

Conceptual Hurdles in the Application of *Atkins v. Virginia*

Lois A. Weithorn*

Introduction

In its 2002 decision in *Atkins v. Virginia*, the United States Supreme Court held that the Eighth Amendment's prohibition against cruel and unusual punishment precludes the execution of "mentally retarded offender[s]."¹ Writing for a six-member majority, Justice Stevens concluded that social attitudes and legal trends had shifted sufficiently in the thirteen years since the Court upheld such executions in *Penry v. Lynaugh*² to justify a reversal of *Penry*.³ Indeed, according to Justice Stevens, during this interval, a social consensus had emerged favoring an absolute ban on the imposition of the death penalty on "a mentally retarded criminal."⁴ To reach this decision, the Court applied the standard set forth in its Eighth Amendment jurisprudence: that "evolving standards of decency [marking] the progress of a maturing society" inform its determination of whether a particular penalty—such as a death sentence imposed on a mentally retarded individual—constitutes cruel and unusual punishment.⁵

In *Atkins*, the Supreme Court took the unusual step of transforming a specific clinical diagnosis into the ultimate legal issue by making a diagnosis of "mental retardation" dispositive of death penalty ineligibility.⁶ Despite the apparent "bright-line" clarity of an absolute ban on the execution of mentally retarded offenders, the determination of which offenders fall within the

* Professor of Law, University of California—Hastings College of the Law; J.D., Stanford University; Ph.D., Psychology, University of Pittsburgh. I am grateful for the insights and suggestions provided by colleagues who read and critiqued this article: Kate Bloch, Richard Bonnie, David Faigman, Stephen Greenspan, John Monahan, Melissa Nelken, Mindy Rosenberg, and Christopher Slobogin.

1. 536 U.S. 304, 321 (2002).

2. 492 U.S. 302 (1989).

3. *Atkins*, 536 U.S. at 306–07, 311–21.

4. *Id.* at 307.

5. *Id.* at 311–12 (quoting *Trop v. Dulles*, 356 U.S. 86, 100–01 (1958)).

6. Richard J. Bonnie & Katherine Gustafson, *The Challenge of Implementing Atkins v. Virginia: How Legislatures and Courts Can Promote Accurate Assessments and Adjudications of Mental Retardation in Death Penalty Cases*, 41 U. Rich. L. Rev. 811, 815 (2007). Alternatively, Bonnie and Gustafson point out, the Court "might have chosen to embrace a principle of diminished responsibility as a constitutional culpability requirement in capital cases," articulating a standard against which the psychological functioning of individual allegedly "mentally retarded" offenders would be measured on a case-by-case basis. *Id.* at 814–15.

protected group is deceptively complex.⁷ The Court's observation that, "[t]o the extent there is serious disagreement about the execution of mentally retarded offenders, it is in determining which offenders are in fact retarded,"⁸ was perhaps even more prescient than Justice Stevens realized. Much post-*Atkins* litigation has involved disputes about whether a particular defendant is or is not "mentally retarded."⁹ This result is not surprising in that the *Atkins* Court elevated the question of whether a defendant is found to be "mentally retarded" to the status of a life or death matter.¹⁰

While the spirit of the Court's decision in *Atkins* is unquestionably humane and compassionate, the implementation of *Atkins* raises new challenges with this singular reliance on clinical judgments about whether an individual is or is not "mentally retarded." The Court provided some guidance, however. It cited to two commonly-accepted definitions of "mental retardation," both of which rely on evaluations of "intellectual functioning" and "adaptive behavior."¹¹ States need not be bound by these definitions, however. The Court explicitly granted states discretion to determine precisely how to comply with the constitutional mandate announced in *Atkins*.¹² Yet, the meanings of the concepts of "mental retardation," "intelligence," and "adaptive behavior" are—like the standards of decency guiding the Court—continually evolving, and are subjects of ongoing reevaluation and debate among scientists, theorists, and professionals.¹³ Shifts in nomenclature illustrate this phenomenon. For example, in 2006 the American Association of Mental Retardation ("AAMR") changed its name to the American Association on Intellectual and Developmental Disabilities ("AAIDD"), and made corresponding changes to the names of its journals as well.¹⁴ AAIDD notes that

7. While the implementation of a "per se rule" excluding offenders from the death penalty on the basis of an easily-measured variable such as the defendant's age is relatively straightforward, *see Roper v. Simmons*, 543 U.S. 551 (2005) (holding unconstitutional the imposition of the death penalty on offenders who were under the age of eighteen at the time they committed the crime), applying a per se death penalty exclusion on the basis of a psychological diagnosis is far more complicated. Bonnie & Gustafson, *supra* note 6, at 814–15. As Bonnie and Gustafson point out:

Whether an adolescent will be constitutionally eligible for the death penalty is easily ascertained by looking at the defendant's birth certificate. In contrast, the constitutionality of a death sentence under *Atkins* turns exclusively on a clinical diagnosis, thereby magnifying the importance, and the stakes, of the clinical assessments of mental retardation and the expert opinions based on those assessments.

Id. at 815.

8. *Atkins*, 536 U.S. at 317.

9. *See, e.g.,* *People v. Superior Court*, 155 P.3d 259, 259 (Cal. 2007); *Commonwealth v. Miller*, 888 A.2d 624, 629 (Pa. 2005); *Ex parte Briseno*, 135 S.W.3d 1, 1 (Tex. Crim. App. 2004).

10. *See* Gerald P. Koocher, *IQ Testing: A Matter of Life or Death*, 13 *Ethics & Beh.* 1, 2 (2003).

11. *Atkins*, 536 U.S. at 308 n.3; *see also infra* notes 34–40 and accompanying text.

12. *Atkins*, 536 U.S. at 317.

13. For a superb examination of current conceptual debates and challenges, *see generally* *What Is Mental Retardation? Ideas for an Evolving Disability in the 21st Century* (Harvey N. Switzky & Stephen Greenspan eds., 2006) [hereinafter *What Is Mental Retardation?*]. *See also*, Robert L. Schalock et al., *The Renaming of Mental Retardation: Understanding the Change to the Term Intellectual Disability*, 45 *Intell. & Developmental Disabilities* 116 (2007).

14. Press Release, Am. Ass'n of Mental Retardation, World's Oldest Organization on Intellectual

the new language reflects a critical shift in the perspectives of researchers, professionals, and others about what is now referred to as “intellectual disability.”¹⁵ Such shifts are not new:

[T]he historical names used for this disability, such as *idiot* and *feeble-minded*, . . . paint a picture of the view of the disability at a particular point in time. Names change as perceptions and attitudes change. For example, the major association that is concerned with mental retardation was founded in 1878 as the Association of Medical Officers of American Institutions for Idiotic and Feeble-minded Persons. That name was changed to “American Association for the Study of the Feeble-minded” in 1906, then to “American Association of Mental Deficiency (AAMD)” in 1933. In 1987, the name was changed again to “American Association on Mental Retardation (AAMR).”¹⁶

If “[t]he field of mental retardation is in a period of great flux and transition”¹⁷ and notions of what constitutes “mental retardation” are changing, we must confront the normative question of which conceptualizations should guide the assessment in *Atkins* evaluations.

In addition, the last several decades have witnessed bold challenges to predominant concepts of “intelligence”¹⁸ which have important implications for how practitioners measure intellectual functioning. And, the concept of “adaptive functioning,” the undervalued companion to “intellectual functioning” in modern definitions of “mental retardation,”¹⁹ is also undergoing reevaluation.²⁰ Given what is at stake in *Atkins* assessments, lawmakers, jurists, practitioners, and others must confront the question of how to comply responsibly with the Court’s mandate in *Atkins* in light of the continual evolution of the knowledge, concepts, and practices relevant to its implementation.

This Article identifies some of the conceptual challenges inherent in determining who is and who is not “mentally retarded” for the purpose of applying the U.S. Supreme Court’s 2002 decision in *Atkins v. Virginia*. In Part I, I examine the Supreme Court’s articulation in *Atkins* of its rationales for

Disability Has a Progressive New Name (Nov. 2, 2006), http://www.aamr.org/About_AAIDD/name_change_PRdreen.htm. The announcement of the organization’s name change was followed by accompanying changes in the names of the two journals published by the newly-named American Association on Intellectual and Developmental Disabilities (AAIDD). See Press Release, Am. Ass’n on Intellectual & Developmental Disabilities, Mental Retardation Is No More—New Name Is Intellectual and Developmental Disabilities (Feb. 20, 2007), http://www.aamr.org/About_AAIDD/MR_name_change.htm.

15. See *infra* notes 41–49 and accompanying text.

16. Ronald L. Taylor et al., Mental Retardation: Historical Perspectives, Current Practices, and Future Directions 36 (2005).

17. See What Is Mental Retardation?, *supra* note 13, at xxiv.

18. See, e.g., Handbook of Intelligence (Robert J. Sternberg ed., 2000); Howard Gardner, Frames of Mind: The Theory of Multiple Intelligences (2d ed. 2004).

19. See *infra* notes 38–44 and accompanying text (discussing the two predominant definitions of “mental retardation”).

20. For a recent examination of the concept and measurement of “adaptive behavior,” see Adaptive Behavior and Its Measurement: Implications for the Field of Mental Retardation (Robert L. Schalock & David L. Braddock eds., 1999) [hereinafter Adaptive Behavior].

excluding “mentally retarded” persons from the reach of the death penalty. In Part II, I briefly discuss current notions as well as some recent reformulations of “mental retardation,” “intelligence,” and “adaptive behavior.” In Part III, I set forth certain key principles of psychological assessment and then, more specifically, *psycholegal* assessments (i.e., psychological assessments conducted with the purpose of informing a legal decision), noting the convergence of the more progressive notions of “mental retardation” and its measurement and modern principles of psycholegal assessment. In Part IV, I contrast the approaches of two states—Florida and California—to implementing *Atkins* and comment on how each approach fares in light of the principles guiding the conduct of valid psycholegal assessments. I conclude that for some defendants—particularly those viewed as “mildly mental retarded”—summary scores on traditional measures of intellectual and adaptive functioning will not constitute the most meaningful and relevant evidence of their disability for the purpose of death-penalty exclusion. Some defendants who are significantly intellectually impaired in ways highlighted by the *Atkins* Court will not be identified as “mentally retarded” on the basis of these scores, and may therefore be sentenced to death. Efforts to develop more appropriate evaluative approaches are necessary and should be grounded in modern models of psycholegal assessment guided by progressive constructions of concepts of “intelligence,” “intellectual disability,” and related psychological variables discussed in this Article. The development and testing of new measurement tools requires several years, however. In the meantime, given the limitations of existing measurement instruments, and in light of the severity and finality of the death penalty, state policies should err on the side of casting a net that is too wide rather than one that is too narrow in defining “mental retardation” for the purpose of *Atkins* compliance. Recommendations as to the appropriate criteria are set forth in this Article’s Conclusion.

