

A More Intelligent and Just *Atkins*: Adjusting for the Flynn Effect in Capital Determinations of Mental Retardation or Intellectual Disability

I.	INTRODUCTION: THE FLYNN EFFECT FOLLOWS <i>ATKINS</i> INTO THE COURTROOM.....	616
II.	INTELLIGENCE IN THE FACE OF DEATH: UNDERSTANDING THE FLYNN EFFECT'S PLACE IN <i>ATKINS</i>	620
A.	<i>It's All Relative: The Measurement of IQ</i>	620
B.	<i>Throwing the Curve: Explanation and Implications of the Flynn Effect</i>	623
C.	<i>Atkins at Work: Determinations of Mental Retardation or Intellectual Disability</i>	626
III.	DEATH ROW DILEMMA: INCONSISTENT TREATMENT OF THE FLYNN EFFECT IN <i>ATKINS</i> CASES.....	629
A.	<i>Validation of and Adjustment for the Flynn Effect</i>	631
B.	<i>Consideration of the Flynn Effect but Not Required Adjustment</i>	633
C.	<i>Indecision Regarding the Flynn Effect</i>	638
D.	<i>Rejection of the Flynn Effect</i>	640
IV.	RESOLVING THE BATTLE OF THE EXPERTS: THE FLYNN EFFECT IS VALID AND SHOULD BE ACCOUNTED FOR IN <i>ATKINS</i> CASES	641
A.	<i>Uncontroverted Acknowledgement That the Flynn Effect Is Real</i>	643
B.	<i>The Flynn Effect Applies to Individual IQ Scores</i>	645
C.	<i>The Dark Side of Bright-Line IQ Cutoffs</i>	646
1.	Approximating IQ as Recommended by Leading Organizations	647
2.	The Blunder of Bright-Line Cutoffs	648
3.	Legislative and Judicial Solutions	650
D.	<i>Validity Despite Variability</i>	652

1.	Research Consistently Measures the Flynn Effect Near 0.3 Points per Year	653
2.	Variability Across Ability Levels May Support Larger Adjustment	654
3.	A Call for More Research but Life and Practice Must Go On	657
E.	<i>Death Is Different: The Flynn Effect Reflects a “Best Practice” for Atkins</i>	658
V.	CONCLUSION: REQUIRING ADJUSTMENT FOR THE FLYNN EFFECT	661
VI.	APPENDIX: STATUTORY DEFINITIONS OF MENTAL RETARDATION OR INTELLECTUAL DISABILITY	664

I. INTRODUCTION: THE FLYNN EFFECT FOLLOWS *ATKINS* INTO THE COURTROOM

In *Atkins v. Virginia*, the U.S. Supreme Court declared a ban on all executions of mentally retarded persons.¹ This declaration, however, rings hollow for those mentally retarded defendants and inmates who continue to face death sentences as a result of the inconsistent enforcement of *Atkins* across jurisdictions. One issue in particular—whether to adjust intelligence-test scores for the phenomenon known as the Flynn Effect—has caused inconsistency among courts and has sparked a contentious battle among experts. It blurs the already-precarious line between life and death.² And yet, the

1. *Atkins v. Virginia*, 536 U.S. 304, 321 (2002). While the Court in *Atkins* used the term “mental retardation” to define its death penalty ban, a recent transition has occurred with several prominent organizations and scholars replacing “mental retardation” with the term “intellectual disability.” See, e.g., Robert L. Schalock et al., *The Renaming of Mental Retardation: Understanding the Change to the Term Intellectual Disability*, 45 INTELL. & DEV. DISABILITIES 116, 116–17 (2007) (“Increasingly, the term *intellectual disability* is being used instead of *mental retardation*. This transition in terminology is exemplified by organization names . . . journal titles, and published research.”). The term “intellectual disability” covers the same population and renders the same individuals eligible for a diagnosis as those covered and rendered by “mental retardation.” *Id.* at 116–17. Additionally, in 2010, Congress passed Rosa’s Law, which ordered that all references in federal laws to “mental retardation” be changed to “intellectual disability.” Rosa’s Law, Pub. L. No. 111-256, 124 Stat. 2643 (2010) (codified as amended at 20 U.S.C. §§ 1140, 1400–01, 7512 (2010); 29 U.S.C. §§ 705, 764, 791 (2010); 42 U.S.C. §§ 217, 247, 280, 285, 291, 294, 300 (2010)). For consistency, however, this Note uses the term “mental retardation” but acknowledges this shift in terminology.

2. As one psychologist explains, the Flynn Effect is “a dire matter with implications we seldom encounter in psychology”—*a matter of life and death*, as echoed by other psychologists. Alan S. Kaufman & Lawrence G. Weiss, *Guest Editors’ Introduction to the Special Issue of JPA on the Flynn Effect*, 28 J. PSYCHOEDUC. ASSESSMENT 379, 380 (2010); see also Stephen J. Ceci &

Flynn Effect captivates capital defendants and inmates with its promise of adjusting intelligence-test scores and of providing a more intelligent and just approach to *Atkins*.

The Flynn Effect (“the Effect”) refers to observed gains in intelligence-quotient (“IQ”) scores over time in the United States and other nations.³ Large-scale studies of the American population show that IQ scores are increasing at an average rate of 0.3 points per year.⁴ Simply put, people today receive higher IQ scores than people of the past. As discussed in Part II, the Flynn Effect’s IQ gains reflect the obsolete norms of outdated IQ tests. Test publishers standardize IQ tests and determine the basis for calculating IQ scores at specific points in time. Therefore, given the rise in IQ scores over time, an IQ test grows outdated and its norms grow obsolete as time passes from the specific point in time when the publisher standardized the test.⁵ Obsolete norms inflate IQ scores because they measure test performances against the scores of test takers from the past, as opposed to the higher scores of test takers from the present.⁶ To adjust American IQ scores for this inflation, the Flynn Effect calls for a score reduction of 0.3 points for every year that has passed between the year the test publisher standardized the test and the year the subject took the test.⁷

Although studied extensively since the 1980s,⁸ the Flynn Effect did not find a place in the courtroom until 2002, when the Supreme

Tomoe Kanaya, “*Apples and Oranges Are Both Round*”: *Furthering the Discussion on the Flynn Effect*, 28 J. PSYCHOEDUC. ASSESSMENT 441, 445–46 (2010) (discussing the consequences of a diagnosis of mental retardation); Alan S. Kaufman, *Looking Through Flynn’s Rose-Colored Scientific Spectacles*, 28 J. PSYCHOEDUC. ASSESSMENT 494, 500–02 (2010) (arguing that adjusting for the Flynn effect is necessary given the life-and-death consequences of not doing so).

3. James R. Flynn, *Massive IQ Gains in 14 Nations: What IQ Tests Really Measure*, 101 PSYCHOL. BULL. 171, 172–77 (1987) [hereinafter Flynn, *Massive IQ Gains in 14 Nations*]; James R. Flynn, *The Mean IQ of Americans: Massive Gains 1932 to 1978*, 95 PSYCHOL. BULL. 29, 32–34 (1984) [hereinafter Flynn, *The Mean IQ of Americans*].

4. Flynn, *The Mean IQ of Americans*, *supra* note 3, at 32, 34.

5. ALAN S. KAUFMAN, IQ TESTING 101 203 (2009); Flynn, *The Mean IQ of Americans*, *supra* note 3, at 39.

6. KAUFMAN, *supra* note 5, at 203.

7. JAMES R. FLYNN, WHAT IS INTELLIGENCE? BEYOND THE FLYNN EFFECT 112–14 (2007) [hereinafter FLYNN, WHAT IS INTELLIGENCE?]; James R. Flynn, *Tethering the Elephant: Capital Cases, IQ, and the Flynn Effect*, 12 PSYCHOL. PUB. POL’Y & L. 170, 173 (2006) [hereinafter Flynn, *Tethering the Elephant*]. While the term “Flynn Effect” refers generally to the phenomenon of IQ gains, this Note, like researchers, experts, and courts, uses the language “adjusting for” the Flynn Effect to specifically refer to the reduction of 0.3 points per year for American IQ scores.

8. See THE RISING CURVE: LONG-TERM GAINS IN IQ AND RELATED MEASURES (Ulric Neisser ed., 1998) [hereinafter THE RISING CURVE] (presenting chapters written by psychologists all addressing the Flynn Effect and its possible causes).

Court decided *Atkins v. Virginia*.⁹ The Court in *Atkins* concluded that the execution of mentally retarded persons constituted excessive punishment in violation of the Eighth Amendment of the U.S. Constitution.¹⁰ Based on “evolving standards of decency” and a national consensus informed primarily by state legislative action,¹¹ the Court found that, while such persons could and should be tried and punished for the crimes they commit, “their disabilities in areas of reasoning, judgment, and control of their impulses” show that “they do not act with the level of moral culpability that characterizes the most serious adult criminal conduct.”¹² Furthermore, the reduced capacity of mentally retarded defendants could jeopardize the fairness of their capital proceedings given the heightened risk of false confessions and wrongful executions and given their inability to provide meaningful assistance to counsel and to present mitigating factors.¹³

In its *Atkins* opinion, the Court referenced definitions of mental retardation set forth by the American Association of Mental Retardation (“AAMR”) and the American Psychiatric Association (“APA”).¹⁴ Despite differences in wording, both definitions generally characterize mental retardation as “significantly subaverage intellectual functioning” coupled with “limitations in adaptive functioning” that manifest before the age of eighteen.¹⁵ Psychologists have “observed that one of the ‘striking aspects’ of the Court’s decision in *Atkins* is that [its] prohibition is framed in the language of a clinical diagnosis” and that “[n]o other class of individuals is constitutionally exempt from the death penalty solely on the basis of a psychological diagnosis.”¹⁶ The Court in *Atkins*, however, acknowledged that “serious disagreement” existed on the issue of “determining which

9. *Atkins v. Virginia*, 536 U.S. 304, 321 (2002); see also LaJuana Davis, *Death-Penalty Appeals and Post-Conviction Proceeding: Intelligence Testing and Atkins: Considerations for Appellate Courts and Appellate Lawyers*, 5 J. APP. PRAC. & PROCESS 297, 309 (2003) (advising appellate practitioners, soon after *Atkins*, to be aware of the Flynn Effect in the *Atkins* context).

10. *Atkins*, 536 U.S. at 321.

11. *Id.* at 312–16. The Court found it significant that, since its decision in *Penry v. Lynaugh*, 492 U.S. 302 (1989), where it upheld the execution of mentally retarded persons, seventeen state legislatures passed legislation prohibiting these executions and states without such prohibitions executed very few to no mentally retarded persons. *Id.* at 310, 314–16.

12. *Id.* at 306.

13. *Id.* at 320–21.

14. *Id.* at 309 n.3.

15. *Id.* (quoting AAMR, MENTAL RETARDATION: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORTS 5 (9th ed. 1992); AM. PSYCHIATRIC ASS’N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 41–43 (4th ed. 2000)).

16. Gilbert S. MacVaugh III & Mark D. Cunningham, *Atkins v. Virginia: Implications and Recommendations for Forensic Practice*, 37 J. PSYCHIATRY & L. 131, 132 (2009) (citation omitted).

offenders are in fact retarded.”¹⁷ It subsequently declined to set forth its own definition of mental retardation, instead, leaving to the individual states “the task of developing appropriate ways to enforce” the *Atkins* decision.¹⁸

Despite the discretion given to states, most statutory definitions of mental retardation or intellectual disability¹⁹ in death penalty states share three key elements, including the requirement that a person have limited intellectual functioning.²⁰ Many of those states further define the intellectual-functioning element by promulgating similar IQ-based definitions.²¹ For example, eleven state statutes establish that a person has the requisite limited intellectual functioning for mental retardation if that person has an IQ of 70 or below.²² Under *Atkins*, a defendant’s IQ score can thus play a significant role in determining whether he faces the death penalty. A lack of consistency exists, however, regarding how courts interpret the IQ-based intellectual-functioning element, including judicial disagreement over whether to adjust IQ scores for the Flynn Effect.²³

Experts in intelligence testing are keenly aware of the Flynn Effect’s serious implications in the capital context.²⁴ Courts openly acknowledge the inconsistent judicial treatment of the Effect by state and federal courts.²⁵ The Flynn Effect, however, reflects a mostly untouched subject in legal scholarship. While many legal scholars write on *Atkins*, few discuss the Flynn Effect in depth.²⁶ This Note fills

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

18. *Id.* (quoting *Ford v. Wainwright*, 477 U.S. 399, 416 (1986)).

19. For an explanation of the interchangeability of these terms, see *supra* note 1.

20. *See infra* Part VI.

21. *Id.*

22. *Id.*

23. *See infra* Part III. While this Note addresses only the inconsistent approaches to the Flynn Effect within the current *Atkins* framework, it acknowledges that *Atkins* raises many other significant debates and controversies. One notable issue is whether IQ scores should be used at all to determine mental retardation. *See, e.g.*, Robert J. Sternberg, *The Flynn Effect: So What?*, 28 J. PSYCHOEDUC. ASSESSMENT 434, 435–37 (2010) (concluding that the use of IQ scores for mental retardation determinations is limited and ignores ethical considerations because those scores only measure cognitive intelligence and not the more significant “ethical intelligence”).

24. *Supra* note 2.

25. *See, e.g.*, *Ybarra v. State*, 247 P.3d 269, 281–82 (Nev. 2011) (“Whether IQ scores should be adjusted to account for the Flynn effect is a matter of great dispute in other jurisdictions.”).

26. Around ten law journal articles, student notes, and case notes address the Flynn Effect, but they only do so briefly as part of a larger discussion of *Atkins*. *See* John H. Blume et al., *Of Atkins and Men: Deviations from Clinical Definitions of Mental Retardation in Death Penalty Cases*, 18 CORNELL J.L. & PUB. POL’Y 689, 700–01 (2009) (briefly discussing failures to account for the Flynn Effect); Penny J. White, *Treated Differently in Life But Not in Death: The Execution*

that gap by advocating for the adjustment of IQ scores based on the Flynn Effect.

Part II of this Note first provides a foundation for understanding the Flynn Effect in the *Atkins* context by explaining intelligence testing, the Effect itself, and state definitions of mental retardation or intellectual disability. Part III identifies the inconsistent treatment of the Flynn Effect across states and the federal circuits. Discussing sources both supporting and opposing IQ adjustments, Part IV argues in favor of validating the Effect in the death penalty context. In the interest of the consistency, accuracy, and fairness of capital proceedings, Parts IV and V conclude that courts should adjust IQ scores for the Flynn Effect in all *Atkins* cases that rely on outdated IQ tests.

II. INTELLIGENCE IN THE FACE OF DEATH: UNDERSTANDING THE FLYNN EFFECT'S PLACE IN *ATKINS*

Given its complicated nature, the Flynn Effect falls solely within the domain of experts, who play the integral role of educating judges and juries on the phenomenon. To facilitate a full understanding of how the Flynn Effect interacts with *Atkins*, the following sections explain intelligence testing, the Flynn Effect, and the use of IQ scores to determine mental retardation.

A. *It's All Relative: The Measurement of IQ*

Understanding the Flynn Effect requires first understanding the set of measurements—IQ tests—that underlie the theory. This Section explains the measurement of IQ on a basic level necessary to appreciate the Flynn Effect, emphasizing three points: (1) IQ scores are relative measurements based on the performance of groups known as standardization samples; (2) the standardization process, which measures the performance of standardization samples, occurs at a specific point in time; and (3) IQ tests are expensive to standardize and to purchase, leading to the use of tests many years after their standardization.

In their current form, IQ scores are standardized.²⁷ Under standardized scoring, a person's IQ score depends on where that

of the Intellectually Disabled After Atkins v. Virginia, 76 TENN. L. REV. 685, 695–96 & nn.71–72 (2009) (mentioning the Flynn Effect as part of a broader discussion of *Atkins*' inconsistencies).

27. KAUFMAN, *supra* note 5, at 107. In their earliest form, IQ scores were quotients that were calculated by dividing the test taker's mental age, or age level at which the test taker

person's test performance falls on a bell-shaped, normal curve or distribution of scores.²⁸ Test developers create these normal curves of IQ scores from the test performances of large groups, known as standardization samples, at specific points in time.²⁹ Different curves and standardization samples exist for different ages or age groups.³⁰ Therefore, the performance of others of the same age or age group on an IQ test at a specific point in time defines an individual's score on the same IQ test.³¹

To avoid bias and to promote accuracy, the standardization samples for an IQ test must be representative of all children of a certain age or all adults in a certain age range.³² After selecting and testing a representative standardization sample, test developers create a normal curve based on the scores of the representative sample with the average of the scores normed at 100, meaning that a score of 100 represents average performance on the IQ test.³³ In addition to a standard mean of 100, IQ tests generally have a standard deviation of fifteen or sixteen points.³⁴ Figure 1 displays a normal distribution of IQ scores, illustrating how scores are distributed based on standard

performs, by the person's actual age. *Id.* at 104–05. These quotients, however, inaccurately portrayed mental ability, especially for adults, because mental age does not increase at the same rate as the yearly increase of a person's actual age. *Id.* at 105–06. Clinical psychologist David Wechsler, recognizing this inaccuracy, began measuring IQ with standardized scores. *Id.* at 107. The Wechsler IQ tests, consisting of different versions for children and adults, represent the most widely used IQ tests by school and clinical psychologists in the education and criminal law contexts. See I. Bruce Frumkin, *Mental Retardation: A Primer to Cope with Expert Testimony*, NATIONAL LEGAL AID & DEFENDER ASSOCIATION CORNERSTONE 2 (Fall 2003), available at <http://www.nlada.org/DMS/Documents/1066919805.15/Mental%20Retardation.pdf> (stating that in criminal mental retardation determinations, the Wechsler tests are “[t]he most accepted”).

28. KAUFMAN, *supra* note 5, at 107.

29. *Id.* at 125.

30. *Id.*

31. *Id.* at 130.

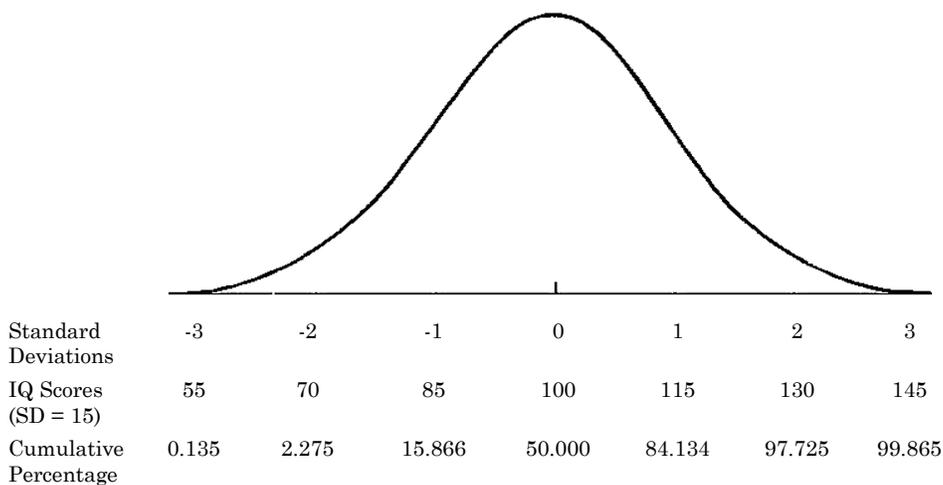
32. *Id.* at 125. To standardize or norm an IQ test, test developers must first identify and track key variables from U.S. Census data, such as gender, ethnicity, socioeconomic status, and geographic region. *Id.* at 126–27. With this data, test developers must then select a standardization sample that matches the census proportions and administer the IQ test to the selected sample. *Id.* at 127.

