

5-1-2007

The Challenge of Implementing *Atkins v. Virginia*: How Legislatures and Courts Can Promote Accurate Assessments and Adjudications of Mental Retardation in Death Penalty Cases

Richard J. Bonnie
University of Virginia

Katherine Gustafson
University of Virginia School of Law

Follow this and additional works at: <https://scholarship.richmond.edu/lawreview>

Part of the [Criminal Law Commons](#), [Criminal Procedure Commons](#), and the [Law and Society Commons](#)

Recommended Citation

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 (2007).
Available at: <https://scholarship.richmond.edu/lawreview/vol41/iss4/3>

This Article is brought to you for free and open access by the Law School Journals at UR Scholarship Repository. It has been accepted for inclusion in University of Richmond Law Review by an authorized editor of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

ARTICLES

THE CHALLENGE OF IMPLEMENTING *ATKINS V. VIRGINIA*: HOW LEGISLATURES AND COURTS CAN PROMOTE ACCURATE ASSESSMENTS AND ADJUDICATIONS OF MENTAL RETARDATION IN DEATH PENALTY CASES

Richard J. Bonnie *

Katherine Gustafson **

INTRODUCTION

In *Penry v. Lynaugh*, decided in 1989, a closely divided United States Supreme Court declined to rule that execution of a person with mental retardation is categorically prohibited by the Eighth Amendment,¹ holding instead that it was constitutionally permissible for the states to regard mental retardation as a mitigating factor rather than a preclusive one.² The Court observed that a national consensus supporting exemption of all defendants with mental retardation from the death penalty had not emerged.³ Of course, this ruling left the states free to adopt a categorical exclusion, and many states subsequently did so.

* John S. Battle Professor of Law, Professor of Psychiatry and Neurobehavioral Sciences, and Director, Institute of Law, Psychiatry and Public Policy, University of Virginia.

** J.D. Candidate 2008, University of Virginia School of Law.

1. 492 U.S. 302, 335 (1989) (5–4 decision) (“[T]here is insufficient evidence of a national consensus against executing mentally retarded people . . . to conclude that it is categorically prohibited by the Eighth Amendment.”).

2. See *id.* at 340 (O’Connor, J., concurring).

3. *Id.* at 335.

Thirteen years later, in *Atkins v. Virginia*,⁴ a 6–3 majority ruled that a national consensus favoring categorical exclusion of defendants with mental retardation from capital punishment had become evident since 1989,⁵ and that the “evolving standards of decency that mark the progress of a maturing society” must be judged by the standards that prevail at the time of the Court’s decision.⁶

Between *Penry* and *Atkins*, eighteen states enacted legislation specifically banning the death penalty for people with mental retardation, and similar bills were passed by at least one house of the legislature in at least three other states.⁷ The *Atkins* Court noted that while the number of states excluding people with mental retardation had increased significantly (only two had so at the time of *Penry*), not a single state had legislated in the opposite direction by reinstating the death penalty for such offenders.⁸ Taking account of the twelve states that had abolished the death penalty altogether, execution of a person with mental retardation was forbidden in thirty states in 2002. The federal death penalty statute also precluded death sentences for offenders with mental retardation.⁹

In *Atkins*, the Supreme Court expressly acknowledged that determining whether a specific individual is mentally retarded is a sometimes controversial task, as it has been in Daryl Atkins’s own case.¹⁰ The Court stated 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.”¹¹ The Court clearly anticipated continuing struggles in the effort to implement its decision, and it is hardly surprising that the Court left to the states the task of putting the new constitutional rule into effect.¹² However, the Court did not leave the states without some definitional guidance: Based on the definitions of mental re-

4. 536 U.S. 304 (2002).

5. *Id.* at 315–16.

6. *See id.* at 311–12 (quoting *Trop v. Dulles*, 356 U.S. 86, 101 (1958)).

7. *Id.* at 314–15.

8. *Id.* at 313–16.

9. *Id.* at 314 (citing Anti-Drug Abuse Act of 1988, Pub. L. No. 100-690, § 7001(l), 102 Stat. 4181, 4390, *repealed by* USA PATRIOT Improvement and Reauthorization Act, Pub. L. No. 109-177, § 221 120 Stat. 192, 231 (2006)).

10. *Id.* at 317.

11. *Id.*

12. *Id.*

tardation enunciated by the American Association on Mental Retardation and by the American Psychiatric Association,¹³ the Court observed that mental retardation has three components: subaverage intellectual functioning, significant limitations in adaptive skills, and onset before age eighteen.¹⁴

The Supreme Court's approach in *Atkins* is noteworthy for many reasons. For example, *Atkins* may have marked the emergence of an approach to an Eighth Amendment excessiveness analysis accepted by a majority of the Justices—a development further evidenced in *Roper v. Simmons*, in which a 5–4 majority declared that the death penalty is constitutionally excessive as applied to adolescent offenders who were under eighteen at the time of the offense.¹⁵

One particularly striking feature of *Atkins* is that it enunciated a constitutional rule that turns explicitly and entirely on a clinical diagnosis. Although clinical diagnoses often serve as a threshold requirement in legal “tests” of incompetence, non-responsibility, and disability, they are almost never sufficient to establish that the legal criteria are satisfied. Instead, the typical legal formulation requires the party seeking to prove incompetence, non-responsibility, or disability to demonstrate a significant level of functional impairment as a result of the diagnosed condition. The most pertinent examples in criminal law are incompetence to proceed to trial or to plead guilty in a criminal case (impaired ability to understand the proceedings or to assist counsel),¹⁶ and insanity or lack of criminal responsibility (impaired ability to appreciate the wrongfulness of one's conduct or to conform one's conduct to the requirements of the law).¹⁷

There is much to be said in favor of a principle of diminished responsibility in capital sentencing—e.g., precluding a death sentence for an offender whose capacity to appreciate the wrongfulness of his conduct, to exercise rational judgment, or to conform his conduct to the requirements of law was significantly impaired

13. *Id.* at 308 n.3.

14. *Id.* at 318.

15. 543 U.S. 551, 568 (2005).

16. See *Godinez v. Moran*, 509 U.S. 389, 402 (1993); *Drope v. Missouri*, 420 U.S. 162, 171 (1975); *Dusky v. United States*, 362 U.S. 402, 402 (1960). See generally Richard J. Bonnie, *The Competence of Criminal Defendants: Beyond Dusky and Drope*, 47 U. MIAMI L. REV. 539 (1993).

17. Model Penal Code § 4.01(1) (1985).

at the time of the offense. Indeed, the American Bar Association, the American Psychiatric Association, and the American Psychological Association recently have urged states to adopt such a principle in cases involving offenders with severe mental illness.¹⁸

In light of the strong legal tradition just described, the Supreme Court might have chosen to embrace a principle of diminished responsibility as a constitutional culpability requirement in capital cases—similar to the minimum level of culpability needed to impose a death sentence on accomplices (“reckless indifference to human life”).¹⁹ What is so striking about *Atkins* is that the Court was persuaded to adopt a per se rule exempting all persons with mental retardation from the death penalty based on diagnosis alone. This is all the more remarkable in light of the fact that the line between being mentally retarded and being of borderline intelligence is operationalized statistically by an admittedly arbitrary line on a bell curve representing performance on an IQ test two standard deviations below the mean. In short, the diagnosis of mental retardation is in large part a statistical construct.

The Supreme Court did not attempt to explain, in constitutional terms, why the Eighth Amendment’s concern with diminished culpability—the concept lying at the heart of the Court’s reasoning in *Atkins*—is best operationalized by exempting offenders diagnosed with mental retardation rather than through a standard test of diminished responsibility. Although the Court can properly be faulted for failing to offer a more elaborate explanation, the decision can justifiably be grounded in either of two juridical judgments—either that *no* person with mental retardation is sufficiently culpable to deserve a death sentence, or that any criterion of diminished responsibility designed to differentiate the few mentally retarded offenders deserving of the death penalty from the many who are not would likely lead, in practice, to death sentences for too many undeserving offenders. Indeed, capital juries may have a tendency to discount the mitigating significance of evidence of mental retardation and severe mental illness because the offenses committed are unusually brutal or raise concerns about future violence. A similar rationale underlies the

18. Task Force on Mental Disability & the Death Penalty, Am. Bar Ass’n, *Recommendation and Report on the Death Penalty and Persons with Mental Disabilities*, 30 MENTAL & PHYSICAL DISABILITY L. REP. 668, 668–69 (2006).

19. *Tison v. Arizona*, 481 U.S. 137, 158 (1987).

subsequent ruling in *Roper v. Simmons* precluding execution of any offender who was less than eighteen at the time of the offense.

The analogy with *Roper v. Simmons* is quite instructive. The criminal law has long drawn responsibility-related lines based on the ages of youthful offenders, including the common-law infancy presumptions and the jurisdictional age lines defining the boundaries between juvenile and criminal courts. The innovation in the mental retardation exclusion statutes enacted between 1988 and 2002 is that they essentially analogized having a diagnosis of mental retardation to being below a statutorily designated age of eligibility, thereby evading the tradition of individualized capacity assessments by forensic mental health professionals that have been customarily required in the context of criminal adjudication. When the Court decided to constitutionalize the exclusion in *Atkins*, it simply embraced the diagnostic approach—a recent innovation in the state legislatures—without any apparent consideration of the diminished responsibility approach.

Three years later, with *Atkins* having laid the groundwork, the *Roper* Court was willing to take the next step, drawing an admittedly arbitrary constitutional line at the age of eighteen. But there is a big practical difference between the otherwise parallel rulings in *Roper* and *Atkins*: 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.

In the usual forensic context, it is well understood that adjudications of competence and responsibility rest ultimately upon value judgments and that there is often no right answer. However, the sole question in an *Atkins* adjudication is whether the offender is or is not mentally retarded. Although the Supreme Court was well aware of the uncertainties at the margins of any mental disorder diagnosis, including the diagnosis of mental retardation, it also must have assumed that the inevitable "disagreement[s] . . . in determining which offenders are in fact re-

tarded"²⁰ can be resolved fairly and with reasonable accuracy based on scientific and clinical knowledge. At the very least, the Court must have assumed that expert disagreements would rest on articulable differences in scientific or clinical judgment, rather than on hidden disagreements about whether the offender deserves a death sentence. In this sense, the very soundness of the *Atkins* decision, and the integrity of *Atkins* adjudications, turns on the effort made by the states to implement it in a scientifically satisfactory manner.

Our goal in this paper is to assist state courts and legislatures as they try to carry out the task that *Atkins* requires of them—promoting fairness and accuracy in the assessment and adjudication of mental retardation. After addressing the definition of mental retardation in Part I, we focus on its assessment in Parts II and III, highlighting several key requirements of a scientifically and clinically adequate assessment.

Part II addresses the assessment of deficits in intellectual functioning, particularly on the measurement of intelligence as represented by an intelligence quotient. Appropriate IQ tests must be used, and the scores must be interpreted in accordance with professional practice, taking into account the scores' statistical meaning, the standard error of measurement, and other factors that affect subjects' performance on these tests and the interpretations of their scores.

Part III addresses deficits in adaptive behavior, the second key component of a diagnosis of mental retardation. Adaptive behavior is difficult to measure and the scientific community's understanding of the construct is not nearly as advanced as its understanding of IQ. We offer guidance on how to operationalize the concept of adaptive behavior and how to give it meaningful consideration in *Atkins* adjudications. Part IV briefly addresses the practical significance of the requirement that mental retardation be manifest before the age of eighteen. Finally, Part V discusses the role of expert testimony and the necessary qualifications for experts testifying on the issue of mental retardation in death penalty cases.

Although the paper is designed to offer guidance to all state courts and legislatures, Virginia is used as a case study. Perhaps

20. *Atkins v. Virginia*, 536 U.S. 304, 317 (2002).

this is fitting in light of the fact that the *Atkins* rule was enunciated in a Virginia case, and that litigation continues concerning its application to Daryl Atkins himself.²¹ In the wake of the Supreme Court's *Atkins* ruling in May 2003, the Virginia General Assembly adopted legislation defining mental retardation, setting requirements for the assessment of mental retardation, and establishing the necessary qualifications for experts.²² The legislation was drafted under the auspices of the Virginia State Crime Commission (the "Crime Commission"), a bipartisan body of legislators from the Senate and House of Delegates. After the Supreme Court handed down *Atkins*, the Crime Commission appointed a Subcommittee to draft legislation to implement the decision (the "*Atkins* Subcommittee") comprised of legislators and representatives from various criminal justice constituencies. The portions of the bill dealing with the definition of mental retardation, its assessment and expert qualifications—the only portions addressed in this paper—were formulated by a Clinical Advisory Group composed of mental health professionals appointed by the *Atkins* Subcommittee.²³ These provisions were embraced by the *Atkins* Subcommittee and the Crime Commission largely as written by the Clinical Advisory Group. The legislation implementing *Atkins* was eventually passed by the Virginia General Assembly with only minor amendments to the version proposed by the Crime Commission.²⁴

Because the provisions in the Virginia statute pertaining to assessment of mental retardation are more detailed than other state laws on this subject, they may offer useful guidance to expert evaluators and courts in other states. In addition, the ques-

21. On June 8, 2006, the Supreme Court of Virginia again remanded Daryl Atkins's case for a new proceeding. *Atkins v. Commonwealth*, 631 S.E.2d 93, 102 (Va. 2006).

