No. 24-872

In the Supreme Court of the United States

JOHN Q. HAMM, COMMISSIONER,
ALABAMA DEPARTMENT OF CORRECTIONS,
Petitioner,

v.

JOSEPH CLIFTON SMITH,
Respondent.

ON WRIT OF CERTIORARI
TO THE UNITED STATES COURT OF APPEALS
FOR THE ELEVENTH CIRCUIT

BRIEF FOR AMICI CURIAE AMERICAN PSYCHOLOGICAL ASSOCIATION, AMERICAN PSYCHIATRIC ASSOCIATION, AND THE ALABAMA PSYCHOLOGICAL ASSOCIATION IN SUPPORT OF RESPONDENT

AARON M. PANNER
BRENNA L. DARLING
KELLOGG, HANSEN,
TODD, FIGEL &
FREDERICK, P.L.L.C.
1615 M Street, NW
Suite 400
Washington, DC 20036

Counsel for Amicus Curiae American Psychiatric Association

MELISSA ARBUS SHERRY
Counsel of Record
PETER A. PRINDIVILLE
LATHAM & WATKINS LLP
555 11th Street, NW
Suite 1000
Washington, DC 20004
(202) 637-3386
melissa.sherry@lw.com

Counsel for Amicus Curiae American Psychological Association

(additional counsel on inside cover)

DEANNE M. OTTAVIANO
General Counsel
AMERICAN
PSYCHOLOGICAL
ASSOCIATION
750 First Street, NE
Washington, DC 20002

Counsel for Amicus Curiae
American Psychological
Association

TABLE OF CONTENTS

Page

TABLE OF CONTENTS—Continued

Page

TABLE OF AUTHORITIES

Page(s)

CASES

OTHER AUTHORITIES

TABLE OF AUTHORITIES—Continued

Page(s)

TABLE OF AUTHORITIES—Continued

Page(s)

TABLE OF AUTHORITIES—Continued

Page(s)

TABLE OF AUTHORITIES—Continued

Page(s)

TABLE OF AUTHORITIES—Continued

Page(s)

INTERESTS OF AMICI CURIAE[1]

The American Psychological Association (APA) is a nonprofit scientific and educational organization dedicated to increasing and disseminating psychological knowledge. With more than 175,000 members and affiliates, APA is the largest organization of psychologists in the world.

Psychology is a diverse discipline grounded in the rigorous application of scientific method to the study of the mind and human behavior. Some psychologists are researchers, developing and testing theories through observation, experimentation, and analysis. Others are practitioners, working in schools, on university campuses, within the judicial system, in corporations, and in private practice to serve patients.

Based on this broad spectrum of research, practice, and application, APA and its members are well-positioned to provide scientific insights regarding many important social phenomena. Among these is the diagnosis of intellectual disability.

Since its founding in 1892, public engagement has been a key element of APA's mission. APA's Bylaws provide that an object of the Association is "to advance . . . the application of research findings to the promotion of health, education and the public

---

[1] No counsel for any party authored this brief in whole or in part; no such counsel or any party made a monetary contribution intended to fund the preparation or submission of this brief. No person or entity, other than amici and their counsel, made a monetary contribution intended to fund the preparation or submission of this brief.

welfare."[2] As part of that mission, APA has submitted over 200 briefs as amicus curiae in this Court and other federal and state courts. This Court has frequently cited APA's amicus briefs in capital cases regarding intellectual disability and the scope of Eighth Amendment protections. *See, e.g.*, *Atkins v. Virginia*, 536 U.S. 304, 316 n.21 (2002); *Graham v. Florida*, 560 U.S. 48, 68 (2010); *Miller v. Alabama*, 567 U.S. 460, 472 n.5 (2012); *Hall v. Florida*, 572 U.S. 701, 710 (2014); *Moore v. Texas*, 581 U.S. 1, 17 (2017).

The question presented is directly trained on matters of scientific study and clinical practice on which APA and its members have considerable expertise. APA has a rigorous approval process for amicus briefs, the touchstone of which is an assessment of whether there is sufficient scientific research, data, and literature on a question such that the APA can usefully contribute to the court's understanding and resolution of that question. The question presented readily satisfies that standard.