33. *Id.* at 108, 122, 127.

34. *Id.* at 119–23. While Wechsler IQ tests use fifteen points as the standard deviation, Stanford-Binet IQ tests have used sixteen points until recently. *Id.* at 107, 119–23. The Stanford-Binet Intelligence Scales, Fifth Edition now uses fifteen as its standard deviation. Gale H. Roid, *Stanford-Binet Intelligence Scales (SB5), Fifth Edition*, RIVERSIDE PUBLISHING COMPANY, <http://www.riversidepublishing.com/products/sb5/details.html> (last visited Mar. 27, 2012). Standard deviation is defined as a statistic that “indicates how closely individual measurements cluster around the mean.” *Evaluation Manual: Glossary Evaluation Resources*, CENTERS FOR DISEASE CONTROL AND PREVENTION, <http://www.cdc.gov/getsmart/program-planner/Glossary-Eval-Res.html> (last visited Nov. 12, 2010).

deviations. On a Wechsler IQ test, with a mean of 100 and a fifteen-point standard deviation, one standard deviation from the mean score, encompassing about two-thirds of test takers, results in a range of scores between 85 and 115 (fifteen points below and above the mean score of 100).³⁵ Two standard deviations from the mean score, encompassing about ninety-five percent of test takers, result in a range of scores between 70 and 130 (thirty points below and above the mean score).³⁶ A score of 130 on Wechsler tests often operates as a lower threshold for giftedness, while a score of 70 often marks the upper threshold for mental retardation.³⁷

FIGURE 1: NORMAL DISTRIBUTION OF IQ SCORES³⁸



35. KAUFMAN, *supra* note 5, at 108.

36. *Id.*

37. *Id.* The same thresholds on older Stanford-Binet tests, which have standard deviations of sixteen points, are scores of 132 and 68, respectively. *Id.* at 119–23. The question of “what do IQ tests measure?” presents an issue worth noting, but not one of significance because this Note assumes, under the current *Atkins* framework, that IQ tests measure “intellectual functioning.” This Note thus focuses only on *how* to interpret IQ scores from IQ tests. While linked inextricably with the notion of “intelligence,” IQ tests actually measure, through several subtests, a variety of different items, such as verbal comprehension, working memory, and perceptual reasoning. See, e.g., Ulric Neisser, *Rising Scores on Intelligence Tests*, 85 AM. SCIENTIST 440 (1997), available at <http://www.americanscientist.org/issues/feature/rising-scores-on-intelligence-tests/1> (describing the various subtests of a Wechsler IQ test). A person may receive a variety of IQ scores on multiple tests; however, “different IQ tests usually result in the same approximate level of mental ability.” KAUFMAN, *supra* note 5, at 146–47, 154.

38. See Rodrigo de la Jara, *IQ Basics*, IQ COMPARISON SITE, <http://www.iqcomparisonsite.com/IQBasics.aspx> (last visited Nov. 14, 2010) (displaying a normal distribution of IQ scores).

The method of standardizing an IQ test is both a time-consuming and expensive process.³⁹ Therefore, test publishers do not restandardize tests every year.⁴⁰ Additionally, given the expense of replacing outdated IQ tests and purchasing up-to-date ones, test administrators will often use IQ tests standardized many years before the subjects actually take the tests.⁴¹

From the brief and basic discussion of intelligence testing presented above, three points stand out: (1) an IQ score is relative to when and how an IQ test was standardized and to the age of the test taker; (2) the standardization of a particular IQ test, based on standardization samples, occurs at a specific point in time; and (3) given the expense and effort entailed, test publishers cannot restandardize a test every year nor can test administrators use the most up-to-date tests, resulting in the administration of IQ tests many years after publishers normed the tests.

B. Throwing the Curve: Explanation and Implications of the Flynn Effect

In 1984, James Flynn published a study documenting American IQ gains across age groups and over a period of more than forty years.⁴² Flynn used IQ-score data that he collected from seventy-three studies, with a combined total of almost 7,500 subjects between

39. KAUFMAN, *supra* note 5, at 125–26.

40. See Mark D. Cunningham & Marc J. Tassé, *Looking to Science Rather Than Convention in Adjusting IQ Scores When Death Is at Issue*, 41 PROF. PSYCHOL.: RES. & PRAC. 413, 416 (2010) (“IQ tests are re-normed at intervals dictated by practical economics rather than optimal accuracy.”).

41. See Tomoe Kanaya et al., *The Flynn Effect and U.S. Policies: The Impact of Rising IQ Scores on American Society Via Mental Retardation Diagnoses*, 58 AM. PSYCHOLOGIST 778, 787 (2003) (explaining that “IQ tests are expensive (approximately \$1,000 per set including a supply of test recorded forms)” and that “[b]efore an old test is completely phased out, different children may be tested on different norms in the same year”).

42. Flynn, *The Mean IQ of Americans*, *supra* note 3. While Flynn was not the first to observe IQ gains, he played an integral role in systematically recording and bringing attention to the trend. Ulric Neisser, *Introduction: Rising Test Scores and What They Mean*, in THE RISING CURVE, *supra* note 8, at 5, 12 (discussing how Wechsler studied test scores of people and discovered a three-point gain per decade, but describing Flynn as the first to systematically document the size and significance of the IQ gains); see also RICHARD J. HERRNSTEIN & CHARLES MURRAY, THE BELL CURVE 307 (1994) (noting that the phenomenon of IQ gains was observed in the 1930s) (citing Maud A. Merrill, *The Significance of IQ's on the Revised Stanford-Binet Scales*, 29 J. EDUC. PSYCHOL. 641 (1938)); Richard Lynn & Susan Hampson, *The Rise of National Intelligence: Evidence from Britain, Japan, and the U.S.A.*, 7 PERSONALITY & INDIVIDUAL DIFFERENCES 23 (1986) (documenting IQ gains per decade).

the ages of two and forty-eight, who had taken various Stanford-Binet and Wechsler IQ tests.⁴³ From this data, he derived a “very rough estimate of the rate of IQ gains prevalent in America since 1932, an estimate that suggested an overall rate of at least 0.30 points per year.”⁴⁴ Flynn discovered that the American subjects, on average, scored higher on earlier-normed IQ tests than on later-normed tests.⁴⁵ Flynn then calculated IQ gains by (1) measuring the difference between the subjects’ mean scores on the two tests—the earlier and later tests, normed at different times but taken around the same time—and (2) dividing that difference by the number of years that had passed between the norming of the earlier test and the norming of the later test, resulting in a figure around 0.3 points per year.⁴⁶

From his findings, Flynn concluded that the IQ gains reflected the obsolete, outdated norms of the earlier IQ tests, as compared to later, more recently normed tests.⁴⁷ Flynn defined obsolete norms as “simply ones that are earlier and easier than later norms.”⁴⁸ Given the relative nature of IQ scores and the observed IQ gains over time, when a person takes an IQ test in 2010 and that test’s norms are based on a standardization sample from 1980, the person receives an inflated score because the score is based on the weaker performance of the 1980 sample, rather than the performance of the person’s peers in 2010.⁴⁹ In other words, a person’s IQ score on an earlier test with obsolete norms may be above average, while the same performance may be average or below average on a later test with updated norms. Psychologist and IQ-test developer Alan S. Kaufman explains this phenomenon with an analogy: while a runner’s time may have won a

43. Flynn, *The Mean IQ of Americans*, *supra* note 3, at 32.

44. *Id.* at 34.

45. *Id.* at 32.

46. *Id.* at 33, tbl.2. For example, based on Flynn’s original empirical findings, Flynn calculated an IQ gain of 0.354 per year by dividing 5.49 points—the point difference between the two test means of 107.13 and 101.64—by 15.5, or the number of years that passed between the standardizations of the earlier test in 1932 and the later test in 1947. *Id.* As for his methodology, Flynn developed a uniform convention of scoring by reconciling different test means and standard deviations and also excluded studies from his data that exhibited weak methodology. *Id.* at 29–31. While Flynn acknowledged that he may not have incorporated every study of American IQ trends, he was confident that his data included most of these studies. *Id.* at 31.

47. *Id.* at 39.

48. *Id.*

49. Because an IQ test is normed at a specific point in time, *supra* note 29, the 2010 test taker cannot be compared to other 2010 test takers on the same test because no 2010 standardization sample exists for that test, normed in 1980.

track meet twenty years ago, the same time may not even qualify for a meet today.⁵⁰

Flynn's research on IQ gains, both in the United States and in other countries,⁵¹ has inspired other empirical studies, all looking to discover more about the Flynn Effect and its impact on how we measure intelligence.⁵² From his initial study to his present-day scholarship, Flynn continues to support his theory that full-scale IQ scores of the American population have increased at an average rate of 0.3 points per year or three points per decade.⁵³ Others in the field of intelligence testing also recognize this rate of IQ gain.⁵⁴

To adjust an individual's IQ score for the inflation caused by obsolete norms, test administrators or interpreters should deduct 0.3 points for every year that lapses between the year the test was normed and the year the individual took the test.⁵⁵ Adjusting for the Flynn Effect, therefore, requires adjusting an IQ score downward, with the magnitude of the adjustment contingent on when the particular test was normed and taken.⁵⁶

50. KAUFMAN, *supra* note 5, at 203–04.

51. Following his initial study, Flynn extended his research on IQ gains beyond the American border to thirteen other nations, where he observed varying degrees of IQ-score increases. Flynn, *Massive IQ Gains in 14 Nations*, *supra* note 3, at 172–84.

52. See Thomas W. Teasdale & David R. Owen, *A Long-Term Rise and Recent Decline in Intelligence Test Performance: The Flynn Effect in Reverse*, 39 PERSONALITY & INDIVIDUAL DIFFERENCES 837 (2005) (studying IQ decline among Danish men); Xiaobin Zhou et al., *Peeking Inside the “Black Box” of the Flynn Effect: Evidence From Three Wechsler Instruments*, 28 J. PSYCHOEDUC. ASSESSMENT 399, 409 (2010) (administering Wechsler tests to samples and observing variation in the magnitude of the Flynn Effect across ability levels).

53. See, e.g., FLYNN, WHAT IS INTELLIGENCE?, *supra* note 7, at 112 (explaining that for every year that passes from when an IQ test was standardized, “obsolescence has inflated their IQs by 0.30 points”); James R. Flynn, *The WAIS-III and WAIS-IV: Daubert Motions Favor the Certainly False over the Approximately True*, 16 APPLIED NEUROPSYCHOL. 98, 100 (2009) [hereinafter Flynn, *The WAIS-III and WAIS-IV*] (advocating for adjusting Wechsler IQ-test scores by 0.3 points for every year of obsolescence).

54. See Neisser, *supra* note 42, at 12–13 (explaining that the Flynn Effect of three points per decade has been a “steady and systematic” gain). Despite the uncontroverted basic definition of the Flynn Effect, active debate surrounds the issue of what causes the IQ gains of the Flynn Effect. The IQ gains, however, exist and pose significant implications regardless of their underlying cause or causes. See *infra* Part IV.A.

55. FLYNN, WHAT IS INTELLIGENCE?, *supra* note 7, at 112–14; Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 100.

56. For example, if a person scores a 75 on an IQ test, normed in 1980 and taken in 2000 (a span of twenty years), the person's score adjusts to 69 under the Flynn Effect: $75 - (20 \text{ years} \times 0.3/\text{year}) = 69$.

*C. Atkins at Work: Determinations of Mental Retardation or
Intellectual Disability*

Thirty-four states, the federal government, and the U.S. military currently authorize capital punishment⁵⁷ and are subject to the Supreme Court's prohibition in *Atkins*.⁵⁸ In 2002, when the Court decided *Atkins*, seventeen states already had statutes banning executions of the mentally retarded.⁵⁹ Other states subsequently passed legislation outlining definitions and procedures for determining mental retardation or intellectual disability.⁶⁰ Some death penalty states, however, still lack legislation on the issue,⁶¹ prompting courts in those jurisdictions to provide judicial guidance in lieu of statutory mandates.⁶²

On the surface, the state definitions, documented in Part VI, closely mirror the definitions set forth by the American Association on Intellectual and Developmental Disabilities ("AAIDD"), formerly the AAMR,⁶³ and the APA—the Court cited both organizations and their definitions in *Atkins*.⁶⁴ Generally, state definitions rely on three requirements for a finding of mental retardation or intellectual disability: (1) significantly subaverage general intellectual functioning; and (2) significant impairment or deficit in adaptive behavior; (3) both manifesting during the person's developmental period.⁶⁵

57. DEATH PENALTY INFO. CTR., FACTS ABOUT THE DEATH PENALTY (2012), available at <http://www.deathpenaltyinfo.org/documents/FactSheet.pdf>.

58. *Atkins v. Virginia*, 536 U.S. 304, 321 (2002).

59. *Id.* at 314–15.

60. *See infra* Part VI.

61. Part VI shows that six states currently have no statute on the books defining mental retardation in the capital context. Julie C. Duvall and Richard J. Morris explain that "[n]ot all death penalty states have enacted statutes, however, primarily because of disagreements over two problems: (a) the definition of mental retardation and (b) the legal and psychological assessment procedures to be specified to determine whether a defendant has mental retardation." Julie C. Duvall & Richard J. Morris, *Assessing Mental Retardation in Death Penalty Cases: Critical Issues for Psychology and Psychological Practice*, 37 PROF. PSYCHOL.: RES. & PRAC. 658, 658 (2006).

62. *See, e.g., Ex parte Briseno*, 135 S.W.3d 1, 7 (Tex. Crim. App. 2004) (establishing the standards and definitions for *Atkins* determinations in Texas).

63. Known as the AAMR at the time the Court decided *Atkins*, the AAIDD represents one prominent organization that now uses the term "intellectual disability," rather than the previously used "mental retardation." *See supra* note 1.

64. *Atkins*, 536 U.S. at 309 n.3; *see also* Blume et al., *supra* note 26, at 690–91 (stating that by citing the AAMR and APA definitions, the Court suggested that state measures of mental retardation would be "appropriate" if they "generally conformed" to the two clinical definitions).

65. *See infra* Part VI.

A majority of the state definitions further define the element of “significantly subaverage general intellectual functioning” as either an IQ score of 70 or below or an IQ score that falls two standard deviations or more below the mean.⁶⁶ The state definitions that do not comport with these definitions of the intellectual-functioning element either offer no definition of the element at all or offer a different IQ threshold.⁶⁷ Therefore, in the majority of death penalty states, a defendant’s performance on an IQ test or tests determines whether the defendant satisfies the “significantly subaverage general intellectual functioning” portion of the mental retardation definition, thereby playing an important and, in some cases, dispositive role in determinations under *Atkins*.⁶⁸

The second element, requiring an individual to have an impairment or deficit in adaptive behavior, varies from state to state and is less structurally defined than the intellectual-functioning element. This element often requires consideration of enumerated, nondispositive factors.⁶⁹ While serious concern exists regarding the adaptive-behavior element,⁷⁰ this Note only addresses the intellectual-functioning element and its interaction with the Flynn Effect. As for the third requirement, that mental retardation manifests during a

66. *Id.* As explained previously, two standard deviations below a mean of 100 on Wechsler IQ tests is a score of 70, while two standard deviations below the mean on Stanford-Binet tests, prior to the most recent Fifth Edition, is a score of 68. *Supra* notes 36–37 and accompanying text.

67. *See, e.g.*, ARK. CODE ANN. § 5-4-618 (2010) (giving a defendant “a rebuttable presumption of mental retardation when a defendant has an intelligence quotient of sixty-five (65) or below”); COLO. REV. STAT. § 18-1.3-1101 (2010) (providing no definition for the “significantly subaverage general intellectual functioning” element).

68. The intellectual-functioning criterion becomes dispositive when courts treat the criterion as “a gateway to a consideration” of the other two elements of adaptive functioning and manifestation during the developmental period. Blume et al., *supra* note 26, at 704 (citing *Cherry v. State*, 959 So. 2d 702 (Fla. 2007), as a case in which the court denied the defendant, whose IQ score was on the border of 70, the opportunity to establish limitations in adaptive functioning).

69. *Compare* ARIZ. REV. STAT. ANN. § 13-753 (LexisNexis 2010) (“‘Adaptive behavior’ means the effectiveness or degree to which the defendant meets the standards of personal independence and social responsibility expected of the defendant’s age and cultural group.”), *with* 2001 N.C. Sess. Laws 346 (requiring “[s]ignificant limitations in two or more of the following adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure skills and work skills”).

70. *See* 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, 846, 848 (2007) (“[S]cientific understanding of adaptive behavior and its measurement is still in its infancy 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.”).

person's developmental period, the majority of states define the "developmental period" as occurring before the age of eighteen.⁷¹

In addition to defining mental retardation, state statutes also establish the procedures through which capital defendants and inmates can bring *Atkins* claims. The procedures vary from state to state.⁷² In some states, the court or judge is the sole decider of whether the defendant is mentally retarded,⁷³ while in others, both the court and the jury play a role, with the court making pretrial determinations and the jury deciding the issue during sentencing.⁷⁴ However, in all states, the defendant bears the burden of proving his mental retardation.⁷⁵

Most state statutes place explicit reliance on expert evaluations as part of the mental retardation determination.⁷⁶ The importance of this reliance becomes clear when reviewing judicial opinions discussing *Atkins* claims, including those opinions that address the Flynn Effect.⁷⁷ No state statute, however, explicitly addresses the

71. See *infra* Part VI. Three states, however, define the period as occurring before the age of twenty-two. IND. CODE. § 35-36-9-2 (LexisNexis 2011); MD. CODE ANN., CRIM. LAW § 2-202(b)(1) (LexisNexis 2011); UTAH CODE ANN. § 77-15a-102 (LexisNexis 2012).

72. According to death penalty scholars John H. Blume, Sheri Lynn Johnson, and Christopher Seeds, procedural differences "could have as much effect on the outcomes of *Atkins* claims, and as much to do with a given claim's success or failure, as the substantive deviations from the clinical definitions." John H. Blume et al., *Implementing (or Nullifying) Atkins?: The Impact of State Procedural Choices on Outcome in Capital Cases Where Intellectual Disability Is at Issue* 5 (Cornell Law Sch. Legal Studies Research Paper Series, Working Paper No. 2010-011, 2010), available at <http://ssrn.com/abstract=1670108>.

73. See, e.g., TENN. CODE ANN. § 39-13-203(c) (2010) ("The determination of whether the defendant had intellectual disability at the time of the offense of first degree murder shall be made by the court.").

74. See, e.g., OKLA. STAT. tit. 21 § 701.10bE-F (2010) (allowing capital defendants to bring their mental retardation claims before the court during a pretrial hearing and if convicted, before the jury during the sentencing phase).

75. Most states require that the defendant show by clear and convincing evidence, see, e.g., ARIZ. REV. STAT. ANN. § 13-753G, or by a preponderance of the evidence, see, e.g., OKLA. STAT. tit. 21 § 701.10bF, that he is mentally retarded. Georgia, an outlier, requires that the defendant prove mental retardation beyond a reasonable doubt. GA. CODE ANN. § 17-7-131(c)(3) (2010). The Eleventh Circuit recently upheld the Georgia Supreme Court's decision that this reasonable doubt standard does not violate the Eighth Amendment, finding the decision not contrary to clearly established federal law under *Atkins*. *Hill v. Humphrey*, 662 F.3d 1335, 1360-61 (11th Cir. 2011).

