey (1999) used in their investigation, regardless of its obsolescence, was developed by Arthur Otis, Lewis Terman’s graduate student, as a group-administered Binet scale (A. S. Kaufman, 2000). In addition, though Wechsler’s first test battery, the Wechsler–Bellevue (Wechsler, 1939), included subtests developed during the early part of the 20th century, his slightly modified current tests remain the most popular clinical tests of intelligence nearly 100 years later within the United States and worldwide (e.g., Camara, Nathan, & Puente, 2000).

Also, the theories of intelligence discussed by Roberts et al. (2001), such as Horn’s (1989) and Guilford’s (1959), are well-respected but are not the most influential for the development of contemporary intelligence tests. Indeed, the theories that actually have a practical effect on IQ testing are often quite different from those that affect cognitive psychology (Sternberg & Kaufman, 1998). Whereas Horn’s theory is important, Roberts et al. (2001) treated it as the dichotomous fluid (Gf)–crystallized (Gc) model that characterized the Horn–Cattell approach, virtually ignoring the 8–10 abilities that define Horn’s (1989; Horn & Noll, 1997) refined and expanded Gf–Gc theory. And, despite the key role that Luria’s (1966) theory has played in contemporary test development, Roberts et al. did not mention this neuropsychological model.

Ultimately, our conclusions about EI as an aspect of intelligence and about the Multi-Factor Emotional Intelligence Scale (MEIS; Mayer et al., 1999) as a...
psychometric instrument are no different from the carefully reasoned conclusions reached by Roberts et al. (2001). Our goal, however, is to extend the context to include IQ measurement, especially the clinical assessment of IQ, from both historical and contemporary viewpoints.

Alfred Binet

To Binet (1886), EI was most surely an aspect of general intelligence. He believed that mental images were fluid and dynamic and that meaning was only comprehensible when interpreted within an experiential context (Binet, 1886; Cunningham, 1995). To Binet (1886), mental images function for each individual “as the life of the hive results from the life of the bees” (p. 168). He was fascinated by schoolchildren’s eyewitness accounts of events and investigated this aspect of contextual memory in a comprehensive research program, well aware of the interaction between children’s emotions and their intellect. The core issue to Binet in understanding memory (and learning and intellect) was “its extraordinary flexibility as a function of changes in mental state, emotional needs, more recent learning, and situational cues” (Cunningham, 1995, p. 960). To Binet, intelligence was complex, and the concepts that underlie EI fit nicely into his perception of its complexity.

David Wechsler

Wechsler (1975) distinguished between intellectual abilities and intelligent behavior or general intelligence. The former are measured by the psychometric tasks that he developed (or, more typically, “borrowed” from others) for his intelligence tests, as well as similar tasks on tests such as Terman’s (1916) Stanford–Binet. In contrast, Wechsler (1950) emphasized, “general intelligence cannot be equated with intellectual ability, but must be regarded as a manifestation of the personality as a whole” (p. 83). However, it is worth noting that Wechsler used the term personality in a different way than personality psychologists use the term today; Wechsler’s usage is a broader and more encompassing term, as he included most domains that involved the person (e.g., cognitive abilities, personality traits, emotional behaviors, and attitudes) within the construct of personality. But Wechsler never implied that the different aspects of personality would correlate so highly that they would be indistinguishable from each other. Like Roberts et al. (2001), he was a firm believer in the need to establish differential validity for each separate aspect of personality, be it intelligence, the Big Five personality factors, or EI.

Wechsler stressed the “conative” or nonintellective factors of intelligence that may either facilitate one’s intelligent behavior (e.g., persistence, curiosity, drive, will, conscientiousness) or inhibit it (e.g., anxiety, emotional insecurity, impulsivity, perseveration). Wechsler (1950) concluded that “factors other than intellectual enter into our concept of general intelligence, and that in everyday practice, we make use of them knowingly or not” (p. 83).