The American Psychiatric Association, with more than 39,000 members, is the Nation's leading organization of physicians who specialize in psychiatry. Its member physicians work to ensure evidence-based treatment and access to care for all persons with mental disorders, including intellectual disabilities. Association members engage in treatment, research, and forensic work, and many of them regularly perform roles in the criminal justice system. The American Psychiatric Association and its members have substantial knowledge and

---

[2] *See Bylaws of the American Psychological Association - Article I: Objects* (2008), http://www.apa.org/about/governance/bylaws/article-1.

experience relevant to the issues in this case. In 2022, the American Psychiatric Association published the Fifth Edition, Text Revision of its *Diagnostic and Statistical Manual of Mental Disorders* ("DSM-5-TR"). DSM-5-TR provides a revised definition for intellectual disability (intellectual developmental disorder) based on expert consensus, review of the scientific literature, and contributions from other professional societies.

The Alabama Psychological Association was incorporated in 1981 for the purpose of advancing the science of psychology, its professional practice, and as a means of promoting welfare. The association strives to maintain high standards in professional psychology and to make available to the public information regarding psychology as a science and as a profession.

INTRODUCTION AND SUMMARY OF ARGUMENT[3]

This Court granted certiorari to decide whether courts should consider multiple IQ test scores when assessing an *Atkins* claim and, if so, how to undertake that assessment. There is broad scientific and professional agreement on these issues. This agreement, and the rigorous scientific study that informs it, are central to the Court's resolution of this case.

In *Atkins v. Virginia*, this Court held that the Eighth Amendment prohibits the execution of individuals with intellectual disability. 536 U.S. 304, 321 (2002). *Atkins* relied on clinical definitions

---

[3] Amici thank Dr. Kevin McGrew, Dr. Cecil Reynolds, and Dr. Joel Schneider, for their assistance in drafting this brief.

promulgated by mental health professionals to identify intellectual disability. *Id.* at 308 n.3, 318. In *Hall v. Florida*, 572 U.S. 701, 719-20, 722-23 (2014), and *Moore v. Texas*, 581 U.S. 1, 20-21 (2017), this Court reiterated that clinical and scientific practice guide the determination of whether an individual is intellectually disabled.

There is consensus among the mental health professions that accurate diagnosis of intellectual disability requires a comprehensive clinical assessment of three diagnostic criteria: general intellectual functioning, adaptive functioning, and onset during the developmental period. This inquiry is not disjunctive nor is it step-by-step. The three criteria are inherently interwoven. And the assessment of these criteria requires clinicians to exercise clinical judgment and undertake a holistic and comprehensive analysis of all relevant data.

Because the diagnosis of intellectual functioning is necessarily holistic, no single criterion or datum is dispositive. Intellectual disability diagnoses based solely on IQ test scores are faulty and invalid. But IQ test scores remain relevant; IQ tests are a scientifically valid means to ascertain estimates of an individual's intellectual ability. The key is to understand both the value of IQ tests and their limits.

IQ test scores are estimates, not certain or absolute measures of intelligence. But these estimates provide important data to clinicians assessing intellectual disability. There are multiple ways to measure IQ, and different IQ tests measure different aspects of human intelligence. These differences mean that IQ test scores are not fungible or interchangeable. While IQ test scores may appear simple on their face (i.e., appearing as a basic

number), they are complex psychometric calculations that must be interpreted and aggregated according to established scientific practice and in light of tests' inherent range of error. And even then, IQ test scores cannot stand alone; they must be considered alongside other data that inform a clinician's assessment both of intellectual functioning and the other diagnostic criteria.

These fundamental scientific understandings directly inform the question presented. Experts testifying in *Atkins* cases should consider multiple IQ test scores in their holistic diagnostic assessment. Because IQ test scores are estimates of intelligence, obtaining multiple estimates from different IQ tests allows clinicians to assess an individual's intellectual functioning with greater accuracy.

There is likewise broad scientific agreement on how (and how not) to consider multiple IQ test scores. There are several scientifically accepted methods to analyze scores from multiple IQ tests—each of which must be undertaken by an expert exercising clinical judgement. But there are some methods all clinicians agree are scientifically unsound—including averaging multiple IQ test scores or selecting the highest observed score. And while obtaining multiple IQ test scores can reduce error (called the Standard Error of Measurement, or SEM), no scientifically sound method comes anywhere close to eliminating it entirely. Experts analyzing multiple IQ test scores therefore must account for latent error when determining the best estimate of an individual's intellectual capacity in light of multiple scores.