76. See, e.g., FLA. STAT. § 921.137(4) ("Upon receipt of the motion, the court shall appoint two experts in the field of mental retardation who shall evaluate the defendant and report their findings to the court and all interested parties prior to the final sentencing hearing.").

77. See MacVaugh & Cunningham, *supra* note 16, at 132 ("Mental health professionals, by necessity, have become primary sources of information and expertise regarding [*Atkins*] assessment and diagnostic determinations.").

Flynn Effect and its place in *Atkins* determinations.⁷⁸ The fate of the Flynn Effect in a particular defendant's case therefore rests in the hands of the court and, in fewer instances, the jury.⁷⁹ The following Part discusses how courts have interpreted the Flynn Effect in varied and inconsistent ways.

III. DEATH ROW DILEMMA: INCONSISTENT TREATMENT OF THE FLYNN EFFECT IN *ATKINS* CASES

In the decade following *Atkins*, capital defendants and inmates have presented mental retardation claims in both federal and state courts at the pretrial, sentencing, and postconviction stages. However, the total number of *Atkins* cases and, specifically, the number involving the Flynn Effect prove difficult to calculate.⁸⁰ Despite this difficulty in measuring the Flynn Effect's exact impact, a review of recent *Atkins* cases shows the Effect's growing presence in *Atkins* litigation: capital defendants and inmates throughout the country incorporate it into their *Atkins* claims, and courts not only discuss the merits of the Flynn Effect but also discuss its inconsistent treatment across jurisdictions.

Blurring the deciding line between life and death, several divergent paths emerge from the body of federal and state judicial opinions addressing the Flynn Effect in the *Atkins* context. This inconsistency among state and federal courts exists on two interrelated issues: whether the Flynn Effect is valid and whether courts should adjust IQ scores for the Flynn Effect in determining the intellectual-functioning element of mental retardation. In recent

78. See Duvall & Morris, *supra* note 61, at 662 (“An assessment issue that no state statute currently addresses is the *Flynn effect*.”).

79. While thirteen death penalty states provide for jury determinations of intellectual disability, such determinations are “relatively infrequent” because judges determine most *Atkins* claims. Blume et al., *supra* note 72, at 19, 30. In Part III, this Note discusses inconsistency only as to judicial determinations of the Flynn Effect given the infrequency of written documentation of jury determinations.

80. According to two empirical studies published in 2009 and 2010, capital defendants had brought approximately 240 cases involving *Atkins* claims since 2002. John H. Blume et al., *An Empirical Look at Atkins v. Virginia and its Application in Capital Cases*, 76 TENN. L. REV. 625, 628 (2009) [hereinafter Blume et al., *An Empirical Look at Atkins*]; Blume et al., *supra* note 72, at 30–31. The authors of those studies acknowledged that, because many cases remain unpublished and many new cases arise each day, such a figure may underestimate the total number of *Atkins* claims raised. Blume et al., *An Empirical Look at Atkins*, *supra*, at 628.

In March 2012, a search for all federal and state opinions including the term “Flynn Effect” resulted in 113 opinions in LexisNexis and 108 in Westlaw. Several of the retrieved opinions arose out of the same cases. I used “Flynn Effect” as the search term because both the psychological and legal fields use the term to describe the phenomenon of IQ gains over time.

decisions, the Eleventh Circuit and several federal district courts have found or affirmed findings that the Flynn Effect is valid and that courts should adjust for the Effect in *Atkins* cases.⁸¹ The Fourth Circuit, a federal district court, and state courts in California and Ohio also recognize the Flynn Effect as valid, but they only require the consideration of, and not necessarily the adjustment for, the Effect.⁸² Having formerly rejected the Flynn Effect, state courts in Tennessee recently changed course and now allow, and perhaps require, the consideration of evidence on the Flynn Effect in *Atkins* cases.⁸³ The Fifth Circuit and state courts in Texas—the federal and state jurisdictions with the highest rate of executions as well as the most *Atkins* cases—remain undecided on whether to validate and account for the Flynn Effect.⁸⁴ Based on strict statutory interpretation, state courts in Kentucky and several federal district courts have rejected the Effect outright, refusing to even consider it.⁸⁵

As discussed above, this Note places cases addressing the Flynn Effect into four categories based on their treatment of the Effect: (1) validation of and subsequent adjustment for the Flynn Effect; (2) validation and required consideration of, but not necessarily adjustment for, the Flynn Effect; (3) indecision on the validity of the Flynn Effect and on the need to adjust for it; and (4) rejection of the Flynn Effect in the *Atkins* context.⁸⁶ Regardless of how one defines these different categories, they exist and pose alarming implications: a defendant may qualify as mentally retarded in one state but not in another state. The constitutionality of executing that defendant thus varies depending on jurisdiction.⁸⁷

81. See *infra* Part III.A.

82. See *infra* Part III.B.

83. *Id.*

84. See *infra* Part III.C.

85. See *infra* Part III.D.

86. This Note recognizes that the different categories are neither completely distinct nor internally consistent.

87. Of the “winning” *Atkins* cases, where the defendants successfully avoided the death penalty, approximately fifteen percent of the defendants’ average IQ scores exceeded 70, showing that those courts took into account adjustments, like the Flynn Effect. Blume et al., *An Empirical Look at Atkins*, *supra* note 80, at 632. On the other hand, of the “losing” *Atkins* cases, eighteen percent of the defendants had IQ scores both below and above 70, and the courts in those cases found the scores above 70 more reliable than the scores below 70 in their determinations against the defendants. *Id.* at 633. These two results show that jurisdictions differ in their interpretations of scores above 70 and that their allowance for adjustments like the Flynn Effect result in some defendants going to death row and others not.

A. *Validation of and Adjustment for the Flynn Effect*

Faced with outdated IQ scores, several U.S. district courts have recognized the Flynn Effect in their mental retardation determinations and have adjusted IQ scores accordingly.

Most recently, two Louisiana federal district courts advocated and adjusted for the Flynn Effect. The U.S. District Court for the Middle District of Louisiana, in *Brumfield v. Cain*, found the Flynn Effect to be “widely accepted as a fact in the scientific community.”⁸⁸ Recognizing the “legal uncertainty” surrounding the Effect, the court nevertheless gave “great weight to the AAIDD’s clinical standards.”⁸⁹ It subsequently adjusted the habeas petitioner’s IQ scores for the Flynn Effect, “plac[ing] those scores solidly and indisputably into the mildly mentally retarded level.”⁹⁰ Based on these adjusted scores, the court concluded that the petitioner met the intellectual-functioning prong of his *Atkins* claim and ultimately granted his federal habeas petition, rendering the petitioner ineligible for the death penalty.⁹¹

In *United States v. Smith*, the U.S. District Court for the Eastern District of Louisiana likewise adjusted the capital defendant’s IQ scores for the Flynn Effect, citing and following its analysis and ruling in a previous case, *United States v. Hardy*.⁹² In *Hardy*, the District Court concluded that the Flynn Effect is “well established scientifically” and that adjusting for it is a “best practice.”⁹³ The *Hardy* decision cited several sources extensively, including the guidelines set forth by the AAIDD and Wechsler test publishers.⁹⁴ The court also referenced the testimony of defense experts Mark D. Cunningham and Victoria Swanson⁹⁵ and the opinions of respected psychologists Stephen Greenspan and J. Gregory Olley—both of whom are members of the APA’s Ad Hoc Committee on Mental Retardation

88. *Brumfield v. Cain*, No. 04-787-JJB-CN, 2012 WL 602163, at *18 (M.D. La. Feb. 23, 2012).

89. *Id.*

90. *Id.*

91. *Id.* at *19, *33.

92. *United States v. Smith*, 790 F. Supp. 2d 482, 491 & n.43 (E.D. La. 2011). In *Smith*, however, the court acknowledged that the defendant’s other unadjusted IQ scores met the intellectual-functioning element without accounting for the Flynn Effect. *Id.* at 491, 501. The court nonetheless recognized the Flynn Effect as both valid and required as an adjustment. *Id.* at 491 & n.43.

93. *United States v. Hardy*, 762 F. Supp. 2d 849, 866–68 (E.D. La. 2010).

94. *Id.* at 859–61.

95. *Id.* at 855.

and the Death Penalty.⁹⁶ The court found unpersuasive the testimony of the government expert, who conceded that the Flynn Effect exists but argued against adjustment for it based on what the court considered irrelevant sources.⁹⁷ While *Hardy* acknowledged that the Fifth Circuit has not yet ruled on the validity of the Flynn Effect,⁹⁸ as discussed below, the court nonetheless found the Flynn Effect valid and adjusted for it, citing persuasive case law from the Fourth and Eleventh Circuits and the fact that no Fifth Circuit precedent explicitly precludes this approach.⁹⁹

Similarly, in *United States v. Lewis*, the U.S. District Court for the Northern District of Ohio, after an extensive discussion of the Flynn Effect's validity, recognized the adjustment for the Flynn Effect as "a best practice for an intellectual disability determination."¹⁰⁰ It then adjusted the defendant's IQ scores to account for the Flynn Effect during a pretrial determination of the defendant's *Atkins* claim under federal death penalty law.¹⁰¹ The *Lewis* court ultimately concluded that the defendant had shown by a preponderance of the evidence that he was mentally retarded and thus ineligible for a death sentence.¹⁰² In reaching its conclusion on the Flynn Effect, the court found that "[i]n circumstances such as these, where a potentially intellectually disabled individual faces the death penalty . . . the Government has not produced evidence to counter the authoritative nature of the AAIDD, persuasive federal case law . . . and the testimony of eminent intellectual disability scholars," such as Stephen Greenspan.¹⁰³

In their respective opinions, the federal district courts in both *Lewis* and *Hardy* referenced other cases that also adjusted for the

96. *Id.* at 858–59.

97. *Id.* at 865–67.

98. *Id.* at 860–61.

99. *Id.* at 862 & n.32.

100. *United States v. Lewis*, No. 1:08 CR 404, 2010 U.S. Dist. LEXIS 138375, at *29–32 (N.D. Ohio Dec. 23, 2010). *But cf.* *Lynch v. Hudson*, No. 2:07-cv-948, 2011 WL 4537890, at *33 (S.D. Ohio Sept. 28, 2011) (finding that the petitioner failed to establish the unreasonableness of the state trial court's decision "that there is no evidence that the Flynn [E]ffect affected Petitioner's test results," because the petitioner only showed that, under the Flynn Effect, a person's IQ score "tends" to increase, not that it actually increases).

101. *Lewis*, 2010 U.S. Dist. LEXIS 138375, at *34–35.

102. *Id.* at *99–100.

103. *Id.* at *33. The court relied heavily on the AAIDD standard, which mandates adjustment for the Flynn Effect, *id.* at *29 (citing AAIDD, *INTELLECTUAL DISABILITY: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORTS* 37 (11th ed. 2009)); federal case law from the Fourth, Fifth, and Eleventh Circuits, *id.* at *29–32; and expert testimony from the defense's witnesses, *id.* at *5, *26–27, *33.

Flynn Effect.¹⁰⁴ In *United States v. Davis*, the U.S. District Court for the District of Maryland adjusted for the Effect, noting its widespread acceptance and the potentially deadly consequences of not accounting for it.¹⁰⁵ In *Thomas v. Allen*, the U.S. District Court for the Northern District of Alabama adjusted the defendant's IQ scores after finding that adjustments based on the Flynn Effect were "well-supported by the accumulation of empirical data over many years."¹⁰⁶ On appeal, after reviewing the evidence cited by the District Court, the Eleventh Circuit affirmed the lower court decision in *Thomas*, finding the lower court's adjustment for the Flynn Effect not clearly erroneous.¹⁰⁷ The Eleventh Circuit, however, expressly stated that its affirmance did not address the issue of whether adjustment for the Flynn Effect is mandatory.¹⁰⁸ The court only stated that evaluators of IQ scores could consider the Effect.¹⁰⁹

B. Consideration of the Flynn Effect but Not Required Adjustment

Compared to the cases above that adjust for the Flynn Effect, courts take a different approach when they require only consideration of the Flynn Effect, as such an approach does not necessarily result in the adjustment of IQ scores for the Effect.¹¹⁰ To support their adjustments, the federal district courts in *Lewis*, *Hardy*, and *Davis* nonetheless cited cases that required only consideration.¹¹¹

104. *United States v. Hardy*, 762 F. Supp. 2d 849, 862 (E.D. La. 2010); *Lewis*, 2010 U.S. Dist. LEXIS 138375, at *29–30.

105. *United States v. Davis*, 611 F. Supp. 2d 472, 486 (D. Md. 2009).

106. *Thomas v. Allen*, 614 F. Supp. 2d 1257, 1280–81 (N.D. Ala. 2009), *aff'd*, 607 F.3d 749 (11th Cir. 2010).

107. *Thomas v. Allen*, 607 F.3d 749, 757 (11th Cir. 2010).

108. *Id.*

109. *Id.* at 753; *see also* *Hill v. Humphrey*, 662 F.3d 1335, 1373 n.15 (11th Cir. 2011) ("[T]his circuit has recognized that the statistical phenomenon known as the Flynn Effect . . . can be applied by a test administrator to an individual's raw IQ test score when arriving at a final IQ score."). The Eleventh Circuit previously acknowledged the Flynn Effect in *Holladay v. Allen*, another case in which it affirmed the district court's finding of the defendant's mental retardation. 555 F.3d 1346, 1364 (11th Cir. 2009). In *Holladay*, however, the Flynn Effect was not used to adjust the defendant's IQ scores as it was in *Thomas*. Instead, the defense expert, whom the district court credited over the government expert, used the Flynn Effect to support her refusal to rely on scores from IQ tests that were outdated by fifteen years. *Id.* at 1351 n.4. Given the Flynn Effect, the expert chose to rely on a score from a more recently normed test. *Id.*

110. *See* *State v. Burke*, No. 04AP-1234, 2005 WL 3557641, at *13 (Ohio App. Dec. 30, 2005) (concluding that a trial court must consider applying the Flynn Effect, but that the trial court maintained the power to decide whether to actually apply the Flynn Effect to IQ scores).

111. *United States v. Lewis*, No. 1:08 CR 404, 2010 U.S. Dist. LEXIS 138375, at *29–30 (N.D. Ohio Dec. 23, 2010); *United States v. Hardy*, 762 F. Supp. 2d 849, 862 (E.D. La. 2010); *United States v. Davis*, 611 F. Supp. 2d 472, 487 (D. Md. 2009).

In the 2009 case of *United States v. Shields*, the U.S. District Court for the Western District of Tennessee, making a pretrial *Atkins* determination under the federal death penalty scheme, concluded that “the Flynn Effect must be considered” and “that it is a valid scientific phenomenon to consider as bolstering the conclusion that the [d]efendant’s IQ truly is 70 or below.”¹¹² In reaching its conclusion about the validity and consideration of the Flynn Effect, the court in *Shields* relied on the testimony of the two defense experts.¹¹³ The court simultaneously found the state expert “especially lacking in credibility,” “at times disingenuous and misleading,” and biased given his “results-driven evaluation designed to deliver the conclusion that [the] [d]efendant is not mentally retarded.”¹¹⁴ The experts from both sides agreed that the Flynn Effect exists but disagreed on its place in the court’s *Atkins* determination. Finding that the defendant met all of the requirements for mental retardation, including limited intellectual functioning, the court ultimately concluded that the defendant was mentally retarded and thus ineligible for the death penalty.¹¹⁵

Like their federal counterpart, state courts in Tennessee recently opened their courtroom doors to the Flynn Effect, impliedly, although not explicitly, requiring consideration of the Effect. In April 2011, while addressing the inmate’s third postconviction petition in *Coleman v. State*, the Tennessee Supreme Court made an important finding: Tenn. Code Ann. § 39-13-203(a)(1), the state statutory provision defining the intellectual-functioning element of an *Atkins* claim, “does not require that raw scores on I.Q. tests be accepted at their face value,” meaning that a trial court “may receive and consider any relevant and admissible evidence” regarding a defendant’s IQ score, including evidence of the Flynn Effect.¹¹⁶ Prior to *Coleman*,

112. *United States v. Shields*, No. 04-20254, 2009 U.S. Dist. LEXIS 130612, at *2, *45 (W.D. Tenn. May 11, 2009).

113. The two defense witnesses were Marc J. Tassé, whom the court described as “exceptionally well-qualified and credible,” and Mark Cunningham, who was “likewise very credible.” *Id.* at *13–14.

114. *Id.* at *15–16.

115. *Id.* at *2.

116. *Coleman v. State*, 341 S.W.3d 221, 224, 242 n.55 (Tenn. 2011). Although the *Coleman* inmate’s briefing to the Tennessee Supreme Court discussed the Flynn Effect extensively, the court in its decision only directly addressed the Flynn Effect in a footnote:

The AAIDD currently recognizes ten potential “challenges” to the reliability and validity of I.Q. test scores. Among these challenges are the standard error of measurement, the *Flynn Effect*, and the practice effect. The *Flynn Effect* refers to the observed phenomenon that I.Q. test scores tend to increase over time. Thus, the most current versions of a test should be used at all times and, when older versions of the test are used, the scores must be correspondingly adjusted downward. . . . Accordingly, if the

Tennessee courts, like those discussed in Part III.D below, interpreted the statutory cutoff score of 70 for the intellectual-functioning element as a strict, bright-line cutoff that precluded adjustment for the Flynn Effect.¹¹⁷ Reviewing its past cases, the statute's legislative history, and opinions of experts in the field, the *Coleman* court interpreted the statutory provision as allowing a trial court to consider expert testimony on "challenges" affecting the validity and reliability of raw IQ scores, explicitly naming the Flynn Effect as one such challenge.¹¹⁸ The court then vacated the decision below and remanded the case to the postconviction trial court for new proceedings on the inmate's *Atkins* claim.¹¹⁹

Two months after *Coleman*, the Court of Criminal Appeals of Tennessee issued a decision interpreting *Coleman*.¹²⁰ In *Howell v. State*, the court explained that, although a trial court could accept an expert's opinion accounting for the Flynn Effect, *Coleman* did not require adjustment for the Flynn Effect, thereby allowing a trial court to reject the Flynn Effect if faced with evidence of its invalidity.¹²¹ The *Howell* court then concluded that, because the postconviction court below "considered" evidence regarding the Flynn Effect, the postconviction court's rejection of the adjustment did not "run afoul of

trial court determines that professionals who assess a person's I.Q. customarily consider a particular test's standard error of measurement, the *Flynn Effect*, the practice effect, or other factors affecting the accuracy, reliability, or fairness of the instrument or instruments used to assess or measure the defendant's I.Q., an expert should be permitted to base his or her assessment of the defendant's "functional intelligence quotient" on a consideration of those factors.

Id. at 242 n.55 (emphasis added) (citations omitted); see also Supplemental Brief of Petitioner-Appellant Michael Angelo Coleman, *State v. Coleman*, No. W2007-02767-CCA-R3-PD (Tenn. 2010) (explaining the Flynn Effect and its application to Coleman's IQ scores). The footnote, however, shows the court's recognition of the Effect and its potential application.