Though he did not refer specifically to the dimensions of EI as articulated by Salovey, Mayer, and other researchers (e.g., Salovey & Mayer, 1990), it is clear that Wechsler considered them as an aspect of intelligence. The dimensions of EI also fit nicely into Wechsler’s (1975) “bottom-line” definition of what intelligence tests measure:

What we measure with tests is not what tests measure—not information, not spatial perception, not reasoning ability. These are only means to an end. What intelligence tests measure, what we hope they measure, is something much more important: the capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges. (p. 139)

Without question, EI reflected aspects of general intelligence to Wechsler. Would he have wanted to measure it? Absolutely. He was impressed, and perturbed, that factor analyses of his tests rarely accounted for more than 60% of the total variance. He assumed that this missing or residual variance was accounted for by nonintellective factors of intelligence, and Wechsler (1950) very much wanted to measure these variables and include them in his tests:

What is needed is that these factors be rigorously appraised. . . . We already have some clues as to what the non-intellective but relevant factors of intelligence may be. What we now need are tests which not only identify but measure them. This in effect demands broadening our concept of general intelligence and calls for a revised psychometric to measure these variables as sub-tests of all general intelligence scales. (p. 83)

Nearly 30 years later, as a man in his early 80s, Wechsler (1979) advanced no further in his quest but continued to pursue it vigorously. At an American Psychological Association address, he reiterated the fact that factor analyses of his IQ tests did not explain nearly all of the variance and his belief that nonintellective factors accounted for the residual variance. But, Wechsler (1979) said,
That was easy, but there was a rub. Where were the four- or five-minute tasks that could be added to an already existing battery of ‘cognitive’ tests, that would correlate sufficiently with the full scale scores, and yet emerge as factorially different? I have been searching for them for the last thirty years. I tried unsuccessfully to find one for the 1955 WAIS. I think I may have been luckier in my latest search. (pp. 6–7)

Indeed, a Level of Aspiration test was standardized with the Wechsler Adult Intelligence Scale—Revised (WAIS-R; Wechsler, 1981), hopefully to be included in the battery. This subtest required adults to estimate how well they would perform in a simple task, then observe how well they actually performed, and, finally, to estimate future performance (Hargus & Wasserman, 1993). Wechsler believed that a comparison of the initial estimate with the second estimate (after getting feedback) would be clinically relevant and hoped the task would measure striving, confidence, and emotional response to the experimental situation.¹

But the Level of Aspiration test failed and was eliminated from the WAIS-R (Hargus & Wasserman, 1993). Wechsler died in 1981, the same year that the WAIS-R was published, and despaired about his failure to achieve his life-long quest to account for some of the residual variance. Without doubt, Wechsler would have applauded the efforts of Mayer et al. (1999) to develop and validate the MEIS and Roberts et al.’s (2001) high-quality cross-validation of the performance-based measure of EI. He would have criticized the need for consensus and expert scoring, because of their impracticality for a clinical measure of intelligence, and his publisher would also have criticized the low reliabilities for several of the EI tasks and the generally poor correlations between expert and consensus ratings. But Wechsler would have read the Roberts et al. article eagerly and would have tracked down all of the background references hungrily (if he had not already known about them), convinced that this line of inquiry showed great promise for explaining some of the residual variance.

Though Wechsler did not write specifically about the constructs of EI, he implicitly considered different aspects of EI as essential aspects of intelligence, most notably in the Comprehension and Picture Arrangement subtests. He refused to yield to the pressures of minority groups who criticized the Comprehension item, “What is the thing to do if a boy (girl) much smaller than yourself starts to fight with you?” The critics argued that in an urban ghetto no one can survive by walking away from any fight. Wechsler held firm, believing that the ability to show restraint, when restraint was called for, was an important aspect of intelligence. The ability to show restraint probably involves several of the branches of EI, including the management and regulation of emotion in oneself, the “highest level” in Mayer et al.’s (1999) hierarchy of EI skills. Other items, especially in Picture Arrangement, likewise called on EI for success, including the nonverbal appraisal of emotions, understanding emotions, and using emotions to facilitate thought. These emotions included the role of embarrassment, anger, fear, frustration, remorse, and joy.