This scientific and professional consensus should inform the resolution of *Atkins* claims, but it is not unique to or tailored for the death penalty context.

"[T]he definition of intellectual disability by skilled professionals has implications far beyond the confines of the death penalty: for it is relevant to education, access to social programs, and medical treatment plans," among other areas. *Hall*, 572 U.S. at 710. Because the same diagnostic criteria are applied broadly in hundreds of thousands of intellectual disability assessments annually, the accuracy and validity of these criteria are of the utmost importance. Scientists and medical professionals develop and test intellectual disability definitions and diagnostic criteria—including those found in the *Diagnostic and Statistical Manual of Mental Disorders* (the DSM)—through a robust process grounded in the scientific method and informed by clinical experience.

ARGUMENT

I. Intellectual Disability Diagnoses Require Clinical Judgment And Holistic Assessment Of All Relevant Data—Not Just IQ Test Scores.

"Intellectual disability is a condition, not a number." *Hall v. Florida*, 572 U.S. 701, 723 (2014). As with other medical conditions, clinical professionals diagnose intellectual disability after undertaking a holistic assessment of all relevant data and applying their clinical judgment. This is because intellectual disability is multifaceted. IQ test scores, while relevant to the diagnostic inquiry, are not dispositive. And a clinician's reliance on only IQ test scores without assessment of other relevant data renders faulty and invalid diagnoses.

As this Court has noted, and as the unanimous consensus among the mental health professions confirms, an accurate diagnosis of intellectual

disability requires assessment of three criteria: general intellectual functioning, adaptive functioning, and onset during the developmental period. *See Atkins v. Virginia*, 536 U.S. 304, 308 n.3, 318 (2002); *Hall*, 572 U.S. at 710-11; *Moore v. Texas*, 581 U.S. 1, 7 (2017). These criteria are not disjunctive. All three criteria are inherently interrelated and are assessed together. As this Court has recognized, "[i]t is not sound to view a single factor as dispositive of a conjunctive and interrelated assessment" like the diagnosis of intellectual disability. *Hall*, 572 U.S. at 723. Mental health professionals must therefore engage in a holistic assessment of all data and apply their clinical judgment rooted in a high level of clinical expertise and experience.[4]

Because the diagnostic inquiry is necessarily holistic and requires the exercise of clinical judgment, no single datum—such as IQ test scores—is dispositive of intellectual functioning. This is especially true in borderline cases. As this Court has noted, evidence of adaptive functioning deficits "can be probative of intellectual disability, including for

---

[4] Robert L. Schalock, Ruth Lukasson & Marc J. Tasse, Am. Ass'n on Intellectual & Developmental Disabilities, *Intellectual Disability: Definition, Classification, and Systems of Support* 7 (12th ed. 2021) ("AAIDD Manual") (clinical judgment "emerges from the clinician's training and experience, specific knowledge of the person and their contexts, analysis of extensive data, and the use of critical thinking skills"); *see also* Robert L. Schalock & Ruth Luckasson, *Clinical Judgment* 1 (Am. Ass'n Mental Retardation 2005) (clinical judgment is "characterized by its being systematic (i.e., organized, sequential, and logical), formal (i.e., explicit and reasoned), and transparent (i.e., apparent and communicated clearly)").

individuals who have an IQ test score *above 70*." *Id.* at 712 (emphasis added). If a clinician focuses solely on IQ test scores, their diagnosis will be flawed, risking both false positives and false negatives.[5]

II. IQ Tests Provide Important Though Not Dispositive Data

While clinical assessment cannot be limited to one diagnostic criterion or datum, IQ test scores are still relevant because they provide important data regarding one diagnostic criterion. The Intellectual Quotient (IQ) is a numerical estimate of intellectual functioning determined on an IQ test. There are many different ways to assess IQ, and different tests assess different elements of intelligence. While not an absolute or certain measurement of intelligence, IQ test scores provide useful data for clinicians diagnosing intellectual disability. Clinicians must consider IQ test scores in the context of professionally and scientifically accepted standards and norms, accounting for error inherent in the scores, and aggregating the scores according to accepted psychometric principles.

