Defining Mental Retardation: A Matter of Life or Death

William Lichten and Elliott W. Simon

Abstract

Because persons with mental retardation cannot be executed for murder, the diagnosis becomes a life and death matter. The American Association on Mental Retardation (now the American Association on Intellectual and Developmental Disabilities) and other associations agree that IQ alone is an insufficient criterion and adaptive functioning also needs to be considered. However, there is no satisfactory quantitative measure including both IQ and adaptive functioning. We propose a solution by defining a total quotient (TQ) scale, a composite of both IQ and standardized adaptive functioning scores. We estimate the margin of error in such scores (IQ, or adaptive functioning, or TQ) is 10 points, four times the usual one *SD* value given by intelligence test developers. The procedure here renders moot the distinction between convergent and divergent validity.

A recent decision (*Atkins v. Virginia,* 2002), in which the Supreme Court ruled against the execution of convicted murderers with mental retardation, has literally raised the stakes of psychological diagnosis to a matter of life and death. In capital cases, the courts have to answer definite yes–no questions. Is the defendant guilty or innocent? If guilty, do we execute or not? In the case of another Supreme Court decision (*Roper v. Simmons,* 2005) not to execute murderers who are children, the latter question is straightforwardly decided by a birth certificate. The Atkins decision, that persons with mental retardation cannot be executed, greatly raises the stakes of the question, "Who has mental retardation?"

Determining a diagnosis of mental retardation is not as simple as reading a birth certificate, and various diagnostic criteria have been proposed. The mental retardation community has accepted the criteria of intelligence quotient (IQ) and adaptive behavior scores (Luckasson et al., 2002), but it has not found a satisfactory method of combining the two. In this paper, we propose a total quotient (TQ) that combines both IQ and adaptive behavior scores. As defined, TQ does not affect the percentage of persons diagnosed with mental retardation. Therefore, it would neither increase nor decrease

the number of defendants excused from the death penalty. To the extent that the use of adaptive behavior scores (and TQ) improves the diagnosis of mental retardation, it would make the court's decisions more accurate and, thus, fairer.

Ever since Goddard (1912) introduced IQ to the mental retardation field, IQ has played an important role in the definition and diagnosis of mental retardation (Trent, 1994). Under the controversial death penalty (American Friends Service Committee, 1998), IQ is now being used to decide the fate of convicted murderers. Given that, the question of "How do mental retardation diagnoses hinge on IQs?" takes on greater significance.

The use of IQs in the definition of mental retardation follows a similar course to that of a professor who makes up a histogram of exam scores and looks for breaks in the distribution to make it easy to separate the *A, B,* and *C* students. As luck usually has it, however, no such natural breaks occur. The professor then has the harder task of arbitrarily assigning scores to divide the different letter grades from each other. The IQ distribution usually falls on the well-known, bell-shaped curve with no convenient gaps between persons with and without mental retardation. Thus, the IQ cutoff score is arbitrary. An IQ of 70, two standard deviations (*SD*s)

below the mean, is now used by definition, with the predetermined outcome that about 2.5% of the population is classed as within the range of mental retardation. However, at one time the cutoff point was set at one *SD*, with the outcome that 15% of the population fell within the range of mental retardation!

As the court, as well as psychiatric, psychological, and educational associations, have pointed out, high stakes decisions never should be made on the basis of a single test score. Nevertheless, intelligence tests are the objective measures of mental ability and carry heavy weight. The main question has been to what extent other criteria should be used to diagnose mental retardation.

An important step was taken by Heber (1961), who introduced adaptive behavior as a definitional component of mental retardation. Although this criterion was accepted by the mental retardation community, it had two drawbacks (Silverstein, 1973; Zigler, Balla, & Hodapp, 1984; for a dissent, see Barnett, 1986). The drawback is connected with an old dilemma in mental testing, the divergent–convergent validity criterion. When one evaluates a new standardized measure, such as an intelligence test, a high correlation with accepted scales may be viewed as a positive attribute (i.e., the new scale has convergent validity). However, a low correlation with accepted measures (divergent validity) may also be viewed as a positive attribute because it would be testing something new that was not probed by existing tests. It is not possible to have both types of validity. A procedure that makes moot the convergent–divergent validity distinction will be discussed later in this paper.

The popular image of the impractical genius and the absent-minded professor remind us that IQ may be a good predictor of academic achievement, but that it alone may not predict success in other aspects of life (Aronson, 1995; Gardner, 1983; Sternberg, 1985). It has long been known that "intelligence tests as now constituted measure effectively only a portion of and not all of the capacities entering into intelligent behavior" (Wechsler, 1943). Similarly, for the purpose of mental retardation definition and diagnosis, tests of adaptive behavior appear to have divergent validity when compared with IQ. For example, the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984, 2005) have a correlation of .2 to .3 with IQ and other mental ability tests, with an overlap variance of less than 10%.

This leads to the second drawback of introducing adaptive functioning into the definition of mental retardation. When *both* IQ and adaptive behavior are imposed as joint definitional criteria for mental retardation, the net effect is to narrow down drastically the number of persons diagnosed with mental retardation.