I. The *ATKINS* Court on Why Death Is an Inappropriate Sentence for “Mentally Retarded” Persons

In his majority opinion, Justice Stevens asserted that mentally retarded defendants who violate the criminal law can be prosecuted, convicted, and punished, but that imposition of the death penalty on such individuals is not constitutionally permissible.²¹ The Court noted that “clinical definitions of mental retardation require not only subaverage intellectual functioning, but also significant limitations in adaptive skills such as communication, self-care, and self-direction that became manifest before age 18.”²² It emphasized that:

Mentally retarded persons . . . [b]ecause of their impairments . . . have diminished capacities to understand and process information, to communicate, to abstract from mistakes and learn from experience, to engage in logical reasoning, to control impulses, and to understand the reactions of others. . . . [T]here is abundant evidence that they often *act* on

21. *Atkins v. Virginia*, 536 U.S. 304, 306 (2001).

22. *Id.* at 318.

impulse rather than pursuant to a premeditated plan, and that in group settings they are followers rather than leaders. Their deficiencies do not warrant an exemption from criminal sanctions, but they do diminish their personal culpability.²³

Because of these limitations, the Court concluded that execution of mentally retarded offenders serves neither the retributive nor deterrent goals of that, most severe, form of punishment.²⁴ The Court characterized mentally retarded offenders as “less morally culpable” than other offenders due to “cognitive and behavioral impairments,” and observed that the same qualities that reduce moral culpability also undercut the penological goals of the death penalty as a deterrent.²⁵ The Court elaborated that “the diminished ability to understand and process information, to learn from experience, to engage in logical reasoning, or to control impulses makes it less likely that [these individuals] can process the information of the possibility of execution as a penalty and, as a result, control their conduct based upon that information.”²⁶

The Court also identified ways in which the psychological limitations of persons with mental retardation may call the fairness of the criminal justice process into question. These individuals may be more likely to make false confessions, less likely to articulate and prove mitigation, less able to assist their attorneys, and more likely to make poor witnesses in their own defense.²⁷ In summary, the Court concluded:

[B]ecause of their disabilities in areas of reasoning, judgment, and control of their impulses, [mentally retarded offenders] do not act with the level of moral culpability that characterizes the most serious adult criminal conduct. Moreover, their impairments can jeopardize the reliability and fairness of capital proceedings against mentally retarded defendants.²⁸

Thus, in justifying its conclusion that mentally retarded offenders should not be subjected to the death penalty, the Court identified a range of deficits, limitations, behaviors, and predispositions frequently associated with this diagnostic category. It expressed concern that these deficits affect cognitive abilities (e.g., information processing, comprehension, and abstract and logical reasoning), behavioral controls (e.g., modulating impulsivity), and social interactions (e.g., social reasoning and judgment, and susceptibility to social pressure or the manipulative conduct of others). Furthermore, it expressed additional concern that these impairments—while not necessarily rendering mentally retarded defendants incompetent to proceed to trial and sentencing—may undercut the procedural protections that typically guard against unfairness to defendants in death penalty cases.

The foregoing suggests that, in the Court’s view, the label of “mental

23. *Id.* (footnotes omitted).

24. *Id.* at 319–20.

25. *Id.*

26. *Id.* at 320.

27. *Id.* at 320–21.

28. *Id.* at 306–07.

retardation” is a *proxy* for the increased likelihood that the limitations cited above will characterize the psychological functioning of particular defendants. And, if the diagnosis of mental retardation could be reliably and accurately made with current assessment techniques in the cases of offenders who commit capital crimes, distinguishing those with the limitations identified by the Court from those without such limitations, the implementation of *Atkins* would be relatively straightforward. Unfortunately, applying *Atkins* is far more difficult. The term “mental retardation” encompasses a large and multifaceted collection of conditions. Not only are there literally hundreds of potential etiological bases for what we call “mental retardation,”²⁹ but the nature and severity of impairment vary dramatically across this diverse category.³⁰ Most of those within the mental retardation classification who function highly enough to engage in criminal activity will fall within the category referred to as “mild mental retardation.”³¹ It is within this subset of those who might be labeled as “mentally retarded” that diagnosis in the *Atkins* context might be most challenging. Depending upon a host of factors that might cause assessment findings to vary across time, situations, measures, or examiners, the answer to the question of whether an individual is “mentally retarded” for the purpose of an *Atkins* hearing may fluctuate. This is a troublesome state of affairs when an individual’s life hangs in the balance.

II. The Concepts of “Mental Retardation,” “Intelligence,” and “Adaptive Behavior”³²

A. Changing Concepts of “Mental Retardation”

In *Atkins*, the Court left in the hands of the states the determination of how to implement the constitutional prohibition against sentencing “mentally retarded” offenders to death.³³ As such, the decision is silent on the normative question of precisely what criteria and assessment methods are most appropriate in diagnosing “mental retardation” in the death penalty context. The Court did, however, tacitly signal its approval of two sources when introducing the concept of “mental retardation”:³⁴ the 1992 American

29. See, e.g., Am. Ass’n of Mental Retardation, *Mental Retardation: Definition, Classification, and Systems of Supports* 81–91 tbl. 7.3 (9th ed. 1992) [hereinafter 1992 AAMR Manual] (listing “disorders in which mental retardation may occur”).

30. Professor J. David Smith characterizes “the universe of human conditions subsumed under the term ‘mental retardation’ [a]s overwhelming,” and notes that the term mental retardation describes people “who are more different than they are alike.” J. David Smith, *Speaking of Mild Mental Retardation: It’s No Box of Chocolates, or Is It?*, 14 *Exceptionality* 191, 201 (2006).

31. Stephen Greenspan, *Functional Concepts in Mental Retardation: Finding the Natural Essence of an Artificial Category*, 14 *Exceptionality* 205, 207 (2006); James R. Patton & Denis W. Keyes, *Death Penalty Issues Following Atkins*, 14 *Exceptionality* 237, 237–38 (2006).

32. It is well beyond the scope of this Article to discuss these complex concepts in a comprehensive manner. Therefore, my goal here is to raise some of the current conceptual and practical issues surrounding these terms that have relevance for their application in the context of *Atkins* evaluations.

33. *Atkins*, 536 U.S. at 317.

34. See *id.* at 308 n.3.

Association on Mental Retardation's Manual ("1992 AAMR Manual")³⁵ and the 2000 American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders ("DSM-IV-TR").³⁶ The 1992 AAMR Manual definition reads:

Mental retardation refers to substantial limitations in present functioning. It is characterized by significantly subaverage intellectual functioning, existing concurrently with related limitations in two or more of the following applicable adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academic, leisure, and work. Mental retardation manifests before age 18.³⁷

The AAMR definition further clarifies that "*significantly subaverage intellectual functioning* . . . is defined as an IQ standard score of approximately 70 to 75 or below" on one of several standardized and individually-administered intelligence tests,³⁸ and provides additional guidance for all phases of the assessment.³⁹ The assessment criteria contained in the DSM-IV-TR are heavily influenced by the AAMR model and therefore quite similar in most respects.⁴⁰

As noted above, however, notions of what constitutes "mental retardation" have evolved over time.⁴¹ Changes in the working definition of "mental retardation" appeared in the tenth edition of the AAMR Manual in 2002 ("2002 AAMR Manual").⁴² The new definition reads: "Mental retardation is a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills. This disability originates before age 18."⁴³ Many in the field were critical of AAMR's failure to incorporate more modern paradigms at a time when "[t]he field of mental retardation is in a period of great flux and transition."⁴⁴

35. 1992 AAMR Manual, *supra* note 29, at 5.

36. Am. Psychiatric Ass'n, Diagnostic and Statistical Manual of Mental Disorders 49 (4th ed. text rev. 2000) [hereinafter DSM-IV-TR]. A third source influential among practitioners and researchers, but not mentioned by the Court, is published by the American Psychological Association's Division of Intellectual and Developmental Disabilities (Division 33). Manual of Diagnosis and Professional Practice in Mental Retardation 13 (John W. Jacobson & James Anton Mulick eds., 1996) [hereinafter Manual of Diagnosis and Professional Practice].

37. 1992 AAMR Manual, *supra* note 29, at 5.

38. *Id.* The range of "approximately 70 to 75 or below" is cited to accommodate the "standard error of measurement," which will be discussed below. *See infra* note 79.

39. 1992 AAMR Manual, *supra* note 29, at 5-7, 23-49.

40. DSM-IV-TR, *supra* note 36, at 48-49. The primary difference in the language of the DSM-IV-TR and the 1992 AAMR Manual is that the former indicates that "significantly subaverage intellectual functioning" is demonstrated by "an IQ of . . . approximately 70 or below" rather than the 70-75 range set forth by the latter. *See id.* at 48; 1992 AAMR Manual *supra* note 29, at 5. As noted below, this distinction can have great practical import in the *Atkins* context. *See infra* notes 152-159 and accompanying text (discussing the rigidity with which Florida courts have interpreted a score of 70 as a fixed cut-off score).

41. *See supra* notes 13-17 and accompanying text.

42. Am. Ass'n of Mental Retardation, Mental Retardation: Definition, Classification, and Systems of Supports 8 (10th ed. 2002) [hereinafter 2002 AAMR Manual].