117. See, e.g., *Black v. State*, No. M2004-01345-CCA-R3-PD, 2005 WL 2662577, at *14 (Tenn. Crim. App. Oct. 19, 2005) (citing *Howell v. State*, 151 S.W.3d 450, 458–59 (Tenn. 2004)) (refusing to apply the Flynn Effect because the threshold IQ score of 70 in the Tennessee Code represented a "bright-line cutoff" that could not be adjusted for factors like the Flynn Effect). Setting the stage for *Coleman*, the court in *Smith v. State* expressed its serious concerns with Tennessee's former bright-line rule, stating that, by refusing to consider other factors, "some mentally retarded defendants are likely to be executed in Tennessee." *Smith v. State*, No. E2007-00719-CCA-R3-PD, 2010 Tenn. Crim. App. LEXIS 793, at *124–25 (Tenn. Crim. App. Sept. 21, 2010), *death sentence vacated*, 357 S.W.3d 322 (Tenn. 2011) (quoting *Cribbs v. State*, No. W2006-01381-CCA-R3-PD, 2009 Tenn. Crim. App. LEXIS 524, at *111 (Tenn. Crim. App. July 1, 2009)).

118. *Coleman*, 341 S.W.3d at 242 n.55.

119. *Id.* at 258.

120. *Howell v. State*, No. W2009-02426-CCA-R3-PD, 2011 Tenn. Crim. App. LEXIS 447, at *36–39 (Tenn. Crim. App. June 14, 2011).

121. *Id.*

Coleman.”¹²² Under *Howell*'s interpretation, *Coleman* appears to require consideration of the Flynn Effect, and, although not required, a court may still adjust for the Effect.

The U.S. Court of Appeals for the Sixth Circuit went one step further in *Black v. Bell*, interpreting *Coleman* to state that “Tennessee courts must also *consider* [Flynn Effect] evidence in assessing a defendant’s ultimate functional I.Q.”¹²³ The court explained that allowing defendants to present evidence on the Effect, without further considering that evidence, is “not enough.”¹²⁴ Although the state courts and federal district court, considering the case pre-*Coleman*, reviewed the inmate’s Flynn Effect evidence, the Sixth Circuit found that they failed to “consider” the evidence because they rejected the Flynn Effect under a bright-line IQ-score cutoff.¹²⁵ The Sixth Circuit subsequently vacated the district court’s rejection of the *Atkins* claim and remanded the case for further review pursuant to both its ruling and to *Coleman*.¹²⁶

Just four days after the *Black* decision, the Tennessee Supreme Court vacated a death sentence in *Smith v. State* because, under *Coleman*’s “applicable legal standard,” the postconviction court below erroneously rejected evidence pertaining to the inmate’s IQ, including that of the Flynn Effect.¹²⁷ In its opinion, however, the court quotes *Coleman* extensively¹²⁸ without providing additional insight into what exactly *Coleman* requires. The *Smith* court stated that its remand for resentencing “provide[s] [the parties] an opportunity to present evidence regarding Smith’s functional I.Q.” pursuant to *Coleman*.¹²⁹ At his prior postconviction hearing, the inmate’s expert testified about the Flynn Effect,¹³⁰ so presumably the inmate had the opportunity already to present such evidence. The lower court, however, did not consider the Effect in determining the inmate’s IQ and denied the inmate’s motion to rehear, in which he reasserted the Flynn Effect.¹³¹ Based on these later cases, *Coleman* appears, although not explicitly, to require consideration of the Flynn Effect in the *Atkins* context.

122. *Id.* at *39–40.

123. *Black v. Bell*, 664 F.3d 81, 96 (6th Cir. 2011).

124. *Id.*

125. *Id.* at 89–90.

126. *Id.* at 100–01.

127. *Smith v. State*, 357 S.W.3d 322, 353–54 (Tenn. 2011).

128. *Id.* at 354.

129. *Id.*

130. *Id.* at 351–52.

131. *Id.* at 353.

The U.S. Court of Appeals for the Fourth Circuit has followed a more explicit approach since 2005, requiring consideration of, but not adjustment for, the Flynn Effect. In *Winston v. Kelly*, the Fourth Circuit vacated the district court's decision, which had denied the petitioner habeas relief without considering evidence on the Flynn Effect.¹³² Remanding the case, the Fourth Circuit instructed the district court to consider the petitioner's evidence on the Flynn Effect in deciding the petitioner's *Atkins* and *Atkins*-related claims.¹³³ Likewise, in 2005, the Fourth Circuit remanded a case, *Walker v. True*, finding that the district court failed to consider the Flynn Effect and expressly directing the district court to consider the Effect on remand.¹³⁴ The Fourth Circuit, however, does not mandate adjustments for the Effect,¹³⁵ so courts within the circuit are free to adjust or not adjust as long as they consider the Effect. California¹³⁶

132. *Winston v. Kelly*, 592 F.3d 535, 539, 557 (4th Cir. 2010). This case arose in the context of the Antiterrorism and Effective Death Penalty Act of 1996 ("AEDPA"), under which a district court cannot consider additional evidence in a federal habeas proceeding unless the petitioner exhausts state remedies. *Id.* at 539. The district court refused to consider the defendant's Flynn Effect evidence, reasoning that it fundamentally altered the claim that the petitioner brought in state court. *Id.* at 557. The Fourth Circuit disagreed, finding that the petitioner exhausted his state court remedies. *Id.* Evidence of the Flynn Effect would not fundamentally alter his claim because he presented the evidence to the Virginia Supreme Court, although that court did not discuss the Effect in its decision. *Id.*

133. *Id.* at 557. On remand, the district court granted the federal habeas petition on the grounds of ineffective assistance of counsel. *Winston v. Kelly*, 784 F. Supp. 2d 623, 626 (W.D. Va. 2011). The court considered the Flynn Effect as part of its evaluation of the defense counsel's failure to review records on the defendant's IQ and mental retardation but did not directly opine on the merits of the Flynn Effect. *Id.* at 630–34. The government in *Winston* can now either conduct a new trial on the issue of mental retardation or seek a noncapital sentence. *Id.* at 626.

134. *Walker v. True*, 399 F.3d 315, 318, 320 (4th Cir. 2005). Following the Fourth Circuit's 2005 decision in *Walker*, the district court, upon remand and without a written decision, denied the defendant's *Atkins* claim. *Walker v. Kelly*, 593 F.3d 319, 322 (4th Cir. 2010). The Fourth Circuit affirmed this ruling against the defendant, without reviewing the district court's analysis of the defendant's IQ scores or the Flynn Effect, because it found that the defendant failed to meet the adaptive-behavior prong. *Id.* at 323. The dissenting opinion, however, stated that the district court erroneously failed to consider the impact of the Flynn Effect, and had it considered the Flynn Effect, the defendant may have satisfied the intellectual-functioning requirement. *Id.* at 336–37 (Gregory, J., dissenting).

135. *See, e.g., Richardson v. Branker*, 668 F.3d 128, 152 (4th Cir. 2012) (finding no requirement under North Carolina law to adjust IQ scores); *Green v. Johnson*, 515 F.3d 290, 300 n.2 (4th Cir. 2008) (concluding that neither *Atkins* nor Virginia law required accounting for the Flynn Effect in mental retardation determinations).

136. *People v. Superior Court*, 155 P.3d 259, 262, 263 n.4, 266 (Cal. 2007). The Supreme Court of California found that the trial court did not err in its interpretation of the defendant's IQ scores, which included the trial court's acceptance of the Flynn Effect. *Id.* at 260–61, 268. Rejecting any absolute rules, the court stated that mental retardation, as a question of fact, was not measured by a fixed IQ score, but instead required a consideration of "all relevant evidence" to determine a person's overall capacity. *Id.* at 266. The court found that the defendant provided

and Ohio¹³⁷ state courts and the U.S. Navy-Marine Corps Court of Criminal Appeals¹³⁸ follow similar approaches, considering the Flynn Effect without necessarily adjusting for it.

This approach, requiring only consideration without mandating adjustment for the Flynn Effect, creates internal inconsistency. The consideration-only requirement can result in two different treatments of the Flynn Effect—with one court considering and adjusting for the Flynn Effect and another court considering the Effect but not adjusting for it. Thus, two courts may reach two different determinations of death penalty eligibility for the same defendant.

C. Indecision Regarding the Flynn Effect

Given the large number of executions in Texas—by far the largest in the nation¹³⁹—developments in death penalty litigation are particularly significant in Texas courts and in the Fifth Circuit. Furthermore, Texas “has as many intellectual disability litigations as most other states combined.”¹⁴⁰ State and federal courts in Texas, however, remain unresolved on the issue of the Flynn Effect’s validity in the capital context and have refused to account for it (albeit often in cases where adjustment for the Flynn Effect would not have affected the outcome of the *Atkins* determination).¹⁴¹ State courts in Alabama,

sufficient expert testimony to support the trial court’s findings, including its acceptance of the Flynn Effect. *Id.* at 267. “All relevant evidence” appeared to include the Flynn Effect. *Id.* at 262, 263 n.4.

137. *State v. Waddy*, No. 09AP-1197, 2011 Ohio App. LEXIS 2664, at *11–12 (Ohio App. June 28, 2011); *State v. Burke*, No. 04AP-1234, 2005 WL 3557641, at *13 (Ohio App. Dec. 30, 2005). The Court of Appeals of Ohio in both *State v. Waddy* and *State v. Burke* concluded that trial courts must consider accounting for the Flynn Effect but are not required to actually adjust for it. In both cases, the trial courts considered evidence on the Flynn Effect but declined to adjust the defendants’ IQ scores—the appellate court found no error with the trial courts’ consideration and subsequent rejection of the Flynn Effect in those cases.

138. *United States v. Parker*, 65 M.J. 626, 629–30 (N-M. Ct. Crim. App. 2007). In *Parker*, the court explicitly stated that it would consider the Flynn Effect, along with other evidence, when deciding whether a defendant is mentally retarded pursuant to the definition set forth by the AAIDD. *Id.* Despite consideration of the Flynn Effect, the defendant’s IQ score only represented one part of the analysis, and the *Parker* court found that the defendant had not satisfied his burden in proving his entire claim by a preponderance of evidence. *Id.* at 629–30.

139. Texas executed thirteen inmates in 2011 with Alabama, the next closest state in executions, executing six inmates. DEATH PENALTY INFO. CTR., FACTS ABOUT THE DEATH PENALTY (2012), available at <http://www.deathpenaltyinfo.org/documents/FactSheet.pdf>. As of March 2012, Texas had executed 480 inmates total since 1976—the next state Virginia had executed 109 since 1976. *Id.*

140. Blume et al., *supra* note 72, at 24.

141. *See, e.g., Maldonado v. Thaler*, 625 F.3d 229, 238 (5th Cir. 2010) (“As the district and state habeas courts recognized, however, neither this court nor the [Texas Court of Criminal

Florida, Mississippi, and Nevada also exhibit the same indecision on the Flynn Effect's validity and role in *Atkins* cases.¹⁴²

In *Ex parte Blue*, the Texas Court of Criminal Appeals, the state's highest criminal court, referred to the Flynn Effect as "an unexamined scientific concept" and declined to address its validity.¹⁴³ Despite its disparaging description of the Flynn Effect, the Texas Court of Criminal Appeals has not explicitly rejected it. In fact, in *Ex parte Cathey*, the court granted the inmate's motion to stay his execution and remanded his application to the trial court for a live hearing on four issues, all regarding the Flynn Effect and whether to account for it in the inmate's case.¹⁴⁴

Similarly, the Fifth Circuit has neither recognized the Flynn Effect as valid nor mandated consideration of or adjustment for it. In three *Atkins* cases, *Maldonado v. Thaler*, *In re Mathis*, and *In re*

Appeals] has recognized the Flynn Effect as scientifically valid."); *In re Salazar*, 443 F.3d 430, 433 n.1 (5th Cir. 2006) (explicitly refusing to opine on the validity of the Flynn Effect, but, for the purposes of the case, noting that even if the Flynn Effect applied, the defendant's score would still exceed the cutoff score of 70); *Ex parte Woods*, 296 S.W.3d 587, 608 (Tex. Crim. App. 2009) (refusing to address the Flynn Effect's validity and explaining that, even with additional evidence, a rational fact finder could find that the defendant's IQ score exceeded 70). *But cf.* *Bell v. State*, 66 So. 3d 90, 103 (Miss. 2011) (Randolph, J., dissenting) (stating that the Fifth Circuit has rejected the scientific validity of the Flynn Effect and describing the Flynn Effect as "predicated upon a dubiously suspect theory").

142. *See, e.g.*, *Daniel v. State*, No. CR-08-0670, 2011 Ala. Crim. App. LEXIS 29, at *63–71 (Ala. Crim. App. Apr. 29, 2011) (discussing the Flynn Effect and how Alabama has not yet addressed its validity); *State v. Herring*, 76 So. 3d 891, 893 n.4 (Fla. 2011) (making "no judgment as to the efficacy of adjusting for the Flynn effect because it is not relevant in this case," as the defendant's adjusted IQ scores would still be above 70); *Thorson v. State*, 76 So. 3d 667, 683 (Miss. 2011) (finding that the trial court did not abuse its discretion in not applying the Flynn Effect but stating that Mississippi's highest court had "not explicitly adopted or rejected the Flynn Effect"); *Ybarra v. State*, 247 P.3d 269, 282 (Nev. 2011) (choosing not to "take sides in the dispute over the Flynn effect" because adjusting for the Flynn Effect would have only resulted in an IQ score of 78). Likewise, in *Sasser v. Hobbs*, the U.S. District Court for the Western District of Arkansas refused to rule on the validity and applicability of the Flynn Effect because the court found it unnecessary to do so—even applying the Flynn Effect, the petitioner still could not prove the required subaverage intellectual functioning. *Sasser v. Hobbs*, 751 F. Supp. 2d 1063, 1081–82 (W.D. Ark. 2010).

143. *Ex parte Blue*, 230 S.W.3d 151, 166 (Tex. Crim. App. 2007); *see also* *Neal v. State*, 256 S.W.3d 264, 273 (Tex. Crim. App. 2008) (refusing to apply the Flynn Effect because it is an "unexamined scientific concept").

144. *Ex parte Cathey*, No. WR-55, 161-02, 2008 Tex. Crim. App. Unpub. LEXIS 850, at *1–3 (Tex. Crim. App. Nov. 18, 2008). Pursuant to the appellate court's order, the trial court must decide: (1) the "scientific validity and reliability" of the Flynn Effect, (2) whether clinical practitioners ordinarily apply the Flynn Effect to IQ-test scores for their mental retardation diagnoses outside of the criminal justice system, (3) whether the relevant professional community, outside of the criminal justice system, generally accepts the application of the Flynn Effect to IQ-test scores, and (4) the known or potential margin of error of the Flynn Effect with regard to IQ-test scores. *Id.* at *2–3. As of the time of this Note's publication, the trial court had held the hearing but had not yet issued its decision.

Salazar, the Fifth Circuit discussed its indecision and refused to rule on the issue.¹⁴⁵ In *Wiley v. Epps*, the Fifth Circuit affirmed a district court's finding that the defendant was mentally retarded and found that the district court's consideration of the Flynn Effect in interpreting the defendant's IQ scores was not clearly erroneous.¹⁴⁶ The court stated that the district court only held that the Flynn Effect "should be considered as one factor when interpreting Wiley's IQ scores in light of the unanimous opinion from the experts," but the district court did not hold that the IQ scores must be adjusted.¹⁴⁷ The Fifth Circuit further qualified that, even without accounting for the Flynn Effect, the evidence showed that Wiley had significant subaverage intellectual functioning.¹⁴⁸ Given the Fifth Circuit's *Wiley* decision and the pending state court decision in *Cathey*, these two overlapping jurisdictions may soon officially recognize the Flynn Effect. But, for now, their indecision fuels the overall inconsistency that characterizes the Flynn Effect in the *Atkins* context.

D. Rejection of the Flynn Effect

Based primarily on strict textual interpretations of state statutory definitions, state courts in Kentucky and Oklahoma and federal district courts in Arizona and Georgia¹⁴⁹ have explicitly rejected the consideration of and adjustment for the Flynn Effect in *Atkins* cases. In *Smith v. State*, the Oklahoma Court of Criminal Appeals stated that the Flynn Effect, "whatever its validity, is not a relevant consideration in the mental retardation determination for capital defendants," because the Oklahoma legislature only included

145. *Maldonado*, 625 F.3d at 238; *In re Mathis*, 483 F.3d 395, 398 n.1 (5th Cir. 2007); *Salazar*, 443 F.3d at 433 n.1. Most recently, the federal district court in the Southern District of Texas refused to apply the Flynn Effect, finding that, although all the testifying experts agreed that the Flynn Effect existed, the Fifth Circuit's lack of adoption and the lack of consensus on its applicability did not support its application. *United States v. Bourgeois*, No. C-02-CR-216, 2011 U.S. Dist. LEXIS 55859, at *90-93 n.37 (S.D. Tex. May 19, 2011).

146. *Wiley v. Epps*, 625 F.3d 199, 214 (5th Cir. 2010).

147. *Id.*

148. *Id.*

149. The rejection of the Flynn Effect by the federal district court in the Northern District of Georgia is mostly "academic" because rejecting the Flynn Effect was not dispositive in the case. *Ledford v. Head*, No. 1:02-CV-1515-JEC, 2008 WL 754486, at *8 (N.D. Ga. Mar. 19, 2008). The court in *Ledford* devoted a separate section in its opinion to discuss and reject the Flynn Effect. *Id.* at *1, *6. The court stated that it was "not impressed" with the proffered evidence advocating for the Flynn Effect, citing how both parties' experts agreed on the lack of clinical use of the effect. *Id.* at *7.

consideration of the standard error of measurement in the statute.¹⁵⁰ Furthermore, the court stated that the Flynn Effect had “not achieved universal acceptance in courts where it has been raised.”¹⁵¹

In *Ramirez v. Ryan*, the U.S. District Court for the District of Arizona cited a number of reasons for not adjusting for the Flynn Effect to reduce the petitioner’s IQ scores: (1) the Arizona mental retardation statute does not mandate adjustment, (2) no Arizona precedent required adjustment, (3) the WAIS-III IQ-test manual did not recommend adjusting for the Flynn Effect, (4) disagreement existed among both sides’ experts on whether to adjust for the Effect, and (5) other courts had decided not to adjust for it.¹⁵²

Like Tennessee state courts prior to *Coleman*,¹⁵³ at least one Kentucky state court has construed the state’s death penalty statute as setting a strict cutoff score of 70 for the intellectual-functioning element, thereby precluding adjustments for the Flynn Effect.¹⁵⁴ The Kentucky court reasoned that, in addition to a lack of scientific support for the Effect, a lack of statutory support existed because the Kentucky legislature deliberately refused to expand the definition of mental retardation to include the Flynn Effect.¹⁵⁵ The rejection of the Flynn Effect in the Kentucky, Oklahoma, and federal courts stands in contrast to the acceptance of and adjustment for the Flynn Effect in other jurisdictions and reflects the spectrum of inconsistency surrounding the Flynn Effect in the *Atkins* context.

IV. RESOLVING THE BATTLE OF THE EXPERTS: THE FLYNN EFFECT IS VALID AND SHOULD BE ACCOUNTED FOR IN *ATKINS* CASES

In courts around the country, a heated “battle of the experts” wages over the Flynn Effect’s proper role in *Atkins* cases.¹⁵⁶ This constant and inevitable battle between defense and government experts contributes to, if not partially causes, the troubling inconsistency in the way courts treat the Flynn Effect when

150. *Smith v. State*, 245 P.3d 1233, 1237 n.6 (Okla. Crim. App. 2010).

151. *Id.* at 1235.

152. *Ramirez v. Ryan*, No. CV-97-1331-PHX-JAT, 2010 U.S. Dist. LEXIS 110131, at *46–47 (D. Ariz. Sept. 28, 2010).

153. *See supra* Part III.B.

154. *Bowling v. Commonwealth*, 163 S.W.3d 361, 375 (Ky. 2005).