A. IQ Is An Estimate, Not An Absolute Or Certain Measurement, Of Intellectual Functioning

As this Court has consistently noted, scientists agree that intellectual disability is characterized, in part, by significant limitations in general intellectual

---

[5] *See* DSM-5-TR at 38 ("The diagnosis of intellectual developmental disorder is based on both clinical assessment and standardized testing of intellectual functions, standardized neuropsychological tests, and standardized tests of adaptive functioning.").

functioning. *See Atkins*, 536 U.S. at 308 n.3, 318; *Hall*, 572 U.S. at 710-11; *Moore*, 581 U.S. at 7. General intelligence is a theoretical construct. But through robust study and research, scientists have developed methods to estimate general intellectual functioning by observing measurable cognitive abilities that correlate with, and are thus proxies for, intelligence. These measurements are scientifically tested and verified. But they are, at bottom, estimates. Because it is not possible to observe intelligence directly, it is not possible to measure intelligence with absolute, certain precision.

IQ is the fruit of a multi-century scientific enterprise to ascertain the best possible estimate of intelligence. IQ is a numerical, statistical estimate of an individual's maximal general intellectual performance. But an individual does not have a single, fixed or immutable IQ. And while IQ is generally stable, some variation is expected and normal over time as individuals interact with their lived environment. An individual therefore does not "have an IQ" that can be conclusively determined. Rather, practitioners assess intellectual capacity through an IQ test, which yields a single estimate of IQ—one datum on which clinical professionals rely to ascertain the best estimate of an individual's intellectual capacity.

B. IQ Is Assessed In Different Ways

There are multiple ways to measure IQ and multiple professionally accepted tests. Because human intelligence is a remarkably complex phenomenon, IQ tests examine different proxies for intellectual capacity. And, as a result, different IQ

tests emphasize different aspects of human intelligence.

There is scientific consensus that "intelligence consists of a large number of diverse but interrelated narrow abilities" that can be clinically observed and empirically assessed as proxies for general intelligence.[6] These abilities are grouped into eight broad domains: fluid reasoning, short-term working memory, learning efficiency, visual-spatial processing, auditory processing, comprehension-knowledge, retrieval fluency, and processing speed.[7] IQ tests assess capacity in these broad domains by observing an individual's performance on tasks related to narrow abilities. There are about 90 narrow-ability tasks that have been scientifically identified and empirically validated.[8] For example, IQ tests can assess "knowledge of vocabulary, ability to rotate mental images quickly, ability to understand speech in a noisy room, [or the] speed at which familiar objects can be named."[9] Other tests assess an individuals' ability to recall stories, recognize pictures, or complete object-number sequences.[10] Others still ask individuals to group objects together

---

[6] Randy G. Floyd, Ryan L. Farmer, W. Joel Schneider & Kevin S. McGrew, *Theories and Measurement of Intelligence*, *in APA Handbook of Intellectual and Developmental Disabilities* 385, 387 (L.M. Glidden ed., 2021).

[7] *Id.* at 387-88.

[8] *Id.* at 387.

[9] *Id.*

[10] Cecil R. Reynolds, Robert A. Altmann & Daniel N. Allen, *Mastering Modern Psychological Testing: Theory and Methods* 363 (2d ed. 2021).

according to similarities or reproduce geometric patterns.[11]

While all IQ tests aim to measure general intelligence, they do not all assess the same domains of intelligence. Nor do IQ tests assess intelligence with the same depth. Some tests are brief and narrow in scope, assessing performance on a limited number of abilities in only a few domains. Such tests have limited utility in making high-stakes determinations regarding intellectual capacity.[12] Other IQ tests are comprehensive, meaning they have, at minimum, six subtests in at least three broad domains.[13]

There is scientific consensus that intellectual functioning is best assessed on an individually administered, comprehensive, standardized IQ test with a contemporary and nationally representative norm or reference sample—such as the Wechsler Adult Intelligence Scale and the Stanford-Binet Intelligence Scale, which were used in this case.[14] Although most comprehensive IQ tests measure

---

[11] *Id.* at 353-54.

[12] *See* DSM-5-TR at 38 ("Invalid scores may result from the use of brief intelligence screening tests or group tests.")