43. *Id.*

44. *See* What Is Mental Retardation?, *supra* note 13, at xxi, xxiv. In the preface to the 2003 edition of this

In response to those critiques, as well as to other factors, the AAMR initiated a more recent and more dramatic example of the conceptual evolution: the 2006 change in the name of the American Association of Mental Retardation to the American Association of Intellectual and Developmental Disabilities.⁴⁵ The AAIDD indicated that the change was motivated in part to conform with international standards promulgated by the World Health Organization, and in part to address the pejorative connotations that the term “mental retardation” has acquired, but also to reflect a critical shift in philosophy.⁴⁶ The change incorporates modern scientific and social conceptualizations of “disability” and “aligns better with current professional practices that are focused on *functional behaviors and contextual factors*.”⁴⁷ Psychologists Stephen Greenspan, Harvey Switzky, and colleagues have consistently stressed the importance of focusing on individuals’ abilities to function in the real-world contexts they confront in “everyday” life when conceptualizing “mental retardation.”⁴⁸ While it is not yet clear to what extent the change in nomenclature will incorporate the contributions of these authors and others who have been critical of traditional frameworks, the AAIDD appears to be laying the groundwork to move in that direction. The following explanation is provided by AAIDD for its embrace of emerging notions of “disability”:

[The] construct of intellectual disability . . . has evolved to emphasize an ecological perspective that focuses on the person-environment interaction The importance of this evolutionary change in the construct of disability is that intellectual disability *is no longer considered entirely an absolute, invariable trait of the person*. Rather the social-ecological construct of disability, and intellectual disability . . .⁴⁹ exemplifies the interaction between the person and . . . environment

This paradigm shift has significant implications for the application of *Atkins*. It focuses attention on the need to examine an individual’s intellectual disability as it manifests within the particular context of interest and with

book, reprinted in the revised and updated 2006 edition, the authors note that the book was written because of the failure of the 2002 AAMR Manual to respond to important input from researchers and professionals regarding needed changes in the conceptualization of mental retardation. *Id.* at xxi–xxv. The various contributions to this edited volume address a range of concerns about the 1992 and 2002 AAMR Manuals. For additional discussion of the critiques of the 1992 definition and the AAMR response, see, for example, 2002 AAMR Manual, *supra* note 42, at 19–48; Stephen Greenspan, *What Is Meant by Mental Retardation?*, 11 *Int’l Rev. Psychiatry* 6, 6 (1999); Stephen Greenspan et al., *Everyday Intelligence and Adaptive Behavior: A Theoretical Framework*, in *Manual of Diagnosis and Professional Practice*, *supra* note 36, at 127–35 [hereinafter Greenspan et al., *Everyday Intelligence*].

45. See *supra* note 14 and accompanying text.

46. Schalock et al., *supra* note 13, at 120. The AAIDD indicated that the new name “covers the same population of individuals who were diagnosed previously with mental retardation.” *Id.*

47. *Id.* (emphasis added).

48. See, e.g., Greenspan et al., *Everyday Intelligence*, *supra* note 44.

49. Schalock et al., *supra* note 13, at 117 (emphasis added). For further discussion of the social-ecological perspective, see generally Urie Bronfenbrenner, *The Ecology of Human Development: Experiments by Nature and Design* (1979); James Garbarino, *Children and Families in the Social Environment* (2d ed. 1992).

respect to the functional demands of that person's social environment.⁵⁰ While persons who are severely disabled intellectually may demonstrate significant impairments across situations and settings, the intellectual functioning of persons whose impairments are less severe—the very persons whose functioning is likely to be at issue in *Atkins*—will vary with the demands of particular situations and with the aspects of functioning that an evaluator measures.⁵¹ Thus, as emphasized below, how an offender performs on tests that tap more academic skills, such as knowledge of vocabulary or ability to execute certain mathematical calculations, may not tell us much about the attributes cited by the *Atkins* Court as justifying its differential treatment of capital offenders who are “mentally retarded” from those who are not.⁵² Precisely how the new model will affect the definition of “intellectual disability” offered in the next edition of the AAIDD Manual (formerly the AAMR Manual) is uncertain.⁵³ The emerging paradigm in the field now known as the study of “intellectual disability,” when applied to a legal context as in *Atkins* assessments, bears much in common with the models articulated by the most highly-respected scholars in the field of *psycholegal* assessment.⁵⁴ This convergence of approaches will be discussed further below.⁵⁵

B. Concepts of “Intelligence” and Its Measurement

In that current definitions of “mental retardation” reference “significant limitations in intellectual functioning *and* in adaptive behavior,”⁵⁶ I will briefly discuss both component concepts. Because the 2002 AAMR Manual and the DSM-IV-TR each require use of standardized intelligence, or IQ, tests to demonstrate levels of intellectual functioning, I also comment on the

50. A humorous illustration of this concept can be found in a scene from the popular 1994 movie *Forrest Gump*. In one scene, the movie parodies the military as a setting in which unquestioning obedience to authority is valued above all. The lead character, Forrest Gump, who is portrayed as intellectually challenged in some contexts, such as in school, is judged to be intellectually gifted by his army drill sergeant:

Drill Sergeant: “Gump! What’s your sole purpose in this army?”

Forrest Gump: “To do whatever you tell me, Drill Sergeant!”

Drill Sergeant: “Gump! You’re a . . . genius! This is the most outstanding answer I have ever heard. You must have [an] I.Q. of 160. You are . . . gifted, Private Gump.”

Forrest Gump (Paramount Pictures 1994).

51. Greenspan distinguishes the frameworks most useful in conceptualizing two subgroups within the larger class of persons identified as intellectually disabled:

- (a) a smaller, more severely impaired sub-category, most of whose members have a known biological etiology and where the physical and behavioral signs of impairment are fairly obvious, and
- (b) a larger, less impaired sub-category, many of whom do not have clearly-established biological etiology and where the physical and behavioral signs of impairment are more subtle.

Greenspan, *supra* note 31, at 206.

52. See *supra* notes 22–28 and accompanying text.

53. In the first of a series of articles by the AAIDD Committee on Terminology and Classification, which is responsible for revision of the AAID manual, the Committee implies that the new model will be influential as it “share[s] . . . thoughts and ask[s] for input from the field prior to the anticipated publication in 2009/2010 of the 11th edition.” Schalock et al., *supra* note 13, at 116.

54. See *infra* notes 140–149.

55. See *infra* note 150 and accompanying text.

56. See *supra* notes 33–43 and accompanying text.

relationship between various notions of intelligence and what is measured by those tests, where appropriate.

The question of what the term “intelligence” means has been the subject of voluminous scientific, professional, and lay commentary. One of today’s most prolific scholars on the subject, psychologist Robert Sternberg, observes: “Looked at in one way, everyone knows what intelligence is; looked at in another way, no one does. Put another way, people all have conceptions—which also are called folk theories or implicit theories—of intelligence, but no one knows for certain what it actually is.”⁵⁷ And, indeed, after over 150 years of theory, research, and assessment practice, there remains a range of expert and lay views as to what precisely constitutes human “intelligence.”

Some of the first discussions about a general “mental faculty” appeared in the writings of certain philosophers.⁵⁸ Among the first attempts to develop a theory of general intellectual abilities with corresponding measurement techniques was that of Francis Galton, a cousin of Charles Darwin, in the latter half of the nineteenth century.⁵⁹ He and other early theorists focused on evaluating “auditory and visual sensory discrimination abilities as well as reaction times to stimuli and the ability to exert hand-squeeze pressure,” believing that these capacities were the foundation of higher cognitive abilities.⁶⁰ Other influential thinkers included British psychologist Charles Spearman, who posited that there was a single general intellectual capacity he referred to as “g,” and that correlational analyses could illuminate the interrelationships of the different components of this general ability.⁶¹ In contrast to this notion of unitary or generalized intellectual abilities, French psychologist Alfred Binet viewed intelligence as the composite of multiple complex cognitive functions,⁶²—among them memory, imagery, imagination, attention, and comprehension.⁶³ American psychologist Edward Thorndike characterized intelligence as comprised of three components: “the ability to understand and manage ideas (abstract intelligence), concrete objects (mechanical intelligence) and people (social intelligence).”⁶⁴ Throughout much of the twentieth century, writers theorized about intelligence, conducted empirical studies, developed assessment techniques, and challenged each others’ points of view. A 1921 symposium on the definition of intelligence revealed that there were many views—and much disagreement—among the prominent thinkers.⁶⁵ Some common themes emerged, however, including the

57. Robert J. Sternberg, *The Concept of Intelligence*, in *Handbook of Intelligence*, *supra* note 18, at 3, 3.

58. Lewis R. Aiken, *Assessment of Intellectual Functioning 1* (2d ed. 1996).

59. Nathan Brody, *History of Theories and Measurements of Intelligence*, in *Handbook of Intelligence*, *supra* note 18, at 16, 16.

60. *Id.*

61. *Id.* at 17–19.

62. *Id.* at 18.

63. Aiken, *supra* note 58, at 11.

64. John F. Kihlstrom & Nancy Cantor, *Social Intelligence*, in *Handbook of Intelligence*, *supra* note 18, at 359, 359.

65. See Aiken, *supra* note 58, at 19–20; Brody, *supra* note 59, at 30.

notion that some core components of intellectual abilities related to: “(1) problem-solving ability, or adaptability to new situations; (2) the ability to deal with symbols, concepts, and relationships; and (3) the ability to learn or profit from experience.”⁶⁶

Binet and his colleague Theophile Simon developed a test to identify school children who were of low intelligence in order to place those children in separate classes.⁶⁷ In 1905, they produced the first formal intelligence test, which measured many of the types of skills generally used by children in school, such as tasks of memory, reasoning ability, and numerical skills.⁶⁸ The Binet-Simon test, as it was called, was revised over the next several years, and ultimately translated into English for use in the United States.⁶⁹ Lewis Terman, a Stanford University based psychologist, continued to revise Binet’s test, now known as the Stanford-Binet Intelligence Scale.⁷⁰ The test is currently in its fifth edition,⁷¹ but its use has waned as the Wechsler family of tests have surpassed it in popularity.⁷² David Wechsler created the Wechsler Adult Intelligence Scale (“WAIS”), the Wechsler Intelligence Scale for Children (“WISC”), and the Wechsler Preschool and Primary Scale of Intelligence (“WPPSI”).⁷³ In developing his tests, he was influenced by his exposure to the widespread use of group-administered IQ tests used during World War I to evaluate the suitability of military recruits for various roles in the service.⁷⁴ The WAIS is currently in its third edition (“WAIS-III”),⁷⁵ and the WISC is in its fourth edition (“WISC-IV”).⁷⁶

Each time a new edition of the Stanford-Binet or the Wechsler Scales is developed, the administration of the test to a large, and ideally heterogeneous and representative, sample of individuals is used to develop norms based on age and/or grade in school.⁷⁷ Those norms are then used to permit standard score conversion formulas so that any single test administrator can report the

66. Aiken, *supra* note 58, at 20.

67. *Id.* at 12–13.