22. VA. CODE ANN. §§ 19.2-264.3:1.1–1.2 (Repl. Vol. 2004).

23. The first author of this paper (RJB) was a member of the *Atkins* Subcommittee and was the Chair of the Clinical Advisory Group. Although the account of the legislative history presented below is based largely on the public record, some observations are based on the first author's papers and recollections. The papers are available in the Special Collections of the University of Virginia Law Library.

24. As will be discussed *infra*, there were some bumps along the way. Although the Senate enacted the Crime Commission bill with little change, the House of Delegates made some substantial modifications. The bill that emerged from the Joint Conference Committee included two changes unacceptable to the Crime Commission drafters. The Governor sent the bill back to the General Assembly recommending that the Crime Commission's approach to each of these issues be restored. The General Assembly accepted one of the Governor's changes but not the other. With the support of the Crime Commission drafters, the Governor signed the bill.

tions that have arisen regarding the proper interpretation of the Virginia statute tend to bear directly on the goal of promoting fair and accurate assessments and adjudications. That interpretive experience may also be of value to experts and courts in other states. Many state supreme courts and lower courts are already confronting the issues addressed in this article,²⁵ underscoring the importance of educating judges and practitioners as soon as possible on how to approach *Atkins* cases in accordance with prevailing scientific understanding and accepted clinical standards. That is our goal.

I. DEFINING MENTAL RETARDATION

Although the Supreme Court in *Atkins* left it to the states to enforce the new constitutional rule, *Atkins* did not leave each state free to define mental retardation.²⁶ As already indicated, it seems clear that the Supreme Court intended for the states to embrace a clinical definition of mental retardation, as a diagnos-

25. See, e.g., *Ex parte Perkins*, 851 So. 2d 453, 456 (Ala. 2002) (finding absence of mental retardation when defendant had an IQ of 76 and had been both married and briefly employed); *In re Hawthorne*, 105 P.3d 552, 557 (Cal. 2005) (declining to adopt an IQ of 70 as a cut-off score); *Trotter v. Florida*, 932 So. 2d 1045, 1049–50 (Fla. 2006) (finding absence of mental retardation when defendant's IQ scores ranged from 72 to 91 and discussing the weight to be given to expert testimony); *Pruitt v. Indiana*, 834 N.E.2d 90, 104–08 (Ind. 2005) (discussing factors to be considered when evaluating IQ scores and assessing adaptive behavior); *Bowling v. Kentucky*, 163 S.W.3d 361, 376, 384 (Ky. 2005) (suggesting that measurement error and the Flynn effect do not need to be considered under the relevant Kentucky statute and holding a bright line cut-off IQ score of 70 to be constitutional); *Louisiana v. Dunn*, 831 So. 2d 862, 884, 887 (La. 2002) (holding that there must be a hearing on the mental retardation issue when the defendant had an IQ score of 71 in order to develop the record and take, for example, standard error of measurement and the defendant's motivation into account); *Goodwin v. Missouri*, 191 S.W.3d 20, 29 (Mo. 2006) (holding that the defendant was not entitled to an evidentiary hearing on mental retardation because three experts testified that he was not retarded and there was no evidence to the contrary); *Pennsylvania v. Miller*, 888 A.2d 624, 629–31 (Pa. 2005) (establishing the definition of mental retardation in Pennsylvania and the issues that must be considered when making the determination); *Howell v. Tennessee*, 151 S.W.3d 450, 458 (Tenn. 2004) (holding that a for a defendant to be considered mentally retarded they must have an IQ of 70 or below); *Howard v. Texas*, 153 S.W.3d 382, 387 (Tex. Crim. App. 2005) (holding that there was insufficient evidence of mental retardation when only portions of an IQ test were administered to the defendant); *Johnson v. Commonwealth*, 591 S.E.2d 47, 59 (Va. 2004) (holding that the defendant's mental retardation claim was frivolous because he had IQ scores of 75 and 78, both of which exceed the threshold score of 70).

26. Any definition of mental retardation used to implement *Atkins* must not cover a smaller group of individuals than the definition adopted by the American Association of Mental Retardation. James W. Ellis, *Mental Retardation and the Death Penalty: A Guide to State Legislative Issues*, 27 MENTAL & PHYSICAL DISABILITY L. REP. 11, 12 (2003).

able disorder, rather than a legally constructed definition focusing on functional impairments and bearing on diminished culpability.²⁷ The Court set out, in full, two well-established clinical definitions of mental retardation, one from the American Association on Mental Retardation (AAMR)²⁸ and one from the American Psychiatric Association (APA).²⁹ Although the language of these two definitions differs somewhat, they are conceptually identical, requiring significant deficits in intellectual functioning and adaptive behavior and onset before age eighteen.

Of the eighteen state statutes mentioned in *Atkins*, seventeen of them use exclusively clinical criteria drawn from these definitions.³⁰ However, some confusion is created by the fact that the Court did not specifically disavow the approach taken in Kansas, the one state that had adopted a diminished capacity approach before the *Atkins* decision.³¹ The Kansas statute defined being mentally retarded as “having significantly subaverage general intellectual functioning . . . to an extent which substantially impairs one’s capacity to appreciate the criminality of one’s conduct or to conform one’s conduct to the requirements of law,” using the Model Penal Code’s criteria for the insanity defense.³² However, a statute using a definition based on diminished capacity to appreciate the wrongfulness of specific conduct or to conform that conduct to the law would run afoul of the rationale for the *Atkins* decision because it would allow a defendant to receive the death penalty as long as he possessed whichever capacities were required by the statute, *even if he is mentally retarded* according to the prevailing professional criteria.³³

27. See *supra* pp. 813–14.

28. *Atkins v. Virginia*, 536 U.S. 304, 308 n.3 (2002). As of January 1, 2007 the American Association on Mental Retardation has changed its name to the American Association on Intellectual and Developmental Disabilities. Press Release, Am. Ass’n on Mental Retardation, World’s Oldest Organization on Intellectual Disability Has a Progressive New Name (Nov. 2, 2006), available at http://www.aamr.org/About_AAIDD/name_change_PR_dreen.htm.

29. *Atkins*, 536 U.S. at 308 n.3 (2002).

30. See 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 & L. 304, 305 (2004).

31. See *Atkins*, 536 U.S. at 314.

32. KAN. STAT. ANN. § 21-4623(e) (Cum. Supp. 2005).

33. Furthermore, a diminished capacity approach would carry a greater risk of expert disagreement and arbitrary application by judges since capacities to appreciate criminality or wrongfulness or to refrain from specific conduct at some time in the past cannot be

Mental retardation has been defined by two prominent professional organizations. At the time of the *Atkins* decision, the AAMR was using its 1992 definition of mental retardation, which the Court quoted in its opinion:

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 academics, leisure, and work. Mental retardation manifests before age 18.³⁴

The fourth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM IV-TR), published in 2000, also quoted in *Atkins*, defines mental retardation as:

[S]ignificantly subaverage general intellectual functioning (Criterion A) that is accompanied by significant limitations in adaptive functioning in at least two of the following skill areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health, and safety (Criterion B). The onset must occur before age 18 years (Criterion C).³⁵

The AAMR definition was updated in 2002 to recognize that the ten specific examples of adaptive skills listed in the 1992 definition and reflected in the APA definition, fall into three groups of cognitive, social, and practical skills. The AAMR now defines mental retardation as follows: "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."³⁶

measured with validated instruments, unlike intelligence and adaptive behavior. Bills to modify the Kansas statute have been introduced but not yet adopted. *See, e.g.*, H.B. 2396, 2007 Sess. (Kan.).

34. *Atkins*, 536 U.S. at 308 n.3 (quoting AM. ASS'N ON MENTAL RETARDATION: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORT 5 (Luckasson et al. eds., 9th ed. 1992)).

35. *Id.* (quoting AM. PSYCHIATRIC ASSN., DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 41 (4th ed., text rev. 2000)).

36. AM. ASS'N ON MENTAL RETARDATION, MENTAL RETARDATION: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORTS 8 (Luckasson et al. eds., 10th ed. 2002).

Both the AAMR and the APA definitions identify three prongs of mental retardation: (1) significantly subaverage intellectual functioning, (2) significant limitations in adaptive behavior, and (3) onset before age eighteen. The current difference between the most recent APA definition and the most recent AAMR definition can be explained by the fact that the APA typically waits for and then follows the AAMR. The APA has not published a new edition of its *Diagnostic and Statistical Manual of Mental Disorders* since the AAMR revised its definition of mental retardation in 2002.³⁷

A state would be on firm ground if it adopted either the AAMR or the APA definition of mental retardation, or if it combined the two.³⁸ Today, there is very little controversy about the concept and definition of mental retardation. Although adaptive behavior was not added by the AAMR until 1959,³⁹ it is now a well-accepted feature of the definition. Most of the disagreement among professionals relates not to the criteria or components of mental retardation, but rather to their measurement.⁴⁰

The drafters of the Virginia statute chose to use the AAMR definition as their model primarily because the AAMR is the principal professional organization in the field of mental retardation. Its definition is highly respected and also reflects the most current research in the field.⁴¹ The Virginia statute provides:

“Mentally retarded” means a disability, originating before the age of 18 years, characterized concurrently by (i) significantly subaverage intellectual functioning as demonstrated by performance on a standardized measure of intellectual functioning administered in conformity with accepted professional practice, that is at least two standard deviations below the mean and (ii) significant limitations in adaptive behavior as expressed in conceptual, social and practical adaptive skills.⁴²

37. Working groups are now just being formed for the drafting of DSM V. Estimated date of completion is 2011. Personal communication from Paul S. Appelbaum (Mar. 1, 2007).

38. Peggy M. Tobolowsky, *Atkins Aftermath: Identifying Mentally Retarded Offenders and Excluding Them From Execution*, 30 J. LEGIS. 77, 99–101 (2003); Bonnie, *supra* note 30, at 305.

39. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 38.

40. Tobolowsky, *supra* note 38, at 101.

41. Tobolowsky urges states generally to opt for the AAMR definition. *Id.* at 100–01.

42. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

The Virginia definition thus includes the three essential components of mental retardation⁴³ and also emphasizes the importance of adhering to professional standards of diagnosis,⁴⁴ a feature that will be discussed in some depth later in the paper.

Recognizing that the constitutional exemption from the death penalty, embraced in *Atkins*, is defined in clinical terms, the question arises whether it might be defined more narrowly for death penalty purposes than the standard clinical definitions adopted by the AAMR and the APA. For example, the overall category of mental retardation has historically been sub-classified as mild, moderate, severe, or profound based on the IQ range in which an individual falls.⁴⁵ Individuals with an IQ between fifty and seventy are typically considered to be "mildly retarded."⁴⁶ While appearing to concede that the Eighth Amendment bars execution of individuals with moderate, severe and profound mental retardation, Justice Scalia objected, in his dissent in *Atkins*, to the majority's decision to extend this protection to offenders with mild mental retardation.⁴⁷ He insisted that people with

43. Like almost every other statute barring the death penalty for people with mental retardation, the Virginia statute includes the component of developmental origin. *Id.* Manifestation of symptoms in childhood is a cardinal feature of the diagnosis; a person whose development into adulthood was normal does not become mentally retarded, as that term is used clinically, if he subsequently experiences the deterioration of cognitive function characteristic of Alzheimer's Disease, for example. However, one can reasonably ask whether the legal principle enunciated in *Atkins* should also be applicable to people who, as adults, acquire deficits in intellectual functioning and adaptive behavior that are equivalent in severity to those experienced by people who are mentally retarded. The two adult-onset conditions that might involve equivalent deficits are traumatic brain injuries and dementia. It has been argued that failure to apply the *Atkins* principle to such people is unfair and irrational if the deficits they experience are equivalent in severity to those experienced by people with mental retardation simply because those deficits developed earlier in life—as long as they were present before the offense was committed. This view has been embraced in identical position statements adopted by the American Bar Association, the American Psychological Association and the American Psychiatric Association. See 30 MENTAL & PHYSICAL DISABILITY L. REP. 668 (2006).

New Mexico appears to be the only state that has omitted the developmental origin component from its statutory definition of mental retardation. See N.M. Stat. Ann. § 31-20A-2.1 (LexisNexis 2000). The Kansas proposal mentioned earlier, H.B. 2396, would omit the requirement of developmental onset by defining cognitive disability (the condition making a person ineligible for the death penalty) as "a disability characterized by significant limitations both in intellectual functioning and deficits in adaptive behavior as expressed in conceptual, social and practical adaptive skills." H.B. 2396, 2007 Sess. (Kan.).

44. VA. CODE ANN. § 19.2-2643:1.1(B) (Repl. Vol. 2004).

45. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36 at 7, 102.

46. *Id.* at 102.