[13] Floyd et al., *Theories and Measurement of Intelligence*, *supra*, at 400-01.

[14] *See* DSM-5-TR at 38 ("Intellectual functioning is typically measured with individually administered and psychometrically valid, comprehensive, culturally appropriate, psychometrically sound tests of intelligence.").

intelligence well, variations in depth and scope mean that "some do it more precisely than others."[15]

C. IQ Test Scores Have Ranges Of Error

An IQ test score is an "approximation, not a final and infallible assessment of intellectual functioning." *Hall*, 572 U.S. at 722. [16] This is because IQ test scores are subject to variability that is external to the abilities of the test taker. This variation reduces the reliability of the scores produced by a testing instrument because it reduces the confidence a clinician has that the score accurately reflects the test taker's abilities.

Scientists have determined a mechanism to estimate the variability in results that would be present if it were practical to test an individual 100, 200, or more times.[17] This mechanism relies on the use of standardized testing procedures and the development of methods to calculate the precision, consistency, and repeatability of the test score—i.e., a test's reliability. Once an IQ test's reliability is calculated based on group data, a simple mathematical calculation produces an estimate of the

---

[15] W. Joel Schneider & Dawn P. Flanagan, *The Relationship Between Theories of Intelligence and Intelligence Tests*, *in Handbook of Intelligence: Evolutionary Theory, Historical Perspective, and Current Concepts*, at 317, 331 (S. Goldstein et al. eds., 2015).

[16] Floyd, et al., *Theories and Measurement of Intelligence*, *supra*, at 411 (IQ test scores "are fallible measures" that "do not produce perfectly precise scores.").

[17] *See, e.g.,* Ross E. Traub & Glenn L. Rowley, *Understanding Reliability*, 10 Educ. Measurement: Issues & Prac. 171, 173-75 (1991).

typical amount of measurement error (i.e., the within-person variability of the observed IQ test scores expected across all persons taking the test). This estimate is called the standard error of measurement, or the SEM. Each IQ test has its own SEM measurements. The SEM allows clinicians to determine the confidence interval of an individual's IQ test score. A confidence interval is a range of test scores that is likely to contain the score an individual might have achieved if a test had perfect reliability.

Because IQ tests have inherent error, an IQ test score is best understood not as a precise score but as a range of confidence with parameters of at least one standard error of measurement. *See Hall*, 572 U.S. at 713 ("For purposes of most IQ tests, the SEM means that an individual's score is best understood as a range of scores on either side of the recorded score."). A confidence interval spanning ±1 SEM equates to a confidence of about 68% that the measured score falls within a given score range. A 95% confidence interval spans ±1.96 SEMs. "There is professional consensus that, in the context of important diagnostic decisions, a 95% confidence interval should be used," as this yields "a reasonably conservative estimate of the amount of potential error in an individual IQ" test score.[18] Across most IQ tests, "the 95% confidence interval is approximately ± 5 IQ points, for a confidence interval spanning 10 points."[19] So, for

---

[18] Floyd et al., *Theories and Measurement of Intelligence*, *supra*, at 411; *see also* AAIDD Manual at 119 ("Best practices recommend interpreting all obtained standard scores using the 95% confidence interval (i.e., obtained score ±2 times the SEM")).

[19] Floyd et al., *Theories and Measurement of Intelligence*, *supra*, at 411.

example, if an individual has an IQ test score of 74, a 95% confidence interval centered on the observed score would range from 69 and 79.

D. IQ Test Scores Are Complicated Psychometric Calculations And Must Be Analyzed And Aggregated According To Established Scientific Standards

Because IQ test scores are indexed to the same, standardized scale—where a score of 100 is the mean, and the standard deviation is 15 points—it may appear that scores from different IQ tests can be easily equated or aggregated. For example, one might think that scores can be compared side-by-side, just like a law school admissions officer would compare LSAT scores from different students. So too it might seem that scores from different IQ tests can be aggregated with simple arithmetic, for example by calculating an average. Such thoughts, while understandable, are fundamentally mistaken.