Silverstein (1973) worked out the math for this case. If the two *SD* rule were applied to IQ, about 2.5% of the United States population, or about 7.5 million persons, would meet this definitional criterion of mental retardation. If the same two *SD* rule were also applied to adaptive behavior scales (assuming a correlation of .2 between the two scales), the number of persons with mental retardation would fall to 411,000, less than one tenth of the previous value. Furthermore, this value would depend strongly on the exact value of the correlation between the two criteria (Silverstein, 1973). Thus, the proposal (Luckasson et al., 2002) to expand the definition of mental retardation to include both intellectual functioning and social adaptation has caused a dilemma. One way out of this perplexing problem is presented later in this paper.

One problem with the adaptive behavior criterion was that until recently it had not won consistent acceptance among mental retardation professionals (Smith & Polloway, 1979). Given current definitional requirements, a psychological evaluation solely measuring IQ should not be accepted by any funding agency to qualify an individual for service. Discussions of professionals about an individual with mental retardation today invariably contain a consideration of the person's adaptive functioning.

Heber (1961) defined *mental retardation* as significant subaverage general intellectual functioning originating during the developmental period that is associated with adaptive behavior impairments. Subaverage intellectual functioning was further defined as one *SD* below the mean of a standardized intelligence test (IQ = 85), adaptive behavior as adaptation to environmental demands and the developmental period was viewed as ending at age 16 years.

A brief historical survey of the modern definitions of mental retardation reveals that although the three hallmarks of Heber's (1961) definition have remained in place, succeeding definitions have greatly changed his rules of application. These changes resulted in major differences in the prevalence of mental retardation as well as in whether a

specific individual, at a single point in time, would meet the mental retardation criterion. As noted above, Grossman (1973) greatly reduced the prevalence of mental retardation by increasing the necessary number of *SD*s below the mean from one to two needed to meet the IQ criterion. He also lengthened the developmental period to 18 years and reworked the adaptive behavior definition to include those abilities needed to meet age and cultural standards of independence and social responsibility.

Grossman's (1977) revision again changed the IQ criterion to allow for individuals with significant adaptive behavior needs to meet the mental retardation criteria if their IQs were as high as 80. Luckasson et al. (1992) revised the concept of mental retardation and specified 10 areas of adaptive behavior (Communication, Self-Care, Home Living, Social Skills, Community Use, Self-Direction, Health and Safety, Functional Academics, Leisure, and Work). To meet mental retardation criteria, deficits in any 2 of these 10 areas were necessary along with an IQ of 70–75 that manifested before the age of 18 years. The 1992 definition also changed the view of mental retardation from a state of "incompetence" to a pattern of support needs in various life activities and domains. (For a history of adaptive behavior, see Schalock and Braddock, 1999.)

Luckasson et al. (2002) redefined the adaptive behavior criterion by dropping the 10 previous areas in favor of three broader life skills domains (Conceptual, Social, and Practical). The supportive orientation of the 1992 definition is strengthened and expanded upon; IQ and developmental period criterion remain the same, but the added proviso of *two SD*s below the norm is now also applied to the adaptive behavior criterion. (Unfortunately, as discussed in an earlier paragraph here, this proviso reduces prevalence by more than a factor of 10. We rectify this unacceptable result in this paper.)

Professional societies and federal law continue to define *mental retardation* using Heber's (1961) three criteria (American Psychiatric Association, 2000; Hawkins-Shepard, 1994; Luckasson, 2002). Criterion A is based on intellectual functioning, currently an IQ below 70 found by means of one or more standardized tests, such as the Wechsler, Stanford-Binet, or Kaufman batteries.

Criterion B is "significant limitations in adaptive functioning" as determined by teacher evaluation; educational, developmental, and medical histories; and/or standardized adaptive behavior scales.

In the American Association on Intellectual and Developmental Disabilities—AAIDD (formerly the American Association on Mental Retardation) current view, Luckasson et al. (2002) stated:

> For the diagnosis of mental retardation, significant limitations in adaptive behavior should be established through the use of standardized measures normed on the general population, including people with disabilities and people without disabilities. On these standardized measures, significant limitations in adaptive behavior are operationally defined as performance that is at least two standard deviations below the mean of either (a) one of the following three types of adaptive behavior: conceptual, social and practical skills. (p. 76)

Criterion C sets the onset of both the IQ and adaptive behavior criteria as having to occur before 18 years of age. This criterion is consistent across both the American Psychiatric Association's (2000) *Diagnostic and Statistical Manual of Mental Disorders*—DSM and the AAIDD (Luckasson, 2002). The federal definition of a developmental disability, important because it determines service eligibility and is public law, selects age 22 years as the developmental cutoff (Developmental Disabilities Assistance and Bill of Rights Act, 2000).

There is no question that IQ will continue to be important for mental retardation diagnoses and, therefore, for death sentence decisions in capital cases. The reason is that intelligence test scores are objective, reliable, psychometrically sound, and easily used.