47. Part II of Scalia's dissent is devoted to arguing why the execution of specifically mildly mentally retarded offenders is not unconstitutional. By crafting his argument in

mild mental retardation have a significant range of abilities and that juries can assess the culpability of such individuals on a case-by-case basis.⁴⁸

It is clear that the Court's majority rejected Justice Scalia's view and meant to embrace the generally accepted clinical definition of mental retardation, including mild retardation; indeed, Atkins himself would be classified as mildly retarded, if he is mentally retarded at all.⁴⁹ Significantly, the Court did not qualify or limit the term mental retardation when it observed that "death is not a suitable punishment for a mentally retarded criminal," and that "execution of mentally retarded criminals" is unconstitutionally excessive under the Eighth Amendment.⁵⁰ All people with mental retardation, whether mild or profound, are ineligible for the death penalty under *Atkins*.⁵¹ Limiting the exemption to people with profound, severe, or moderate mental retardation is analogous to requiring proof of deficiencies in capacities (such as moral understanding or impulse control) in addition to the deficiencies in intellectual functioning and adaptive behavior that are embodied in the AAMR and APA definitions and is constitutionally impermissible for the same reason.⁵²

this way, Scalia implies that perhaps the execution of severely mentally retarded individuals is prohibited by the Eighth Amendment. *Atkins v. Virginia*, 536 U.S. 304 339–40 (2002) (Scalia, J., dissenting).

48. *Id.* at 349–50 (Scalia, J., dissenting).

49. *Id.* at 308 (noting that Atkins had been evaluated by a forensic psychologist who concluded he was "mildly mentally retarded.").

50. *Id.* at 321.

51. *Id.*

52. Although the Crime Commission embraced the AAMR definition, VA. STATE CRIME COMM'N, *Atkins v. Virginia*, A STUDY TO THE GENERAL ASSEMBLY OF VIRGINIA (Jan. 2003), a competing bill, S.B. 1239, included a definition based on the DSM IV language and supplemented that language with a diminished capacity prong; the bill would have defined mental retardation as follows:

"Mentally retarded" means a disability, documented before the age of 18 years, characterized concurrently by (i) significantly subaverage intellectual functioning as demonstrated by a full scale intelligence quotient (IQ) on a standardized measure of intellectual functioning carried out in conformity with accepted professional practice, that is more than 2 standard deviations below the mean; (ii) significant limitations in adaptive functioning indicated by a significant inability to perform in at least 2 of the following skill areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety; and (iii) a diminished capacity to understand and process information, communicate, abstract from mistakes and learn from experience, engage in logical reasoning, control impulses and understand the reaction of others.

One final question to be addressed by state legislatures in defining mental retardation is whether they should limit the definition exclusively to intellectual deficits, excluding the requirement of deficits in adaptive behavior. In effect, states adopting this approach would be barring the death penalty in more cases than *Atkins* requires. Although no state has omitted the adaptive behavior element of the definition as far as we know, a good argument can be made for taking this course. As will be discussed below, the measures for adaptive behavior are not as well developed as those for intellectual ability, and assessments are more challenging and less reliable, and therefore entail a greater risk of error (in either direction). In defining mental retardation and in prescribing policies to guide assessment and adjudication, state legislatures and courts must decide whether to err in the direction of an erroneous finding that a mentally retarded defendant is not mentally retarded, or vice versa. Focusing the assessment and adjudication exclusively on whether the defendant exhibits substantial intellectual deficits will remove a significant risk of mistake in *Atkins* adjudications that could lead to executions of defendants with mental retardation. On the other hand, as prosecutors often point out, the adaptive behavior requirement helps to differentiate people who are significantly impaired in their ability to function in social settings or in everyday life, from people who are savvy enough to carry out criminal activities, notwithstanding their significant intellectual deficits. When seen from this perspective, omitting deficits in adaptive behavior from the definition unduly broadens the reach of the *Atkins* exemption and compromises the retributive function of the death penalty.

S.B. 1239, 2003 Sess. (Va.) (emphasis added).

Although more subtle than the Kansas formulation of diminished capacity, this proposal suffers from the same constitutional flaw: by adding additional criteria to the accepted clinical definition of mental retardation, it would permit execution of some defendants who are mentally retarded.

The Senate Committee for Courts of Justice rejected this approach in favor of the Crime Commission's proposal. S. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003). However, the House Committee for Courts of Justice embraced the Kansas diminished capacity approach that had been specifically rejected by the Crime Commission. H. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003). Those two bills were then approved by the Senate and House respectively. JOINT CONF. COMM. S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003). The Crime Commission's diagnostic approach based on the AAMR prevailed in the conference committee, although the Crime Commission's definition of adaptive behavior was modified in several respects and is discussed in Part IV below.

II. ASSESSMENT OF INTELLECTUAL FUNCTIONING

Intellectual functioning, typically measured by standardized tests of IQ, has traditionally been at the core of diagnosing mental retardation. Under the 2002 AAMR definition of mental retardation, embraced by the Virginia statute, mental retardation requires significant limitations in intellectual functioning “as demonstrated by performance on a standardized measure of intellectual functioning . . . that is at least two standard deviations below the mean.”⁵³ The Virginia statute further provides:

Assessment of intellectual functioning shall include administration of at least one standardized measure generally accepted by the field of psychological testing and appropriate for administration to the particular defendant being assessed, taking into account cultural, linguistic, sensory, motor, behavioral and other individual factors. Testing of intellectual functioning shall be carried out in conformity with accepted professional practice, and whenever indicated, the assessment shall include information from multiple sources. The Commissioner of Mental Health, Mental Retardation and Substance Abuse Services shall maintain an exclusive list of standardized measures of intellectual functioning generally accepted by the field of psychological testing.⁵⁴

Courts must pay close attention to which IQ test was administered to a particular defendant and to how the scores are interpreted. The diagnosis of mental retardation always involves significant clinical judgment, and thus, the Virginia statute correctly emphasizes conformity with standards of practice generally accepted by the field. The IQ test chosen to be administered must be a “standardized measure generally accepted by the field of psychological testing,” and the testing process must be “carried out in conformity with accepted professional practice.”⁵⁵ These requirements should ensure fair assessments of defendants and reliable results that courts can use with confidence.

By embracing a clinical definition of mental retardation, the United States Supreme Court put professional standards of measurement, assessment, and diagnosis at the center of *Atkins* adjudications. Fair implementation of *Atkins* requires states to take

53. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004); AM. ASS’N ON MENTAL RETARDATION, *supra* note 36, at 8.

54. VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

55. *Id.*

appropriate steps to ensure that these assessments are held to high standards of quality. States should go to great lengths at all levels—in statutory provisions, administrative guidelines, and in the training of judges, experts, and other court personnel—to ensure that capital defendants are evaluated fairly and accurately. In this Part, guidance is offered in relation to the selection of IQ tests, the interpretation of IQ scores, and the advisability of prescribing a cut-off score in the definition of mental retardation.

A. IQ Tests

A wide range of standardized IQ tests have been developed over the years, many of which are now considered generally accepted within the field of psychological testing. The Wechsler scales for children and adults and the Stanford-Binet scale are the most commonly administered and most highly respected measures of IQ.⁵⁶ Administration of either the Stanford-Binet or the Wechsler scale is appropriate for most individuals, unless they have particular characteristics such as language impairment that would influence performance on these tests. Other IQ tests have been designed to suit a variety of specific patient characteristics, including age, race and cultural background, limited verbal ability, and profound cognitive impairment.

The Wechsler Adult Intelligence Scale is also in its third revision (WAIS-III). Like the WISC-III, it is designed to assess intelligence as a global, multi-faceted construct. It contains verbal and perceptual-motor subtests and yields a verbal score, a performance score, and a full-scale score.⁵⁷ The WAIS-III can be used for individuals aged sixteen to eighty-nine years.⁵⁸ The WAIS-III has a mean score of 100, with a standard deviation of 15.⁵⁹

56. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 59.

57. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 61.

58. *Id.*; MARY BEIRNE-SMITH, RICHARD F. ITTENBACH & JAMES R. PATTON, MENTAL RETARDATION 90 (6th ed. 2002) (1998). The Wechsler Intelligence Scale for children is currently in its fourth revision (WISC-IV). It is designed to test the intelligence of children between the ages of six years and sixteen years, eleven months. The instrument assesses the ability of an individual to act purposefully, think rationally, and deal effectively with his or her environment. *Id.* at 89. However, after the Supreme Court's decision in *Roper v. Simmons*, 543 U.S. 551, 578 (2005), there is no longer any need to assess mental retardation in defendants younger than 18 in the context of a capital case.

59. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 61.

The Stanford-Binet Intelligence Scale is in its fifth revision (SB5). It tests verbal reasoning, quantitative reasoning, and short-term memory, which together comprise intelligence as a general concept. The instrument is designed for children aged two years through adults.⁶⁰ Like the WAIS-III, the SB5 has a mean score of 100.⁶¹

In addition to the Wechsler and Stanford-Binet scales, other general measures of IQ, such as the Cognitive Assessment System and the Kaufman Adolescent and Adult Intelligence Test, are suitable for administration to most individuals. For individuals with special characteristics, tests tailored to those characteristics should be used.

Selection of a measure of IQ that is appropriate for a particular defendant is crucial in death penalty cases. As required by the Virginia statute, a defendant should always be administered at least one standardized measure of IQ,⁶² meaning that the test has been normed on a representative sample of the population. The standardized measure must also be generally accepted as accurate and reliable within the field of psychological testing.⁶³ The Wechsler, the Stanford-Binet, and the Kaufman Adult and Adolescent Intelligence Test, among others, fit both criteria for the typical defendant. However, statutes and courts should require IQ tests to be selected with due consideration of all the pertinent individual characteristics of the defendant.⁶⁴ Once a standardized measure generally accepted by the field has been selected, it must be administered in conformity with accepted professional practice.⁶⁵ This typically includes administering the entire measure rather than just parts of it, and avoiding of repeat administrations of the same measure within a short period of time.

State statutes should not require the use of any specific measure for the assessment of intellectual functioning of defendants in death penalty cases. Mandating the administration of one particular IQ test in all cases may be appealing to legislatures wish-

60. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 62; BEIRNE-SMITH ET AL., *supra* note 58, at 86–87.

61. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 62.

62. VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

63. *See id.*

64. *See id.*

65. *See id.*

ing to promote efficiency and consistency in assessment and adjudication of mental retardation. Requiring a single test might also seem to promote equal application of the *Atkins* exemption by facilitating judicial familiarity with the required test. However, the benefits of uniformity are greatly outweighed by its disadvantages.

First, and most important, IQ tests must be appropriate for the particular individual being assessed in order for the tests to provide an accurate measure of IQ. Clinicians are expected to consider cultural, linguistic, sensory, motor, and behavioral factors in deciding which test to use and in administering and interpreting it.⁶⁶ Examiners should also choose a test normed on a sample containing individuals with characteristics similar to those possessed by the defendant.⁶⁷ Different measures will be better suited to different defendants, and a statutory requirement regarding test selection would unduly interfere with the exercise of clinical judgment.

Second, in instances when one of the standard tests like the Wechsler or Stanford-Binet would be appropriate, clinicians should be allowed to administer the IQ test with which they personally are most experienced. When clinicians use the test with which they are most familiar, the results are likely to be most accurate.⁶⁸

Third, it would be impractical to specify the use of one measure in all cases because IQ tests change significantly over time in response to scientific developments and clinical research, rendering older versions of the tests outdated. The speed of change in IQ tests does not seem to be decelerating. The Stanford-Binet scale was first published in 1905, and it was revised in 1916, 1937, 1960, 1972, 1986, and 2003.⁶⁹ The first Wechsler scale was published in 1939, and was revised in 1955, 1981, and 1997.⁷⁰ Because future revisions of common IQ tests can be expected, state

66. AM. PSYCHIATRIC ASS'N, *supra* note 35, at 42; AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 8; *see* VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

67. AM. PSYCHIATRIC ASS'N, *supra* note 35, at 46.

68. *See id.* at 42.

69. GALE H. ROID & R. ANDREW BARRAM, *ESSENTIALS OF STANFORD-BINET INTELLIGENCE SCALES (SB5) ASSESSMENT 3-9* (2004).

70. David S. Tulsky et al., *Revising a Standard: An Evaluation of the Origin and Development of the WAIS-III*, in *CLINICAL INTERPRETATION OF THE WAIS-III AND WMS-III* 43, 43 (David S. Tulsky et al. eds., 2003).

legislatures should allow sufficient flexibility for practice in forensic settings to keep pace with scientific and clinical advances in the field.

Consistent with this analysis, the post-*Atkins* legislation in Virginia emphasizes the need for clinical discretion in selecting the appropriate test for a particular defendant, while requiring use of at least one “standardized measure generally accepted by the field of psychological testing.”⁷¹ The Code language pertaining to assessment of intellectual functioning quoted above is virtually identical to the language recommended by the Clinical Advisory Group,⁷² except for the last sentence requiring the state to maintain an “exclusive list” of accepted measures of intellectual functioning for use in the context of *Atkins* evaluations.⁷³ Adoption of such a requirement by other states is strongly discouraged. Because the “exclusive list” requirement was subject to ongoing dispute during legislative consideration of the bill, further discussion is warranted.