IQ test scores are not fungible or interchangeable.[20] As explained, different IQ tests measure different aspects of intelligence and thus "represent fundamentally different latent intelligence factors."[21] As a result, it is both common and expected for IQ test scores to vary across different test

---

[20] *See* Kevin S. McGrew, *Intellectual Functioning*, *in The Death Penalty and Intellectual Disability* 85, 85-111 (Ed Polloway ed., 2015); Dale G. Watson, *Intelligence Testing*, *in The Death Penalty and Intellectual Disability* 113, 113-40 (Ed Polloway ed. 2015).

[21] Katarzyna Uzieblo, Jan Winter, Johan Vanderfaeillie, Gina Rossi & Walter Magez, *Intelligence Diagnosing of Intellectual Disabilities in Offenders: Food for Thought*, 30 Behav. Sci. Law. 28, 34 (2012).

administrations. In one study, researchers found that there was up to a 10-point variation in IQ test scores between different tests for 25% of the population.[22] And score variations can be even more extreme between brief, limited-scope tests and comprehensive tests. Another study, examining differences in IQ test scores in a sample group, found a remarkable difference of *50-63 points* between administrations of the Kaufman Brief Intelligence Test and Wechsler Intelligence Scale for Children–Third Edition in some individuals.[23]

IQ test scores also do not statistically behave like other measurements. Unlike height or weight, IQ test scores are not susceptible to easy manipulation with simple arithmetic. As complex psychometric calculations, clinicians must apply scientifically accepted methods to aggregate scores.

III. There Is Scientific Consensus On Whether, How, And—Perhaps Most Importantly—How Not To Assess Multiple IQ Test Scores.

Clinicians should consider multiple IQ test scores when diagnosing intellectual disability because multiple test scores offer nuance and allow for greater accuracy in the assessment of the intellectual functioning diagnostic criterion. But the analysis of

---

[22] Randy G. Floyd, M.H. Clark & William R. Shadish, *The Exchangeability of IQs: Implications for Professional Psychology*, *in* 39 *Professional Psychology: Research and Practice* 414, 414-23 (2008).

[23] Anthony Thompson, Janet Browne, Fred Schmidt & Marian Boer, *Validity of the Kaufman Brief Intelligence Test and a Four-Subtest Form With Adolescent Offenders*, *in* 4 *Assessment* 385, 390-93 (1997).

multiple test scores is a complicated endeavor that must be consistent with accepted scientific practice and grounded in the exercise of clinical judgment. And not all methods of aggregation are scientifically sound.

A. Experts Should Consider Multiple IQ Test Scores In Assessing An *Atkins* Claim

There is scientific consensus—and perhaps now even consensus among the parties in this case—that mental health professionals should consider multiple IQ test scores holistically in their assessment of the intellectual functioning criterion in high-stakes intellectual disability diagnoses, such as in the context of an *Atkins* claim. *See* Pet. Br. 23; U.S. Amicus Br. 18; Resp. Br. 20-21, 25.

This consensus is grounded in common sense. Because each IQ test score is an estimate, multiple IQ test scores provide additional data on which clinical professionals can rely. And because different IQ tests measure different domains of intelligence in different ways, these additional data provide nuance and depth that one score lacks, providing a more complete and holistic picture of an individual's intellectual capacity. Multiple IQ test scores therefore permit professionals to make diagnostic determinations with more accuracy and confidence, leading to more accurate diagnoses.

Obtaining and assessing multiple IQ test scores is particularly important in borderline cases where statistical accuracy is critical. One test can be sufficient to determine whether an individual is within the average distribution (i.e., near the population's statistical mean). But determining where, precisely, an individual falls on the upper and

lower ends of the distribution requires more focused assessment. Relying on a single test can yield inaccurate results when an individual's IQ is near either end of the distribution. In borderline cases, more tests are almost always better because more data allows for a more refined, accurate assessment. Obtaining multiple IQ estimates also addresses error and bias in particular test administrations. By obtaining multiple IQ test scores, experts are better able to observe convergence and consistency among different scores, important indicia on which experts rely to estimate intelligence with greater accuracy.[24]

In the end, though, even when IQ test scores are multiple or aggregated, IQ test scores remain just one datum in a multi-faceted, holistic inquiry that is guided by the exercise of professional judgment.