Wechsler embraced the inclusion of items with clinical content in his test batteries, believing that they enhanced the more complete measurement of intelligent behavior and improved the value of the psychometric instrument as a clinical tool. Subsequent to Wechsler’s death, the publisher of revisions of his batteries yielded to political correctness and removed virtually all of the clinically charged items, the very ones that Wechsler believed would assess aspects of EI. A. S. Kaufman (1994) discussed his years of working with Wechsler in the early 1970s during the revision of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) and the standardization of the WISC-R (Wechsler, 1974). In this excerpt, which focuses on deleted or modified Picture Arrangement items, A. S. Kaufman (1994) described Wechsler’s probable reaction to the changes from the WISC-R to the WISC-III:

He loved to expose examinees to emotion-laden situations, to watch how they solved the problems, to listen to their spontaneous comments, to study their reactions to danger, to conflict, to authority, to violence. “Where’s the boxing match?” he would have stormed. “Replaced by a girl on a slide! And what happened to the burglar? And look what they did to the fire item! Instead of burning down his house, the kid’s a damned hero!” And he would have been incensed at the emasculation of the bench item. In the old item, “Some poor sap gets hit by the bench and then gets clobbered in a fight. Great stuff! But now they just kiss and make up.” (p. xiv)

It is interesting that outstanding clinicians, including those who are trained to administer intelligence tests, must have highly developed EI to be successful.

¹ The Level of Aspiration test is closer to measures of metacognition, or knowledge-monitoring ability, which has since been found to correlate with such performance criterion as undergraduate grade point average (Everson & Tobias, 1998).
As N. L. Kaufman (1980) noted: “Psychoeducational diagnosis involves sharp-eyed detective work. Each child presents a new mystery to be solved” (p. 398). As trainers of clinical test examiners know axiomatically, this kind of detective work requires abilities that accord well with Mayer et al.'s (1999) four branches of EI, notably the nonverbal appraisal and verbal expression of emotions both in the client and in the examiner; the use of emotion to facilitate action and self-monitoring; understanding and reasoning about emotions, especially the recognition and understanding of emotional problems; and the regulation of emotion in both the examiner and the client (e.g., to establish and maintain rapport). Expert clinicians, as well as members of other professions (e.g., playwrights, police detectives) for which success seems highly dependent on most of the EI branches, may provide good samples of subjects for future construct validation studies of the MEIS and other measures of EI.

And no one was as much a consummate clinician, replete with highly developed EI, as David Wechsler. As A. S. Kaufman (1994) stated when describing his weekly one-on-one meetings at Wechsler’s East Side Manhattan apartment to revise the WISC:

After the first meeting, he told me to put down my pen. He then talked for about 20 minutes, recounting his version of what had just transpired; he revealed every one of my ‘secret’ feelings and perceptions, unraveling in intricate detail my attitudes and emotions about every issue we discussed and each decision we made. (p. x)

However, Wechsler’s EI was not a profile of uniform strengths in the four branches. Wechsler may have had a significant weakness in Branch 4, the management and regulation of emotions. A. S. Kaufman (1994) continued:

From that point on, I never held back anything. He would usually respond calmly, but occasionally I’d strike a raw nerve, and his grandfatherly smile would evaporate. His temples would start to pulse, and his entire face and scalp would turn crimson... When I suggested... that he just had to get rid of the item, “Why should women and children be saved first in a shipwreck?” or incite the wrath of the new wave of militant feminists, his response was instant. With red face and pulsing head, he stood up, leaned on his desk with extended arms, and said as if he were firing a semiautomatic, “Chivalry may be dying. Chivalry may be dead. But it will not die on the WISC.” (p. x)

Contemporary Intelligence Tests

Roberts et al. (2001) compared scores on the MEIS with scores on the Armed Services Vocational Aptitude Battery (ASVAB). They found consistently significant—but low—correlations between the four MEIS subscales and the overall MEIS score and the five ASVAB scores (Mechanics, Administration, General, Electric, and Air Force Qualifying Test [AFQT], a measure of “g”). Roberts et al.’s findings were interesting and valid within the limitation of the criterion they used, but we propose that a truly complete validation of the MEIS would use intelligence tests that cover a broader spectrum of theories. Indeed, the ASVAB does not stem from any particular theory and is more of a hodgepodge of subtests than an organized array of abilities. Moreover, it is more of an achievement test than an intelligence or ability test, placing too much emphasis on reading, mathematics, specialized knowledge (General Science), and idiosyncratic knowledge (Electrical Information). Despite the virtues of the ASVAB outlined by Roberts et al., including its relatively high g loadings (Ree & Earles, 1990), the authors greatly overstated its value as a criterion for evaluating EI, especially from the perspective of current theoretical approaches to the measurement of intelligence.