155. *Id.*

156. *See Flynn, The WAIS-III and WAIS-IV, supra* note 53, at 104 (discussing the battle of the experts over the Flynn Effect); *see also Bowling*, 163 S.W.3d at 375 (describing how states have chosen not to define “significantly subaverage intellectual functioning,” “presumably relegating the issue to a ‘battle of the experts’”).

determining death penalty eligibility. This Note does not seek to undermine or remedy the adversarial approach to expert testimony but rather seeks to provide an answer to the following question: Who is right? Empirical evidence and American death penalty jurisprudence suggest that the Flynn Effect is valid and should be adjusted for in *Atkins* cases.

To honor the *Atkins* holding that no state shall execute a mentally retarded person, courts must attempt to make accurate determinations of who falls within this categorical exclusion. Whether a state can constitutionally execute a person should not vary by which experts testify or by when the cited IQ tests were normed. Recognizing the moral tension inherent in an expert's role in *Atkins* determinations,¹⁵⁷ psychologist and IQ-test developer Lawrence G. Weiss states that expert witnesses should not "stake out" positions on the Flynn Effect nor should they fail to "acknowledge the legitimacy of any competing point of view."¹⁵⁸ Rather, according to Weiss, experts should inform courts or juries fully of research on the Flynn Effect.¹⁵⁹

Although such neutrality is difficult to achieve in the adversarial system, this Note takes to heart Weiss's words and reaches a conclusion about the Flynn Effect only after reviewing all information concerning the Effect and after considering the arguments from both sides of the debate. Based on a full record of the evidence, this Note finds that: (1) the Flynn Effect represents a widely accepted phenomenon; (2) although based on aggregate IQ gains, the Flynn Effect applies to individual IQ scores; (3) given the relative nature of IQ testing and scoring, state legislatures and courts should not impose a strict IQ-cutoff requirement for *Atkins* determinations but should allow for flexible interpretation of IQ scores that includes adjustment for the Flynn Effect; (4) the Flynn Effect, while not perfectly precise, still provides a well-supported and valid basis for adjusting IQ scores; and (5) as distinguished from other contexts, the death penalty

157. See, e.g., Duvall & Morris, *supra* note 61, at 664 (recognizing the "ethical dilemma" psychologists face when testifying in *Atkins* cases in states where "some procedures mandated by statute appear to defy sound psychological measurement principles," such as the bright-line IQ score cutoff, forcing psychologists into positions where they "may violate the APA ethics code"); Linda Knauss & Joshua Kutinsky, *Into the Briar Patch: Ethical Dilemmas Facing Psychologists Following Atkins v. Virginia*, 11 WIDENER L. REV. 121 (2004) ("[*Atkins*] raises a host of nascent ethical concerns for those psychologists who choose to become involved. Principled psychological practitioners cannot possibly ignore the potentially lethal consequences of their determinations.").

158. Lawrence G. Weiss, *Considerations on the Flynn Effect*, 28 J. PSYCHOEDUC. ASSESSMENT 482, 491 (2010).

159. *Id.*

context necessitates consideration of and adjustment for the Flynn Effect in *Atkins* cases given the severity of the Effect's life and death implications. Throughout Part IV and as summarized in Part V, this Note also sets forth solutions to the problems posed by the inconsistent treatment of the Flynn Effect.

A. Uncontroverted Acknowledgement That the Flynn Effect Is Real

Experts and researchers generally agree that the Flynn Effect exists.¹⁶⁰ Since the Flynn Effect surfaced as a phenomenon in the 1980s, many researchers have studied it, and test publishers have revised and renormed IQ tests more frequently than before to account for the Effect.¹⁶¹ A search in the APA's database found 112 peer-reviewed articles, books, book chapters, and dissertations, all addressing the "long-recognized and empirically demonstrated" phenomenon of the Flynn Effect.¹⁶²

Several debates have arisen regarding issues implicated by the Flynn Effect. These disagreements, however, do not affect the accepted existence of the phenomenon. For example, although a heated debate surrounds the question of what causes the Flynn

160. See Ceci & Kanaya, *supra* note 2, at 441 (stating that "few question the existence of the effect"); Cecil Reynolds et al., *Failure to Apply the Flynn Correction in Death Penalty Litigation: Standard Practice of Today Maybe, But Certainly Malpractice of Tomorrow*, 28 J. PSYCHOEDUC. ASSESSMENT 477, 478 (2010) ("[T]he existence of the effect has no significant scholarly challenges of which we are aware. The FE, whatever its cause, is as real as virtually any effect can be in the social sciences. Studies have observed an increase of 0.3 points per year in average IQs"); Weiss, *supra* note 158, at 489 (stating that all contributors to the special issue of the *Journal of Psychoeducational Assessment* agree that the generational effects of the Flynn Effect exist); Zhou et al., *supra* note 52, at 399 ("Research in the past two decades has shown cumulative support for the existence of the Flynn effect (FE)—the massive IQ gain at about 0.3 points per year.") (internal citations omitted). *But see* Kevin S. McGrew, *The Flynn Effect and Its Critics: Rusty Linchpins and 'Lookin' for g and Gf in Some of the Wrong Places*, 28 J. PSYCHOEDUC. ASSESSMENT 448, 449 (2010) ("The consensus of most (but not all) intelligence scholars is that the Flynn effect (FE), at the level of the global IQ score, is real.")

In an article published in 1999, Joseph L. Rodgers questioned the existence of the Flynn Effect, stating that he was "not sure what the Flynn Effect really is." Joseph L. Rodgers, *A Critique of the Flynn Effect: Massive IQ Gains, Methodological Artifacts, or Both?*, 26 INTELLIGENCE 337, 339 (1999). Rodgers, however, also stated that "[e]ven with a healthy dose of skepticism, the effect rises above purely methodological interpretation, and appears to have substantive import." *Id.* at 354. Less than a decade later, Rodgers stated in a 2007 article that "[t]he basic empirical status of the Flynn Effect is well-established" and that "measured IQ has been increasing at a rate of approximately .33 IQ points per year for most of the past century, a pattern consistent across both time and geography." Joseph L. Rodgers & Linda Wanstrom, *Identification of a Flynn Effect in the NLSY: Moving from the Center to the Boundaries*, 35 INTELLIGENCE 187, 187, 193 (2007).

161. Reynolds et al., *supra* note 160, at 478.

162. Cunningham & Tassé, *supra* note 40, at 416.

Effect's IQ gains,¹⁶³ this causality debate does not affect the Flynn Effect's accepted existence, its measured rate of gain, or its applicability to IQ scores—the observed increase in IQ scores is real regardless of the reason behind the increase.¹⁶⁴

Given the largely unquestioned existence of the Flynn Effect, a natural question arises: Why do some experts and courts refuse to adjust for the Flynn Effect in *Atkins* cases? Opponents to Flynn Effect adjustments in the *Atkins* context view the issue of the Effect's validity and applicability as distinct from the issue of its existence.¹⁶⁵ To support their position, those against accounting for the Flynn Effect offer four primary arguments, as addressed below: (1) the Effect's inapplicability to individuals, (2) the lack of statutory mandates for adjustments based on the Effect, (3) the Effect's variability and imprecise measure of gain, and (4) the conflict with routine, standard practices posed by adjusting scores for the Effect.

163. See KAUFMAN, *supra* note 5, at 207–08 (providing an overview of proposed causes of the Flynn Effect, including globalization, improved nutrition, and education); Richard Lynn, *In Support of the Nutrition Theory*, in THE RISING CURVE, *supra* note 8, at 207, 207–12 (regarding nutrition as the sole factor for secular increases in intelligence); Carmi Schooler, *Environmental Complexity and the Flynn Effect*, in THE RISING CURVE, *supra* note 8, at 67, 73–74, 76–77 (attributing the IQ gains to changes in environmental complexity and work culture).

164. See Kanaya et al., *supra* note 41, at 779 (“Although there is not a consensus among professionals as to why these gains are occurring or what these gains actually mean (e.g., are we really getting smarter?), all are in agreement that the gains occur and that they hold great theoretical and practical importance.” (citations omitted)); Alan S. Kaufman, “*In What Way Are Apples and Oranges Alike?*” *A Critique of Flynn’s Interpretation of the Flynn Effect*, 28 J. PSYCHOEDUC. ASSESSMENT 382, 384–92 (2010) (noting that his criticism of Flynn’s interpretation of the Flynn Effect challenges Flynn’s theory of causality, not the existence or the magnitude of the Flynn Effect); Kaufman, *supra* note 2, at 503 (noting that the controversy about what causes the Flynn Effect “makes no difference for the scientific validity of the FE or the strong recommendation by myself, Flynn, Reynolds, Fletcher, and others to adjust the IQs in death penalty cases”). In recent literature, Flynn and other researchers also voice their disagreement on the issue of whether the norms on a specific IQ test, the Wechsler WAIS-III test, are substandard and require additional adjustments. See, e.g., Weiss, *supra* note 158, at 486 (discussing his and others’ criticism of Flynn for arguing that the norms on the WAIS-III test are substandard solely because they do not conform to predictions based on the Flynn Effect). Like the causality debate, this disagreement is separate from the debate on which this Note focuses and does not affect the existence or validity of the Flynn Effect. The WAIS-III debate revolves around the issue of whether to adjust WAIS-III scores for substandard, not obsolete, norms beyond the Flynn Effect’s 0.3 points per year; whereas, this Note only advocates for adjusting scores to correct for obsolete norms based on the 0.3-point-per-year figure. See Weiss, *supra* note 158, at 486 (stating that the WAIS-III debate “is a different issue than adjusting scores for obsolete tests”).

165. See, e.g., Leigh D. Hagan et al., *Adjusting IQ Scores for the Flynn Effect: Consistent With the Standard of Practice?*, 39 PROF. PSYCHOL.: RES. & PRAC. 619, 620 (2008) (“Although mainstream recognition of the [Flynn Effect] as an authentic psychometric consideration has increased, the question of how to accurately represent its impact for a particular individual’s earned scores on IQ tests is a different question altogether.”).

B. The Flynn Effect Applies to Individual IQ Scores

Courts and such experts as Leigh Hagan have expressed the concern that the “probabilistic nature of group data and potential inconsistency when applied to individuals” render the Flynn Effect inapplicable to individual IQ scores because the Effect reflects changes in the means of aggregated or group IQ scores.¹⁶⁶ Based on the well-established tenets of IQ testing, this Section debunks this false premise, showing instead that the Flynn Effect is applicable to individual IQ scores, including those of capital defendants.

An IQ score is not absolute nor can it be interpreted in isolation. Rather, one can only determine an IQ score by comparing it to the IQ scores of a group—the standardization sample. The Flynn Effect applies to individual scores because it updates the norms produced by the standardization sample.¹⁶⁷ As discussed in Part II.A, all IQ scores reflect relative measurements determined from the performance of representative groups, known as standardization samples.¹⁶⁸ Test publishers use these samples to create normal distributions of scores that show where an individual test taker’s performance falls relative to the group mean score of 100.¹⁶⁹ Group data therefore define and serve as the bases for all individual IQ-score measurements. The Flynn Effect actually improves that body of group data by updating the norms set by the original standardization sample; it also improves the measurement of individual IQ scores by

166. *Id.* (“Of particular importance to the evaluating psychologist is whether the observed changes in group mean scores over time apply reliably to a specific individual.”); Leigh D. Hagan et al., *IQ Scores Should Not Be Adjusted for the Flynn Effect in Capital Punishment Cases*, 28 J. PSYCHOEDUC. ASSESSMENT 474, 475 (2010) [hereinafter Hagan et al., *IQ Scores Should Not Be Adjusted*]; see also *Neal v. State*, 256 S.W.3d 264, 274 (Tex. Crim. App. 2008) (referencing the testimony of one of the state’s expert witnesses, who noted that the Flynn Effect applies only to groups and not individuals). *But see United States v. Davis*, 611 F. Supp. 2d 472, 487–88 (D. Md. 2009) (rejecting the government’s claim or “senseless mantra” that the Flynn Effect only applies to groups and not individuals).

167. See Cunningham & Tassé, *supra* note 40, at 417 (“The interpretation of any IQ score involves utilizing information from the standardization group (which almost never contained the individual being assessed) to interpret the performance of a specific individual. Indeed, this application of group data to the individual constitutes virtually the entirety of the field of psychometrics”); Reynolds et al., *supra* note 160, at 479 (“[T]o argue the [Flynn Effect] should not be applied to individuals belies the fact that all IQs, obtained or otherwise, are to a significant extent based on a group effect and derived from aggregated data.”).

168. Flynn explains that “[n]o individual ever got an IQ score except by comparison with the performance of a group, namely, a standardization sample.” Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 103.

169. For a more detailed explanation of standardized scoring for IQ tests and the standardization process, see Part II.A.

comparing them to the scores of their contemporaries, rather than the lower scores of past test takers.¹⁷⁰

Given the relationship between individual IQ scores and group data, adjusting an individual's IQ score for the Flynn Effect achieves the same effect as adjusting group data, such as the mean score for an IQ test or the cutoff score for mental retardation.¹⁷¹ Simply put, to account for the Flynn Effect, one could either adjust the individual test taker's IQ score or adjust the mean score for the entire IQ test, a value derived from group data.¹⁷² No real distinction exists between individual and group IQ values. The "group" phenomenon of the Flynn Effect thus applies with equal force to individual IQ scores.

C. *The Dark Side of Bright-Line IQ Cutoffs*

Courts that strictly construe the intellectual-functioning element as a bright-line cutoff with no room for adjustments like the Flynn Effect misunderstand and distort the use of IQ scores to determine mental retardation, potentially producing perverse results. IQ scores represent relative measurements that "should always be interpreted within a range of scores" and not as single, absolute values.¹⁷³

170. See Reynolds et al., *supra* note 160, at 479 ("As the group used to provide these statistics ages and becomes less like the current target population, applying any correction that can improve the accuracy of the placement of the individual on this continuum (e.g., the Flynn correction) improves IQ estimation for the individual.").

171. See, e.g., Kaufman, *supra* note 2, at 501 (describing the application of the Flynn Effect as a "group-to-group application," not a "group-to-individual application"); Weiss, *supra* note 158, at 489 (describing the adjustment of individual IQ scores as "really a back-door method to adjusting the cut-off score for intellectual deficiency based on a more current reference sample").

172. As an example, if a defendant receives a score of 75 on an IQ test taken twenty years after the test was normed, an adjustment to the individual IQ score for the Flynn Effect results in an adjusted score of 69. See *supra* note 56 (explaining the mathematics behind the adjustment for the Flynn Effect for these specific values). Assuming the IQ test has a mean score of 100 and a standard deviation of fifteen points, the adjusted score of 69 now falls more than two standard deviations below the mean score (below 70). One could achieve the same result by adjusting the aggregate mean score, instead of adjusting the individual IQ score, because the rate of gain of 0.3 points per year applies to all scores on the distribution, including the mean and individual scores. Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 104. Thus, for the example above, twenty years after the IQ test was normed, the original mean score of 100 would have inflated to a score of 106, given the 0.3-point-per-year increase over twenty years. With an adjusted mean score of 106, two standard deviations (thirty points) below the mean score becomes a score of 76. By adjusting the group mean and not the individual IQ score, the defendant's score of 75 again falls more than two standard deviations below the mean. The two methods above both effectively update a test's norms, but as Weiss notes, because many states define mental retardation as a score of 70 or below, attorneys must argue for adjusting individual scores, rather than the mean, even though the adjustments accomplish the same result. Weiss, *supra* note 158, at 489.

173. Frumkin, *supra* note 27, at 2.

1. Approximating IQ as Recommended by Leading Organizations

The AAIDD, a leader in the intellectual disability area,¹⁷⁴ states that “an IQ test score of around 70 or as high as 75 indicates a limitation in intellectual functioning.”¹⁷⁵ The APA defines significantly subaverage intellectual functioning for purposes of diagnosing mental retardation as “an IQ of *approximately* 70 or below.”¹⁷⁶ The Supreme Court itself in *Atkins* noted that “between 1 and 3 percent of the population has an IQ between 70 and 75 or lower, which is typically considered the cutoff IQ score for the intellectual function prong of the mental retardation definition.”¹⁷⁷ Both the AAIDD and APA definitions show that, in determining mental retardation, one should consider a range of IQ scores, as well as the approximate nature of a person’s IQ. A bright-line test—one that would preclude a successful *Atkins* claim if the defendant’s IQ score did not fall below an inflexible threshold—presents a requirement contrary to the positions of the AAIDD and APA because it allows for neither a range “around 70” nor an IQ of “approximately 70.”

In a 2010 letter written to its APA counterpart, the AAIDD Terminology and Classification Committee, which publishes the AAIDD definition manuals, further defined the intellectual-functioning criterion as “approximately two standard deviations below the mean, considering the standard error of measurement for the specific assessment instruments used and the strengths and limitations of the instrument.”¹⁷⁸ The AAIDD explained its use of the term “approximately” in its definition:

It has become increasingly clear to us in both our academic and clinical work that measures of human functioning contain error and that a major role of a well trained clinician is to assist others in understanding that any measure of intelligence or adaptive behavior is only an approximation. To underscore the importance of recognizing and using “approximately” in best clinical practices we have stressed . . . [t]o

174. The Supreme Court in *Atkins* cited and recognized the mental retardation definition of the AAIDD’s predecessor, the AAMR. *Atkins v. Virginia*, 536 U.S. 304, 309 n.3 (2002).

175. *Definition of Intellectual Disability*, AAIDD, http://www.aaid.org/content_100.cfm?navID=21 (last visited Mar. 11, 2012).

176. APA, *DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS* 49 (4th ed. 2000) (emphasis added).

177. *Atkins*, 536 U.S. at 309 n.5 (citing 2 BENJAMIN J. SADOCK & VIRGINIA A. SADOCK, *COMPREHENSIVE TEXTBOOK OF PSYCHIATRY* 2952 (7th ed. 2000)).

178. AAIDD 11th Edition Implementation Committee, *Public Comments Regarding Draft Definition of Intellectual Disability*, AAIDD 2, Feb. 22, 2010, <http://www.aaid.org/media/PDFs/DSMV.pdf>.

not specify a hard and fast cutoff point/score for meeting the significant limitations criteria for both intelligence and adaptive behavior scores.¹⁷⁹

While *Atkins* did not require states to adopt the AAIDD and APA definitions, those definitions serve as persuasive authority, supported by articulated clinical rationales, for how states should define the *Atkins* categorical exclusion. Many state definitions of mental retardation or intellectual disability set forth the intellectual-functioning prong as an IQ score of 70 or below.¹⁸⁰ This “70 or below” definition is the “most common statutory deviation from the definitions of mental retardation” because the strict cutoff not only directly conflicts with the AAIDD and APA definitions, but also “mistakenly suggests that a 1-point difference in two people’s scores reflects a significant difference in their cognitive capacities.”¹⁸¹ In actuality, no real difference exists between IQ scores of 69 and 71.¹⁸²

2. The Blunder of Bright-Line Cutoffs

Despite the strong guidance offered by the AAIDD and APA, some state and federal courts continue to adhere to strict IQ cutoffs in rejecting the Flynn Effect. For instance, the Supreme Court of Kentucky, in *Bowling v. Commonwealth*, found that the state legislature “chose not to expand the mental retardation ceiling” and “chose a bright-line cutoff ceiling of an IQ of 70, a generally recognized level at which persons are considered mentally retarded.”¹⁸³ The *Bowling* court asserted that the legislature knew about the Flynn Effect, as well as margins of error, at the time it enacted the statutory definition and thus deliberately chose not to incorporate the Flynn Effect into the intellectual-functioning element.¹⁸⁴

The *Bowling* court’s reasoning relies on two unsupported assumptions. First, the court’s bright-line approach construes IQ scores as absolute values. This assumption runs counter to the basic premise of IQ scoring and mental retardation determinations. As stated throughout this Note and as expressly recognized by the AAIDD and APA, IQ scores are relative, not absolute, measurements that serve as “approximate” indicators of a person’s IQ.¹⁸⁵

179. *Id.* at 3.