A crucial question in using an IQ as a pass–fail criterion is the possible error of that number. In all sciences, error estimates are notoriously fallible (Lichten, 1999). History has shown that errors are typically underestimated because investigators have left out unsuspected sources of inaccuracy. The mathematical Appendix to this paper details that Luckasson et al.'s (2002) quoted error of 5 points (presumably a margin of error, double of manufacturers' values) is only half of the present value. The AAIDD value only considers test–retest reliability with the exclusion of other sources of error.

Under the No Child Left Behind Act of Congress (2002), achievement test scores for a school are serious matters; IQ and closely related SAT and ACT tests (Frey & Detterman, 2004) are critical factors in educational assignment and admission decisions, which can have major lifelong outcomes.

The death penalty and assignment to special education involve much the same issues. Court records on special education have been totally inconsistent. In the words of Scarr in Elliott (1987),

No more troubling examples of the inadequate union of law and social sciences can be found . . . two reputable judges can reach opposite conclusions [on use of IQs to make special education assignments] from essentially the same evidence. (p. v)

Such blatant inconsistencies in court decisions in the use of IQ to diagnose mental retardation, and thus to decide on the death penalty, have seldom occurred (Virginia and Texas are exceptions; see Perske, 2005). Thus, we may expect IQ to remain the mainstay of such decisions.

A decision or characterization that will have a major impact on a student should not be made on the basis of a single test score. Other relevant information should also be taken into account (American Educational Research Association, 1999, Standard 13.7). This statement, made for education, is equally valid for other important decisions (*Atkins v. Virginia*, 2002). It is true that intelligence tests are psychometrically sound, within the limits that we have evaluated in the mathematical Appendix.

Nevertheless, regardless of the reliability of the IQ, its validity is limited. It is fair to say that the consensus of the mental retardation community is that a life and death decision should involve more than a single score. Thus, we turn to other criteria.

Historically, compared to IQ, the measurement of adaptive behavior has been vague, less precise, and unspecific. However, recently Luckasson et al. (2002) emphasized standardized, adaptive behavior scales, which use the same measuring stick as IQ, with a mean of 100, an *SD* of 15, and a two *SD* cutoff score for mental retardation. Adopting these measures will be a major step in the direction of making adaptive behavior measures quantitative. However, as already discussed, simply piggybacking such a score at a value of 70 has been criticized for its major, unrealistic effects on the prevalence of mental retardation (Silverstein, 1973; Zigler, Balla, & Hodapp, 1984).

If a court is using a psychological quantity to decide capital punishment, that quantity should have a stable meaning that does not change from one year to another. This is generally true for intelligence test scores. Although there is some drift (Flynn, 1984), it is slow, easily correctable, and covered by the error estimates for IQ (see Appendix). However, even if the definition of mental retardation relies on both IQ and adaptive functioning scales, there is at present no acceptable way of combining the two scores (Silverstein, 1973; Zigler et al., 1984). In view of the importance of mental retardation, these are serious lacunae in definitions.

It is important to note the consensus among the mental retardation community: Adaptive behavior should be included in the definition and diagnosis of mental retardation. The problem at hand is, then, not whether to consider adaptive functioning in diagnosing mental retardation; rather, it is *how* to include it. In the latest adaptive behavior criterion, Luckasson et al. (2002) failed to specify how adaptive functioning is to be combined with IQ without drastic changes in the prevalence of mental retardation. That is our task in the present paper.

However, before we start, we point out an inescapable implication that follows from any mental retardation definition that includes adaptive behavior: *IQ can no longer be the king of the hill*. For example, it is well-known that persons with IQs less than 70–75 and adaptive behavior abilities that are less than *two SD*s below the mean often get and hold jobs, raise families, etc., and no longer should be considered appropriate for a diagnosis of mental retardation. (For a numerical example, see a later discussion.)

Method

The task at hand is to keep the advantages of IQ, which is a single, standardized scale that can be easily interpreted by specialists and nonspecialists alike. However, this scale should include both IQ and adaptive behavior. We name this scale TQ (for total quotient = IQ + adaptive behavior score); TQ combines two well-standardized measures: cognitive (such as an IQ obtained on the Wechsler scales) and adaptive behavior (such as the Adaptive Behavior Composite on the Vineland scales). The simplest way to combine these two scales is to take the average of both. In doing this, we follow the example of mental testers, who routinely combine subtest scores. We do this to preserve the deviation definition (*M* = 100, *SD* = 15), which both the IQ and adaptive functioning scales follow. In addition to simply adding the two scores, we adjust the resulting scale to keep the deviation definition (*SD* = 15), which is universally used today. (Mathematical details are shown in the Appendix.) On the new, composite TQ scale, mental retardation is diagnosed with a TQ of 70 or less. This approach has the advantage of not changing the percentage of the total predicted population of persons with

mental retardation (approximately 2.5%) nor does it require changing the accepted severity subclassifications of mental retardation.