68. *Id.* at 13.

69. *Id.* at 13–16.

70. Anne Anastasi, *Psychological Testing* 238–48 (6th ed. 1988).

71. The current version of the test is described at the website of its publisher. *See* Riverside Publishing, *Stanford-Binet Intelligence Scales (SB5), Fifth Edition*, <http://www.riverpub.com/products/sb5/details.html> (last visited Apr. 1, 2008); *see also* Gale H. Roid et al., *Essentials of Stanford-Binet Intelligence Scales (SB5) Assessment* (2004).

72. Alan S. Kaufman, *Tests of Intelligence*, in *Handbook of Intelligence*, *supra* note 18, at 445, 448.

73. Aiken, *supra* note 58, at 19; Anastasi, *supra* note 70, at 248–63.

74. Kaufman, *supra* note 72, at 446–47.

75. The website of the publisher of WAIS-III, Pearson, offers a description of the test. *See* Pearson, *Wechsler Adult Intelligence Scale—Third Edition (WAIS-III)*, <http://harcourtassessment.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8980-727&Mode=summary> (last visited Apr. 20, 2008); *see also* Alan S. Kaufman et al., *Essentials of WAIS-III Assessment* (1999).

76. The website of the publisher of WISC-IV, Pearson, offers a description of the test. *See* Pearson, *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV)*, <http://harcourtassessment.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8980-727&Mode=summary> (last visited Apr. 20, 2008); *see also* Alan S. Kaufman et al., *Essentials of WISC-IV Assessment* (2004).

77. *See* Anastasi, *supra* note 70, at 245–46, 252, 258.

examinee's scores with reference to the appropriate comparison group. Thus, the Wechsler tests permit the report of the Full Scale score, as well as the two composite Verbal and Performance Scale scores, for which the mean is 100, and the standard deviation is fifteen.⁷⁸ As noted above, the 1992 and 2002 AAMR Manuals and the DSM-IV-TR generally specify a score of approximately 70 or below to satisfy the "intellectual functioning" prong of the diagnostic criteria for "mental retardation," which is two standard deviations below the mean of 100.⁷⁹ The AAMR Manuals, as noted above, refer to an IQ score *range* of approximately 70 to 75 points, in order to factor in what is referred to as the "standard error of measurement."⁸⁰

The widespread use of intelligence tests for a range of purposes in our country has led to criticisms, including the following objections: some test developers have failed to ground the tests in relevant theory;⁸¹ the tests focus disproportionately on skills relevant to academic pursuits (such as verbal, numerical, and spatial abilities) and are poor predictors of abilities to perform in contexts requiring different skill sets;⁸² the tests disadvantage test-takers from minority groups in society, thus reinforcing stereotypic biases about intelligence hierarchies among groups within society;⁸³ and the tests and the concept of IQ conveys the notion that intelligence is "highly heritable, and therefore passed on through the genes from one generation to the next."⁸⁴

Recent decades, however, have brought with them new theories of intelligence and, in some instances, new tests designed to avoid some of the perceived problems of earlier measures. Sternberg reports the results of a symposium he co-sponsored in 1986—sixty-five years after the 1921 symposium⁸⁵—soliciting the perspectives of expert theorists and researchers in response to the question: What is intelligence?⁸⁶ The symposium yielded "roughly two dozen definitions," each different from the next.⁸⁷ Among the

78. *Id.* The Wechsler tests are comprised of individual subtests, each of which focuses on a more narrow set of abilities. More recent editions of the tests allow for the calculation of several "index" scores, based on individuals' performance on particular clusters of subtests. All of these scores are converted into "standard scores" so that the evaluator can compare the scores to general population norms. See, e.g., Paul E. Williams et al., WISC-IV Technical Report #1: Theoretical Model and Test Blueprint, June 1, 2003 (available at: <http://harcourtassessment.com/hai/Images/pdf/wisciv/WISCIVTechReport1.pdf>).

79. See *supra* notes 35–43 and accompanying text.

80. See *supra* notes 38–40 and accompanying text.

81. Kaufman, *supra* note 72, at 448.

82. See, e.g., Sternberg, *supra* note 57, at 46.

83. See, e.g., Etienne Benson, *Intelligent Intelligence Testing*, Monitor on Psychology, Feb. 2003, at 48, available at <http://www.apa.org/monitor/feb03/intelligent.html>.

84. Robert J. Sternberg & James C. Kaufman, *Human Abilities*, 49 Ann. Rev. Psychol. 479, 491 (1998) (discussing critiques of the controversial book by Richard J. Herrnstein & Charles Murray, *The Bell Curve: Intelligence and Class Structure in American Life* (1994)).

85. See *supra* note 65 and accompanying text.

86. What Is Intelligence?: Contemporary Viewpoints on its Nature and Definition, at vii (Robert J. Sternberg & Douglas K. Detterman eds., 1986). The symposium papers are compiled and published in this book.

87. Sternberg, *supra* note 57, at 46.

“common threads,” several definitions focused on the ability to function in everyday life and to learn.⁸⁸ Sternberg notes that the traditional standardized intelligence tests do not evaluate either of these aspects of human functioning.⁸⁹ Sternberg’s own theories of intelligence emphasize analytical, creative, and practical abilities, and differ substantially from that which is typically measured on standardized IQ tests.⁹⁰ Psychologist Stephen Greenspan, an expert in the field of intellectual disabilities, builds on the work of early twentieth century psychologist Edward Thorndike, as well as Robert Sternberg and others in formulating a “tripartite model of adaptive intelligence.”⁹¹ He applies his model to an understanding of the limitations and needs of persons with intellectual disabilities, and views intelligence as comprised of conceptual or academic intelligence (the focus of most IQ tests), social intelligence, and practical intelligence.⁹²

Psychologist Howard Gardner recently published the twentieth anniversary edition of his popular book, *Frames of Mind: The Theory of Multiple Intelligences*.⁹³ A harsh critic of traditional views of intelligence and its testing as overly narrow, Gardner focuses on individuals’ abilities

*to resolve genuine problems or difficulties that he or she encounters and, when appropriate, to create an effective product—and must also entail the potential for finding or creating problems—thereby laying the groundwork for the acquisition of . . . those intellectual strengths that prove of some importance within a cultural context.*⁹⁴

88. *Id.*

89. *Id.*

90. Sternberg & Kaufman, *supra* note 84, at 494. Sternberg & Kaufman elaborate on the three components of what they refer to as “successful intelligence”:

Analytical abilities are required to analyze and evaluate the options available to oneself in life. They include things such as identifying the existence of a problem, defining the nature of the problem, setting up a strategy for solving the problem, and monitoring one’s solution processes.

Creative abilities are required to generate problem-solving options in the first place. . . . Research shows that these abilities are at least partially distinct from conventional IQ, and that they are moderately domain-specific, meaning that creativity in one domain (such as art) does not necessarily imply creativity in another (such as writing).

Practical abilities are required to implement options and to make them work. Practical abilities are involved when intelligence is applied to real-world contexts. A key aspect of practical intelligence is the acquisition and use of tacit knowledge, which is knowledge that is not explicitly taught and that usually is not verbalized. Research shows that tacit knowledge is acquired through mindful utilization of experience that is relatively domain specific

Id.

91. Stephen Greenspan, *Mental Retardation in the Real World: Why the AAMR Definition Is Not There Yet, in What Is Mental Retardation?* *supra* note 13, at 165, 172.

92. *Id.*

93. Gardner, *supra* note 18.

94. *Id.* at 60–61. In *Frames of Mind*, Gardner delineated several different categories of intelligence, referred to as linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, and two personal intelligences. *Id.* at 73–76. While some of these areas of intellectual functioning—such as musical intelligence—are unlikely to be relevant to the death penalty context, Gardner’s writings underscore that discussions of intelligence as a single global trait may be inaccurate. Gardner’s writings on the personal intelligences, and particularly interpersonal intelligence (which involves one’s ability to perceive, notice, and make distinctions about the moods, temperaments, motivations, and intentions of others), *id.* at 239, are not unrelated to the Court’s. See *supra* note 22 and accompanying text; see also *supra* notes 124–129 and

His view of intelligence is far broader and more differentiated than are more traditional notions, which focus primarily on verbal (or linguistic) and mathematical skills. In recent years, psychologist Daniel Goleman has compiled substantial empirical support for the proposition that there are facets of intellectual functioning he describes as “emotional intelligence,”⁹⁵ and “social intelligence,”⁹⁶ and that capacities in these areas have significant impact on individuals’ real-world functioning. Other psychologists have proposed new approaches, and some have developed measures grounded more heavily in theory.⁹⁷

And what about lay perspectives? The research of Sternberg and his colleagues reveals that laypersons in the United States generally view intelligence as comprised of three different factors: verbal abilities, practical problem-solving abilities, and social competence abilities.⁹⁸ The commonly-used intelligence tests, however, incorporate measurement of skills falling within the first category, but not the second or third. Sternberg has also found cross-cultural differences in prevalent views of what constitutes intelligence, further underscoring the importance of looking at intelligence contextually.⁹⁹ In other words, the question of an individual’s intelligence may be best answered with reference to the demands and challenges placed upon him or her in a particular setting. The particular setting might be defined more globally in terms of wider societal expectations (hence, cross-cultural differences) or more narrowly in terms of a particular sphere of functioning within a more circumscribed context (e.g., “naiveté” and “gullibility” in social interactions more generally, and in the context of collaborative criminal offending in particular).¹⁰⁰

Thus, there is a conceptual convergence of sorts with respect to evolving notions of “mental retardation” and “intelligence” that are relevant to the application of *Atkins*. In adopting new terminology and apparently embracing the social-ecological and functional perspectives, and modern notions of “disability,” AAIDD—the group whose manual and definition of “mental retardation” is the standard in the field—has moved closer to some of the more modern theories of intelligence. How this convergence will express itself when

accompanying text (discussing Professor Greenspan’s concerns about the social vulnerability of intellectually disabled individuals). Gardner’s notions of intelligence were updated in 1999. Howard Gardner, *Intelligence Reframed: Multiple Intelligences for the 21st Century* 47 (1999).