180. *See infra* Part VI.

181. Duvall & Morris, *supra* note 61, at 659–61.

182. MacVaugh & Cunningham, *supra* note 16, at 157.

183. *Bowling v. Commonwealth*, 163 S.W.3d 361, 375 (Ky. 2005).

184. *Id.*

185. *See supra* Part II.A; Part IV.C.1.

Second, the court in *Bowling* presumed that its state legislature knew and understood the concepts of IQ testing and scoring, as well as the merits of the Flynn Effect, when enacting the statutory definitions for mental retardation. The complicated nature of mental retardation determinations and the need for expert testimony show that state legislatures most likely do not possess any specific intent as to the Flynn Effect. The Supreme Court in *Atkins* openly recognized the difficulty in determining mental retardation and cited many outside sources in the opinion to indicate the general consensus on the criteria for mental retardation.¹⁸⁶ Given the specialized knowledge and expertise necessary, courts around the country rely on the testimony of not one but several experts to make the determination of whether a defendant possesses significantly subaverage intellectual functioning.¹⁸⁷ The claim that a legislature knew of, understood, and had specific intent toward the Flynn Effect when enacting statutory provisions to effectuate *Atkins* appears both unsupported and disingenuous in light of the recognized complexity of the Flynn Effect and mental retardation determinations as a whole.

Despite its straightforward application, a bright-line rule does not work in the *Atkins* context and potentially disserves those individuals that *Atkins* intended to protect. In many *Atkins* cases, the defendant has taken multiple IQ tests resulting in multiple IQ scores, often scores both below and above 70.¹⁸⁸ A bright-line requirement of “an IQ score of 70 or below” makes no sense in a multiple-score context and even fails at providing a bright line.

The most disturbing implication of a strict IQ cutoff, however, is the possibility that a state, rejecting the Flynn Effect and other margins of error, will execute a mentally retarded defendant because

186. *Atkins v. Virginia*, 536 U.S. 304, 309 n.5, 317 (2002).

187. Even in *Bowling*, the Supreme Court of Kentucky described how, during the determination of the intellectual-functioning criterion, the judge must consider evidence that might “come in the form of expert testimony as to the margin of error of a given testing procedure or as to the so-called ‘Flynn effect.’” *Bowling*, 163 S.W.3d at 388.

188. See, e.g., *Maldonado v. Thaler*, 662 F. Supp. 2d 684, 708–09 (S.D. Tex. 2009) (reporting defendant’s five IQ scores, which, unadjusted, ranged from 61 to 83); *Thomas v. Allen*, 614 F. Supp. 2d 1257, 1293–98, 1300–02 (N.D. Ala. 2009) (reporting seven IQ scores, ranging from 56 to 74), *aff’d*, 607 F.3d 749 (11th Cir. 2010); Stephen Greenspan, *Issues in the Use of the “Flynn Effect” to Adjust IQ Scores When Diagnosing MR*, PSYCHOL. IN MENTAL RETARDATION & DEVELOPMENTAL DISABILITIES: OFFICIAL PUBLICATION OF DIVISION 33, AM. PSYCHOL. ASS’N, Spring 2006, at 5 (“Typically, there are multiple IQ tests in the record of an *Atkins* applicant . . . and these fluctuate around the upper IQ boundary of the mild [mental retardation] range.”).

he scored above a 70 on an outdated IQ test.¹⁸⁹ The Tennessee Court of Criminal Appeals in *Smith v. State*, decided prior to *Coleman*,¹⁹⁰ articulated the troubling tension between Tennessee's former bright-line test and the risks inherent in drawing a definitive line for death penalty eligibility that is unsupported by clinical practice and sound legal doctrine.¹⁹¹ The inflexibility of such a cutoff, the court stated, restricts courts from evaluating the evolving nature of IQ scores and risks the execution of a mentally retarded defendant.¹⁹² By allowing the consideration of factors like the Flynn Effect, the Tennessee Supreme Court in *Coleman* essentially withdrew its hard, unsubstantiated bright-line rule. As demonstrated in *Coleman* and this Section, bright-line cutoffs have no place in the *Atkins* framework. IQ scores constitute approximate values subject to error, for which the Flynn Effect is one adjustment.

3. Legislative and Judicial Solutions

In addition to addressing problems with the bright-line IQ cutoffs currently employed by several courts, this Note also offers an alternative to these cutoffs. Both state legislatures and judiciaries possess the power to establish more accurate requirements and guidelines to better serve *Atkins*.

State legislatures should amend their *Atkins* mental retardation requirements or, for those states without such existing statutes, enact new provisions to reflect the widely accepted consensus that IQ scores are relative, nonabsolute measurements that are most properly construed as approximations or ranges. Although one uniform definition in all death penalty states would best ameliorate the inconsistency among states, *Atkins* does not require such uniformity from the states. In drafting definitions for the intellectual-functioning element, states should nevertheless follow two unifying principles.

First, the definitions should allow for a flexible interpretation of IQ scores. Currently, five state statutes define "significantly subaverage general intellectual functioning" as two or more standard

189. Some courts will stop their *Atkins* determinations at the intellectual-functioning prong and will not move on to address the adaptive-functioning criterion if they do not find a requisite IQ score—making the defendant's IQ score dispositive. *See supra* note 68.

190. *See supra* Part III.B.

191. *Smith v. State*, No. E2007-00719-CCA-R3-PD, 2010 Tenn. Crim. App. LEXIS 793, at *123–125 (Tenn. Crim. App. Sept. 21, 2010), *death sentence vacated*, 357 S.W.3d 322 (Tenn. 2011).

192. *Id.* at *124–125.

deviations from the mean score of a standardized IQ test.¹⁹³ As compared to the “70 or below” bright-line definition, a “standard deviation” definition allows for more flexible interpretation and thus more accurately reflects the nature of IQ scores. It permits the decisionmaker to consider the appropriate “standard deviation” and “mean score” based on the specific IQ tests presented. Bright-line IQ cutoffs, on the other hand, only permit the decisionmaker to apply a set threshold that ignores the fact that different tests may have different standard deviations and mean scores.¹⁹⁴ Furthermore, flexibility in the interpretation of “mean score” allows a decisionmaker to adjust for the Flynn Effect and update norms, as discussed in Part IV.B, by raising the mean score 0.3 points for every year the test is outdated.¹⁹⁵

Those opposing a flexible interpretation of IQ scores may argue that this approach inserts too much discretion into *Atkins* determinations, but this argument fails to recognize that flexibility is inherent in a mental retardation determination and cannot be set aside for arbitrary rigidity. Moreover, the Supreme Court, in its death penalty jurisprudence, has long preferred discretion and individualized consideration over hard, unsubstantiated lines.¹⁹⁶

The adaptive-behavior element of mental retardation determinations also requires individualized treatment, raising the second unifying principle that state legislatures should heed: state

193. See *infra* Part VI. States without such statutory definitions have promulgated a similar definition through judicial precedent. See, e.g., *Ex parte Briseno*, 135 S.W.3d 1, 7 n.24 (Tex. Crim. App. 2004) (defining significantly subaverage general intellectual functioning as a score “approximately [two] standard deviations below the mean”).

194. A cutoff of “70 or below” only serves Wechsler IQ tests and the most recent Fifth Edition Stanford-Binet IQ test to the extent that those tests have up-to-date norms. See *supra* note 34.

195. Flexible statutory interpretation of IQ scores would also allow for other necessary adjustments—notably standard errors of measurement and the practice effect. While not discussed in this Note, these adjustments should also apply in *Atkins* cases to adjust IQ scores, because such concepts are statistically sound and aim to improve the accuracy of IQ measurements. Standard error of measurement (“SEM”) relies on the “fundamental assumption in the field of psychological assessment . . . that all tests have error” and “provides an estimate of the amount of error in a person’s observed test score.” MacVaugh & Cunningham, *supra* note 16, at 146. “The SEM is simply another way of expressing the reliability of a test; as the reliability of the instrument increases, the SEM decreases, which gives the examiner more confidence in the accuracy of an observed score.” *Id.* The SEM is usually construed as a range of error. *Id.* Practice effects, on the other hand, are “caused by repeated administrations of the same intelligence test in a short period of time,” which “may be problematic in *Atkins* cases should multiple experts administer the same intelligence test to offenders within a relatively brief timeframe.” *Id.* at 147.

196. See *Woodson v. North Carolina*, 428 U.S. 280, 303–04 (1976) (striking down North Carolina’s mandatory death penalty statute and clarifying that *Furman v. Georgia* requires individualized consideration of each defendant’s case).

statutes should instruct decisionmakers, whether judge or jury, to not treat borderline IQ scores as dispositive of the entire *Atkins* claim and to consider adaptive behavior, in conjunction with intellectual functioning, before making an *Atkins* determination. Given the approximate and imprecise nature of IQ scores, such psychologists as Stephen Greenspan and Mark Cunningham emphasize the need to consider both intellectual functioning and adaptive behavior when making mental retardation determinations.¹⁹⁷ An *Atkins* decision cannot rely solely on IQ scores, and thus strict IQ cutoffs do not belong in *Atkins* determinations.

Like state legislatures, both federal and state courts can also effectuate the two principles discussed above. Whether interpreting existing state statutory definitions or filling a statutory void by judicially establishing a definition, courts should approach IQ scores as approximate measures of the intellectual-functioning element. They should adjust IQ scores for the Flynn Effect when presented with scores from outdated IQ tests. Courts should also consider adaptive behavior before deciding the overall *Atkins* claim. When faced with precedential bright-line IQ cutoffs, courts, like the court in *Coleman*, should reject those bright-line rules. They should also go one step further than *Coleman*, replacing such rules with precedent that requires a flexible interpretation of IQ scores, including the practice of adjusting scores for the Flynn Effect.

D. Validity Despite Variability

Experts and researchers generally acknowledge that the Flynn Effect's IQ gain of 0.3 points per year is an approximate measure of average IQ gains that could vary across ability levels or different ranges of IQ scores.¹⁹⁸ However, differences of opinion exist regarding whether, despite this variability, the Flynn Effect should apply in *Atkins* cases to adjust for obsolete norms. Opponents to Flynn Effect adjustments set forth two arguments regarding the Effect's variability: (1) the amount of the Flynn Effect's average IQ gain is inexact and may change, and (2) the amount of gain is not constant for

197. Greenspan, *supra* note 188, at 6; *see also* MacVaugh & Cunningham, *supra* note 16, at 143 ("This psychometric imprecision [of IQ scores] underscores the importance of considering IQ scores within the context of other evidence of adaptive functioning.").

198. *See, e.g.*, Jack M. Fletcher et al., *IQ Scores Should Be Corrected for the Flynn Effect in High-Stakes Decisions*, 28 J. PSYCHOEDUC. ASSESSMENT 469, 471–72 (2010) ("There is variability across studies, and age/ability level . . ."); Weiss, *supra* note 158, at 490 (stating that most of the causal theories of the Flynn Effect imply that "the rate of gain is not normally distributed in the population").

all ages and ability levels or test scores—for example, the argument posits that, although the average rate of gain hovers around 0.3 points per year, people who score in one range, such as 120 to 130, experience a different rate of gain than people who score in another range, such as 60 to 70. As discussed below, empirical evidence still weighs in favor of adjusting for the Flynn Effect, despite these arguments.

1. Research Consistently Measures the Flynn Effect Near 0.3 Points per Year

Citing the different values offered for the Flynn Effect over the years, Hagan and coauthors describe the magnitude of the Flynn Effect as a “moving target” and argue that IQ scores should not be adjusted in *Atkins* determinations.¹⁹⁹ Although researchers have set forth multiple figures for the Flynn Effect’s IQ gains since Flynn’s 1984 study, these figures have all hovered around the 0.3-point figure, as to not make a significant difference.²⁰⁰ Flynn himself presently sets his Flynn Effect adjustments at 0.3 points per year for each year between the test norming and test administration.²⁰¹ Therefore, although Hagan and coauthors correctly state that researchers have presented several figures for the rate of IQ gain, those figures have not wandered far from the 0.3-point figure advocated by this Note.

Hagan and coauthors also cite two studies that found that the Flynn Effect may have ceased or reversed in Norway²⁰² and

199. Hagan et al., *supra* note 165, at 620; Hagan et al., *IQ Scores Should Not Be Adjusted*, *supra* note 166, at 474–75.

200. *See, e.g.*, Kaufman, *supra* note 2, at 495 (calling the three-point-per-decade figure “robust” and proven among several populations); Weiss, *supra* note 158, at 491 (“The FE has been shown to be near 3 IQ points per decade on average across a large number of studies . . .”).

Hagan and coauthors claim that Flynn previously identified the IQ gain at 0.25 and 0.29 points per year. Hagan et al., *IQ Scores Should Not Be Adjusted*, *supra* note 166, at 474–75. Experts testifying in several cases have recommended and some courts have applied Flynn Effect adjustments of 0.33 points per year. *See, e.g.*, *United States v. Lewis*, No. 1:08 CR 404, 2010 U.S. Dist. LEXIS 138375, at *17 (N.D. Ohio Dec. 23, 2010) (stating that the Flynn Effect entails an adjustment of “0.33 points for each year between the time the test was normed and the time the test was administered”). From his empirical research published in the *Journal of Psychoeducational Assessment* (“JPA”) special issue, Jack Fletcher measured the average rate of IQ gain for his data set at 0.28 points per year and after removing outliers, at 0.296 and 0.286. Fletcher et al., *supra* note 198, at 471–72. Fletcher reported the amounts in point gains per decade, so, to maintain consistent measurements, this Note translated the points-per-decade amount to points per year by dividing the figures by ten.

201. *See supra* Part II.B.

202. Jon Martin Sundet et al., *The End of the Flynn Effect? A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts During Half a Century*, 32 *INTELLIGENCE* 349 (2004).

Denmark.²⁰³ These Norwegian and Danish studies, however, present data irrelevant to American IQ gains; no researchers have reported such data for the American population,²⁰⁴ which serves as the relevant population for American capital defendants and inmates. Furthermore, the two Scandinavian studies used unrepresentative data sets composed of eighteen- to nineteen-year-old males and excluded females, children, and other males.²⁰⁵ Flynn's studies over the years, as well as those of others, use sound research methods and maintain a position of high regard in the field of intelligence testing.²⁰⁶ The exact magnitude of the Flynn Effect nevertheless continues to warrant more research to update and improve the existing body of knowledge.²⁰⁷ However, given the current body of knowledge, along with the irreversible implications of an incorrect *Atkins* determination, courts should use the 0.3-point-per-year gain as a basis to adjust for obsolete norms in current *Atkins* cases.

2. Variability Across Ability Levels May Support Larger Adjustment

Two studies—a 2003 study by Tomoe Kanaya, Matthew H. Scullin, and Stephen J. Ceci and a recently published study by Xiaobin Zhou, Jianjun Zhu, and Lawrence G. Weiss—suggest that the magnitude of the Flynn Effect may vary across ability groups or test scores. This variability issue raises the concern that the average rate

203. Thomas W. Teasdale & David R. Owen, *A Long-Term Rise and Recent Decline in Intelligence Test Performance: The Flynn Effect in Reverse*, 39 PERSONALITY & INDIVIDUAL DIFFERENCES 837 (2005); Thomas W. Teasdale & David R. Owen, *Secular Declines in Cognitive Test Scores: A Reversal of the Flynn Effect*, 36 INTELLIGENCE 121 (2008).

204. See Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 101 (“If someone has a case that IQ gains in America should cease, let them bring it forward. Their data should be about America and not about, say, Scandinavia . . . We do not predict temperatures in America on the basis of data from the North Pole.”).

205. See Kaufman, *supra* note 2, at 503 (“These studies of 18- to 19-year-old males did not include females, children, or men older than 20 years, yet conclusions are reached about entire countries.”).

206. See *id.* at 501 (“[W]ith the [Flynn Effect], in the United States, a multitude of studies converge around 3 points per decade and the variability around 3 points is relatively minor.” (citations omitted)); Weiss, *supra* note 158, at 491 (citing conclusions that “the field” can agree on: “Dr. Flynn is to be commended for a meaningful life’s work,” “[t]he Flynn effect (FE) is real,” and “[t]he FE has been shown to be near 3 IQ points per decade on average across a large number of studies, countries, and tests”); Zhou et al., *supra* note 52, at 399 (“Research in the past two decades has shown cumulative support for the existence of the Flynn effect (FE)—the massive IQ gain at about 0.3 points per year.” (citations omitted)).

207. See Weiss, *supra* note 158, at 492 (“[Flynn Effect] research should continue. We will know more about the [Flynn Effect] in 10 years than we know today.”).

of 0.3 points per year may not apply to people in particular ability groups.

The Kanaya study explored the Flynn Effect's impact on the IQ-test performances of school-age children, which were in or below the mental retardation borderline range.²⁰⁸ Before describing their empirical findings, Kanaya and coauthors explained that past research "suggest[ed] that the Flynn effect and changing IQ norms not only affect individuals in the average range of the IQ distribution but also individuals with [mental retardation], and that the magnitude of the Effect may be even larger among [mentally retarded] children than among children at the mean."²⁰⁹ The results from their data reflected what past studies suggested: a "far larger magnitude effect" existed among children in the 55 to 70 and 71 to 85 IQ ranges than the effect on children in the middle of the IQ distribution.²¹⁰ Furthermore, Kanaya and coauthors found that "the Flynn Effect has an impact on which individuals are diagnosed [mentally retarded] and which are not, regardless of their actual cognitive ability"—specifically, three times as many children who took a recently updated IQ test and received scores of 66 to 70 qualified for mental retardation placement than "cognitively equivalent" children who took an outdated IQ test and received scores of 71 to 75.²¹¹

Similar to the Kanaya study, the Zhou study, using two methods of analysis, found that the average rate of gain in IQ scores appeared larger at the middle and lower portions of the distribution (lower IQ scores), while the rate appeared smaller at the upper portion of the distribution (higher IQ scores).²¹²

208. Kanaya et al., *supra* note 41, at 778. The Kanaya study included data from 8,944 special education IQ assessments from nine geographically and socioeconomically diverse school districts, lending to the study's generalizability. *Id.* at 782, 789.

209. *Id.* at 780 (citing L.M. Bolen et al., *A Comparison of the Performance of Cognitively Disabled Children on the WISC-R and WISC-III*, 51 *J. CLINICAL PSYCHOL.* 89 (1995); J.R. Slate & D.A. Saarnio, *Differences Between WISC-III and WISC-R IQs: A Preliminary Investigation*, 13 *J. PSYCHOEDUC. ASSESSMENT* 340 (1995); H. Vance et al., *A Longitudinal Comparison of WISC-III and WISC-R Scores of Special Education Students*, 33 *PSYCHOL. SCHS.* 113 (1996)).

210. *Id.* at 787.

211. *Id.*

212. Zhou et al., *supra* note 52, at 408. Specifically, they found the following average IQ gains per year: for scores at or below 79, the gain was 0.31; for scores 80 to 90, the gain was 0.36; for scores 90 to 109, the gain was 0.37; for scores 110 to 119, the gain was 0.06; and for scores at or above 120, the gain was 0.15. *Id.* at 404 tbl.3. While they concluded that the magnitude of the Flynn Effect varies across ability groups, the authors acknowledged that their findings on the direction of the variability were inconclusive and did not address the potential variability for ability levels on different tests or of different age groups. *Id.* at 406, 409.