As originally drafted by the Clinical Advisory Group and proposed to the *Atkins* Subcommittee, the bill did not include the last sentence.⁷⁴ However, prosecutors were concerned that the requirement that experts use “at least one standardized measure generally accepted by the field” and that testing be “carried out in conformity with accepted professional practice” did not ade-

71. VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

72. Draft Bill, Nov. 8, 2002: *Atkins v. Virginia* Materials, The Papers of Richard J. Bonnie (unpublished bill, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p). As proposed by the CAG in its report on October 17, 2002, the bill provided:

Assessment of intellectual functioning must include administration of at least one standardized measure generally accepted by the field and appropriate for administration to the particular person being assessed, taking into account cultural, linguistic, sensory motor, behavioral and other individual factors. Testing of intellectual functioning should be carried out in conformity with accepted professional practice, and, whenever indicated, the assessment should include information from multiple sources.

CLINICAL ADVISORY GROUP, REPORT TO SUBCOMMITTEE ON THE DEATH PENALTY AND DEFENDANTS WITH MENTAL RETARDATION, Oct. 17, 2002: *Atkins v. Virginia* Materials, The Papers of Richard J. Bonnie (unpublished report, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p) [hereinafter CAG REPORT, Oct. 17, 2002]. The single change in this language was to specify that the relevant “field” of professional practice is “psychological testing.” VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

73. *Id.*

74. See VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004); CAG REPORT, Oct. 17, 2002, *supra* note 72.

quately narrow the types of tests that can be used and would allow experts retained by defendants to administer idiosyncratic tests purposely selected to produce low scores. The *Atkins* Subcommittee shared the concern raised by prosecutors and concluded that the Commissioner of Mental Health, Mental Retardation and Substance Abuse Services should be directed to keep a list of qualifying instruments.

The key point of disagreement that emerged in the subcommittee deliberations and persisted all the way through the legislative process was whether the list of generally accepted tests should serve solely as a resource for experts conducting *Atkins* evaluations, as proposed by the CAG representatives,⁷⁵ or whether it should have prescriptive force, as urged by the prosecutors. This disagreement was captured in the question of whether the list should be a non-exclusive list or an exclusive one. The CAG representatives attempted to dissuade the Subcommittee and the Crime Commission from requiring the list to be exclusive, arguing that a prescriptive approach was not needed and that it would likely cause more problems than it would solve.

First, the CAG representatives insisted that the language requiring the use of standardized tests "generally accepted by the field," as well as "conformity with accepted professional practice," provided adequate safeguards against idiosyncratic selection of homemade, unproven, or unsuitable tests, particularly if a resource list of such tests is maintained by the Commissioner. Moreover, expert testimony as to what constitutes accepted professional practice can be expected to discredit testing and opinion by outliers.

Second, it was argued that conferring authority on a government official to prescribe how mental health professionals should carry out clinical testing would be institutionally inappropriate. If the list maintained by the Commissioner is given prescriptive effect, it implies that an opinion based on an unlisted test would be inadmissible. Such a regulatory requirement implies that the Commissioner should adhere to legal norms governing administrative rule-making. Criteria for determining whether a measure

75. The phrase "CAG representatives" refers to the Chair of the CAG (the first author) and other members of the CAG who were consulted on the drafting of the bill on an ongoing basis, whether or not they were actually present at the meetings of the *Atkins* Subcommittee or the Crime Commission.

should be on the list would have to be developed. Guidelines would also have to be established stating when the list would be updated and at what point in time an instrument must be on the list for it to qualify. A brand new, validated measure of intelligence might have to be disqualified simply because it does not appear on the state's list.

It is highly unusual for the government to prescribe which of a variety of acceptable psychometric measures should be used in conducting a clinical evaluation, including a forensic assessment.⁷⁶ Professional organizations like the AAMR, as well as peer-reviewed professional literature, can and should be relied upon to guide clinicians in exercising their professional judgment. It is also institutionally awkward for the legislative branch to direct the executive branch to keep a list of measures designed to displace judicial judgment as to what evidence should be admissible as expert opinion bearing on proof of whether a defendant is mentally retarded—a constitutionally required finding.

During the course of deliberations by the *Atkins* Subcommittee and the Crime Commission, CAG representatives successfully warded off language that would give the Commissioner's list of generally accepted measures prescriptive effect. The November 22, 2002 draft of the Crime Commission's bill included the language that "[t]he Commissioner of Mental Health, Mental Retardation and Substance Abuse Services shall maintain a list of standardized measures of intellectual functioning generally accepted by the field."⁷⁷ The January 8, 2003 draft required the Commissioner to maintain a "nonexclusive list."⁷⁸ This was the

76. CAG REPORT, Oct 17, 2002, *supra* note 72, at 2. To our knowledge, the only context in which the Commonwealth of Virginia has directed use of a particular scientific instrument is the screening of "prisoners incarcerated for sexually violent offenses . . . who are scheduled for release from prison within 10 months" where the Director of the Department of Corrections is directed identify those prisoners "who receive a score of five or more on the Static-99 or a like score on a comparable, scientifically validated instrument designated by the Commissioner," VA. CODE ANN. § 37.2-903(C) (Supp. 2006). Even in this context, though, use of "comparable, scientifically validated instruments" is permitted.

77. Draft Bill, Nov. 22, 2002: *Atkins v. Virginia* Materials, The Papers of Richard J. Bonnie (unpublished bill, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p).

78. Draft Bill, Jan. 8, 2003: *Atkins v. Virginia* Materials, The Papers of Richard J. Bonnie (unpublished bill, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p).

phrase that was included in the Crime Commission bills introduced in the General Assembly.⁷⁹

In February 2003, the House of Delegates amended the Crime Commission bill to require an "exclusive list" of acceptable measures of intellectual functioning,⁸⁰ and that language was retained in the bill approved by the Joint Conference Committee.⁸¹ After consultation with the drafters of the Crime Commission bill, including the CAG chair, the Governor recommended that "reference list" be substituted for "exclusive list."⁸² However, his suggestion was not adopted by the General Assembly,⁸³ and, after renewed consultation with the CAG chair, the Governor signed the bill rather than abandon what was, overall, a very good bill. As a result, Virginia law requires the Commissioner of Mental Health, Mental Retardation and Substance Abuse Services to maintain an "exclusive list" of IQ tests "generally accepted" within the field of clinical psychology and therefore acceptable for use in *Atkins* evaluations.⁸⁴

In light of this legislative history, it should come as no surprise that litigants in *Atkins* adjudications have sparred over the significance of the fact that a particular test is not on the Commissioner's list,⁸⁵ and have lobbied to have particular tests added to the list. It is, of course, necessary for the list to be open to additions based on developments in the field of intelligence testing. Until recently, the process of reviewing new tests has been an informal one: the Director of the Office of Forensic Services of the

79. S.B. 1141 was transmuted into S.B. 1239 when it was approved as an amendment in the nature of a substitute by the Senate Committee for Courts of Justice. S. 1141, 2003 Gen. Assem., Reg. Sess. (Va. 2003); S. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003).

80. H. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003).

81. J. CONF. COMM., AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003).

82. See Virginia General Assembly Legislative Information System, S.B. 1239, <http://leg1.state.va.us/cgi-bin/legp504.exe?ses=031&typ=bil&val=sb1239> (last visited Apr. 10, 2007).

83. S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003) (codified as amended at VA. CODE ANN. § 19.2-264.3:1.1 (Repl. Vol. 2004)).

84. VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

85. Personal communication from James Morris, Director of the Office of Forensic Services, Department of Mental Health, Mental Retardation and Substance Abuse Services, April 4, 2007 (discussing a case where the list was changed after the testing, and the measure used was removed. The court retroactively applied the edited list, denying admissibility of the test used, even though it was on the list at the time it was administered).

Department has recommended changes to the list as information about previously unlisted tests has emerged. When such an addition is made, the list has been updated.⁸⁶

In light of the prescriptive effect of the Commissioner's list, the process of reviewing proposed additions (or deletions) should be formalized, whether or not a more formal process is required by the Administrative Process Act.⁸⁷ The list itself should be officially published and easily available on the Department's website. The website should set forth the procedure to be used by anyone who seeks to add a new measure or to challenge the continued listing of measures already on the list. Measures under review at any time, either by the Office of Forensic Services on its own initiative or on the request of someone else, should be identified on the website. The Office of Forensic Services should appoint a panel of experts to review all proposed changes to the list and should identify the panel members on the website. After review, the panel should make recommendations to the Commissioner. The Commissioner's decisions should be specifically communicated on the website.⁸⁸

B. *Interpretation of IQ Scores*

According to the definition of mental retardation in the Virginia statute, "significantly subaverage intellectual functioning" must be "demonstrated by performance on a standardized measure of intellectual functioning administered in conformity with accepted professional practice, that is at least two standard de-

86. The most recent list is dated October 31, 2005. See "List of Standardized Measures of Intellectual Functioning, www.dmhmrsas.virginia.gov/documents/OFO-standardized-measures.pdf (last visited Apr. 10, 2007).

87. We do not believe that the maintenance of this list constitutes formal rule-making within the meaning of Virginia's Administrative Process Act. VA. CODE ANN. §§ 2.2-4000-4031 (Repl. Vol. 2005). The APA only governs agency actions concerning rules or regulations which "hav[e] the force of law." VA. CODE ANN. § 2.2-4001. Because the Commissioner does not have generic authority to adopt regulations and § 19.2-264.3:1.1 does use the terms "rule" or "regulation" to refer to the list, maintaining the standardized measures of intellectual functioning is not properly characterized as promulgating a "rule or regulation" that has "the force of law," and is not subject to the rule-making procedures found at § 2.2-4007.

88. When this article was in press, the Commissioner established such a committee. See Memorandum from James S. Reinhard, M.D., on the subject of "Commissioner's Review Panel for the selection of measures of intellectual functioning pursuant to § 19.2-264.3:1.1 of the Virginia Code" (April 4, 2007) (on file with the authors).

viations below the mean.”⁸⁹ In addition, the statute directs that “[t]esting of intellectual functioning shall be carried out in conformity with accepted professional practice.”⁹⁰ These two provisions are perhaps the most important provisions in the statute because they import standard practices of administering, scoring, and interpreting IQ tests into the law. Four clinically and legally important issues affecting the meaning of an IQ score will be addressed below: the standard error of measurement in psychological testing; the increase in scores on standardized tests among the general population over time, known as the Flynn effect; practice effects; and the presence of mental illness.

For the purposes of this section regarding the interpretation of IQ scores, we will assume that “significantly subaverage intellectual functioning”⁹¹ is operationalized as a score of 70 or below on a specific IQ test. That is often the case, given that a score of 70 is typically two standard deviations below the mean on a test designating a score of 100 as the mean. This assumption, and the question of whether the law should explicitly establish a cut-off IQ score for mental retardation (as many states do), will be discussed later in this paper.

1. Standard Error of Measurement

All measurement, both physical and psychological, has some potential for error. For example, when someone’s height is being measured, the result will be influenced by many factors including the particular tool being used, the eyesight of the measurer, the care taken by the measurer, and whether the person being measured is wearing shoes or slouching. Psychological testing has even greater potential for error because it is more subjective. Error may be introduced by the examiner making a timing mistake, failing to record responses, over-prompting, mishandling stimuli objects, or neglecting to repeat parts of the instructions.⁹² Error may also be introduced by the defendant’s mood and general health, luck, or other undetermined factors.⁹³ In any kind of

89. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

90. VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

91. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

92. ALAN S. KAUFMAN & ELIZABETH O. LICHTENBERGER, *ASSESSING ADOLESCENT AND ADULT INTELLIGENCE* 197 (3d ed. 2006).

93. James R. Flynn, *Tethering the Elephant: Capital Cases, IQ, and the Flynn Effect*,

measurement there are always tradeoffs between cost and accuracy.

Standard error of measurement (SEM) helps to quantify the errors in intelligence tests in order to facilitate the most accurate interpretation and presentation of scores.⁹⁴ Both the AAMR and the APA definitions of mental retardation stress the importance of considering SEM when evaluating IQ scores.⁹⁵ SEM varies between measures and between age groups within each measure. Each measure is accompanied by a table of calculated SEMs by age group. Generally, SEM is estimated to be between three and five points for well-standardized IQ tests. This means that if a person were retested using the same measure, there would be a two-thirds likelihood that he would score within three or four points above or below his previous score.⁹⁶ Consequently, an IQ of 70 is not accurately understood as a precise score, but rather as a range of confidence with parameters of at least one SEM (meaning scores between 66 and 74 would be expected 67% of the time) or with parameters of two SEMs (meaning scores between 62 and 78 would be expected 95% of the time).⁹⁷ In order to be *highly* confident that a person is *not* mentally retarded, therefore, an IQ score of at least 79 might be required, whereas a score of 76 might be required to rule out mental retardation if a somewhat lower level of confidence (“reasonable confidence”) were acceptable. Conversely, one could be reasonably confident that a person *is* mentally retarded if the person had an IQ score of 64,⁹⁸ and a score of 66 would be very strong, though not conclusive, evidence of mental retardation. Any score in between these illustrative scores (i.e., between 67 and 75) would be consistent with the existence of mental retardation.

SEM can thus either increase or decrease a defendant’s score, and this uncertainty in the direction of the effect may be complicated for courts to deal with. Courts, based on a societal value

12 PSYCHOL. PUB. POL’Y & L. 170, 171 (2006); see AM. ASS’N ON MENTAL RETARDATION, *supra* note 36, at 57.

94. AM. ASS’N ON MENTAL RETARDATION, *supra* note 36, at 57–58.

95. *Id.* at 58 (with respect to intellectual functioning, the criterion for a diagnosis of mental retardation is performance on an IQ test “approximately two standard deviations below the mean, considering the SEM for the specific assessment instruments used”); AM. PSYCHIATRIC ASS’N, *supra* note 35, at 41–42.