Note that this procedure amounts to giving IQ and the adaptive behavior score equal weights. We have considered choosing unequal weights but have yet to find a satisfactory alternative. In fact, IQ and achievement testers, such as Terman and Wechsler, almost invariably formed composites from equally weighted subtests. For example, consider the Wechsler intelligence tests. Coding (a speed subtest involving rapid copying of numbers) correlates poorly with more intellectual tasks such as vocabulary. Yet Wechsler gave all equal weight. To try to do otherwise, such as giving vocabulary twice the weight of coding, leads to questions that psychometricians found unanswerable.

To further address the question of unequal weights, we compared the structure of IQ (such as Wechsler) and adaptive (Vineland Adaptive Behavior) scales, and we found both to have comparable psychometric quality. Both have such similar correlations and factor structures that it is not possible to distinguish one from the other statistically.

To illustrate, we compare the fraction of total variance in IQ and adaptive functioning measures due to the g-factor. The g^2 for the Stanford Binet 5 (11 to 16-11 years) is .54, for the WISC (6 to 16-11 years) is .43, and for the Vineland (7 to 13 years) is .55. Both intelligence tests are stable from one edition to another.

Likewise, the fact that adaptive behavior scales have a low correlation with IQ begs the question as to which score deserves greater weight. In conclusion, in lieu of a better solution, we adopt equal weights, but do not rule out other choices. If AAIDD could reach a consensus as to a specific alternative (which we consider unlikely), it would be a simple matter to adjust the TQ scale to reflect the modified weighting.

Results

Table 1 presents the results of the computation of the TQ scale. The Appendix presents the mathematical details.

Discussion

In looking at the TQ scale as it applies to the diagnosis of mental retardation, one might ask the questions, "Does it make sense?" particularly for a scale that equally weights adaptive functioning and IQ, "How do the new results compare with past experience in diagnosing mental retardation?" To answer these questions, we select a few examples that probe the effect of adaptive functioning.

Examples of TQ

Consider a person with an IQ of 70 and an adaptive behavior score of 70. Table 1 indicates that the sum of subscores would be 70 + 70 = 140. With an r of .2, this person would have a TQ of

Table 1 Combination of Two IQ-Like Subscores Into a Composite Score as a Function of r

Sum of subscores	Composite score as a function of r						
	0	0.1	**0.2**	0.3	0.5	0.7	1.0
60	1.0	5.6	**9.6**	13.2	19.2	24.1	30
80	15.1	19.1	**22.5**	25.6	30.7	35.0	40
100	29.3	32.6	**35.5**	38.0	42.3	45.8	50
120	43.4	46.1	**48.4**	50.4	53.8	56.6	60
140	57.6	59.5	**61.3**	62.8	65.4	67.5	70
160	71.7	73.0	**74.2**	75.2	76.9	78.3	80
180	85.9	86.5	**87.1**	87.6	88.5	89.2	90
200	100	100	**100**	100	100	100	100
220	114.1	113.5	**112.9**	112.4	111.5	110.8	110
240	128.3	127.0	**125.8**	124.8	123.1	121.7	120
260	142.4	140.5	**138.7**	137.2	134.6	132.5	130

Note. Mean = 100 and *SD* = 15. The general expression for combining *N* IQ-like subscores is $IQ = 100 + \sum_{i=1}^{N} IQ_i - 100N / \sqrt{\sum_{i,j=1}^{N} r_{ij}}$, where $r_{ii} = 1$.

61, which would be within the range of mild mental retardation, but further under the "magic" 70 cutoff.

Consider next a person with a TQ of 70, but with an adaptive functioning score of 100. We then work backwards in Table 1 to find the sum of subscores. We interpolate between 61.3 and 74.2 in the $r = .2$ column, which gives a sum of subscores of $140 + [(70 - 61.3)/(74.2 - 61.3)] \times (160 - 140) = 153.4 = \text{IQ} + 100$, which gives an IQ of 53. This case illustrates the possibility of a person who would be diagnosed on the basis of IQ alone as having mental retardation yet is able to function at the "average" level of adaptive functioning.

Next, consider a person with an IQ of 100. What adaptive score would result in a TQ that would meet the mental retardation criterion? The math is the same as the previous example, except that IQ and adaptive functioning scores are switched. It would take an adaptive functioning score of 53 or less to produce a TQ of 70 or less. An adaptive score of 53, based on the Vineland norms, would correspond to the overall adaptive skills of an "average" 3-year-old. Such a person would likely be in need of significant supports.

It is instructive to compare how TQ "handles" these two cases with values of 53 and 100 for IQ and adaptive functioning or vice versa. Both give a TQ value of 70, the cutoff for mental retardation. In such cases clinical judgment plays a major role and the diagnoses could conceivably differ: The person with lower IQ would be capable of living without intervention while the person with higher IQ would in all likelihood require support. However, in both examples, a diagnosis on a strictly cognitive or adaptive criterion would be totally misleading. Parenthetically, note that the overall prevalence of mental retardation is the same in both sets of diagnoses, which is built into the TQ scale.