JPA special-issue contributor Kevin McGrew further criticized the methodology and results of the Zhou study. McGrew, *supra* note 160, at 462 ("Unfortunately, the confidence placed in

The Kanaya and Zhou studies present a legitimate concern regarding the issue of whether to adjust IQ scores for the Flynn Effect. The studies also serve as ammunition for both sides of the debate. Opponents can point to the general risks of adjusting IQ scores for the Flynn Effect,²¹³ namely that the average rate may not apply equally to all. Advocates for the Flynn Effect can cite the studies' findings of a higher rate of gain for lower IQ scores as evidence that the 0.3-point per-year figure is actually a conservative amount to adjust for.²¹⁴

The authors themselves appear conflicted: Stephen Ceci and Tomoe Kanaya have stated that "it is not appropriate to merely subtract 0.3 points for every year that a norm has aged until we know that everyone experiences the same gains on the same subtests at the same time."²¹⁵ However, the Kanaya study recognized that "[n]owhere are the consequences of IQ score fluctuations due to the Flynn [E]ffect more critical than in the determination of whether a death row inmate . . . can be considered mentally retarded."²¹⁶ Zhou and coauthors caution that "forcing an IQ adjustment using a fixed rate . . . could cause misleading results and potentially misclassify a proportion of examinees," but they also recognize that "early indications appear to favor slightly larger adjustments in the lower range of scores where high-stakes legal evaluations are most likely to occur."²¹⁷

If the Kanaya and Zhou findings are valid and an annual gain greater than 0.3 points exists for lower scores, *not* forcing an IQ adjustment, even one at the average 0.3 points per year, could potentially misclassify defendants as eligible for the death penalty when, in fact, they are not.²¹⁸ The issue comes down to what type of

Zhou et al.'s admittedly tentative conclusions is weak given the variety of methodological complications . . . Zhou et al.'s inconsistent and contradictory results (across and within methodologies) are difficult to explain; and I believe these results render their IQ Ability Level x FE interaction findings inconclusive.").

213. See Weiss, *supra* note 158, at 490 ("[I]t is likely that the Zhou et al. study will be used by others as evidence of the risks attending adjustments to individual scores . . .").

214. See, e.g., Reynolds et al., *supra* note 160, at 479–80 ("Taken as a whole, and noting some of the inconsistencies in the results, the Zhou et al. analyses support the idea that an even larger correction may need to be applied to low IQs, especially given the paucity of individuals scoring at the extremes in the samples evaluated . . .").

215. Ceci & Kanaya, *supra* note 2, at 446.

216. Kanaya et al., *supra* note 41, at 789 ("[S]ome borderline death row inmates or capital murder defendants who were not classified as MR in childhood because they were administered an older version of an IQ test will qualify as MR if they are administered a more recent test.").

217. Zhou et al., *supra* note 52, at 409.

218. See Weiss, *supra* note 158, at 490 ("As practice must continue, however, Flynn (2010) and Reynolds et al. (2010) may be correct that 3 points per decade is conservative in the low ability range.").

misclassification is more tolerable. Is it more tolerable to misclassify a person as mentally retarded, sparing him from the death penalty? Or is it more tolerable to misclassify a mentally retarded person as eligible for the death penalty, subjecting him to a death sentence contrary to the holding of *Atkins*?

This Note argues that the latter misclassification is intolerable given existing research on the Flynn Effect. As James Flynn explains:

[W]e have enough evidence to know that we must make a rough adjustment to avoid a great injustice. . . . The law must not lose touch with reality by yearning after the unattainable. . . . Ideally, we would know exactly how to adjust IQ scores for obsolescence But one thing we know for certain: IQ gains have not been nil.²¹⁹

In other words, the “approximately true”—recognizing that American IQ scores become inflated by an average rate of 0.3 points for every year that an IQ test grows outdated—is “preferable to the certainly false”—treating unadjusted IQ scores from outdated tests as accurate.²²⁰

3. A Call for More Research but Life and Practice Must Go On

Commentators from both sides agree that the variability of the Flynn Effect across ability levels and other classifications, like age groups and test versions, warrants more research.²²¹ During this interim, however, clinical practice cannot halt at a standstill nor wait passively for future research to perfect an inexact measure.²²² While research on the magnitude of the Flynn Effect progresses, courts should adjust IQ scores for the Flynn Effect at 0.3 points per year in the *Atkins* context because that figure represents a value substantiated by empirical research. Although approximate, it also honors the purpose of *Atkins*—avoiding the unconstitutional execution of the mentally retarded—better than the alternative of failing to adjust scores for known obsolete norms.

Given the imprecise nature of and potential for change in the Flynn Effect, this Note does not suggest a statutory mandate that “all IQ scores should be reduced by 0.3 points for every year an IQ test is

219. Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 100, 104.

220. Cunningham & Tassé, *supra* note 40, at 416.

221. See Ceci & Kanaya, *supra* note 2, at 446 (“It is clear that we still have a long way to go before we understand the exact nature and magnitude of the FE for all types of individuals, across time, and across tests.”); Fletcher et al., *supra* note 198, at 470 (“This variation is important to understand, and it is surprising that more effort has not been expended toward evaluating the precision of the correction.”).

222. See Zhou et al., *supra* note 52, at 409 (“[C]linical practice must continue even as research continues that may affect practice.”).

outdated.” Rather, a statutory definition that allows for flexible interpretation of IQ scores accommodates both the Flynn Effect’s rate of 0.3 points per year as supported by current research and the possibility of change in the Flynn Effect’s magnitude as informed by future research.

As emphasized above, the Flynn Effect’s role in *Atkins* calls for a continual effort by researchers to study the Effect’s IQ gains. The Court in *Atkins* identified the AAMR, now the AAIDD, as a guiding light in the field of mental retardation and intellectual disability.²²³ The AAIDD should also serve as a guide for future research. This Note suggests that the AAIDD convene a centralized body of certified and recognized experts in the fields of psychology and intelligence testing to counsel courts and parties on the Flynn Effect, as well as on other *Atkins* issues, through a regular flow of research and publications. A model currently exists in Division 33, a professional organization of APA-member psychologists.²²⁴ Division 33’s regularly published newsletter has included articles discussing and advocating for adjustments based on the Flynn Effect.²²⁵

E. Death Is Different: The Flynn Effect Reflects a “Best Practice” for Atkins

Since its pivotal decision in *Furman v. Georgia*, the Supreme Court has consistently recognized that “the penalty of death differs from all other forms of criminal punishment, not in degree but in kind.”²²⁶ Indeed, a death sentence is unique in its severity, finality, and irrevocability, and, thus, imposing such a sentence requires “that every safeguard is observed” to ensure reliability and accuracy.²²⁷

More than twenty years of research supports the Flynn Effect, but this Note acknowledges that the Flynn Effect’s measurement of IQ gains is neither exactly precise nor accepted by all for adjustment purposes in capital cases. Regardless, the Flynn Effect should play a

223. *Atkins v. Virginia*, 536 U.S. 304, 308 n.3 (2002).

224. DIVISION 33 AM. PSYCHOL. ASS’N, <http://www.apa.org/divisions/div33/homepage.html> (last visited Mar. 11, 2012).

225. *E.g.*, James R. Flynn, *Capital Offenders and the Death Sentence: A Scandal That Must Be Addressed*, PSYCHOL. MENTAL RETARDATION & DEVELOPMENTAL DISABILITIES: OFFICIAL PUBL’N OF DIV. 33, AM. PSYCHOL. ASS’N, Spring 2007, at 3–7; Greenspan, *supra* note 188, at 3–7.

226. *Furman v. Georgia*, 408 U.S. 238, 306 (1972) (Stewart, J., concurring); *see also* *Gregg v. Georgia*, 428 U.S. 153, 188 (1976) (“[T]he penalty of death is different in kind from any other punishment imposed under our system of criminal justice.”).

227. *Gardner v. Florida*, 430 U.S. 349, 357–58 (1977); *Gregg*, 428 U.S. at 187; *Woodson v. North Carolina*, 428 U.S. 280, 305 (1976).

role in determining mental retardation in *Atkins* cases because adjustment for the Effect comports with standard practice, and, even if it did not comport, the values underlying this country's death penalty jurisprudence require adjustment for the Flynn Effect.

The practice of adjusting for the Flynn Effect is consistent with the recommendations of the AAIDD and many distinguished experts in the field of intelligence testing,²²⁸ as well as with the recommendations of test publishers who advocate the use of updated norms.²²⁹ On the other hand, some researchers, notably Leigh D. Hagan, Eric Y. Drogin, and Thomas J. Guilmette, assert that adjusting IQ scores based on the Flynn Effect "does not comport with the standard of forensic psychological practice."²³⁰ To support their conclusion, Hagan and coauthors present empirical research, published in 2008, that surveyed program directors of APA-approved psychology programs, graduate faculty, and clinicians who were certified school psychologists.²³¹ The results of their survey showed that most of their sample did not adjust for the Flynn Effect in their practices.²³² Hagan and coauthors also found a lack of reference to the Flynn Effect and adjustments for it in IQ-test manuals, a Social Security Administration report and manual, and APA ethical and testing guidelines.²³³ Those sources, however, all recommended that clinicians and psychologists rely on up-to-date test norms and use regularly updated tests.²³⁴

Several psychologists have written in response to the argument that adjusting for the Flynn Effect does not fall within standard practice. Adjusting an IQ score for the Effect accomplishes the same goal and corrects the same inflation from obsolete norms that the recommended updating of norms for an IQ test accomplishes.²³⁵ When

228. See, e.g., *United States v. Hardy*, 762 F. Supp. 2d 849, 858–59 (E.D. La. 2010) (citing the testimony of "respected psychologists" Stephen Greenspan, also an expert in *Lewis*, and J. Gregory Olley, who both recommended applying the Flynn Effect); Fletcher et al., *supra* note 198, at 470 ("Since 2002, [the AAIDD manual for intellectual disability diagnosis] has explicitly recommended correcting IQ scores for norms obsolescence, with other researchers agreeing.").

229. See Reynolds et al., *supra* note 160, at 478 ("Because it is at this time a practical impossibility to renorm tests annually to maintain a more appropriate reference group, to the extent corrections are available and valid, they should be applied to obtained scores so the most accurate estimate of standing possible is obtained.").

230. Hagan et al., *IQ Scores Should Not Be Adjusted*, *supra* note 166, at 475.

231. Hagan et al., *supra* note 165, at 620–21.

232. *Id.*

233. *Id.* at 622–23.

234. *Id.*

235. See, e.g., Fletcher et al., *supra* note 198, at 470 ("Correcting an IQ score is not a violation of test administration. Rather, it is selecting an appropriate normative comparison. . . .

courts have before them childhood IQ scores from tests with obsolete norms at the time the defendant took the tests, as is often the case in the *Atkins* context,²³⁶ a psychologist can neither present up-to-date norms nor present a childhood IQ score from an updated test. In these situations, the only option is to adjust for the Flynn Effect, while still comporting with standard practice. Although IQ-test manuals do not mention Flynn Effect adjustments, test developer and psychologist Alan S. Kaufman and psychologist Jack Fletcher explain that test publishers have acknowledged the Flynn Effect by renorming IQ tests more frequently than before and by providing validity studies and conversion tables.²³⁷ Moreover, test publishers do not seek to address every application or use of an IQ test. For example, the WAIS-IV manual does not mention the diagnosis of mental retardation, but mental retardation determinations clearly constitute one use for the IQ test.²³⁸ Furthermore, consensus reports by experts set standards of practice, and one such set of guidelines, the most recent AAIDD manual, recommends correcting IQ scores for the obsolete norms identified by the Flynn Effect.²³⁹

Even if one thought that adjustments for the Flynn Effect did not reflect standard practice, the life and death implications distinguish *Atkins* adjustments from routine standard practice. Flynn himself clarifies that he has never advocated for altering IQs in routine clinical practice but advocates for it only in the death penalty context.²⁴⁰ Others echo this recognition of the special nature of and necessity for adjusting IQ scores for the Flynn Effect when lives are at stake.²⁴¹ Psychologist Cecil Reynolds asks, “[W]hat possible

Correcting an IQ score is a simple procedure that avoids having to change standards.” (internal citations omitted)).

236. See Weiss, *supra* note 158, at 488 (“The finding of intellectual deficiency in these cases often rests on historical records of IQ test scores such as when the felon was in school, and those tests may not have been current at the time they were administered.”).

237. Fletcher et al., *supra* note 198, at 472; Kaufman, *supra* note 2, at 502.

238. Fletcher et al., *supra* note 198, at 472.

239. AAIDD, *INTELLECTUAL DISABILITY: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORTS* 37 (11th ed. 2010).

240. James R. Flynn, *Problems with IQ Gains: The Huge Vocabulary Gap*, 28 J. PSYCHOEDUC. ASSESSMENT 412, 412 (2010) (“I have not advocated altering IQs in routine clinical practice. A psychologist whose clinical judgment wavers because of a few IQ points is incompetent. But courts that use inflated IQs to kill people are morally remiss.”).

241. See Matthew H. Scullin, *Large State-Level Fluctuations in Mental Retardation Classifications Related to Introduction of Renormed Intelligence Test*, 111 AM. J. MENTAL RETARDATION 322, 333 (2006) (“In capital murder cases . . . fairness would dictate that IQs be adjusted to compensate for IQ gains.”); Weiss, *supra* note 158, at 488–89 (recognizing that an adjustment for the Flynn Effect could make or break the decision of whether a defendant faces the death penalty).

justification could there be for issuing estimates of general intelligence in a death penalty case that are less than the most accurate estimates obtainable?”²⁴² Both Flynn and Reynolds analogize a decision to *not* adjust for the Flynn Effect to essentially a lottery of life and death that depends solely and arbitrarily on when the IQ test was normed.²⁴³

In *Furman*, as one of five Justices rejecting the death penalty schemes existing in 1972 for their arbitrary and “freakishly imposed” sentences, Justice Stewart stated that “[t]hese death sentences are cruel and unusual in the same way that being struck by lightning is cruel and unusual.”²⁴⁴ While not quite lightning, this lottery system, driven by indecision and refusal by many courts to validate and account for the Flynn Effect, nonetheless exhibits its own troubling inconsistency, unpredictability, and unfairness in its imposition of death sentences in cases of known outdated norms.

V. CONCLUSION: REQUIRING ADJUSTMENT FOR THE FLYNN EFFECT

Since the Supreme Court decided *Atkins v. Virginia* in 2002, numerous inconsistencies have arisen regarding how federal and state courts should apply the Court’s mandate against executing the mentally retarded. One such inconsistency, whether to adjust a defendant’s IQ score for the Flynn Effect, creates a dire matter of life and death that is heavily debated among experts and courts. This Note contributes to the debate by increasing legal awareness on the issue. Citing several reasons, listed below, this Note also advocates for the mandatory adjustment of IQ scores for the Flynn Effect in all *Atkins* cases that rely on outdated IQ tests.

First, empirical evidence and consensus in the field of intelligence testing show that the Flynn Effect is real and exists regardless of the debate about its cause. Second, given the relative nature of IQ scores, the Flynn Effect, as a group phenomenon, applies directly to the individual IQ scores of capital defendants and inmates. Third, legislatures and courts should reject bright-line IQ cutoffs that preclude adjustments for the Flynn Effect because these cutoffs disserve the basic premise of IQ scoring and the holding of *Atkins*. Fourth, despite the variability in the Flynn Effect’s magnitude, its average rate of gain of 0.3 points per year should still apply in current

242. Reynolds et al., *supra* note 160, at 478.

243. Flynn, *The WAIS-III and WAIS-IV*, *supra* note 53, at 100; Reynolds et al., *supra* note 160, at 480.

244. *Furman v. Georgia*, 408 U.S. 238, 309 (1972) (Stewart, J., concurring).

Atkins cases while future research expands knowledge on the exact magnitude. Finally, the Flynn Effect comports with standard clinical practice, but its life and death implications also set it apart from standard practice and require, for the accurate and constitutional determination of death penalty eligibility, adjustment for the Effect in all *Atkins* cases involving outdated IQ tests.

This Note also makes several specific recommendations. First, state legislatures should redraft or enact new provisions that allow for a flexible interpretation of IQ scores for the purpose of determining the intellectual-functioning element of mental retardation. This would make room for IQ-score adjustments based on the Flynn Effect. Courts should likewise interpret IQ scores flexibly and adjust for the Flynn Effect when faced with outdated IQ tests. State legislatures, as well as courts, should also mandate consideration of the adaptive-behavior prong alongside the intellectual-functioning prong because the imprecise nature of IQ scores counsels against treating them as dispositive of a defendant's entire *Atkins* claim. Further, this Note recognizes that, while current research supports the Flynn Effect, the phenomenon may change in the future. Thus, in adjusting for the Flynn Effect, courts must stay abreast of research in the fields of psychology and intelligence testing. A centralized committee organized by the AAIDD, recognized by the Court in *Atkins*, could serve the integral role of advising courts on this matter.

The above discussion of the Flynn Effect also implicates areas beyond the reach of *Atkins*, including the areas of special education, social security benefits, disability insurance, and military service and other occupations that consider mental retardation in hiring.²⁴⁵ Special education represents a field traditionally associated with IQ testing and has produced much of the empirical research on the Flynn Effect, including the conclusion that the Flynn Effect poses serious consequences for special education assessments.²⁴⁶ Notably, "two children in the same classroom with the same cognitive ability could be diagnosed differently simply because different test norms were used for each child."²⁴⁷

The import of the Flynn Effect in special education makes the phenomenon's place in *Atkins* capital proceedings all the more

245. Kanaya et al., *supra* note 41, at 778; Scullin, *supra* note 241, at 331.

246. See Scullin, *supra* note 241, at 331 ("[T]he enormous disparities at any one point in time among states in terms of the percentage of enrolled students receiving mental retardation services raises a number of questions with regard to whether mental retardation is adequately conceptualized.")

247. Kanaya et al., *supra* note 41.

striking. The fear of misdiagnosing a school child based on an outdated IQ test is compounded by the fear that the same inflated IQ score may wrongly subject that same child, years later, to a death sentence. Despite its lack of proffered definitions, the Court in *Atkins* unequivocally concluded under the Eighth Amendment that no mentally retarded person may be subject to death. Adjusting for the Flynn Effect reflects a practice consistent with both *Atkins* and the known world of IQ measurements. While a freakish strike of lightning is difficult to avoid, the potentially deadly and unconstitutional consequences of refusing to account for the Flynn Effect are wholly preventable. Thus, for the intelligent and just enforcement of *Atkins*, courts and juries should adjust IQ scores from outdated tests for the Flynn Effect.

*Geraldine W. Young**

* Candidate for Doctor of Jurisprudence, May 2012, Vanderbilt University Law School. For introducing me to the Flynn Effect, I thank Layne Kruse and Sumera Khan. I also thank Bradley MacLean and Mike Passino for their help and valuable insight. As always, I am grateful for the hard work and support of my friends and colleagues at the VANDERBILT LAW REVIEW. Finally, for their unwavering love, inspiration, and encouragement, I extend a special thank you to John, Brian, and my parents Jerry and Cindy.