95. Daniel Goleman, *Emotional Intelligence: Why It Can Matter More than IQ* (10th anniv. ed. 2005).

96. Daniel Goleman, *Social Intelligence: The New Science of Human Relationships* (2006).

97. See, e.g., Kaufman, *supra* note 72, at 462–70 (discussing the Woodcock Johnson, Kaufman, and Das-Naglieri assessment tools).

98. Robert J. Sternberg et al., *Intelligence, Race, and Genetics*, 60 *Am. Psychologist* 46, 47 (2005).

99. Sternberg & Kaufman, *supra* note 84, at 481–84.

100. Stephen Greenspan asserts that the crux of the vulnerability of intellectually disabled persons in the context of capital offending is grounded in the naiveté and gullibility they bring to social situations. Stephen Greenspan, *Execution Exemption Should Be Based on Actual Vulnerability, Not Disability Label*, 13 *Ethics & Behav.* 19, 23 (2003); Stephen Greenspan & Harvey N. Switzky, *Lessons from the Atkins Decision for the Next AAMR Manual, in What Is Mental Retardation?*, *supra* note 13, at 281, 300; see also *infra* notes 129–131 and accompanying text.

AAIDD publishes the eleventh edition of its manual in the next few years is not yet known.

C. Concepts of “Adaptive Functioning” and Its Measurement

In the 1992 AAMR Manual, the AAMR defined the second of its three criteria for the diagnosis of “mental retardation” as the existence, concurrent with “significantly subaverage intellectual functioning,” of “related limitations in two or more of the following applicable adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academic, leisure, and work.”¹⁰¹ The AAMR’s rationale for including the adaptive skills criterion is that such measures provide confirmation of the diagnosis of “mental retardation” obtained with standardized IQ tests, that is, as a check against measurement error rather than as a way of measuring aspects of functioning not tapped by the IQ tests.¹⁰² In response to criticism that this definition minimized the importance of adaptive functioning, the AAMR altered the diagnostic criteria to read: “significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social, and practical adaptive skills.”¹⁰³ This change, according to Stephen Greenspan, misses the point.¹⁰⁴ Greenspan’s “tripartite model” of intelligence views conceptual or academic intelligence, which is what the traditional IQ tests generally measure, as but one facet of intellectual functioning.¹⁰⁵ Social intelligence and practical intelligence are two other domains which, taken together, comprise *everyday intelligence*, that is, the ability to apply one’s intellectual abilities to real-world settings and problems.¹⁰⁶ Thus, according to Greenspan, measurement of adaptive behavior makes sense only if it is conceptualized as measurement of these two other domains of intelligence, rather than as an add-on reflecting certain “skills,” that are not identified as part of the complex of intellectual abilities.¹⁰⁷ Therefore, Greenspan asserts that the 2002 definition “still views adaptive behavior as something different from intelligence and thus of less centrality to the diagnosis.”¹⁰⁸

With this controversy in mind, what is meant by the AAMR in its reference to “adaptive behavior”? Given the array of views on the nature of “mental retardation” and “intelligence” discussed above, as well as the controversy introduced in the preceding paragraph, it is probably not surprising that there is no single, commonly-accepted conceptualization of “adaptive

101. See 1992 AAMR Manual, *supra* note 29 and accompanying text.

102. The Manual states: “Evidence of adaptive skill limitations is necessary because [the] impact on functioning of these limitations must be sufficiently comprehensive to encompass at least two adaptive skill areas, thus showing a generalized limitation and reducing the possibility of measurement error.” 1992 AAMR Manual, *supra* note 29, at 6.

103. See 2002 AAMR Manual, *supra* note 42, at 23.

104. Greenspan, *supra* note 91, at 177.

105. *Id.* at 172.

106. *Id.* at 172–73.

107. *Id.* at 173–79.

108. *Id.* at 177.

behavior.” “Adaptive behavior” was first included among the criteria for diagnosing “mental retardation” in the 1961 Manual of the American Association of Mental Deficiency (AAMD),¹⁰⁹ and thus efforts to define and measure adaptive behavior picked up steam in the 1960s and thereafter.¹¹⁰ One team of writers analogizes the theoretical and empirical efforts examining the concept of adaptive behavior in the 1980s and 1990s to scholarly efforts to understand “intelligence” in the first half of the twentieth century.¹¹¹ Thus, as with our discussions of concepts of “mental retardation,” and “intelligence,” this Article’s treatment will barely scratch the surface in representing the breadth and depth of the scientific and professional literature.

According to the 2002 AAMR Manual: “*Adaptive behavior* is the collection of conceptual, social, and practical skills that have been learned by people in order to function in their everyday lives.”¹¹² Various formulations of *conceptual* adaptive functioning focus on “ability to solve abstract ‘intellectual’ problems and to use and understand symbolic processes, including language.”¹¹³ Attempts to measure these abilities focus on behavior relating to communication (e.g., expressive and receptive language skills and nonverbal communication) and academic skills (e.g., reading, writing, and numerical skills, as in handling money).¹¹⁴ *Social* adaptive functioning involves abilities to “understand and deal effectively with social and interpersonal objects and events, including the ability to act wisely in human relations, to exhibit appropriate social skills, be empathetic and self-reflective, and achieve desired interpersonal outcomes.”¹¹⁵ Some of the ways in which the adaptive behavior scales try to measure such abilities is through evaluating various facets of individuals’ interactions and relationships with others, their social problem-solving skills, and their responses to certain social situations.¹¹⁶ *Practical* adaptive functioning involves abilities to perform tasks of daily living and self-care and vocational activities.¹¹⁷ Thus, measures would typically evaluate the individual’s capacities for “independent living,” as demonstrated by competencies in performing daily routines such as dressing, personal hygiene, eating, basic housekeeping, managing in the community (such as taking public transportation), as well as the skills required by

109. The AAMD was the name of the AAMR prior to its 1987 name change.

110. Kazuo Nihira, *Adaptive Behavior: A Historical Overview*, in *Adaptive Behavior*, *supra* note 20, at 7, 7. It is noteworthy that Nihira refers to 1959 as the date on which this version of the AAMD Manual was published. Greenspan, however, indicates that the Manual was published in 1961, and that while “a preliminary draft of the manual was first published in 1959” and is often cited by authors, that preliminary draft did not refer to adaptive behavior. Greenspan, *supra* note 31, at 213.

111. James R. Thompson et al., *Adaptive and Maladaptive Behavior: Functional and Structural Characteristics*, in *Adaptive Behavior*, *supra* note 20, at 15, 16.

112. 2002 AAMR Manual, *supra* note 42, at 73.

113. Robert L. Schalock, *The Merging of Adaptive Behavior and Intelligence: Implications for the Field of Mental Retardation*, in *Adaptive Behavior*, *supra* note 20, at 43, 45.

114. *Id.* at 51.

115. *Id.* at 45.

116. *Id.* at 51.

117. *Id.* at 45.

particular vocational or occupational endeavors.¹¹⁸ Because it is not possible for an evaluator to observe a representative-enough spectrum of an individual's functioning in many of these areas in testing sessions, information is typically gleaned from those who have observed the individual's behavior in the relevant settings.¹¹⁹ Thus, parents and teachers would be the likely sources of ratings for children. For adults, family members and others who work closely and spend time with the individual are likely information providers.

In recent decades, there has been increasing attention paid to the psychometric properties of various adaptive behavior scales. Although there are reportedly over 200 scales that seek to measure various aspects of adaptive behavior, only a few have gained acceptance in the field.¹²⁰ While not endorsing one measure of adaptive behavior over another, the 2002 AAMR Manual describes several measures that meet the standards in the field relative to evidence of reliability and validity¹²¹ and with sufficient empirical research to establish norms against which individuals' performance on the measures can be compared.¹²² The 2002 AAMR Manual mentions the following measures as scales that have met these standards: Vineland Adaptive Behavior Scales; the AAMR Adaptive Behavior Scales; the Scales of Independent Behavior (of the Woodcock-Johnson Psycho-Educational Battery); the Comprehensive Test of Adaptive Behavior-Revised; and the Adaptive Behavior Assessment System.¹²³

The concept of adaptive behavior, because it focuses on the individual's actual functioning in real-world situations, initially appears to be well-suited to the social-ecological concept of "intellectual disability" as a relationship between an individual's functioning and the demands of particular contexts. Unfortunately, according to Greenspan, developers of the currently available adaptive behavior measures never adequately articulated "a clear theoretical understanding of the construct" *before* the development of measures.¹²⁴ Instead, once the tests were in use, post hoc constructs of adaptive behavior were defined by looking at what areas the tests covered.¹²⁵ Furthermore, initial tests were developed by focusing on institutional residents whose functioning was greatly impaired and for whom mastering the basics of tasks like dressing and self-feeding is significant.¹²⁶ Yet, for individuals with mild "mental retardation," many of the tasks measured on the scales are not particularly

118. *Id.* at 51.

119. Sharon A. Borthwick-Duffy, *Adaptive Behavior*, in *Handbook of Intellectual and Developmental Disabilities* 279, 286 (John W. Jacobson et al. eds., 2007).

120. 2002 AAMR Manual, *supra* note 42, at 87.

121. For a discussion of reliability and validity, see *infra* Part III.A and accompanying text.

122. 2002 AAMR Manual, *supra* note 42, at 87.

123. *Id.* at 88-90.

124. Greenspan, *supra* note 31, at 214; *see also* Stephen Greenspan, *A Contextualist Perspective on Adaptive Behavior*, in *Adaptive Behavior*, *supra* note 20, at 64 [hereinafter Greenspan, *Contextualist Perspective*].

125. Greenspan, *supra* note 31, at 214.

126. *Id.*

challenging. By contrast, tasks that require complex cognitive and social abilities are more difficult for persons with mild intellectual deficits.