Roberts et al. (2001) referenced Horn’s (1989) theory, which has been quite influential in providing the framework for several contemporary intelligence tests, notably the Stanford–Binet: Fourth edition, Woodcock–Johnson—Revised Tests of Cognitive Ability (WJ-R), Woodcock–Johnson—Third Edition Tests of Cognitive Ability (WJ III; Woodcock, McGrew, & Mather, 2001), and Kaufman Adolescent and Adult Intelligence Test (KAIT; A. S. Kaufman & Kaufman, 1993). In fact, Roberts et al. and Davies, Stankov, and Roberts (1998) argued that EI is part of Gc, or crystallized intelligence. Their analysis and arguments, however, are based primarily on the initial Gf–Gc theory of fluid and crystallized intelligence, and not on Horn’s expanded model of 8–10 abilities (Horn & Noll, 1997). It is the refined and expanded model that provides the basis of the WJ-R; furthermore, the merger of the Horn–Cattell and Carroll (1993) theories into the Cattell–Horn–Carroll (CHC) theory has provided the framework for the WJ III as well as the basis for state-of-the-art interpretation of virtually all intelligence tests, including Wechsler’s measures and newer tests such as the KAIT (Flanagan, McGrew, & Ortiz, 2000).

In the expanded Horn and CHC models, it becomes less clear which abilities would be linked to EI. Certainly, there would still be a significant link between the more specific definition of Gc in the revised model and EI scores on the MEIS. The vocabulary
needed to solve MEIS problems certainly requires Gc abilities—subtests like Complex Blends and Progres-
sions are intermixed with one’s knowledge of the En-
glish language. But visuospatial thinking, or Gv, and auditory thinking, or Ga, would also be relevant abili-
ties. The ability to analyze both visual and auditory stimuli would extend to the ability to analyze faces, designs, or music for various emotions. Similarly, long-term retrieval (Glr) would seem to be needed for understanding and regulating emotions, and fluid reason-
ing as it is defined in the expanded Horn model may underlie reasoning about emotions and regulating emotions in others. Although tests of Gc are easy to find (subtests such as Vocabulary and Comprehen-
sion), and tests of Gv are likewise plentiful on Wechs-
sler’s Performance scales (Flanagan et al., 2000), measures of Ga, Glr, and Gf are less obvious. Yet such measures are readily found in the WJ III (Ga, Glr, Gf) and KAIT (Glr) and would be useful measures for further validating the MEIS.

A different theory that has not been studied in re-
lation to EI is the Planning, Attention, Simultaneous, and Successive theory (PASS), based on Luria’s (1966) neuropsychological model that focuses on the brain’s supposed three functional units. Planning involves determining, choosing, applying, and evaluating problem solutions. Attention is the ability to selectively attend to specific stimuli. Simultaneous processing requires the integration of stimuli into a larger whole to solve problems, and successive process-
ing is a mental process by which the individual integrates stimuli into a specific serial order that forms a chain-like progression (Das, Naglieri, & Kirby, 1994). The PASS theory is the basis of Na-
glieri and Das’s (1997) cognitive assessment system (CAS).

One obvious connection of the PASS theory to EI is that both Planning and Attention abilities have been theoretically linked to creative performance (see Naglieri & Kaufman, 2001, for an overview). In the ini-
tial conception of EI, the “Utilizing Emotional Intelli-
gence” factor included creative thinking as one of its four components (Salovey & Mayer, 1990). The cur-
rent model uses creativity in Branch 2; one aspect of this branch would be to use one’s emotions to think more creatively (Salovey, Woolery, & Mayer, 2001).

Simultaneous processing abilities, however, may be the most conceptually linked to measures of EI, par-
ticularly to the “Understanding Emotions” factor. In the MEIS, Understanding Emotions includes such tasks as analyzing how two emotions may blend into a more complex one. This skill involves processing two emotional concepts simultaneously—perhaps not a very different ability from being able to identify a figure that is embedded in a more complex figure, as required in the Figure Memory subtest of the CAS. In addition, simultaneous processing is also undoubtedly involved to a considerable extent in the nonverbal appraisal of emotions in others. Like the CAS, the Kaufman Assessment Battery for Children (K-ABC; A. S. Kaufman & Kaufman, 1983) is based on Luria’s (1966) neuropsychological theory and also provides a Simultaneous Processing scale that is conceivably related to the understanding of emotions.