96. AM. ASS’N ON MENTAL RETARDATION *supra* note 36, at 57.

97. *Id.*

98. Flynn, *supra* note 93, at 186.

judgment, may decide to consider SEM in only one direction. If courts consider it desirable to err on the side of finding mental retardation in *Atkins* cases in order to avoid the risk of mistakenly executing a defendant who actually qualified for Eighth Amendment protection, perhaps they would decide only to use SEM to decrease a score. If so, SEM could only be used to put a defendant into the mentally retarded range but not to take a defendant out of that range. Using the example above, a recorded score of 73 might be a score of 69 taking one SEM into account, placing the defendant in the zone of mental retardation. However, a score of 68 might be regarded as definitive evidence of mental retardation even though it might actually represent a score of 72 or 76 (outside the range of mental retardation) if one or two SEM were taken into account.

The main point here is that the SEM must *always* be taken into account when interpreting scores on IQ tests; failing to do so would be a clear departure from accepted professional practice in scoring and interpreting any kind of psychological test, including IQ tests. The importance of the SEM is so well-established in the field that it would be superfluous to direct experts to take it into account in a statute governing *Atkins* evaluations and adjudications, and most state laws say nothing about it.⁹⁹ Nonetheless, in its effort to provide as much guidance as possible to courts, the proposal drafted by the Clinical Advisory Group explicitly stated that SEM must be considered in *Atkins* cases,¹⁰⁰ as did the initial drafts of the bill in the *Atkins* subcommittee.¹⁰¹ In the course of

99. See State Statutes Prohibiting Death Penalty for People with Mental Retardation, Death Penalty Information Center, <http://www.deathpenaltyinfo.org/article.php?scid=28&did=138> (last visited Apr. 10, 2007) (listing all state statutes that prohibit death penalty for people with mental retardation); State that Have Changed Their Statutes to Comply with the Supreme Court's Decision in *Atkins v. Virginia*, Death Penalty Information Center, <http://www.deathpenaltyinfo.org/article.php?scid=28&did=668> (last visited Apr. 10, 2007) (listing changes in state statutes since *Atkins* was decided by the Supreme Court).

100. See CAG REPORT, Oct. 17, 2002, *supra* note 72 ("Substantial limitation in intellectual functioning means performance that is at least two standard deviations below the mean, considering the standard error of measurement for the specific instruments used, as well as their strengths and limitations in the context of the particular assessment.") This language was drawn directly from the AAMR Manual. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 58 (with respect to intellectual functioning, the criterion for a diagnosis of mental retardation is performance on an IQ test "approximately two standard deviations below the mean, considering the standard error of measurement for the specific assessment instruments used").

101. Draft Bill, Nov. 1, 2002: *Atkins v. Virginia* Materials, The Papers of Richard J. Bonnie (unpublished bill, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p).

the subcommittee and Crime Commission deliberations, prosecutors proposed that specific reference to SEM be omitted on the ground that this would simplify the scientific language of the bill. However, the CAG representatives urged the subcommittee and the Crime Commission to retain the SEM language in order to emphasize that any IQ score actually represents a range of possible scores. The Crime Commission proposal as passed by the Senate included the CAG's reference to the SEM in the definition of mentally retarded:

"Mentally retarded" means a disability . . . characterized . . . by (i) significantly subaverage intellectual functioning as demonstrated by performance on a standardized measure of intellectual functioning carried out in conformity with accepted professional practice, that is at least 2 standard deviations below the mean, *considering the standard error of measurement for the specific instruments used*. . . .¹⁰²

However, the reference to the SEM was deleted from the version of the *Atkins* bill passed by the House of Delegates¹⁰³ and was then omitted in the bill approved by the joint conference committee.¹⁰⁴ The Crime Commission drafters decided to acquiesce in this amendment on the ground that the omitted language, though desirable, was not necessary. As noted above, the requirement that intellectual functioning be assessed in conformity with accepted professional practice mandates the consideration of the SEM. As a matter of professional practice, experts will have to testify about why they think a particular score indicates that a defendant's performance on a specific measure was at least two standard deviations below the mean. Ensuring that forensic experts and lawyers sufficiently understand the importance and effects of SEM thus becomes a matter of training.

2. The Flynn Effect

Courts interpreting IQ scores must take the Flynn effect into account if they are to reach accurate understandings of the mean-

102. S. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003) (emphasis added).

103. See H. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003); see also Draft Bill, Feb. 3, 2003: *Atkins v. Virginia Materials*, The Papers of Richard J. Bonnie (unpublished bill, on file with the University of Virginia Law Library, Special Collections, Box 1, Mss 81-9p).

104. S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003) (codified as amended at VA. CODE ANN. 19.2-264.3:1.2 (Repl. Vol. 2004)).

ing of an individual's score. Broadly speaking, the Flynn effect is the finding that IQ scores tend to increase in the general population over time and thus IQ tests will yield different scores purely based on when they were normed.¹⁰⁵ The rate of increase in test scores varies somewhat across measures, but the Wechsler and Stanford-Binet scales have a roughly uniform rate of increase.¹⁰⁶ James R. Flynn, whom first noted this effect and whose name it carries, has found the rate of increase to be 0.3 IQ points per year between the time the test was normed and when the test was taken.¹⁰⁷ In order to calculate an accurate IQ score taking the Flynn effect into consideration, courts and clinicians must multiply the number of years that have elapsed between the norming and the administration of the measure by 0.3 and subtract that number from the defendant's score.¹⁰⁸ The explanation for constant gains in IQ over time is unclear, but the data are highly convincing and the 0.3 point rate of increase "holds true both at the mean and for low IQ" scores.¹⁰⁹

A number of state and federal courts have considered the relevance of the Flynn effect to IQ scores admitted as evidence over the last few years.¹¹⁰ In *Walker v. True*, for example, the Fourth Circuit held that the Flynn effect must be applied whenever a defendant can show that it affected his IQ score.¹¹¹ Given the robust nature of the data supporting the Flynn effect, failing to adjust scores accordingly in capital punishment cases leaves eligibility for the death penalty to chance, since the accuracy of the reported score otherwise depends on how recently the IQ test that happened to be administered was normed.¹¹²

105. Flynn, *supra* note 93, at 170. See also, Tomoe Kanaya, Matthew Scullin & Stephen Ceci, *The Flynn Effect and U.S. Policies: The Impact of Rising IQ Scores on American Society via Mental Retardation Diagnoses*, 58 AMERICAN PSYCHOLOGIST 778, 786–87 (2003).

106. Flynn, *supra* note 93, at 173.

107. *Id.*

108. The formula to account for the Flynn effect is: $IQ = \text{Test Score} - (\text{Years} \times 0.3)$. *Id.* at 187.

109. *Id.* at 184. See also, Kanaya et al., *supra* note 105, at 779.

110. See, e.g., *Moore v. Quarterman*, 454 F.3d 484, 499 & n.5 (5th Cir. 2006); *Walton v. Johnson*, 407 F.3d 285, 296 & n.7 (4th Cir. 2005); *Walker v. True*, 399 F.3d 315, 322–23 (4th Cir. 2005); *In re Hicks*, 375 F.3d 1237, 1242–43 (11th Cir. 2004); *Bowling v. Commonwealth*, 163 S.W.3d 361, 374 & n.12 (Ky. 2005); *Myers v. State*, 130 P.3d 262, 267–68 n.11 (Okla. Crim. App. 2005).

111. *Walker*, 399 F.3d at 322–23.

112. Flynn, *supra* note 93, at 174–75.

3. Practice Effects

Practice effects refer to the impact on later test scores of previously administered IQ tests.¹¹³ Practice effects nearly always inflate IQ scores where the subject had a previous opportunity to practice the tasks required by the tests.¹¹⁴ Practice effects are strongest when there is a short time between a test and retest.¹¹⁵ On the WAIS-III, verbal IQ scores are expected, on average, to increase by about 2.5 points for individuals between the ages of sixteen and eighty-nine in retest situations.¹¹⁶ Performance IQ scores are to be expected, on average, to increase by 8–8.5 points for 16–54 year-olds, by six points for 55–74 year-olds, and by four points for 75–89 year-olds.¹¹⁷ When two IQ tests must be administered within a short timeframe, practice effects can be avoided by administering a different test the second time. A common rule of thumb is to avoid using the same test twice within one year.¹¹⁸ Six months is probably sufficient, but there appears to be no research identifying the point at which short-term practice effects dissipate.

The need to take practice effects into account is complicated by several factors.¹¹⁹ First, the above-referenced figures are just averages, and are usually based on the original test norms and data collected by the test publisher. Obviously, there are important individual variations. Some subjects gain more than others, and some subtests are more amenable to learning than others. For example, once the object assembly puzzles are solved, they are more easily solved the next time the test is administered. Particular trouble with this subtest would result in an above-average practice effect.

Practice effects vary according to the person's age, their learning ability and the time interval between testing. For subjects with significant intellectual deficits, such as mental retardation,

113. See KAUFMAN & LICHTENBERGER, *supra* note 92, at 202.

114. *Id.*

115. See *id.* at 207–08.

116. *Id.* at 209.

117. *Id.*

118. See *id.*

119. The observations on practice effects in the next few paragraphs are based on personal e-mail correspondence from Dewey Cornell, Ph.D., Clinical Psychologist, Curry School of Education, University of Virginia (March 6, 2007) (on file with author).

short term practice effects are likely to be smaller than for persons with greater learning ability. However, there appears to be little research demonstrating this effect.¹²⁰ Clinically, adults who have been in mental retardation educational programs for many years will have been tested numerous times. It is common that such persons show an overall increase in IQ (sometimes now scoring above 70), despite continuing deficits in social intelligence and adaptive behavior. In these cases, there may be an accumulated practice effect which suggests—based on scores alone—that the adult is no longer mentally retarded.

It should also be emphasized that motivation may vary among subjects and across settings. People with mental retardation are often reluctant to engage and put forth their best effort in testing because of frustration, and as a result, their ability may be underestimated.¹²¹ A skilled examiner who establishes rapport with the subject can obtain markedly higher scores which might be misconstrued as “practice effects” or even gains in intelligence.

Practice effects are likely to be a particular problem in the context of the criminal justice system. The contested nature of *Atkins* adjudications means that both prosecutors and defendants are likely to retain their own experts. Each expert is likely in turn to want to administer his own IQ test. In the post-*Atkins* assessments of Daryl Atkins, the defense expert administered the WAIS-III, and the government’s expert, Dr. Stanton Samenow, administered the same test two days later.¹²² On cross-examination, Dr. Samenow testified that he knew the WAIS-III had been administered two days prior and that ordinarily it would not be accepted professional practice to readminister the same test because of the resulting practice effects.¹²³ However, Dr. Samenow testified that he decided to readminister the WAIS-III because it was the only IQ test he had brought to the prison.¹²⁴ The court held that, while it did not affect the admissibility of the expert’s opinion, readministering the test was highly relevant to the weight to be accorded to that opinion.¹²⁵

120. *Id.*

121. *Id.*

122. *Atkins v. Commonwealth*, 631 S.E.2d 93, 96 (Va. 2006).

123. *Id.*

124. *Id.* at 96 n.4.

125. *Id.* at 96 & n.6.

4. Mental Illness

Mental illness, as well as general mood and physical health, adds a situational component to performance on measures of IQ.¹²⁶ However, since mental retardation is an enduring condition, courts should be aware that whether a defendant performed in the mentally retarded range at the time of the IQ test does not necessarily demonstrate that he is mentally retarded. Rather, it is important to have an IQ score that is representative of the defendant's typical intellectual functioning. If a defendant is affected by acute symptoms of mental illness at the time of an IQ test, it is well known that his cognitive functioning may be depressed.¹²⁷ This would result in a low IQ score and perhaps, a diagnosis of mental retardation although the defendant is not in fact mentally retarded. In an ironic reversal of the usual roles, well-informed prosecutors may therefore want to rely upon the defendant's mental illness (at the time of testing) in support of the argument that the defendant's actual IQ is higher than suggested by scores on tests administered during acute exacerbations of mental illness. Courts should be aware of this problem and regard scores obtained when a defendant was affected by acute symptoms of mental illness with skepticism.

5. Substandard Normative Sample for the WAIS-III

All IQ scores from the WAIS-III require some adjustment because the normative sample was substandard and thus unrepresentative of the population as a whole.¹²⁸ Scores on the WAIS-III were inflated by 2.34 points even at the time it was normed and thus courts and clinicians should deduct 2.34 points from all WAIS-III scores.¹²⁹

C. Cut-Off Scores

For purposes of the preceding discussion of the role of standard error of measurement, the Flynn effect, and other factors affect-

126. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 182.

127. Thomas Grisso and Paul S. Appelbaum, *The MacArthur Treatment Competence Study, III: Abilities of Patients to Consent to Psychiatric and Medical Treatments*, 19 J. LAW & HUMAN BEHAVIOR 149, 154, 159 (1995).