Greenspan postulates that the core limitations and vulnerabilities of those who are mildly “mentally retarded”—which is the category within which most *Atkins* defendants would fall—are not adequately tapped by modern adaptive behavior measures:

Most of [the challenges faced by those with mild mental retardation] involve dealing with other people and the games, manipulations, and deceptions that, unfortunately, characterize much of human behavior, particularly toward people who make easy targets. Thus if there is a universal quality that all people with mild MR possess and that defines its natural essence, it is vulnerability to social exploitation owing to an inability to understand other people, especially when their motives are malevolent but disguised as benevolent.¹²⁷

Greenspan posits that the core deficits that place persons with mild “mental retardation” at risk in society are “credulity” (“inability to see through untruthful assertions”) and “gullibility” (“ease with which one can be duped”).¹²⁸ These deficits make such individuals vulnerable to social exploitation. Greenspan observes that constructs reflecting deficits in *social* intelligence are all but absent from existing adaptive behavior measures.¹²⁹ This absence is problematic when one attempts to draw inferences from scores on these measures to the question of death-penalty eligibility in the *Atkins* context. Patton and Keyes list a range of characteristics that place mildly mentally retarded individuals at risk as criminal offenders, several of which reflect limitations in social intelligence: gullibility, acquiescence, naiveté or suggestibility, desire to please, and a desire to “pass as normal.”¹³⁰ Other limitations identified by these authors relate to the more “conceptual” intellectual realm, such as concrete thinking, memory issues, and language problems.¹³¹

Thus, in the final analysis, despite the intuitive appeal of the idea of measuring adaptive behavior, particularly within the social-ecological framework, the existing measurement instruments are inadequate. They fail to examine several key realms of intellectual functioning in which mildly intellectually disabled persons are likely to be deficient. In the *Atkins* context, the existing measures may have very little utility. Indeed, if the social-ecological model requires us to examine the individual’s abilities with reference to the demands of the *particular social context* in which he or she must function, we must then treat all inquiries about adaptive behavior as requiring us to first specify the nature of the particular context and situational demands to which the person must “adapt.” In other words, we can’t possibly

127. *Id.* at 215.

128. Greenspan, *Contextualist Perspective*, *supra* note 124, at 69.

129. *Id.* at 67–71; *see also* Greenspan, *supra* note 31, at 215.

130. Patton & Keyes, *supra* note 31, at 241.

131. *Id.*

begin to evaluate an individual's functioning in a given real-world situation without first identifying the specific cognitive, social, and practical demands that must be confronted successfully by those in that situation. Arguably, we must strive for a far more focused and situation-specific mode of assessment in order to determine whether an individual is truly disabled with reference to a particular context.

III. Striving for Valid *Atkins*' Evaluations

A. Basic Principles of Psychological Assessment

Tests, measuring scales, and other psychological assessment techniques must meet at least two fundamental conditions before they can be deemed useful:

First, the measuring instrument which is used on a given occasion and for a given purpose [must be *valid*. That is, it] must really measure the trait it is intended to measure. Second, the instrument must give a *reliable* measurement, so that we obtain the same result if we remeasure the trait under similar conditions Data should thus be dependable from two points of view—they should be *meaningful* and they should be *reproducible*.¹³²

The principle that measurement techniques must be both *reliable* and *valid* in order to satisfy the most minimal and basic scientific requirements is one of the first concepts learned by graduate students in psychology.

Although issues related to reliability may also affect the usefulness of *Atkins* assessments, my focus here is on validity. The most recent edition of the *Standards for Educational and Psychological Testing*, prepared by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, refers to validity as:

[T]he degree to which evidence and theory support the interpretations of test scores entailed by *the proposed uses of the tests*. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores required by *proposed uses* that are evaluated, not the test itself.¹³³

In other words, even if a test has been demonstrated to be valid for one type of use, it may not be valid for a different proposed use. This principle is central to the issues at stake in *Atkins* implementation. The tests that are typically used to measure “mental retardation” have been used in educational contexts (e.g., to address questions about the student's academic strengths and limitations and what types of special educational services must be provided) and eligibility for services and benefits (e.g., to address whether the individual qualifies for a

132. David Magnusson, *Test Theory* 59 (1967) (emphasis added).

133. Am. Educ. Research Ass'n et al., *Standards for Educational and Psychological Testing* 9 (1999) [hereinafter *AERA Standards*] (emphasis added).

vocational training or sheltered-living program, or for social security disability benefits). These tests were not developed for the purpose of distinguishing between capital offenders whose deficits in intellectual functioning render them ineligible for the death penalty and capital offenders without such deficits. Therefore, use of these measures for that purpose requires some persuasive evidence of the validity of the test for that application.

The analysis of the validity of a measure for use in a particular situation begins with what is referred to as *construct* validity. One must carefully articulate the nature and scope of the construct that the proposed use of a measure is expected to test.¹³⁴ “The detailed description [of the construct] provides a conceptual framework for the test, delineating the knowledge, skills, abilities, processes, or characteristics to be assessed.”¹³⁵ One form of evidence for validity focuses on the *content* of the measure and how that content relates to the construct of interest. Thus, in the *Atkins* context, if one accepts Professor Greenspan’s proposition that gullibility and credulity are among the deficits that place mildly mentally retarded individuals at risk of social exploitation in collaborative criminal endeavors, the absence of any content on adaptive behavior scales measuring these constructs is highly problematic.

Indeed, in the development of valid measures, one must move from one’s theory about the construct of interest to the development of test items that are expected to tap the overt manifestations of the construct. Ideally, the theory itself is derived from prior research findings, and the process of developing test items is guided by the scientific method. In one of the most influential articles written on the subject of psychological measurement, Professors Lee Cronbach and Paul Meehl emphasized the centrality of the scientific method to notions of *construct validity*.¹³⁶ The teachings of Cronbach and Meehl were recently restated by Professors Clark and Watson, who observed that construct validity requires the “(a) articul[ation of] a set of theoretical concepts and their interrelations, (b) develop[ment of] ways to measure the hypothetical constructs proposed by the theory, and (c) empirically testing the hypothesized relations among constructs with their observable manifestations.”¹³⁷ Thus, the process of testing the construct validity of a measure is analogous to the way in which the scientific method applies theory, hypothesis-testing through experimentation, and interpretation of findings as means of confirming or rejecting hypotheses.¹³⁸ Psychologist Samuel Messick states:

This comprehensive view of validity integrates considerations of content,

134. *Id.*

135. *Id.*

136. Lee J. Cronbach & Paul E. Meehl, *Construct Validity in Psychological Tests*, 52 *Psychol. Bull.* 281, 291 (1955).

137. Lee Anna Clark & David Watson, *Constructing Validity: Basic Issues in Objective Scale Development*, in *Methodological Issues & Strategies in Clinical Research* 215–17 (Alan E. Kazdin ed., 2d ed. 1998).

138. For further discussion of Cronbach and Meehl’s legacy and subsequent development of concepts of construct validity, see Gregory T. Smith, *On Construct Validity: Issues of Method and Measurement*, 17 *Psychol. Bull.* 396 (2005).

criteria, and consequences into a construct framework for empirically testing rational hypotheses about score meaning and utility. Therefore, it is fundamental that score validation is an empirical evaluation of the meaning and consequences of measurement. As such, validation combines scientific inquiry with rational argument to justify (or nullify) score interpretation and use.¹³⁹

The commonly-applied scientific methods relevant to the development, evaluation, and refining of diagnostic tools require us to examine the relationship between the real-world phenomena that are the target of measurement and what we use to try to assess those phenomena. If the target of measurement is of particular relevance to the law, perhaps defined by a legal standard, the tasks for theorists, researchers, and evaluators are to identify the component aspects of psychological functioning that are relevant to the phenomenon of interest to the law, and to develop operational definitions of those components. Because operational definitions translate the construct of relevance into observable and purportedly-measurable phenomena, they form the bridge between the constructs of interest to the law and the psychological methods intended to assess them. The development and use of those methods in the particular instance should reflect a form of hypothesis-testing which, in turn, further contributes to our knowledge about the phenomena that are the subject of evaluation and how to best measure them.

B. Principles of “Psycholegal Assessment”

The past thirty years have witnessed a virtual revolution in the ways in which we think about psychological evaluations in legal contexts. “Psycholegal assessment” or “forensic psychological assessment” refers to evaluations of individuals conducted by professionals such as psychologists, psychiatrists, educational specialists or others, for the purpose of assisting the courts or other legal actors (e.g., legislators).¹⁴⁰ The theoretical and empirical work in the specialty of psycholegal assessment grew out of dissatisfaction

139. Samuel Messick, *Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning*, 50 *Am. Psychologist* 741, 742 (1995). In the death penalty context, it is difficult to apply traditional notions of criterion validity. Criterion-relevant evidence can be *concurrent* or *predictive*. See AERA Standards, *supra* note 133, at 14–15. Concurrent methods compare the findings of the measure in question with findings obtained by another test deemed to assess the same construct. *Id.* If the results of both tests are highly correlated, that correlation provides support for the validity of the new measure. Predictive methods evaluate the degree to which the measure in question accurately predicts some future aspect of performance or functioning. *Id.* Thus, for example, studies of how well college admission examinations (such as the SAT) predict grade point averages provide evidence for the validity of the SAT as a test used to *predict* success in college. In the *Atkins* context, however, criterion comparisons are difficult. One might view the 1992 AAMR Manual's reference to adaptive behavior measures as a check on the concurrent validity of the IQ scores used to diagnose mental retardation. See *supra* note 102 and accompanying text. But, as noted above, it is not clear that either traditional IQ measures or standardized scales of adaptive behavior tap some of the most central ways in which intellectually disabled capital defendants might be impaired. Thus, there really are no suitable concurrent measures or predictive outcomes against which to evaluate any proposed assessment methods.

140. See Gary B. Melton et al., *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers* 43 (3d ed. 2007); see also Kirk Heilbrun, *Principles of Forensic Mental Health Assessment* 3 (2001).

with the sources of data used by expert witnesses who testify about their clinical assessments of individual behavior. The courtroom testimony of many mental health experts failed to focus on the precise question of relevance to the specific legal inquiry.¹⁴¹ Thus, for example, when asked to render an opinion about whether a particular defendant is competent to stand trial, a professional might conduct a general evaluation of the defendant's functioning, including diagnosis of mental disorders and/or limitations of intellectual abilities with standardized IQ tests. Experts would then make giant inferential leaps and opine about the defendant's competence to stand trial or lack thereof.¹⁴² Such evaluations were problematic, however, because they did not examine the precise aspects of psychological functioning relevant to the legal inquiry.¹⁴³ Thus, for example, in *Dusky v. United States*, the Supreme Court set forth a test of "competency to stand trial" that requires a defendant to be able to "consult with his attorney with a reasonable degree of rational understanding" and to have a "rational" and "factual understanding of the proceedings against him."¹⁴⁴ A meaningful forensic assessment should attempt to focus more directly on the facets of the individual's abilities at the core of the legal standard. While diagnoses of mental disorders or intellectual disabilities might ultimately help explain *why* a particular defendant could not understand the nature of the charges against him or could not communicate effectively with the lawyer, the diagnoses do not directly inform the initial question of whether the prongs of the Supreme Court's test are satisfied.