B. Experts Should Apply Established Scientific Methods To Obtain The Best Estimate Of IQ Across Multiple IQ Test Scores

"[T]he analysis of multiple IQ scores jointly is a complicated endeavor." *Hall*, 572 U.S. at 714. But there are scientifically accepted methods to aggregate multiple IQ test scores. There are also methods that scientists agree are unsound. There is likewise consensus that the aggregation of multiple scores is not rote or formulaic; it requires professional and clinical judgment. And critically, all agree IQ test scores only have value as part of a holistic analysis that considers adaptive functioning and other factors.

1. In *Hall*, this Court cited favorably a chapter by Professor Joel Schneider that proposes one such

---

[24] *See* Watson, *Intelligence Testing*, *supra*, at 123-24.

method. *See id.* (citing W. Joel Schneider, *Principles of Assessment of Aptitude and Achievement*, *in The Oxford Handbook of Psychological Assessment of Children and Adolescents* 286, 289-91, 318 (Donald H. Saklofske, Cecil R. Reynolds & Vicki L. Schwean eds., 2013)). The Solicitor General likewise cites Professor Schneider's chapter favorably in its amicus brief. *See* U.S. Amicus Br. 19. Professor Schneider's approach, which has been further developed in subsequent publications, applies accepted psychometric principles that are used to calculate single-test IQ test scores from individual subtests to aggregate scores from different test administrations. Key to this assessment is the statistical application of test intercorrelations and calculation of a revised confidence interval (i.e., SEM) for the composite score.[25]

2. Clinical judgment must be exercised within the bounds of accepted scientific consensus. And some approaches to assessing multiple scores are statistically or scientifically invalid.

Averaging scores is an invalid approach. Averaging IQ test scores produces statistical error that renders the average "incorrect and biased."[26] This is because an IQ test score is indexed, meaning the score is produced by comparing and ranking an individual's performance against a normative sample that represents the entire population's performance

---

[25] *See* Schneider, *Principles of Assessment of Aptitude and Achievement*, *supra*, at 287, 289-291; *see also* Floyd, et al., *Theories and Measurement of Intelligence*, *supra*, at 416.

[26] Floyd et al., *Theories and Measurement of Intelligence*, *supra*, at 414-15.

on the same test. Averaging such comparative rankings always yields statistical error, both in the psychometric context and elsewhere.

Consider, for example, an Olympic decathlete who places second in every event. The athlete's average ranking is second. But, in the final scoring, the athlete will most likely rank first and win the gold medal. So too with a competitor who places second-to-last in every event. They most likely will rank *last* in the decathlon—not second-to-last. Simply averaging rankings across each event yields error because it takes no account of *other athletes'* differing rankings; averaging assumes that each athlete ranked the same in every event, which is highly unlikely.

The same is true for IQ test scores. An average of multiple IQ test scores "will be misleadingly high if most of the scores are below average and misleadingly low if most of the scores are above average."[27] This is because averaging IQ test scores yields an averaged score with a smaller standard deviation than a single score—i.e., the standard deviation is no longer 15.[28] Like Olympic decathlon medals, IQ test scores are meaningful only when they properly report an individual's relative ranking against the group, whether competitor athletes or an IQ test's normed sample population. In statistics, a normed standard deviation defines the relative ranking. Because an average of multiple IQ test scores has a different

---

[27] *Id.* at 415.

[28] *See* Schneider, *Principles of Assessment of Aptitude and Achievement*, *supra*, at 290.

standard deviation, the average loses its descriptive value and thus is inaccurate.[29]

Selecting the highest or lowest observed score is also unsound. Contrary to petitioner's suggestion (at 41-42), selecting the highest observed score is also unsound—as is selecting the lowest.[30] IQ test scores are estimates. And due to differences in test devices and the chance that some test questions align better or worse with an individuals' knowledge base, variation in observed scores is common and expected. That a single observed score is higher or lower than others does not mean it is better or more accurate. The opposite could very well be true. Clinicians look to *consistency* among observed scores to develop a more nuanced understanding of intellectual functioning. Where a score is inconsistent with other scores (i.e., it is an outlier), an expert exercising clinical judgment could reasonably determine that the outlying score—whether high or low—is not a reliable indicator of an individual's estimated intellectual functioning. While grounded in scientific study, such a conclusion comports with common sense. Because each IQ test score is an estimate, consistency among multiple estimates provides compelling evidence that the best estimate lies within the convergent range.

Eliminating the SEM is invalid too. In addition to relying on accepted methodologies to

---

[29] *See id.* at 318.