128. See Flynn, *supra* note 93, at 178.

129. See *id.*

ing the interpretation of IQ scores, we have assumed that “significantly subaverage intellectual functioning” is operationalized as scoring 70 or below on a specific IQ test. While this is a useful assumption to make when explaining the effects of factors like standard error of measurement, a score of 70 or below (or any other specific cut-off score) should not be embraced by courts or state legislatures as a part of the definition of mental retardation.¹³⁰ Instead, a significant limitation in intellectual functioning should be defined as performance that is at least two standard deviations below the mean, as both the AAMR and the APA recommended,¹³¹ and the Virginia statute provides.¹³²

The standard deviation measures the variation of scores in comparison to the mean score of the population on which the test has been normed. Two-thirds of the population will have scores falling within one standard deviation on either side of the mean, and 95 percent of the population will have scores falling within two standard deviations on either side of the mean. The WAIS-III and the SB5 both have a mean score of 100 and a standard deviation of 15.¹³³ Two-thirds of the population will thus have a score between 85 and 115 (one standard deviation) and 95 percent of the population will have a score between 70 and 130 (two standard deviations).¹³⁴ Consequently, an IQ score of 70 is sometimes used as a proxy for a score two standard deviations below the mean, but the score of 70 should not be reified.

130. See Ellis, *supra* note 26, at 13; 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 & L. 304, 305–06 (2004); Kay B. Stevens & J. Randall Price, *Adaptive Behavior, Mental Retardation, and the Death Penalty*, 6 J. FORENSIC PSYCHOL. PRAC. 1, 21 (2006). Unfortunately, at least thirteen states use a specific score as a cut-off, typically 70, although Arkansas specifies 65 and Illinois specifies 75. See State Statutes Prohibiting Death Penalty for People with Mental Retardation, Death Penalty Information Center, <http://www.deathpenaltyinfo.org/article/php?scid=28&did=138> (last visited Apr. 10, 2007); States that Have Changed Their Statutes to Comply With the Supreme Court's Decision in *Atkins v. Virginia*, Death Penalty Information Center, <http://www.deathpenaltyinfo.org/article.php?scid=28&did=668> (last visited Apr. 10, 2007).

131. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 13; AM. PSYCHIATRIC ASS'N, *supra* note 35, at 41–42 (specifying *approximately* 70 as a cut-off score, but making clear that 70 is intended to be an approximation of a score two standard deviations below the mean); see also Bonnie, *supra* note 130, at 305–06.

132. See VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

133. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 61; ROID & BARRAM, *supra* note 69, at 3.

134. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 61; ROID & BARRAM, *supra* note 69, at 3.

Standard deviations should be used in the definition and diagnosis of mental retardation instead of cut-off scores for a number of reasons. First, different IQ tests use different scoring norms, meaning that the mean score does not necessarily have to be set at 100. Setting a cut-off score of 70 in the definition would therefore be nonsensical since 70 means different things on different tests, unless the law also directed evaluators to use a particular test normed with a mean of 70. As previously discussed, however, it would be unwise to mandate the use of one particular measure because it may not be well suited to all defendants and examiners.

Second, different IQ tests may have different standard deviations. For example, the fourth revision of the Stanford-Binet scale (SB4) had a standard deviation of 16.¹³⁵ The diagnosis of mental retardation refers to intellectual functioning that is in a low percentile *compared to the rest of the population*, not to the fact that a person scored less than a certain number on an IQ test. The score has no inherent meaning; it refers to the subject's performance in comparison to the relevant population. A person who scored 70 on the SB4 would not have performed in the bottom 2.5 percent of the population (the percentage of the population that falls more than two standard deviations below the mean) because only a score of 68 was two standard deviations below the mean on that test.

Third, a fixed cut-off score would ignore the fact that different IQ tests have different standard errors of measurement (SEM).¹³⁶ Given the inevitable presence of some measurement error, each IQ score should actually be viewed as a range of possible scores. Like differences in standard deviations, differences in SEM between measures suggest that a score of 70 does not necessarily mean the same thing on different instruments.

Fourth, IQ tests are generally most accurate with respect to people who fall within two standard deviations of the mean. Since people with mental retardation by definition fall outside this group, their scores are somewhat less trustworthy than those

135. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 62.

136. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 58; Brian R. Bryant, *Intelligence Testing*, in ASSESSMENT OF INDIVIDUALS WITH MENTAL RETARDATION 13, 13 (Ronald L. Taylor ed., 1997).

closer to the mean.¹³⁷ Although this decrease in confidence as IQ scores approach high or low extremes certainly does not discredit the validity of extreme scores, it does highlight the importance of viewing the diagnosis of mental retardation as involving clinical judgment and evaluation of all three diagnostic criteria (intellectual functioning, adaptive behavior, and age of onset). Establishing a fixed cut-off score would ignore the role of clinical judgment in the diagnosis of mental retardation.

In sum, a cut-off score for intellectual functioning would only be plausible if it were defined as 70 or below on a specific test with a mean of 100 and a standard deviation of 15 and if it incorporated some way of taking individualized factors such as the Flynn effect and practice effects into account. This is not an advisable approach. Not only should clinicians have freedom to select the IQ test that in their judgment is most appropriate for the defendant being examined, it would be very cumbersome to develop a standard protocol for incorporating important individualized factors "into" the cut-off score.

In addition to these scientific objections to using a cut-off score, such scores are also objectionable on legal grounds. Focusing on a "number" implies that the diagnosis of mental retardation is more mechanical and more objective than it really is, and tends to obscure the inevitable clinical factors that affect performance on standardized cognitive measures. These clinical factors must be taken into account in judging whether a particular person exhibits "significantly subaverage intellectual functioning," meriting a diagnosis of mental retardation. Ultimately, in cases on the margin, experts need to exercise their own judgment, as do judges and juries. De-emphasizing a cut-off score is another way of highlighting the challenge of deciding each *Atkins* adjudication on its own merits. Courts and juries can hear testimony from experts on both sides to determine what a particular IQ score actually means. Then, the court should determine for themselves whether a defendant has the requisite intellectual deficits and whether he is mentally retarded when all three components of mental retardation are considered.

In *Johnson v. Virginia*, the Supreme Court of Virginia incorrectly observed that the statutory requirement that a defendant's

137. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 56.

IQ score fall at least two standard deviations below the mean is interchangeable with scoring 70 or less on an IQ test.¹³⁸ The court stated that the defendant's "scores of 75 and 78 on these I.Q. tests exceed the score of 70 that the General Assembly has chosen as the threshold score below which one may be classified as being mentally retarded."¹³⁹ The General Assembly did not establish 70 as a threshold score.¹⁴⁰ It referred to a score that is at least two standard deviations below the mean as a threshold score.¹⁴¹ As we have explained above, this is not a trivial distinction.

III. ASSESSMENT OF ADAPTIVE BEHAVIOR

A. *The Definition and Measurement of Adaptive Behavior*

In addition to having substantial limitations in intellectual functioning, an individual must have significant limitations in adaptive behavior in order to be diagnosed as mentally retarded.¹⁴² Adaptive behavior is included in all major classification systems and there is a professional consensus that it is an essential component of mental retardation.¹⁴³ Adaptive behavior refers to how effectively an individual can meet the demands of daily life and to what extent an individual can live independently, in comparison to what would be expected of his age group, sociocultural background, and community.¹⁴⁴ For the purposes of measurement, adaptive behavior is a combination of conceptual, social, and practical skills that people learn in order to function in everyday life.¹⁴⁵ Conceptual skills include cognitive abilities, communication, and academic skills such as language, the use of

138. *Johnson v. Commonwealth*, 591 S.E.2d 47, 59 (Va. 2004).

139. *Id.*

140. *See* VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

141. *Id.*

142. *Id.*; AM. PSYCHIATRIC ASS'N, *supra* note 35, at 41; AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 8.

143. Robert L. Schalock, *Introduction to ADAPTIVE BEHAVIOR AND ITS MEASUREMENTS* 2 (Robert L. Schalock ed., 1999).

144. AM. PSYCHIATRIC ASS'N, *supra* note 35, at 42; JAMES R. THOMPSON ET AL., *Adaptive and Maladaptive Behavior: Functional and Structural Characteristics*, in *ADAPTIVE BEHAVIOR AND ITS MEASUREMENTS* 15, 16 (Robert L. Schalock ed., 1999).

145. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 14; *see also* VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

money, and self-direction.¹⁴⁶ Social skills include interpersonal relationships, self-esteem, lack of gullibility, and the ability to follow rules.¹⁴⁷ Practical skills are independent living skills such as toileting, eating, housekeeping, transportation, and occupational skills.¹⁴⁸ Limitations in adaptive behavior may result from not knowing how to perform a skill (acquisition deficit) or not knowing when to use a learned skill (performance deficit).¹⁴⁹ Either type of deficit can contribute to mental retardation.¹⁵⁰ Adaptive behavioral strengths are likely to coexist with weaknesses. Maladaptive or problem behavior is not considered a limitation in adaptive behavior itself, but it may affect the performance or acquisition of adaptive behavioral skills.¹⁵¹

Adaptive behavior was added to the AAMR definition in 1959 in order to reflect the social characteristics of mental retardation, to reduce reliance on IQ scores in diagnoses and incorporate evidence of functioning in the real world, and to decrease the number of people falsely identified as having mental retardation.¹⁵² Although the importance of adaptive behavior is universally recognized, scientific understanding of adaptive behavior and its measurement is still in its infancy, especially in comparison to the conceptualization and measurement of intelligence. Notwithstanding general agreement that limitations in adaptive behavior should be part of the definition of mental retardation, researchers continue to investigate its structure and the best methods to assess it.¹⁵³ The lack of clarity in the construct of adaptive behavior is evidenced by the fact that while over 200 measures of adaptive behavior have been created, they differ markedly from one another.¹⁵⁴ This calls into question the validity of the measures—i.e., whether the tests actually measure what scientists intend them to measure and whether useful inferences regarding the di-

146. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 42.

147. *Id.*

148. *Id.* at 42.

149. *Id.* at 73–74.

150. *Id.* at 74.

151. *Id.* at 75.

152. *Id.* at 24; Stephen Greenspan, *A Contextualist Perspective on Adaptive Behavior*, in *ADAPTIVE BEHAVIOR AND ITS MEASUREMENTS* 61, 61 (Robert L. Schalock ed., 1999).

153. See AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 24; Schalock, *supra* note 143, at 2; THOMPSON ET AL., *supra* note 144, at 16; Stevens & Price, *supra* note 130, at 5.

154. Scott Spreat, *Psychometric Standards for Adaptive Behavior Assessment*, in *ADAPTIVE BEHAVIOR AND ITS MEASUREMENTS* 103, 103 (Robert L. Schalock ed., 1999).

agnosis of mental retardation can be drawn from the subjects' performance on these tests.¹⁵⁵

The concept of adaptive behavior is far more important to diagnoses of mild mental retardation than to diagnoses of moderate, severe, or profound mental retardation. A diagnosis of moderate to profound mental retardation will rarely turn on standardized measurement of adaptive behavior because significant limitations will be so clearly apparent in those individuals.¹⁵⁶ Precise and accurate assessment of adaptive behavior, however, is crucial for diagnoses of mild mental retardation because in borderline cases it may well tip the scales either towards or away from a diagnosis.¹⁵⁷

Instruments for assessing adaptive behavior typically involve interviews with (or questionnaires completed by) third-party informants, such as parents or teachers, who have significant experience interacting with the individual being evaluated.¹⁵⁸ Informants should satisfy four criteria: (1) they should have contact with the defendant almost every day, (2) their contacts should generally last for at least several hours at a time, (3) they should have had these contacts within the last several months, and (4) they should have had opportunities to observe the variety of skills measured by the adaptive behavior test.¹⁵⁹ Informants rate the individual's typical behavior in settings with which they are familiar and they must report the individual's actual performance as opposed to ability to perform.¹⁶⁰

Three main criteria should be considered by clinicians when selecting a measure of adaptive behavior: (1) the reliability and validity of the measure, (2) the reliability of the informant, and (3) the scope, structure, and clinical utility of the measure.¹⁶¹ Measures should be chosen based on the purpose of the assessment, the characteristics of the individual being tested, and the technical adequacy of the measure (for example, the normative sample

155. *Id.* at 110.

156. Stevens & Price, *supra* note 130, at 6.

157. *Id.*

158. *Id.*

159. *Id.* at 24.

160. *Id.* at 6.