Likewise, other combinations of IQ and the adaptive behavior score easily illustrate the overall statistical distribution of mental retardation that would result from a TQ approach that gives IQ and adaptive behavior equal weight. For example an individual with an IQ of 52 and an adaptive behavior score of 100 would meet the TQ criterion for mental retardation. What is attractive about TQ, compared to other possible approaches, such as requiring an individual to meet both criteria simultaneously, is that a TQ approach keeps the predicted statistical distribution at approximately 2.5% of the population.

Of course, no mechanically applied measure can be trusted to substitute for informed clinical judgment. Nevertheless, it is apparent that TQ, by building on both cognitive and adaptive skills, gives a closer approximation to an acceptable clinical finding as defined by Luckasson et al.'s (2002) criteria. In the final example above, examination of the person's support needs might lead to a conclusion that no supports are needed. Given that, the individual would not meet the Luckasson et al. (2002) definition of mental retardation. It is also conceivable that the individual with the reverse scores (adaptive behavior = 52, IQ = 100) does have significant support needs and would, therefore, meet Luckasson et al.'s (2002) definitional requirements.

In the last analysis, the courts will decide what evidence they will accept and how they will interpret the data. Generally, courts will use a measurement, such as IQ, at face value if it is decisive. In borderline cases, clinicians could draw the attention of the courts to the meaning of the errors and to other considerations.

In conclusion, it would be simpler if courts' decisions hinged on a single test score, but as the Supreme Court, psychological and educational research organizations, and federal law have said, justice would not be done (Ellis, 2002). As the AAIDD now specifically includes the use of standardized adaptive functioning scales with a two SD criterion in its definition of mental retardation (Luckasson et al., 2002), it should now also be possible and preferable to use TQ for the determination of mental retardation.

A Remark on Single Scores

One could argue that TQ is also a single number. What is the difference between it and IQ? One answer is that any indicator, whether a single number or not, is superior if it shows divergent validity. Thus, the Wechsler Performance Scale, which includes the average of five subtest scores, is superior to a single subtest scale, such as Coding. Likewise, the Full Scale score is preferable to either the Verbal or Performance score.

All definitions of mental retardation include detailed instructions in the appropriate procedures for administering standardized tests, taking into account cultural, communication, psychiatric, and sensory status. We do not mean to suggest a blind application of the joint TQ criteria without taking into account other psychological/sociological factors. Total Quotient also does not obviate the use

of sound clinical judgment by trained professionals in such cases as depression, physical and sensory challenges, or the inappropriate use of standardized tests, which might result in mental retardation misdiagnoses. As is now the case, trained clinicians need to differentiate mental retardation from other instances in which the mental retardation criteria may be met through spurious means.

Diagnosis Before the Age of 18

One reason for the courts to accept this criterion is that it objectifies the diagnosis of adaptive functioning and avoids attempts to dredge up last-minute evidence. This is well and good but, conversely, can we expect the courts to try to analyze historical teacher reports? Furthermore, it does not solve the problem of a lack of records, which often are missing or hard to trace.

If the courts accept both the IQ and adaptive behavior criteria, it may require an expert to review historical information and give an opinion as to age of onset. The court (not the attorneys) would need to hire a qualified clinician to critically examine the accused and his or her records, when and if such information exists.

Implications for Social Policy

Once again a Virginia court has ruled that a convicted murderer did not meet the criteria for mental retardation and sentenced him to death. He was none other than Daryl Atkins, whose life was previously spared by the Supreme Court's 2002 decision. Perske (2005) questioned the judgment of the prosecution's expert witness for once more dismissing the defendant's low IQ and once again basing his recommendation for the death penalty on what Perske considered an erroneous evaluation of adaptive functioning. The TQ, which is advocated here, based on standardized scores for both IQ and adaptive functioning, would leave less leeway for subjective judgment in such a case.

Psychometrics

Table 2 provides error estimates for an individual's IQ based on a single test. The margin of error (5% level, two SDs) is 10 points, four times those in test manuals. For the adaptive behavior and composite scales, or for any combination of these scores including TQ, the errors are unavoidably the same, no matter how important the decision. In the interests of accuracy, the courts should allow for this margin of error in making decisions.

Table 2 Estimates of Errors for a Person Taking a Single Test

Error source	Error
Reliability[a]	2–3
Flynn Effect[b]	2–3
Stanford Binet–Wechsler differences[c]	3–4
Rms[d] sum (SD) of all effects	5
Margin of error[e]	10

Note. SD unless otherwise stated. Obtained from differences between various Stanford-Binet and Wechsler Tests.
[a]Test–retest differences. [b]For typical age of test version. [c]From Table 3. [d]Root mean square. [e]2 SDs.