Although none of the major clinical tests of intel-
ligence include measures of EI, it is noteworthy that A. S. Kaufman and N. L. Kaufman (1983) clearly considered certain aspects of EI, particularly Branch 1, to be aspects of general intelligence. While devel-
oping the K-ABC, the authors piloted an experimental subtest that required children to identify the emotions depicted in photos of men and women. In addition, the authors administered a Concept Formation subtest during a national tryout in which several items in-
volved emotions in people and animals (e.g., happy faces, angry faces; A. S. Kaufman, Kaufman, Kam-
phaus, & Naglieri, 1982). Finally, in the development of the KAIT, a Fluid subtest called Story Completion required the examinee to interpret emotions—and to some extent, reason with emotions and use emotions to facilitate action—to solve the picture-sequencing problems. (The items depicted real-life situations on emotional topics such as an adolescent being caught stealing money from a parent’s wallet.) Although none of these subtests were included in the published versions of the tests, the KAIT Auditory Comprehen-
sion subtest measures verbal perception and expres-
sion of emotions in several items that require exam-
ines to identify the mood or tone of a news story or the likely emotional responses of people in the stories.

EI also plays a role in WJ III interpretation, spe-
cifically in the Cognitive and Achievement Per-
mance models (CA-PM), a set of frameworks rooted in Snow’s (1989) research that can be used to interpret WJ III cognitive and achievement test scores (Mc-
grew, Ford, & Woodcock, 2001). The CA-PM are based more on theory than data but play an integral part in the interpretive process. According to the CA-
PM, there are several important factors that can be either facilitators or inhibitors of performance on cog-
nitive tasks; three of these primary facilitators—inhibitors are motivation, cognitive style, and emo-
tional state (McGrew et al., 2001).
Summary

EI has its roots in Binet’s notion of intelligence, and, especially, in Wechsler’s transformation of psychometric intelligence testing into the field of clinical assessment. Today’s clinical assessment tools, like Wechsler’s tests that trace back to World War I, have never embraced EI as a formal construct but have included measures of EI and aspects of EI in their development and interpretation. The occasional EI items in conventional tests, however, have barely skimmed the surface of the intricate and well-thought-out branches of EI tests and theory (Mayer et al., 1999). Nonetheless, the EI researchers would do well to incorporate some of the theories associated with contemporary measures of intelligence, and the tasks that are included in these new measures, to better understand their EI constructs; for example, Horn’s Ga, Gv, Gf, and Glr abilities all may relate meaningfully to some branches of EI, as may the Attention, Planning, and Simultaneous components of the Luria (1966)-inspired PASS model.

From a clinical assessment perspective, the EI tasks in the MEIS generally lack enough reliability to provide meaningful profile interpretation, and the consensus- and expert-scoring systems are impractical for clinical assessment. From a research and validity perspective, however, these kinds of scoring systems seem to be an essential aspect of EI measurement. We support Roberts et al.’s (2001) suggestion of developing a “hybrid scoring protocol” (p. 227), and, indeed, the forthcoming Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2001) uses such expert consensus (Salovey, Mayer, Caruso, & Lopes, in press). The use of expert consensus is a common technique in creativity research. Amabile and her colleagues (see Amabile, 1996, for a review of her work) have refined a consensus assessment technique that has consistently provided reliable creativity scores. This approach involves soliciting raters who have significant experience in the domain in question, and then having these raters make their judgments independently to avoid being influenced by other raters. These judgments compare different products against each other, not against an absolute ideal, and the products are judged in a random order (Amabile, 1996). Although this technique is at present mostly applied to the evaluation of products, it could also be applied to performance measures.

Overall, we applaud the efforts of the EI test developers and theorists (Mayer et al., 1999) for making important contributions both to theory and assessment. With the caveats noted in this article, we agree with the general conclusions reached by Roberts et al. (2001) about the state of the art. We look forward to the research on the new MSCEIT that will undoubtedly be equally as rigorous and hope that Roberts et al. will conduct a similar study with this instrument, incorporating some additional IQ measures for further insight. The future of EI assessment is surely in the continued refinement of these performance-based measures and not in self-report inventories, at least if the goal is to incorporate EI into the broader domain of intelligence.

References