161. Candace H. Boan & Patti L. Harrison, *Adaptive Behavior Assessment and Individuals with Mental Retardation*, in ASSESSMENT OF INDIVIDUALS WITH MENTAL RETARDATION 33, 43 (Ronald L. Taylor ed., 1997).

must include people without disabilities).¹⁶² Three commonly used tests are the Vineland Adaptive Behavior Scales, the AAMR Adaptive Behavior Scales, and the Scales of Independent Behavior.¹⁶³

The reliability of adaptive behavior measures remains a serious concern. Reliability refers to the repeatability of a measurement when administered by the same person (test-retest reliability) and the degree of agreement when the measure is administered by different people (inter-rater reliability). Most published measures of adaptive behavior have acceptable levels of reliability, as shown by mathematically calculated reliability coefficients of at least 0.90.¹⁶⁴ However, measures of adaptive behavior assume that it is a trait rather than a temporary state, like mood.¹⁶⁵ As a trait, it is assumed to be accurately measurable by administering the measure once. In contrast, if adaptive behavior depends on environmental variables, a single administration would not yield a reliable score.¹⁶⁶

Specifically, the administration of measures of adaptive behavior to a defendant who is incarcerated poses significant problems. Many of the skills in the operational definition of adaptive behavior are not relevant in prisons, such as self-direction, community resources, and leisure skills. A mentally retarded person is also likely to show stronger adaptive behavior in the structured environment of a correctional facility than in society, thus possibly inflating scores that would have been indicative of mental retardation in the community environment.¹⁶⁷ Unfortunately, there will

162. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 13, 83–84.

163. *Id.* at 88–90. A relatively new measure that can be used to assess adaptive behavior in adults up to 89 years old and that has received favorable reviews is the Adaptive Behavior Assessment System – Second Edition (ABAS-II). Personal Communication from James Morris, Director of Office of Forensic Services, Department Mental Health, Mental Retardation and Substance Abuse Services, Apr. 6, 2007.

164. *See id.* at 87–88.

165. *Id.*

166. Spreat, *supra* note 154, at 107. It is not uncommon for an individual to be able to demonstrate a particular skill to one examiner but not another. *Id.* at 107–08.

167. Melissa Piaseki, *handout, Mental Retardation in Capital Cases, Managing the Capital Case in North Carolina Conference, The University of North Carolina, School of Government and the National Judicial College (Mar.-Apr. 2006) available at [http://www.judges.unc.edu/200603CapitalCases/Bell%20-%20Pretrial%20Issues%20\(Mental%20Retardation\).doc](http://www.judges.unc.edu/200603CapitalCases/Bell%20-%20Pretrial%20Issues%20(Mental%20Retardation).doc); Stevens & Price, *supra* note 130, at 19; Stanley L. Brodsky & Virginia A. Galloway, *Ethical and Professional Demands for Forensic Mental Health Professionals in the Post-Atkins Era*, 13 J. ETHICS & BEHAVIOR 3, 7 (2003).*

usually be no adaptive behavior test on record prior to the offense. For this reason, it is important for experts conducting *Atkins* evaluations to obtain information relating to the defendant's adaptive skills before the offense occurred and prior to incarceration to augment whatever recent information is provided by informants. Lawyers should make every effort to locate as many records as possible from schools, employers, health providers, and government agencies so that evaluations of adaptive behavior take into account information from all available sources.¹⁶⁸

The measures of adaptive behavior that are currently available have several further limitations. Unless the informant is a parent, he or she is unlikely to have extensive knowledge of the individual's behavior in different settings. A teacher, for example, would be able to speak to the individual's abilities at school but not at home. Third-party informants may also have an inaccurate perception of what is age-appropriate. For example, a teacher of disabled children and a teacher of gifted children of the same age would likely have different perceptions. Particularly important in the context of the criminal justice system, parents or other informants may be influenced in their descriptions by their relationship with the defendant and by their desire to help the defendant avoid a death sentence. If an informant who has a close relationship to the defendant realizes that the defendant may escape the death penalty if the informant downplays the defendant's adaptive behavior, it is unlikely that the informant would be objective. A final limitation of adaptive behavior measurement is that they cannot be administered retrospectively and thus can only measure the defendant's current functioning.¹⁶⁹ Generally, measures of adaptive behavior are not as objective as measures of intelligence; consequently, clinical judgment matters even more in the context of adaptive behavior than it does in the assessment of deficits in intellectual functioning.

B. Definition and Assessment of Adaptive Behavior under the Virginia Statute

The definition of mental retardation in the Virginia statute requires "significant limitations in adaptive behavior as expressed

168. Stevens & Price, *supra* note 130, at 14.

169. *Id.* at 15-16.

in conceptual, social and practical adaptive skills.”¹⁷⁰ With respect to the assessment of adaptive behavior, the statute states that:

[A]daptive behavior shall be based on multiple sources of information, including clinical interview, psychological testing and educational, correctional and vocational records. The assessment shall include at least one standardized measure generally accepted by the field of psychological testing for assessing adaptive behavior and appropriate for administration to the particular defendant being assessed, unless not feasible. In reaching a clinical judgment regarding whether the defendant exhibits significant limitations in adaptive behavior, the examiner shall give performance on standardized measures whatever weight is clinically appropriate in light of the defendant's history and characteristics and the context of the assessment.¹⁷¹

This language was substantively identical to the language originally proposed by the CAG.¹⁷² The drafters envisioned that a standardized measure of adaptive behavior would ordinarily be administered in *Atkins* cases in the absence of a compelling clinical or scientific demonstration of infeasibility, such as the absence of a measure taking into account certain disabilities or a virtually life-long history of confinement.¹⁷³ The leeway provided in the statute for a decision to forgo an adaptive behavior test contrasts with the statute's clear requirement that a test of intellectual functioning must always be administered.¹⁷⁴ This difference in language was created to reflect the differences between standard-

170. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004).

171. VA. CODE ANN. § 19.2-264.3:1.1(B)(2) (Repl. Vol. 2004).

172. The CAG report dated Oct. 17, 2002, recommended:

Assessment of adaptive behavior should be based on multiple sources of information, including clinical interview, psychological testing and educational, correctional, and vocational records, and should, whenever feasible, include at least one standardized measure for assessing adaptive behavior, administered in accordance with methods generally accepted by the field and appropriate for the particular person being assessed, taking into account the environments in which the person has lived as well as cultural, linguistic, sensory, motor, behavioral and other individual factors. In reaching a clinical judgment regarding whether the person exhibits “substantial limitations in adaptive behavior,” the examiner should give performance on standardized measures whatever weight is clinically appropriate in light of the person's history and characteristics and the context of the assessment.

CAG Report, Oct. 17, 2002, *supra* note 72.

173. *See id.*

174. “Assessment of intellectual functioning shall include administration of at least one standardized measure generally accepted by the field of psychological testing” VA. CODE ANN. § 19.2-264.3:1.1(B)(1) (Repl. Vol. 2004).

ized measures of intelligence and adaptive behavior. In particular, it was designed to reflect scientists' comparatively less developed understanding of adaptive behavior and its measurement, as well as the more subjective nature of measures of adaptive behavior which may be difficult to administer accurately in a prison setting.

Based on the language in the AAMR definition, the statute's definition of mental retardation requires that an individual possess "significant limitations in adaptive behavior as expressed in conceptual, social *and* practical adaptive skills."¹⁷⁵ This language reflects an important ambiguity: Does the use of the conjunction "and" imply that significant limitations must be found in all three domains? The answer is no; the AAMR Manual explicitly states that in the context of measuring adaptive behavior, an individual need only score at least two standard deviations below the mean *in at least one domain*.¹⁷⁶ If that is so, why didn't the AAMR definition use "or" instead of "and"?

The confusion created by the AAMR definition can be dispelled by highlighting the important distinction between *the concept of adaptive behavior* on the one hand, and *the tests used to measure deficits in adaptive behavior* on the other. Although the construct of adaptive behavior includes significant deficits in all three domains, the AAMR *operationalizes* significant limitations in adaptive behavior as a score at least two standard deviations below the mean either on a test of one type of adaptive behavior or in a composite test of all three types.¹⁷⁷ In other words, people diagnosed with mental retardation must have significant limitations in adaptive behavior by definition and must have some meaningful limitation in all three domains. A score of at least two standard deviations below the mean on a standardized test is required in only one domain to document and support the diagnosis.

When the problem is understood as a measurement problem rather than a conceptual problem, one wonders why a performance at least two standard deviations below the mean in one domain is regarded as sufficient to support the diagnosis. Several reasons have been offered. First and most important, adaptive

175. VA. CODE ANN. § 19.2-264.3:1.1(A) (Repl. Vol. 2004) (emphasis added).

176. AM. ASS'N ON MENTAL RETARDATION, *supra* note 36, at 78.

177. *Id.* at 13.

behavior is viewed as a unitary concept.¹⁷⁸ Thus, generally speaking, substantial limitations in one domain are likely to be accompanied by substantial limitations in the other two domains.¹⁷⁹ However, at the present time, correlations in performance across the three domains vary significantly among different instruments.¹⁸⁰ As a result, if a score at least two standard deviations below the mean were required in more than one domain, the choice of instrument would have a significant impact on whether a person is found to be significantly limited in their adaptive behavior.

In addition, when an individual has a score at least two standard deviations below the mean in one domain, that limitation in itself will usually have such a major adverse effect on the person's adaptive functioning to qualify them as having a substantial limitation in adaptive behavior generally.¹⁸¹ A final reason for requiring a score in only one domain is that the probability of someone scoring at least two standard deviations below the mean on more than one of the three domains is very low at the present time, given the tests that are now in use.¹⁸² Requiring scores more than two standard deviations below the mean in all three domains would likely exclude nearly all individuals with "mild" retardation (i.e., with scores in the range of 60–70 on the customary IQ tests) from qualifying for the diagnosis. In other words, when viewed cumulatively, the available tests are not valid.

The Clinical Advisory Group discussed this problem at length. Recognizing that the practical significance of the definitional language lies in its impact on expert testimony and arguments by counsel and jury instructions, the CAG considered the possibility of altering the confusing AAMR language to change "and" to "or," in congruence with the AAMR's directions regarding testing and diagnosis. However, the CAG decided that making such a change would likely be regarded as an attempt to change rather than clarify the AAMR definition and would raise suspicion about the CAG's motives. The CAG therefore decided to leave the AAMR definition intact. Preserving the AAMR definition also carries im-

178. *Id.* at 74.

179. *Id.* at 74.

180. *Id.* at 78.

181. *Id.*

182. *Id.*

portant symbolic value since it embodies a commitment to scientific objectivity in the definition and diagnosis of mental retardation—a key implication of the *Atkins* decision.

That being said, it is important for witnesses, attorneys, and courts to understand the particular subtlety in the definition of mental retardation. The definition envisions significant deficits in all three domains. However, when it comes to the *testing* of deficits in adaptive behavior, the diagnosis is supported by significant limitations in adaptive behavior as expressed in conceptual, social, or practical skills. Courts must be careful to avoid an interpretation of the statute that requires scores at least two standard deviations below the mean in all three domains. Otherwise, courts will impose far more demanding requirements for a diagnosis of mental retardation than were intended by the scientists who drafted the AAMR definition and are used in clinical practice.

In light of the drafting history just described, it will come as no surprise that the CAG and the Crime Commission drafters strongly resisted proposals to resolve the ambiguity in the AAMR definition in the other direction—i.e., by requiring proof that the defendant had substantial deficits in *all three* domains. The AAMR definition was embraced unchanged in the bill that emerged from the Senate.¹⁸³ However, as indicated earlier, the bill that emerged from the House had added a “diminished capacity” requirement to the AAMR definition.¹⁸⁴ When the two bills were reconciled by the Joint Conference Committee on February 22, 2003, the compromise entailed dropping the House’s “diminished capacity” language while modifying the AAMR definition of adaptive behavior to require “significant limitations in adaptive behavior as expressed in *all of the following*: conceptual adaptive skills, social adaptive skills and practical adaptive skills.”¹⁸⁵ After

183. S. 1239, as originally introduced in the Senate by Senator Norment, discarded the Crime Commission’s definition of adaptive behavior in favor of a definition based on the language of DSM IV. S. 1239 offered January 8, 2003, § 19.2-264.3:1.1(A). However, the Senate Courts of Justice Committee rejected the Norment bill in favor of the Crime Commission’s proposal introduced by Senator Stolle (originally S. 1141). Technically, the Crime Commission’s bill became an Amendment in the Nature of a Substitute for S. 1239 reported out of Committee on January 22, 2003.

184. H. COMM. FOR COURTS OF JUSTICE, AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003).

185. JOINT CONF. COMM., AMENDMENT IN THE NATURE OF A SUBSTITUTE, S. 1239, 2003 Gen. Assem., Reg. Sess. (Va. 2003) (emphasis added).

consultation with the CAG chair, the Governor recommended that the AAMR definition be restored, and this recommendation was accepted by the Senate (39–0) and by the House (47–44).¹⁸⁶ This was a critical victory for the Crime Commission proposal.

IV. DEVELOPMENTAL ORIGIN

The requirement in the definition of mental retardation that the disability become manifest before the age of eighteen is intended to differentiate between mental retardation and intellectual deficits “acquired” after the developmental period, typically due to brain injuries or brain diseases such as dementia. As discussed earlier, it can be argued that people with “acquired” deficits in intellectual functioning and adaptive behavior equivalent to those associated with mental retardation should also be exempted from the death penalty because there is no reason to think that the age at which the disability becomes manifest is relevant to the disabled person’s culpability at the time of the offense.¹⁸⁷ Based on that reasoning, the American Bar Association and the mental health professional associations have endorsed a broadened statement of the *Atkins* principle.¹⁸⁸ Be that as it may, however, most current statutes, including Virginia’s, confine the exemption to “mental retardation” and therefore include developmental origin as part of the definition.¹⁸⁹

The developmental origin limitation has some practical advantages in an *Atkins* adjudication because requiring onset before age eighteen also reduces the danger of malingering.¹⁹⁰ The disability must have manifested during the defendant’s childhood and thus prior to the creation of any incentives to appear mentally retarded in order to avoid the death penalty. Although malingering that escapes detection by clinicians has not been found to be a significant concern in the diagnosis of mental retardation, the age-of-onset criterion should eliminate any concern that de-

186. Legislative Information Systems Bill Tracking, Summary of S. 1239 (2003 Reg. Sess.).

187. See *supra* note 43.

188. *Id.*

189. See, e.g., ARIZ. REV. STAT. ANN. § 13-703.02 (2001); ARK. CODE ANN. § 5-4-618 (2006); COLO. REV. STAT. § 18-1.3-1102 (2006); FLA. STAT. § 921.137 (2006).