Intelligence Quotient, Crime, and the Courts

Although persons with low IQs have decreased analytic ability, they can at the same time have high adaptive abilities or "street smarts." It is also peculiar that failure or a low score on an intelligence test could be life-saving. Given this, in court situations the temptation to malinger on an intelligence test is strong. If the stakes are life and death, the temptation becomes irresistible (Esposito, 2004). Although it is difficult to raise IQ, one could be coached on how to fail the test and, thus to, live. Although malingering can be detected, because the stakes are high, one expects more expert coaching. Finally, one's attorney could use the Fifth Amendment, "No person . . . shall be compelled in any criminal case to be a witness against himself," which may apply to intelligence tests used to decide on lethal punishment. How could you be compelled to take a test that could sentence you to the chair? The test might be worth the chance, because taking it and flunking could save your life! Furthermore, the Supreme Court in *Roper v. Simmons* (2005) ruled out the execution of persons who committed murder as minors (persons less than 18 years of age). Thus, the court in capital cases would only be concerned with adult test scores, which change very little in the aggregate after age 18. Such changes are small and inconsequential.

Remarks on Intelligence Research

One of our major purposes in this article is to furnish a diagnostic approach to mental retardation that both combines IQ and adaptive functioning and uses only standardized measures. However, Garner (1972) pointed out that applied research results

sometimes could feed back on basic laws. For the sake of completeness, we note that a similar TQ scale could be obtained by combining IQ with other psychological scales, such as tests of emotional, social, and practical intelligence. The only requirement of these other scales is that of standardization, as is the case of adaptive behavior scales.

Furthermore, we note that the proposed method of combining scores (which is commonplace in psychological test construction) renders moot the distinction between convergent and divergent validity in this context. In particular, the drawback of combining IQ and adaptive functioning is nonexistent.

Thus, the TQ scale, or similar scales formed by combining IQ with other psychological scales, could improve selection of persons for special education, college admission, jobs, etc., for any application for which intelligence tests are presently used. The advantage of such a composite scale would be that it would respond to broader measures of ability other than purely cognitive skills. It might then address the desirability, mentioned at the beginning of this paper, for extending the concept of intelligence.

Summary

It is now the consensus that both IQ and adaptive behavior measures are needed in diagnosing mental retardation. However, at present there is no generally accepted, single unambiguous, combinatory measure of IQ and adaptive behavior that preserves a predicted mental retardation prevalence of approximately 2.5% in the general population. Furthermore, we find IQ error estimates to be inadequate. We have alleviated these shortcomings by combining standardized IQ and adaptive behavior scales into a new TQ scale with an appropriate normalization ($SD = 15$) and a margin of error (5% significance) of 10 points. In current American law, a diagnosis of mental retardation precludes the death sentence for convicted murderers. In the past, courts used IQ because it measured mental ability in a simple, easily understood number. Our new measure clearly would affect such decisions. We note in passing that Garner's (1972) symbiotic relationship between basic and applied knowledge implies our results could be useful in other areas of intelligence research.

References

American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.

American Friends Service Committee. (1998). *The death penalty: The religious community calls for abolition*. Philadelphia: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Arlington, VA: Author.

Aronson, E. (1995). *The social animal* (7th ed.). New York: Freeman.

Atkins v. Virginia. No. 00-8452 U.S. (2002).

Barnett, W. S. (1986). Definition and classification of mental retardation: A reply to Zigler, Balla, and Hodapp. *American Journal of Mental Deficiency, 91*, 111–116.

Developmental Disabilities Assistance and Bill of Rights Act of 2000. Pub. L. No. 106-402, 114 Stat. 1678.

Doppelt, J. E., & Kaufman, A. S. (1977). Estimation of the differences between WISC-R and WISC IQs. *Educational and Psychological Measurement, 37*, 417–424.

Elliott, R. (1987). *Litigating intelligence. IQ tests, special education, and social science in the classroom*. Dover, MA: Auburn House.

Ellis, J. (2002). Mental retardation and the death penalty: A guide to state legislative issues. http://www.aamr.org/Reading_Room/pdf/statelegislatures_guide.pdf

Esposito, S. (2004, April 26). Defense attorneys say expert changed killer's I.Q. test. *Chicago Sun Times*, p. 9.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Frey, M., & Detterman, D. K. (2004). Scholastic assessment or g? The relation between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*, 373–378.

Gardner, H. (1983). *Frames of mind. The theory of multiple intelligences*. New York: Basic.

Garner, W. R. (1972). The acquisition and appli-

cation of knowledge: A symbiotic relation. *American Psychologist, 27,* 941–946.

Goddard, H. H. (1912). *The Kallikak family. A study in the heredity of feeblemindedness.* New York: MacMillan.

Grossman, H. J. (Ed.). (1973). *Manual on terminology in mental retardation* (rev.). Washington, DC: American Association on Mental Deficiency.

Grossman, H. J. (Ed.). (1977). *Manual on terminology in mental retardation* (Rev.). Washington, DC: American Association on Mental Deficiency.

Hawkins-Shepard, C. (1994). Mental retardation. ERIC Digest E 528, ED 372593 1994-09-00. http://www.ericfacility.net/ericdigests/

Heber, R. (1961). A manual on terminology and classification in mental retardation. *American Journal of Mental Deficiency, 65,* Monograph Suppl.