Psychologist Thomas Grisso, a pioneer in the measurement of legally-relevant competencies, emphasizes that the observations of, and data collected by, the forensic evaluator must be logically linked to the "*specific abilities and capacities with which the law is concerned.*"¹⁴⁵ Consistent with our preceding discussion of validity of assessment instruments more generally, Grisso asserts that legal competencies are themselves *constructs*.¹⁴⁶ Grisso notes that despite "the elusive quality of legal competence constructs," there are systematic ways to proceed in attempting to provide assessments relevant in particular legal contexts.¹⁴⁷ And, indeed, in the past several decades, researchers have focused substantial efforts on elucidating the precise questions about human functioning and behavior that are of relevance to the law in a range of contexts.¹⁴⁸

141. Thomas Grisso, *Evaluating Competencies: Forensic Assessments and Instruments* 12 (2d ed. 2003).

142. See, e.g., Ronald Roesch & Stephen L. Golding, *Competency to Stand Trial* 69–82 (1980) (discussing diagnostically-oriented testimony).

143. *Id.* at 10–43.

144. Melton et al., *supra* note 140, at 127–28 (citing *Dusky v. United States*, 352 U.S. 402 (1960)).

145. Grisso, *supra* note 141, at 13 (emphasis added).

146. *Id.* at 22.

147. *Id.* at 23.

148. See generally Thomas Grisso & Paul Appelbaum, *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals* (1998); Thomas Grisso, *Clinical Evaluations for Juveniles' Competence to Stand Trial: A Guide for Legal Professionals* (2005); Grisso, *supra* note 141; Thomas Grisso, *Evaluating Juveniles' Adjudicative Competence* (2005); John Monahan et al., *Rethinking Risk*

There appears to be a strong convergence of the principles that guide this functional, contextual approach to psycholegal assessment just described and the tenets of the social-ecological disability model discussed in Parts II and III.A of this Article. The work of both groups is informed by rigorous application of fundamental principles of psychological assessment. The extension of those fundamental principles to *Atkins* assessments mandates a close relationship between the measures used to evaluate offenders and the underlying constructs that guided¹⁴⁹ the Court in exempting persons referred to as “mentally retarded” from the reach of the death penalty. As I hope the prior analysis demonstrates, it is unlikely that the currently-available measures used to diagnose “mental retardation” meet these rigorous standards, particularly in light of the high stakes.

IV. A Tale of Two Jurisdictions: Florida and California

The Supreme Courts of Florida and California have each had the opportunity to consider precisely how to apply the *Atkins*’ mandate. The statutes of both states are conceptually similar, although Florida’s statute is more specific in delineating what constitutes subaverage general intellectual functioning (i.e., scores that are two standard deviations below the mean) and adaptive behavior (i.e., “the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected of his or her age, cultural group, and community.”)¹⁵⁰ The California statute states only that, “[a]s used in this section, ‘mentally retarded’ means the condition of significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested before the age of 18.”¹⁵¹

Assessment: The MacArthur Study of Mental Disorder and Violence (2001); Norman G. Poythress, Jr. et al., *Adjudicative Competence: The MacArthur Studies* (2002); Roesch & Golding, *supra* note 142; Lois A. Weithorn & Susan B. Campbell, *Children’s Capacities to Render Informed Treatment Decisions*, 53 *Child Dev.* 1589 (1982); Lois A. Weithorn & Thomas Grisso, *Psychological Evaluations in Divorce Custody: Problems, Principles, and Procedures*, in *Psychology and Child Custody Determinations: Knowledge, Roles, and Expertise* 157 (Lois A. Weithorn ed., 1987).

149. It is not always clear precisely what psychological constructs guide the Court’s (or any other legal body’s) decisions regarding psycholegal phenomena. Perhaps the Court (or legislature) did not articulate the underlying constructs clearly (or at all). Thus, the process of determining which aspects of psychological functioning are at the core of the inquiry may require substantial analysis and inference. In the absence of evidence as to what psychological concepts guided the relevant legal body, one can look elsewhere for guidance, such as to scholarly or other sources.

150. Fla. Stat. Ann. § 921.137(1) (West 2008) reads:

As used in this section, the term “mental retardation” means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the period from conception to age 18. The term “significantly subaverage general intellectual functioning,” for the purpose of this section, means performance that is two or more standard deviations from the mean score on a standardized intelligence test specified in the rules of the Agency for Persons with Disabilities. The term “adaptive behavior,” for the purpose of this definition, means the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected of his or her age, cultural group, and community.

151. Cal. Penal Code § 1376(a) (West 2008).

The approaches of the two states in implementing these statutes, however, differ dramatically. In *Cherry v. State*, the Florida Supreme Court considered an offender's claim for post-conviction relief under *Atkins*.¹⁵² In support of his claim, the defendant's expert witness, Dr. Bursten, gathered history on the defendant's background, administered the WAIS-III, and interviewed three people who knew the defendant before the offense.¹⁵³ The defendant had scored as follows on various IQ tests throughout his life: 71 in 1968, 85 in 1972, 79 in 1976, 86 in 1979, 68 in 1987, 72 in 1992, 78 in 1996, and a Full Scale IQ score of 72 when the WAIS-III was administered by Dr. Bursten in 2005.¹⁵⁴ Dr. Bursten and a second defense expert testified that proper interpretation of the WAIS score required a consideration of the standard error of measurement.¹⁵⁵ Dr. Bursten stated:

The concept of mental retardation is considered to be a range or band of scores, not just one score or a specific cutoff for mental retardation. The idea behind that is there's recognition that no one IQ is exact or succinct, that there's always some variability and some error built in. . . . The [DSM] guides us to look at IQ scores as being a range rather than absolute. And, the manual talks about a score from 65, a band, so to speak, from 65 and 75—and of course, lower than 65—comprising mental retardation.¹⁵⁶

The court considered whether the statute creates a "strict cutoff" of an IQ score at 70 "in order to establish significantly subaverage intellectual functioning," or whether scores presented by evaluators need to be interpreted in light of the standard error of measurement.¹⁵⁷ The court answered the question as follows: "the statute does not use the word approximate, nor does it reference the [standard error of measurement]. Thus, the language of the statute and the corresponding rule are clear. We defer to the plain meaning of the statute."¹⁵⁸ The court thus established a firm and unyielding cutoff of 70 as the score above which defendant's claims to be mentally retarded would be rejected. As such, it rejected the recently-obtained score of 72 as indicative of mental retardation.¹⁵⁹

The California Supreme Court's approach is quite different. *In re Hawthorne* addressed the same question before the *Cherry* court: Can the state adopt an IQ score of 70 as the upper cut-off for a prima facie showing of mental retardation?¹⁶⁰ Noting legislative silence as to the appropriateness of any particular numerical cut-off score, recognizing the problems of measurement error and other factors, the court refrained from establishing a

152. 959 So. 2d 702 (Fla. 2007). The Florida Supreme Court also upheld lower courts' denials of defendants' petitions for post-conviction relief in *Brown v. State*, 959 So. 2d 146 (Fla. 2007), and *Jones v. State*, 966 So. 2d 319 (2007).

153. *Cherry*, 959 So. 2d at 711.

154. *Id.*

155. *Id.*

156. *Id.* at 711–12.

157. *Id.* at 712.

158. *Id.* at 713.

159. *Id.* at 714.

160. 105 P.3d 552, 557 (Cal. 2005).

priori interpretive rules.¹⁶¹ In the case before it, the court upheld the lower court's finding that the defendant was mentally retarded and therefore not eligible for the death penalty.¹⁶² In addition, the court in *Hawthorne* left the door open to more wide-ranging testimony by experts on the question of mental retardation.

The California Supreme Court's decision in *People v. Vidal*¹⁶³ in 2007 revealed the flexibility of this type of open-ended standard. Multiple sets of intelligence test scores in *Vidal* revealed a consistent pattern: significant discrepancies between the Verbal and Performance Scale scores on the Wechsler tests.¹⁶⁴ The Verbal Scale scores ranged from 59 to 77; the Performance Scale scores ranged from 96 to 126.¹⁶⁵ The Full Scale scores ranged from 81 to 96.¹⁶⁶ The expert for the defense testified that, although the defendant's Full Scale IQ scores were above the levels typically considered to be in the range of mental retardation, his Verbal Scale scores were quite low.¹⁶⁷ The trial court had found convincing the expert's testimony that the defendant manifested "significantly subaverage general intellectual functioning" because

[h]is "very low scores in terms of verbal I.Q.," even if due to a deficit in auditory processing rather than to low intellectual functioning "across the board," demonstrated a significant deficit in his "ability to process information and handle it adequately and to think logically." . . . The court further observed that Verbal IQ was particularly relevant in applying *Atkins* because "[w]e are talking about issues of premeditation, deliberation, appreciation of concepts of wrongful conduct, ability to think and weigh reasons for and not for doing things and logic, foresight, and all of those are related to verbal I.Q." Accepting the existence of the Flynn effect, the court also noted that "one or two point" gaps between IQ scores and the theoretical cutoff were not persuasive. Finding *Vidal* also met the remainder of the statutory definition of mental retardation, the court ordered the prosecution precluded, under section 1376, from seeking the death penalty.¹⁶⁸

The contrast of the Florida and California approaches is reminiscent of the familiar juxtaposition of rules and standards.¹⁶⁹ Each approach, in the abstract, has features that may be advantageous or problematic, depending upon the circumstances.¹⁷⁰ Rigid rules that result in arbitrary decisions—such as

161. *Id.* at 557–58.

162. *Id.* at 559.

163. 155 P.3d 259 (Cal. 2007).

164. *Id.* at 261.

165. *Id.*

166. *Id.*

167. *Id.*

168. *Id.* at 263 (citation omitted).

169. The rules versus standards debate has been the subject of much scholarly commentary. *See, e.g.*, Eric A. Posner, *Standards, Rules, and Social Norms*, 21 Harv. J.L. & Pub. Pol'y 10 (1997); Pierre Schlag, *Rules and Standards*, 33 UCLA L. Rev. 379 (1985); Kathleen M. Sullivan, *The Supreme Court 1991 Term: Foreword: The Justices of Rules and Standards*, 106 Harv. L. Rev. 22 (1992).