190. See *supra* note 43; Tobolowsky, *supra* note 38, at 99.

endants may somehow be able to feign impaired cognitive functioning.¹⁹¹

It must be emphasized, however, that requiring a finding of developmental onset does not require that the diagnosis have been made before the age of eighteen or that standardized testing used to support the diagnosis have been administered before the age of eighteen.¹⁹² Such a requirement would be unconstitutional because it would amount to discrimination against people whose need for special education was overlooked and who did not have access to adequate clinical or social services as a child. The age-of-onset requirement therefore only requires that there is evidence, not necessarily test scores, that intellectual and adaptive deficits became manifest before the age of eighteen.¹⁹³ Furthermore, if a defendant is shown by testing and history to have mental retardation, and there is no suspicion of malingering,¹⁹⁴ it should be presumed that its onset was before the age of eighteen in the absence of any evidence of later onset, such as brain injury sustained in adulthood or the dementia.

In sum, the developmental onset requirement, though diagnostically essential, does very little work in the ordinary *Atkins* adjudication. Scores on standardized tests before age eighteen should, all things being equal, be accorded greater weight than tests administered in adulthood. Courts should carefully consider any other evidence from the developmental period. However, courts should not require a diagnosis before the age of eighteen or scores in the range of mental retardation from IQ tests administered before the age of eighteen. Courts should regard tests administered during adulthood (even after the capital offense) as highly probative on the diagnosis of mental retardation. Finally, courts should presume that currently diagnosed mental retardation had a developmental onset in the absence of clear evidence of post-childhood onset of the defendant's disability.

191. Ellis, *supra* note 26, at 13–14. Any suspicions of malingering in relation to current testing can be pursued by assembling evidence regarding manifestations of deficits in intellectual functioning and adaptive behavior *before* the offense, especially during childhood. Such evidence may substantiate malingering, and therefore undermine the diagnosis. Because malingering is so difficult to carry out successfully, such cases should be rare.

192. See Tobolowsky, *supra* note 38, at 99.

193. See *id.*

194. See *supra* note 191 and accompanying text.

V. EXPERTS IN *ATKINS* CASES

Expert testimony plays a crucial role in *Atkins* determinations. It is important that courts have access to high quality professional opinions, but also to a sufficient breadth of expert opinion. The Virginia statute requires experts appointed by the court to conduct *Atkins* evaluations (at the request of either the defense or the Commonwealth) have a combination of (i) specialized expertise in the assessment of mental retardation and (ii) specialized training in conducting forensic evaluations.¹⁹⁵ The mental health expert must be,

(a) a psychiatrist, a clinical psychologist or an individual with a doctorate degree in clinical psychology, (b) skilled in the administration, scoring and interpretation of intelligence tests and measures of adaptive behavior and (c) qualified by experience and by specialized training, approved by the Commissioner of Mental Health, Mental Retardation and Substance Abuse Services, to perform forensic evaluations.¹⁹⁶

This language was designed to accomplish several things. First, by requiring expertise in administering, scoring, and interpreting intelligence tests and measures of adaptive behavior, the state will limit the pool of experts qualified for appointment for state-subsidized evaluations to experts in the diagnosis of mental retardation, thereby ruling out the mental health clinicians, especially psychiatrists, who carry out the vast majority of forensic assessments in criminal cases. The expertise envisioned by the statute encompasses the range of topics reviewed in this paper, including technical issues such as the Flynn effect and the measurement of deficits in adaptive behavior that are outside the ken of many mental health clinicians.¹⁹⁷ Second, the requirement of forensic expertise, modeled after analogous provisions governing other mental health evaluations in criminal cases, was designed to assure that the appointed experts are familiar with the legal context of the evaluations.¹⁹⁸ The practical effect of this requirement is that experts in the mental retardation services system

195. See VA. CODE ANN. § 19.2-264.3:1.2(A) (Repl. Vol. 2004).

196. *Id.*

197. See *supra* Part II.B.2 and text accompanying notes 177–82 (discussing the Flynn effect and explaining measurement of deficit in adaptive behavior).

198. Compare VA. CODE ANN. § 19.2-264.3:1.2(A) (Repl. Vol. 2004) with VA. CODE ANN. § 19.2-169.5 (Repl. Vol. 2004).

who want to conduct *Atkins* evaluations must enroll in an intensive training course approved by the Commissioner of Mental Health, Mental Retardation, and Substance Abuse Services.¹⁹⁹

It is important to emphasize that this provision governs the *appointment of experts* to conduct state-subsidized evaluations in connection with *Atkins* adjudications.²⁰⁰ It was not designed to govern or limit the pool of witnesses who are qualified to offer expert opinions in *Atkins* adjudications, a task that belongs to the judiciary.²⁰¹ For example, an expert retained by the defense who has recognized expertise in special education and experience in diagnosing mental retardation may be highly qualified to evaluate a capital defendant for purposes of an *Atkins* adjudication and to offer an opinion on that subject, even though he or she does not have a degree in clinical psychology or specialized forensic training. Exclusion of an otherwise admissible opinion by an otherwise qualified expert would not only constitute a misinterpretation of the statute, but would violate the defendant's right to present his case in mitigation under the Eighth and Fourteenth Amendments.

These observations are relevant to the Virginia Supreme Court's decision in its most recent review of Daryl Atkins's case. In *Atkins v. Commonwealth (Atkins V)*, a clinical psychologist, Stanton Samenow, was appointed by the court on the Commonwealth's motion to conduct an evaluation of Atkins under section 19.2-264.3:1.2(A).²⁰² Dr. Samenow conceded that he had never administered a standardized measure of adaptive behavior, although he testified that he routinely assessed adaptive behavior in other ways.²⁰³ The court held that his opinion on the diagnosis of mental retardation was inadmissible because he was not "skilled" in the administration of measures of adaptive behavior

199. The training is conducted by the University of Virginia Institute of Law, Psychiatry and Public Policy under contract with the Department of Mental Health, Mental Retardation and Substance Abuse Services and the Office of the Attorney General. Training Calendar, Institute of Law, Psychiatry & Public Policy, University of Virginia School of Law & School of Medicine, <http://www.ilpp.virginia.edu/trainingcalendar.html> (last visited Apr. 10, 2007).

200. See VA. CODE ANN. § 19.2-264.3:1.2 (Repl. Vol. 2004).

201. See *Tarmac Mid-Atlantic, Inc. v. Smiley Block Co.*, 458 S.E.2d 462, 465 (Va. 1995) ("The admission of expert testimony is committed to the sound discretion of the trial judge.") (quoting *Brown v. Corbin*, 423 S.E.2d 176, 178 (Va. 1992)).

202. *Atkins v. Commonwealth*, 631 S.E.2d 93, 96 (Va. 2006).

203. *Id.* at 97-98.

. . . [and] lack[ed] the requisite expertise in scoring and interpreting such tests," and thus not qualified to be appointed as a state expert.²⁰⁴ The court made clear that in the context of adaptive behavior, the expert does not actually have to administer a standardized measure of adaptive behavior to the defendant, but only needs to be skilled in scoring and interpreting such measures.²⁰⁵ If the expert has these necessary skills and determines that administering a standardized measure would be infeasible or inappropriate in a given case, he can testify as to the defendant's mental retardation and explain why an adaptive behavior test was not given.²⁰⁶ While the basic qualifications possessed by the expert may go to the admissibility of his opinion, the court said, the decision by a qualified expert about which test to administer only goes to the weight of his opinion.²⁰⁷

Although the court reached the correct result in *Atkins V* regarding the admissibility of Dr. Samenow's testimony, the court's rationale needs a little elaboration. As indicated above, the qualifications specified in section 19.2-264.3:1.2(A) relate to the appointment of experts, not to the qualifications of witnesses to offer expert opinions, a judgment that should be governed by the customary criteria regarding the qualification of a person as an expert.²⁰⁸ Accordingly, if Dr. Samenow had been called to testify, not as an expert appointed under section 19.2-264.3:1.2(A), but rather as an expert retained by one of the parties, his opinion would have been admissible if he had been found to be skilled in the diagnosis of mental retardation.²⁰⁹ In this hypothesized situation, the alleged deficiencies of his evaluation—including failing to administer a test of adaptive behavior and administering a WAIS-III two days after *Atkins* had taken the test in another evaluation—would bear on the weight that should be given his opinion.²¹⁰ However, the Supreme Court of Virginia properly held

204. *Id.* at 98.

205. *Id.*

206. *Id.* at 98–99.

207. *Id.* at 99. The Court also noted that Dr. Samenow's decision to administer the WAIS-III two days after another psychologist had given *Atkins* the same test would not render his opinion inadmissible either; it would bear on the weight his opinion should be given. See *supra* text accompanying notes 124–26.

208. See *supra* note 201.

209. See *id.*

210. Dr. Samenow actually had a plausible reason for questioning the usefulness of a test of adaptive behavior in *Atkins*'s case. He testified, "I don't think it would have been a

that Dr. Samenow's opinion was inadmissible in *Atkins V* because he was erroneously appointed by the court to conduct the evaluation in the first place, and there is no other way, aside from exclusion, to enforce the restrictions on appointment of experts set forth in section 19.2-264.3:1.2(A).²¹¹

In sum, the ultimate aim of the training and assessment requirements specified in the Virginia statute is to assure that *Atkins* evaluations are of the highest possible quality. Training programs for experts in *Atkins* cases should be developed by all states, even in the absence of an explicit legislative directive. Such programs should be jointly developed by the state supreme courts and the executive departments responsible for delivering and regulating mental retardation health services. The program should include the requisite training in forensic evaluation so that non-forensic specialists will understand the legal context of the assessment and the expectations courts will have of the evidence they provide. Training should further ensure that experts have an understanding of the features of mental retardation assessment that are particularly pertinent to *Atkins* evaluations and are particularly susceptible to challenges based on lack of validity or reliability. Experts need to be aware of the boundaries of scientific consensus in relation to the assessment of mental retardation. In particular, the construct of adaptive behavior should be emphasized in light of the uncertainties surrounding its measurement.²¹² Finally, training programs should explicitly address the potential for bias in conducting forensic evaluations, especially in capital cases, and the ways in which experts can try to maintain their objectivity.²¹³

CONCLUSION

Implementing the Supreme Court's decision in *Atkins* presents a host of challenges to the states. In its unprecedented ruling, *Atkins* makes the diagnosis of mental retardation a matter of life or

valid or reliable measure because [Atkins] had been incarcerated for quite some time. It would have had to have been retroactive. I think there are many ways to assess adaptive functioning that would be more informative." See *Atkins*, 631 S.E.2d at 97.

211. *Id.* at 98.

212. Currently, there is no accepted standard of practice for evaluating the adaptive behavior of defendants in death penalty cases. See Stevens & Price, *supra* note 130, at 20.

213. Brodsky & Galloway, *supra* note 167, at 5.

death. It is therefore imperative that state courts and legislatures make every effort to assure the scientific integrity and fairness of *Atkins* evaluations and adjudications. These diagnoses involve scientific and clinical nuance, particularly in the mild range of disability. In order to effectuate the Supreme Court's mandate faithfully and fairly, legislatures should conform their legislation as closely as possible to current scientific understanding, and should prescribe requirements for assessments designed to assure that evaluations are of the highest possible quality. Lawyers and judges should make a concerted effort to understand the definition of mental retardation and its measurement from a clinical perspective. In particular, courts must become familiar with scientific knowledge and clinical practice in the field of mental retardation.

Special attention should be paid to the interpretation of IQ scores and the meaning and measurement of adaptive behavior. With respect to intellectual functioning, standard error of measurement, the Flynn effect, practice effects, and other factors—especially when taken together—can influence IQ scores significantly. Courts should recognize and help juries understand that IQ scores should rarely be taken at face value and that professional practice standards play a crucial part in the interpretation of test scores. With respect to adaptive behavior—a construct more difficult to understand and apply than intelligence—courts need to understand its importance within the diagnosis of mental retardation. Courts should also recognize that it will be of particular relevance, as well as most challenging, in cases involving defendants with IQ scores in the upper range of mental retardation.

While careful application of current clinical understanding of mental retardation is important in all *Atkins* cases, the subtleties and the difficulties of using scientific standards in a legal setting are most pronounced when the defendant arguably falls within the mild range of mental retardation. In the end, clinical judgment plays an inescapable role in diagnosis in these cases. The challenge facing the courts, acknowledged by the Supreme Court in *Atkins*, is to adjudicate these borderline cases in a fair and consistent manner.