VI. APPENDIX: STATUTORY DEFINITIONS OF MENTAL RETARDATION OR
INTELLECTUAL DISABILITY

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Alabama	"A person with significant subaverage general intellectual functioning resulting in or associated with concurrent impairments in adaptive behavior and manifested during the developmental period, as measured by appropriate standardized testing instruments." ²⁴⁸	No specific definition	No specific definition	No specific definition
Arizona	"'Intellectual disability' means a condition based on a mental deficit that involves significantly subaverage general intellectual functioning, existing concurrently with significant impairment in adaptive behavior, where the onset of the foregoing conditions occurred before the defendant reached the age of eighteen." ²⁴⁹	"[A] full scale intelligence quotient of seventy or lower. The court in determining the intelligence quotient shall take into account the margin of error for the test administered." ²⁵⁰	"[T]he effectiveness or degree to which the defendant meets the standards of personal independence and social responsibility expected of the defendant's age and cultural group." ²⁵¹	Before the age of eighteen ²⁵²
Arkansas	" '[M]ental retardation' means: (A) Significantly subaverage general intellectual functioning accompanied by a significant deficit or impairment in adaptive functioning manifest in the developmental period . . . and (B) A deficit in adaptive behavior." ²⁵³	"There is a rebuttable presumption of mental retardation when a defendant has an intelligence quotient of sixty-five (65) or below." ²⁵⁴	No specific definition	No later than the age of eighteen ²⁵⁵

248. ALA. CODE § 15-24-2(3) (LexisNexis 2011).

249. ARIZ. REV. STAT. ANN. § 13-753(K)(3) (LexisNexis 2011).

250. *Id.* § 13-753(K)(5).

251. *Id.* § 13-753(K)(1).

252. *Id.* § 13-753(K)(3).

253. ARK. CODE ANN. § 5-4-618(a)(1) (2011).

254. *Id.* § 5-4-618(a)(2).

255. *Id.* § 5-4-618(a)(1).

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
California	“‘[M]entally retarded’ means the condition of significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested before the age of 18.” ²⁵⁶	No specific definition	No specific definition	Before the age of eighteen ²⁵⁷
Colorado	“‘Mentally retarded defendant’ means any defendant with significantly subaverage general intellectual functioning existing concurrently with substantial deficits in adaptive behavior and manifested and documented during the developmental period.” ²⁵⁸	No specific definition	No specific definition	No specific definition
Connecticut	“‘[M]ental retardation means a significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period.’” ²⁵⁹	“‘[S]ignificantly subaverage’ means an intelligence quotient more than two standard deviations below the mean for the test” ²⁶⁰	“‘[A]daptive behavior’ means the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected for the individual’s age and cultural group” ²⁶¹	“‘[T]he period of time between birth and the eighteenth birthday’” ²⁶²

256. CAL. PEN. CODE. ANN. § 1376(a) (Deering 2010).

257. *Id.*

258. COLO. REV. STAT. § 18-1.3-1101(2) (2011).

259. CONN. GEN. STAT. § 1-1g(a) (2011).

260. *Id.* § 1-1g(b).

261. *Id.*

262. *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Delaware	“ ‘Seriously mentally retarded’ or ‘serious mental retardation’ means that an individual has significantly subaverage intellectual functioning that exists concurrently with substantial deficits in adaptive behavior and both the significantly subaverage intellectual functioning and the deficits in adaptive behavior were manifested before the individual became 18 years of age.” ²⁶³	“ ‘Significantly subaverage intellectual functioning’ means an intelligent quotient of 70 or below obtained by assessment with 1 or more of the standardized, individually administered general intelligence tests developed for the purpose of assessing intellectual functioning.” ²⁶⁴	“ ‘Adaptive behavior’ means the effectiveness or degree to which the individual meets the standards of personal independence expected of the individual’s age group, sociocultural background and community setting, as evidenced by significant limitations in not less than 2 . . . adaptive skill areas.” ²⁶⁵	Before the age of eighteen ²⁶⁶
Florida	“ ‘[M]ental retardation’ means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the period from conception to age 18.” ²⁶⁷	“The term ‘significantly subaverage general intellectual functioning,’ . . . means performance that is two or more standard deviations from the mean score on a standardized intelligence test specified in the rules of the Agency for Persons with Disabilities.” ²⁶⁸	“The term ‘adaptive behavior,’ . . . means the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected of his or her age, cultural group, and community.” ²⁶⁹	“[T]he period from conception to age 18.” ²⁷⁰

263. DEL. CODE ANN. tit. 11, § 4209(d)(3)(d)(2) (2011).

264. *Id.* § 4209(d)(3)(d)(3).

265. *Id.* § 4209(d)(3)(d)(1).

266. *Id.* § 4209(d)(3)(d)(2).

267. FLA. STAT. ANN. § 921.137(1) (LexisNexis 2011).

268. *Id.*

269. *Id.*

270. *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Georgia	“‘Mentally retarded’ means having significantly subaverage general intellectual functioning resulting in or associated with impairments in adaptive behavior which manifested during the developmental period.” ²⁷¹	No specific definition ²⁷²	No specific definition	No specific definition ²⁷³
Idaho	“‘Mentally retarded’ means significantly subaverage general intellectual functioning that is accompanied by significant limitations in adaptive functioning The onset of significant subaverage general intelligence functioning and significant limitations in adaptive functioning must occur before age eighteen (18) years.” ²⁷⁴	“‘Significantly subaverage general intellectual functioning’ means an intelligence quotient of seventy (70) or below” ²⁷⁵	“[S]ignificant limitations in adaptive functioning in at least two (2) of the following skill areas: communication, self-care, home living, social or interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety” ²⁷⁶	Before the age of eighteen ²⁷⁷

271. GA. CODE ANN. § 17-7-131(a)(3) (2011).

272. The Supreme Court of Georgia states that “[w]hile not by itself conclusive, the generally accepted IQ score for an indication of mental retardation is approximately 70 or below.” *Head v. Stripling*, 590 S.E.2d 122, 124 n.1 (Ga. 2003) (citing *Stripling v. State*, 401 S.E.2d 500, 504 (Ga. 1991)).

273. The Supreme Court of Georgia defines this element as “before age 18.” *Id.*

274. IDAHO CODE ANN. § 19-2515A(1)(a) (2011).

275. *Id.* § 19-2515A(1)(b).

276. *Id.* § 19-2515A(1)(a).

277. *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Indiana	“‘[I]ndividual with mental retardation’ means an individual who, before becoming twenty-two (22) years of age, manifests: (1) significantly subaverage intellectual functioning; and (2) substantial impairment of adaptive behavior; that is documented in a court ordered evaluative report.” ²⁷⁸	No specific definition	No specific definition	Before the age of twenty-two ²⁷⁹
Kansas	“‘[M]entally retarded’ means 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.” ²⁸⁰	“‘Significantly subaverage general intellectual functioning’ means performance which is two or more standard deviations from the mean score on a standardized intelligence test specified by the secretary.” ²⁸¹	“‘Adaptive behavior’ means the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected of that person’s age, cultural group and community.” ²⁸²	“[D]uring the period from birth to age 18.” ²⁸³
Kentucky	“‘A defendant with significant subaverage intellectual functioning existing concurrently with substantial deficits in adaptive behavior and manifested during the developmental period is referred to . . . as a seriously mentally retarded defendant.’” ²⁸⁴	“‘Significantly subaverage general intellectual functioning’ is defined as an intelligence quotient (I.Q.) of seventy (70) or below.” ²⁸⁵	No specific definition	No specific definition

278. IND. CODE ANN. § 35-36-9-2 (LexisNexis 2011).

279. *Id.*

280. KAN. STAT. ANN. § 21-4623(e) (2011).

281. *Id.* § 76-12b01(i).

282. *Id.* § 76-12b01(a).

283. *Id.* § 76-12b01(d).

284. KY. REV. STAT. ANN. § 532.130(2) (LexisNexis 2011).

285. *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Louisiana	“‘Mental retardation’ means a disability characterized by significant limitations in both intellectual functioning and adaptive behavior as expressed in conceptual, social, and practical adaptive skills.” ²⁸⁶	No specific definition ²⁸⁷	No specific definition ²⁸⁸	Before the age of eighteen ²⁸⁹
Maryland	“[A] defendant is ‘mentally retarded’ if: (i) the defendant had significantly below average intellectual functioning . . . and an impairment in adaptive behavior; and (ii) the mental retardation was manifested before the age of 22 years.” ²⁹⁰	“[S]ignificantly below average intellectual functioning, as shown by an intelligence quotient of 70 or below on an individually administered intelligence quotient test” ²⁹¹	No specific definition	Before the age of twenty-two ²⁹²

286. LA. CODE CRIM. PROC. ANN. art. 905.5.1(H)(1) (2011).

287. In the absence of a specific statutory definition, the Supreme Court of Louisiana has relied on the definition set forth by the AAMR, now the AAIDD, defining the intellectual-functioning criterion as “approximately two standard deviations below the mean.” *See, e.g., State v. Dunn*, 41 So. 3d 454, 461–62 (La. 2010) (citing the AAMR’s 2002 definition).

288. For this element, the Supreme Court of Louisiana has also cited the definitions set forth by the AAMR, the AAIDD, and the APA. *See, e.g., id.* at 458–59 (quoting the AAMR and APA definitions and discussing the AAIDD’s most recent definition).

289. LA. CODE CRIM. PROC. ANN. art. 905.5.1(H)(1).

290. MD. CODE ANN., CRIM. LAW § 2-202(b)(1) (LexisNexis 2011).

291. *Id.*

292. *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Mississippi	<p>“ ‘Person with an intellectual disability’ means any person (i) who has been diagnosed as having substantial limitations in present functioning, manifested before age eighteen (18), characterized by significantly subaverage intellectual functioning, existing concurrently with related limitations in two (2) or more of the . . . applicable adaptive skill areas . . . and (ii) whose recent conduct is a result of having an intellectual disability and poses a substantial likelihood of physical harm to himself or others in that there has been (A) a recent attempt or threat to physically harm himself or others, or (B) a failure and inability to provide necessary food, clothing, shelter, safety, or medical care for himself.”²⁹³</p>	No specific definition	<p>“[L]imitations in two (2) 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”²⁹⁴</p>	Before the age of eighteen ²⁹⁵
Missouri	<p>“ ‘[M]ental retardation’ or ‘mentally retarded’ refer to a condition involving substantial limitations in general functioning characterized by significantly subaverage intellectual functioning with continual extensive related deficits and limitations in two or more adaptive behaviors . . . which conditions are manifested and documented before eighteen years of age.”²⁹⁶</p>	No specific definition	<p>“[C]ontinual extensive related deficits and limitations in two or more adaptive behaviors such as communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure and work”²⁹⁷</p>	Before the age of eighteen ²⁹⁸

293. MISS. CODE ANN. § 41-21-61(f) (2011).

294. *Id.*

295. *Id.*

296. MO. REV. STAT. § 565.030(6) (2010).

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Montana	No statutory definition ²⁹⁹			
Nebraska	“[M]ental retardation means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior” ³⁰⁰	“An intelligence quotient of seventy or below on a reliably administered intelligence quotient test shall be presumptive evidence of mental retardation.” ³⁰¹	No specific definition	No specific definition
Nevada	“‘[M]entally retarded’ means significant subaverage general intellectual functioning which exists concurrently with deficits in adaptive behavior and manifested during the developmental period.” ³⁰²	No specific definition	No specific definition	No specific definition
New Hampshire	No statutory definition ³⁰³			

297. *Id.*

298. *Id.*

299. Notably, many states without statutory definitions of mental retardation or intellectual disability have very few inmates on death row and rarely execute individuals. For example, as of 2011, Montana had two inmates on its death row and had executed a total of three individuals since 1976. DEATH PENALTY INFO. CTR., FACTS ABOUT THE DEATH PENALTY (2012), available at <http://www.deathpenaltyinfo.org/documents/FactSheet.pdf>.

300. NEB. REV. STAT. § 28-105.01(3) (LexisNexis 2011).

301. *Id.*

302. NEV. REV. STAT. ANN. §§ 174.098(7), 433.174 (LexisNexis 2010).

303. *See Atkins v. Virginia*, 536 U.S. 304, 316 (2002) (“Some States, for example New Hampshire and New Jersey, continue to authorize executions, but none have been carried out in decades. Thus there is little need to pursue legislation barring the execution of the mentally retarded in those States.”).

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
North Carolina	“Mentally retarded.— Significantly subaverage general intellectual functioning, existing concurrently with significant limitations in adaptive functioning, both of which were manifested before the age of 18.” ³⁰⁴	“Significantly subaverage general intellectual functioning.—An intelligence quotient of 70 or below.” ³⁰⁵	“Significant limitations in two or more of the following adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure skills and work skills.” ³⁰⁶	Before the age of eighteen ³⁰⁷
Ohio	No statutory definition ³⁰⁸			

304. N.C. GEN. STAT. § 15A-2005(a)(1) (2011).

305. *Id.*

306. *Id.*

307. *Id.*

308. The Supreme Court of Ohio follows the AAMR and APA definitions of mental retardation, as cited in *Atkins*. *State v. Lott*, 779 N.E.2d 1011, 1014 (Ohio 2002). With regard to the intellectual-functioning requirement, the court specifically states that IQ scores “are one of the many factors that need to be considered, they alone are not sufficient to make a final determination on this issue,” and “there is a rebuttable presumption that a defendant is not mentally retarded if his or her IQ is above 70.” *Id.*

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Oklahoma	“ ‘Mental retardation’ or ‘mentally retarded’ means significantly subaverage general intellectual functioning, existing concurrently with significant limitations in adaptive functioning” ³⁰⁹	“ ‘Significantly subaverage general intellectual functioning’ means an intelligence quotient of seventy (70) or below. ³¹⁰ . . . However, in no event shall a defendant who has received an intelligence quotient of seventy-six (76) or above . . . be considered mentally retarded” ³¹¹	“ ‘Significant limitations in adaptive functioning’ means significant limitations in two or more of the following adaptive skill areas; communication, self-care, home living, social skills, community use, self-direction, health, safety, functional academics, leisure skills and work skills” ³¹²	Before the age of eighteen ³¹³
Oregon	“ ‘Mental retardation’ means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period. . . Definitions and classifications shall be consistent with the ‘Manual on Terminology and Classification in Mental Retardation’ of the American Association on Mental Deficiency.” ³¹⁴	“ ‘Significantly subaverage’ means a score on a test of intellectual functioning that is two or more standard deviations below the mean for the test.” ³¹⁵	“ ‘Adaptive behavior’ means the effectiveness or degree with which an individual meets the standards of personal independence and social responsibility expected for age and cultural group.” ³¹⁶	“[T]he period of time between birth and the 18th birthday.” ³¹⁷
Pennsylvania	No statutory definition			

309. OKLA. STAT. tit. 21, § 701.10bA1 (2011).

310. *Id.* § 701.10bA3.

311. *Id.* § 701.10bC.

312. *Id.* § 701.10bA2.

313. *Id.* § 701.10bB.

314. OR. REV. STAT. § 427.005(11) (2009) (defining mental retardation in the statutory title addressing mental health and developmental disabilities, as no such definition appears in the criminal law-related titles; however, this statutory section is referenced in OR. REV. STAT. § 132.090).

315. *Id.* § 427.005(16).

316. *Id.* § 427.005(1).

317. *Id.* § 427.005(4).

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
South Carolina	“ ‘Mental retardation’ means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period.” ³¹⁸	No specific definition	No specific definition	No specific definition
South Dakota	“[M]ental retardation means significant subaverage general intellectual functioning existing concurrently with substantial related deficits in applicable adaptive skill areas.” ³¹⁹	“An intelligence quotient exceeding seventy on a reliable standardized measure of intelligence is presumptive evidence that the defendant does not have significant subaverage general intellectual functioning.” ³²⁰	No specific definition	Before the age of eighteen ³²¹
Tennessee	“ ‘[I]ntellectual disability’ means: (1) Significantly subaverage general intellectual functioning . . . ; (2) Deficits in adaptive behavior; and (3) The intellectual disability must have been manifested during the developmental period” ³²²	“Significantly subaverage general intellectual functioning as evidenced by a functional intelligence quotient (I.Q.) of seventy (70) or below” ³²³	No specific definition	“[B]y eighteen (18) years of age” ³²⁴
Texas	No statutory definition ³²⁵			

318. S.C. CODE ANN. § 16-3-20(C)(b)(10) (2010).

319. S.D. CODIFIED LAWS § 23A-27A-26.2 (2011).

320. *Id.*

321. *Id.* § 23A-27A-26.1.

322. TENN. CODE ANN. § 39-13-203(a) (2011).

323. *Id.* § 39-13-203(a)(1).

324. *Id.* § 39-13-203(a)(3).

325. The Texas Court of Criminal Appeals established the relevant standard for Texas in *Ex parte Briseno*, 135 S.W.3d 1, 7 (Tex. Crim. App. 2004). *Briseno* held that mental retardation claims should be adjudicated under the framework of the AAMR (now AAIDD), in conjunction with the standard supplied by the Texas Persons with Mental Retardation Act: “ ‘Mental retardation’ means significantly subaverage general intellectual functioning that is concurrent with deficits in adaptive behavior and originates during the developmental period.” TEX. HEALTH & SAFETY CODE ANN. § 591.003(13) (West 2011). “ ‘Subaverage general intellectual functioning’ refers to measured intelligence on standardized psychometric instruments of two or more standard deviations below the age-group mean for the tests used.” *Id.* § 591.003(20). “ ‘Adaptive

State	Definition of Mental Retardation or Intellectual Disability	Intellectual-Functioning Prong	Adaptive-Behavior Prong	Developmental-Period Prong
Utah	“[A] defendant is ‘mentally retarded’ if: (1) the defendant has significant subaverage general intellectual functioning that results in and exists concurrently with significant deficiencies in adaptive functioning” ³²⁶	No specific definition	“[S]ignificant deficiencies in adaptive functioning that exist primarily in the areas of reasoning or impulse control, or in both of these areas” ³²⁷	Before the age of twenty-two ³²⁸
Virginia	“ ‘Mentally retarded’ means a disability, originating before the age of 18 years, characterized concurrently by (i) significantly subaverage intellectual functioning . . . and (ii) significant limitations in adaptive behavior” ³²⁹	“[S]ignificantly subaverage intellectual functioning . . . that is at least two standard deviations below the mean” ³³⁰	“[S]ignificant limitations in adaptive behavior as expressed in conceptual, social and practical adaptive skills” ³³¹	Before the age of eighteen ³³²
Washington	“ ‘Intellectual disability’ means the individual has: (i) Significantly subaverage general intellectual functioning; (ii) existing concurrently with deficits in adaptive behavior; and (iii) both significantly subaverage general intellectual functioning and deficits in adaptive behavior were manifested during the developmental period.” ³³³	“ ‘Significantly subaverage general intellectual functioning’ means intelligence quotient seventy or below.” ³³⁴	“ ‘Adaptive behavior’ means the effectiveness or degree with which individuals meet the standards of personal independence and social responsibility expected for his or her age.” ³³⁵	“[T]he period of time between conception and the eighteenth birthday.” ³³⁶
Wyoming	No statutory definition			

behavior’ means the effectiveness with or degree to which a person meets the standards of personal independence and social responsibility expected of the person’s age and cultural group.” *Id.* § 591.003(1).

326. UTAH CODE ANN. § 77-15a-102 (LexisNexis 2012).

327. *Id.*

328. *Id.*

329. VA. CODE ANN. § 19.2-264.3:1.1A (2011).

330. *Id.*

331. *Id.*

332. *Id.*

333. WASH. REV. CODE § 10.95.030(2)(a) (2011).

334. *Id.* § 10.95.030(2)(c).

335. *Id.* § 10.95.030(2)(d).

336. *Id.* § 10.95.030(2)(e).