Kanaya, T., Ceci, S. J., & Scullin, M. H. (2003). The rise and fall of IQ in special education. *Journal of School Psychology, 41,* 453–465.

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies. *American Psychologist, 58,* 778–790.

Lichten, W. (1999). *Data and error analysis* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Lichten, W., & Simon, E. W. (2006, May). *Beyond IQ: Total Quotient (TQ).* Paper presented at the annual convention of the American Association for Psychological Science, New York.

Lichten, W., & Wainer, H. (2004, May). *IQ: A matter of life and death.* Paper presented at the annual convention of the American Association for Psychological Science, New Orleans.

Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A., Schalock, R. L., Snell, M. E., Spitalnik, D. M., Spreat, S., & Tassé, M. J. (2002). *Mental retardation: Definition, classification, and systems of supports* (10th ed.). Washington, DC: American Association on Mental Retardation.

Luckasson, R., Coulter, D. L., Polloway, E. A., Reiss, S., Schalock, R. L., Snell, M. E., Spitalnik, D. M., & Stark, J. A. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.). Washington, DC: American Association on Mental Retardation.

No Child Left Behind Act. Pub. L 107-110. 115 Stat. 1425. (2002).

Perske, R. (2005). Strange shift in the case of Daryl Atkins. *Mental Retardation, 43,* 451–455.

Roper v. Simmons. U.S. 03-633 (2005).

Schalock, R. L., & Braddock, D. L. (Eds.). (1999). *Adaptive behavior and its measurement.* Washington, DC: American Association on Mental Retardation.

Silverstein, A. B. (1973). Note on prevalence. *American Journal of Mental Deficiency, 77,* 380–382.

Smith, J. D., & Polloway, E. A. (1979). The dimension of adaptive behavior in mental retardation research: An analysis of recent practices. *American Journal of Mental Deficiency, 84,* 203–206.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales. Manual.* Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2005). *Vineland Adaptive Behavior Scales. Manual* (2nd ed.). Circle Pines, MN: American Guidance Service.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Trent, J. W. (1994). *Inventing the feeble mind. A history of mental retardation in the United States.* Berkeley: University of California.

Wechsler, D. (1943). Nonintellective factors in general intelligence. *Journal of Abnormal and Social Psychology, 38,* 101–103.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children.* New York: Psychological Corp.

Wechsler, D. (1974) *Manual for the Wechsler Intelligence Scale for Children-Revised.* San Antonio: Psychological Corp.

Zigler, E., Balla, D., & Hodapp, R. (1984). On the definition and classification of mental retardation. *American Journal of Mental Deficiency, 89,* 215–230.

Authors:

William Lichten, PhD, *Professor Emeritus, Koerner Center for Emeritus Professors, Yale University, 145 Elm St., New Haven, CT 06520-8368.* **Elliott W. Simon, PhD** *(E-mail:* elliott_simon@elwyn.org*), Executive Director, Elwyn, 111 Elwyn Rd, Elwyn, PA 19063.*

Mathematical Appendix

Intelligence Quotient (IQ), Adaptive Functioning, Total Quotient (TQ), and Errors

This Appendix provides details for those readers who care to delve more deeply into the mathematics of TQ, IQ, and adaptive behavior scores.

TQ

Here we construct a composite score (TQ) of intelligence (IQ) and adaptive functioning (AF), which has the same prevalence of mental retardation as the component scores (i.e., has a mean of 100 and *SD* of 15) (see Table 2). The computation goes as follows, in the example of an *r* of .2. The mean of the raw sum of the two scaled scores is the sum of the means, namely, $100 + 100 = 200$. The *SD* of the raw sum is given by the expression $\sigma(IQ + AF) = 15\sqrt{2(1 + r)} = 15\sqrt{2.4} = 23.2$. For the composite to have a mean of 100 and a *SD* of 15, we first align the TQ mean of 100 with the raw score sum of 200. To set the *SD* to 15, we multiply the raw score deviations in Column 1 of Table 2 by 0.645, which gives a step of 12.9 for *r* = .2. This result holds for the Vineland Adaptive Behavior Scale. If a different measure of adaptive functioning was used, simply put its correlation coefficient in the expression for $\sigma(IQ + AF)$.

The column in boldface holds for the current edition (Sparrow et al., 2005) of the Vineland, *r* = .2. Note that the values for TQ approximately 70 differ negligibly (about 1 point) from the values from the previous edition (Sparrow et al., 1984), for which *r* was typically about .3. This fact, in addition to the total constancy of prevalence, shows that TQ is as mathematically sound as IQ or standardized adaptive functioning scores, which addresses the concerns of Silverstein (1973) and Zigler et al. (1984).