170. For example, Professor Kathleen Sullivan identifies some of the distinguishing features of rules versus standards:

Rules, once formulated, afford decisionmakers less discretion than do standards . . .

Florida's categorical refusal to consider the standard error of measurement—promote unfairness and undercut the public's trust in our system of justice. Yet, unguided discretion, particularly as it concerns the testimony of expert witnesses, carries its own risks. *Hawthorne* may invite substantial inconsistency in application of the statute from one case to the next, given the breadth of interpretations of data about which it allows experts to testify.¹⁷¹ That said, the substance of the expert testimony in the *Vidal* case has greater intuitive appeal, and seems more consistent with the principles of validity articulated in Part III above, irrespective of whether it ultimately constituted a correct interpretation of the test findings. The expert in *Vidal* not only acknowledged some of the psychometric limitations of the test administered, but he also tried to provide a more functionally-oriented interpretation of the meaning of the test scores in light of the particular question of relevance to the court. He focused on the specific psychological capacities which, if impaired, might limit an individual's ability to make wise choices and refrain from socially harmful conduct. Given the lack of theoretical and empirical analysis on the validity of using standardized intelligence tests for *Atkins* determinations, neither I or the expert in *Vidal* know whether a nuanced interpretation of particular WISC or WAIS subtest scores might provide more relevant information than do the Full Scale scores most typically entered into evidence. Conceptually, an inquiry that focuses on the defendant's capacities and limitations in those areas of functioning articulated by the Court in *Atkins* holds the greatest likelihood of providing meaningful information about the appropriateness of precluding application of the death penalty on a case by case basis.

Yet, as the comparison between Florida's and California's use of standardized IQ tests suggests, there are noteworthy inconsistencies in the ways in which state courts are using these tests. This result is disturbing in light of the dramatic real-world consequences of the application of these tests in the *Atkins* context. A defendant with Full Scale IQ scores ranging from 68 to 86 was determined to be eligible for the death penalty in Florida, while a

(a) *Rules*.—A legal directive is “rule”-like when it binds a decisionmaker to respond in a determinate way to the presence of delimited triggering facts. Rules aim to confine the decisionmaker to facts, leaving irreducibly arbitrary and subjective value choices to be worked out elsewhere. . . . A rule necessarily captures the background principle or policy incompletely and so produces errors of over- or under-inclusiveness. But the rule's force as a rule is that decisionmakers follow it, even when direct application of the background principle or policy to the facts would produce a different result.

(b) *Standards*.—A legal directive is “standard”-like when it tends to collapse decisionmaking back into the direct application of the background principle or policy to a fact situation. Standards allow for the decrease of errors of under- and over-inclusiveness by giving the decisionmaker more discretion than do rules. Standards allow the decisionmaker to take into account all relevant factors or the totality of the circumstances. Thus, the application of a standard in one case ties the decisionmaker's hand in the next case less than does a rule—the more facts one may take into account, the more likely that some of them will be different the next time.

Sullivan, *supra* note 169, at 57–59 (footnotes omitted).

171. For a further discussion of *Hawthorne*, see Christopher Scott Tarbell, Note, *Implementing Atkins: The Strengths and Weaknesses of Hawthorne*, 56 *Hastings L.J.* 1249 (2005).

defendant with Full Scale scores of 81 to 96 was found to be ineligible in California. We do not have additional information about the adaptive functioning of the two defendants in question, or about behavior that would have been indicative of constructs such as social or practical intelligence. Nor do we have background information about the offenders, about the nature and circumstances of their offenses, or about other potential indicia of criminal culpability. Thus, we do not know how the results in these cases were affected by other information before the trial courts. But, this cursory jurisdictional comparison suggests troubling disparities in the ways in which the tests are used.

Conclusion

There are many topics relating to *Atkins* evaluations that this Article did not address. It did not address the Flynn Effect—a poorly-understood phenomenon that causes population IQ scores to rise over time—and whether and how examiners should “correct” for this effect when reporting IQ scores in *Atkins* evaluations.¹⁷² It did not address inflation of IQ scores resulting from the “practice effect,” that is, from repeated administrations of the same IQ test on the same subject.¹⁷³ It did not address questions of whether retroactive reports on an individual’s adaptive behavior from five, ten, fifteen, or twenty years ago constitute valid data sources for *Atkins* evaluations, or whether measurement of the current adaptive behavior of a person who has lived on Death Row for ten years is at all relevant to questions about that person’s level of adaptive behavior at the time of the offense.¹⁷⁴ This Article did not address these and many other challenges that confront those who perform assessments of “mental retardation” in the *Atkins* context because all of those issues relate to the standards and challenges that should govern *administration* and *interpretation* of the current IQ tests and adaptive behavior scales in the *Atkins* context. My focus in this Article, however, is on the more fundamental question of whether the currently applied definitions of mental retardation and currently-available measures of intellectual and adaptive functioning are appropriate for the *Atkins* context.¹⁷⁵

172. See James R. Flynn, *Tethering the Elephant: Capital Cases, IQ, and the Flynn Effect*, 12 Psychol. Pub. Pol’y & L. 170 (2006); see also Tomoe Kanaya et al., *The Flynn Effect and U.S. Policies: The Impact of Rising IQ Scores on American Society via Mental Retardation Diagnoses*, 58 Am. Psychologist 778 (2003).

173. See LaJuana Davis, *Intelligence Testing and Atkins: Considerations for Appellate Courts and Appellate Lawyers*, 5 J. App. Prac. & Process 297, 309–10 (2003); Julie C. Duvall & Richard J. Morris, *Assessing Mental Retardation in Death Penalty Cases: Critical Issues for Psychology and Psychological Practice*, 37 Prof. Psychol.: Res. & Prac. 658, 663 (2006).

174. See Greenspan & Switzky, *supra* note 100, at 294–95; see also Bethany Young et al., *Four Practical and Conceptual Assessment Issues That Evaluators Should Address in Capital Case Mental Retardation Evaluations*, 38 Prof. Psychol.: Res. & Prac. 169, 171, 174–75 (2007).

175. Of course, those practitioners performing *Atkins* evaluations with only the tools presently available should heed the advice of those writers who have identified the most responsible methods of administering and interpreting currently available measures. For thoughtful analyses addressing these issues, see Bonnie & Gustafson, *supra* note 6; Richard J. Bonnie, *The American Psychiatric Association’s Resource Document on Mental Retardation and Capital Sentencing: Implementing Atkins v. Virginia*, 32 J. Am. Acad. Psychiatry &

As this Article demonstrates, the use of IQ or adaptive behavior scale *summary* scores in the *Atkins* context is not supported by theory or empirical studies. State legislatures and courts—guided to some extent by the Supreme Court’s reference to the AAMR and DSM-IV-TR definitions in Footnote 3 in *Atkins*¹⁷⁶—have presumed that the assessment methods used to evaluate “mental retardation” in educational and social service settings can be employed in the death penalty context without modification, reservation, or additional scrutiny. Yet, modern theory and research in the field of intellectual disability argue against such a presumption. The social-ecological model of intellectual disability recognizes that the functioning of intellectually-challenged individuals will vary from one situation to the next, as environmental demands interact with each individual’s particular skills, abilities, and deficits. Traditional IQ tests and adaptive behavior scales leave entire domains of functioning relating to social and practical intellectual skills untapped. Indeed, many of the manifestations of intellectual disability cited by Justice Stevens as the basis for the *Atkins* holding are not tapped by these measures. Thus, use of summary scores on these traditional measures to determine whether or not a particular criminal defendant is “mentally retarded” for the purpose of *Atkins* will yield an underinclusive group of excluded defendants.

Deleted: is likely to

Deleted:

There is a critical need for research identifying valid assessment methods for *Atkins* evaluations. Attempts to develop such methods must focus first on the specific constructs underlying the Court’s concerns in *Atkins*, supplemented by relevant theoretical and scientific work relevant to those constructs, and to modern notions of “intellectual disability.” Such efforts should be informed by: (1) basic psychometric principles, particularly those addressing the need to demonstrate the validity of measures *with respect to the particular applications or uses of the measures*; (2) modern strategies and models of psycholegal assessment; and (3) progressive theories defining intellectual disability, intelligence, and adaptive functioning. Given the dramatic consequences of a finding of “mental retardation” in the death penalty context, it is of paramount importance that clinical experts ground their testimony in firm scientific foundations.

Even if vigorous efforts to develop *Atkins*-relevant assessment tools begin today, it is likely that useful measures would not be available to practitioners for years. Thus, in light of my conclusions, how should legislatures, courts, lawyers, and clinical experts address the question of death penalty eligibility under *Atkins*? Given that the death penalty is the most severe punishment available in our criminal justice system and—once carried out—it is irrevocable, state policies must err on the side of casting a net that is too wide rather than one that is too narrow in defining “mental retardation” for the purpose of death penalty eligibility. Many (if not most) of the individuals who score approximately two standard deviations below the mean on *either* IQ tests

L. 304 (2004); Duvall & Morris, *supra* note 173; Young et al., *supra* note 174.

¹⁷⁶ See, e.g., *supra* notes 33-39.

or adaptive behavior scales are likely to manifest many of the deficits described by the Court in *Atkins* and should therefore be exempt from the death penalty's reach. Yet, given the failure of these scales to evaluate some of the manifestations of intellectual disabilities emphasized by the Court and of particular relevance to criminal offending, below par scores on either of these traditional measures should be sufficient, *but not necessary*, for exclusion from the death penalty. Defendants whose scores on either measure do not make the "cut" should have the opportunity to demonstrate "mental retardation" using a combination of data sources, alternative data sources, and/or alternative interpretation of data,¹⁷⁷ where the evidence proffered is: (1) consistent with the Court's concerns in *Atkins* and (2) grounded in a functional analysis of a defendant's specific deficits in the context of the particular capital offense committed. This third manner of demonstrating "mental retardation" is wholly consistent with modern theory and research on "mental retardation" (i.e., "intellectual disability") and, more fundamentally, on the nature and measurement of the construct of intelligence.

¹⁷⁷ The challenged testimony in *Vidal*, see supra notes 163-168, provides an example of an "alternative interpretation of data."