Errors

IQ. We give a detailed error analysis for IQ here, with a briefer treatment later for adaptive functioning and a composite of the two. Manufacturers' test score errors estimates (standard error approximately 2–3 points) are based on reliability, which is only part of the story. Validity studies usually involve correlations between different tests. What can one do with these numbers? Here again, if the *r* is too small, the test is not consistent with other tests and is presumably invalid. If an *r* is too large, the test merely duplicates its competitors and adds nothing new. We use both reliability and validity data to give a definite number for errors.

The *Flynn effect* (Doppelt & Kaufman, 1977; Flynn, 1984, 1987) is an IQ drift, which increases with time up to the point when a new version of the intelligence test comes out. In the past this drift could be as large as 5 points over the life of a Wechsler test (Kanaya, Ceci, & Scullin, 2003; Kanaya, Scullin, & Ceci, 2003; Wechsler, 1949, 1974). Manufacturers now minimize it to a few points by shortening the time between editions. However, corrections are fallible. Too often, the correction has the wrong sign, thereby doubling the error (e.g., see Lichten & Wainer, 2004); thus, we recommend treating the Flynn effect as a random standard error of about 2–3 points.

Possible Wechsler–Stanford-Binet Discrepancies

To ask for the real value of a person's intelligence measured by an intelligence test leads to well-known circularities. To minimize the error by choosing the best test leads to further circularities. We know of no

Table 3 Stanford-Binet (S-B) and Wechsler (W) Means Compared

| Test | | | | Test scores | | Diff | Year of test | | Flynn effect | Total shift |
W[a]	S-B	Age group	N	W	S-B	W−S-B	W	S-B	W−S-B	W−S-B
WPPSI-R	5	3–5	71	99	105	−6	1989	2002	4	−2
WISC-III	5	6–16	66	106	101	+5	1991	2002	3	+8
WAIS-III	5	16–84	87	107	101	+6	1997	2002	2	+8
WAIS-III	4	16–45	26	113	115	−2	1997	1985	−3	−5
WISC-R	4	14 ± 4	61	67	66	+1	1974	1985	3	+4
WAIS-R	4	19.5 ± 2	21	73	64	9	1981	1985	1	+10
Rms dev.[b]										7

Note. Corrected for Flynn effect.
[a]WPPSI-R = Wechsler Preschool and Primary Scale of Intelligence-Revised, WAIS-III = Wechsler Adult Intelligence Scale-Third Edition, WISC-R = Wechsler Intelligence Scale for Children-Revised, WISC-III = Wechsler Intelligence Scale for Children-Third Edition.
[b]Rms dev. = Root Mean Square Deviation.

Table 4 Stanford-Binet and Wechsler Tests: Direct Comparison in the Range of Mental Retardation

Topic	WISC-III	S-B4	Diff	WISC-IV	S-B4	Diff
IQ mean	60.6	63.2	2.6	59.3	62.9	3.6
IQ *SD*	7.0	6.9	5.1	6.5	6.7	4.8
Correlation	.73			.74		
Corrected correlation	.92			.93		

Note. Age range in years. WISC = III = Wechsler Intelligence Scale for Children-III, 6.6 to 12.4; WISC = IV = Wechsler Intelligence Scale for Children, 6.1 to 10.5.

objective evidence that proves one test to be superior to others. Instead, we compare the results of two widely used tests, the Stanford-Binet and the Wechsler tests.

Table 4 shows disagreements between the two test score means from manufacturer's manuals. Because we cannot say one exam is right and the other is wrong, we estimate the standard error in each test from the discrepancy between them. We use the result from statistics that the *SD* of the average of two quantities is half the *SD* of the difference between the two.

These values are often larger than manufacturers' quoted errors (Table 3). If we assume random deviations, the standard error for a given test is 3–4 IQ points (half of Rms deviation in Table 3).

Table 3 refers to *means* (mainly for an IQ of approximately 100). There are few norming data on the IQs of persons with mental retardation. Consider the Wechsler tests, which typically involve norming samples of 200 persons in each age group. Because a person with an IQ under 70 is *two SDs* or more below the average, only about 2.5% of the sample (five persons) have mental retardation, which is hardly adequate.

The current deviation definition of intelligence is a linear approximation to the data. The difficulty comes in extrapolating. The longer the extrapolation, the more questionable is the IQ (American Educational Research Association, 1999, Standard 4.1). We made a test of the extrapolation by directly measuring the IQ of people with mental retardation.

Mental retardation. David DeLucia, a school psychologist in Portland, CT, has kindly furnished Stanford-Binet and WISC IQs of 123 school-age children with mental retardation (Table 4). These data show that, in the aggregate, the two tests correspond quite well in IQ.

Errors

Adaptive functioning and composite (TQ). Compared to IQ, adaptive functioning scales have slightly larger errors caused by the lower correlation between scales but lack the Flynn effect. The net result is that the error for adaptive functioning is about the same as for IQ. If one follows through the math for the composite TQ, it also happens to have the same error. Therefore, we have for all three scales, IQ, adaptive functioning, and composite, an *SD* of 5, and a margin of (significant) error of 10. From the standpoint of reliability (errors), all three scales are equivalent; from the standpoint of validity, the composite is the scale of choice.