[30] *See* Watson, *Intelligence Testing*, *supra*, at 124 ("[I]t is inappropriate to simply accept, in a rote fashion, the higher score in the false belief that one can never score higher than their true IQ but can always score poorer in the face of limited effort.").

consider cumulative IQ test scores, clinicians must account for error. While multiple IQ tests can reduce error, they cannot eliminate it. This is a critical point on which there is often misunderstanding. Petitioner's brief, and the Solicitor General's amicus brief, cite several articles for the proposition that if there are multiple IQ test scores the standard error of measurement (SEM) is effectively irrelevant. *See, e.g.*, Pet. Br. 40-41; U.S. Amicus Br. 18-19. That is wrong. None of the accepted methodologies for aggregating multiple IQ test scores renders the SEM inapplicable. At most they *reduce* it, providing a better estimate of intelligence.[31] But a better estimate does not create scientific or statistical certainty. Estimates are still estimates While more precise, aggregate scores remain an "approximation, [and] not a final and infallible assessment of intellectual functioning." *Hall*, 572 U.S. at 722.

3. Aggregating multiple IQ test scores to determine the best estimate of an individual's intellectual functioning is not a rote or formulaic enterprise. There are many important assessments that require the exercise of clinical judgment. As such, the aggregation and consideration of multiple IQ test scores remains a question of scientific and clinical fact on which experts must testify and on which courts must make factual findings.

For example, a clinician must make important threshold determinations regarding the scope and caliber of test instruments to determine whether and how a particular observed score should be properly

---

[31] *See, e.g.*, Floyd et al., *Theories and Measurement of Intelligence*, *supra*, at 416 (noting Schneider method "gives an estimate with a *narrower* confidence interval than a single IQ").

considered. So too must a clinician determine whether particular test administrations rendered a valid and reliable estimate or were tainted by administrative error, practice effects, or lack of effort. If the explanation of how to aggregate IQ test scores sounds complicated, that is because it is: simple arithmetic performed by courts or parties is no substitute for expert testimony.

4. The judgment below is fully consistent with all of these principles. The IQ test scores in this case fall comfortably within the range within which a clinician could reasonably conclude, based on the totality of the evidence, that Smith satisfied the intellectual deficiency criterion.

Amicus Criminal Justice Legal Foundation purports to apply Professor Schneider's aggregation method to suggest that the IQ test scores at issue indicate a so-called "true" IQ score that is probably 70 or above. But its brief, which notes that it was not drafted by an expert, makes fundamental errors no reasonable expert would make, including by mistakenly assuming that the correlation between two measurements using the same test will be perfect. *See* Criminal Justice Legal Foundation Amicus Curiae Br. 6-11. That only underscores the complex, technical nature of this aspect of the diagnostic inquiry—and why it must remain squarely within the domain of experts (and trial court factfinding) in *Atkins* cases.

5. Finally, but critically, just as an IQ test score cannot be the sole datum on which a clinician assesses intellectual functioning, an aggregation of multiple scores can likewise not form the sole basis for a clinical diagnosis. Again, "[i]t is not sound to view a single factor as dispositive of a conjunctive and

interrelated assessment." *Hall*, 572 U.S. at 723. And it remains a "reality that an individual's intellectual functioning cannot be reduced to a single numerical score," even when that score is the amalgamated product of multiple assessments. *Id.* at 713.

IV. Diagnostic Criteria Are Not Used Only In Death Penalty Cases And Are Based On Scientific Study And Clinical Experience

Amicus America First Legal Foundation (America First) launches a direct attack on the credibility of diagnostic criteria and amici's clinical and scientific work. It is tempting to dismiss those attacks and focus exclusively on the actual question before this Court. But the charge that diagnostic definitions and criteria are informed not by rigorous science and clinical experience, but rather by ideological opposition to the death penalty, is irrational, unfounded, and wrong. That cannot be left unsaid.

Cognizant of the profound importance of accurate diagnoses, scientists and medical professionals develop and test intellectual disability definitions and diagnostic criteria—including those found in the *Diagnostic and Statistical Manual of Mental Disorders* (the DSM)—through a robust process grounded in the scientific method and informed by clinical experience. Psychiatrists and psychologists apply these criteria in hundreds of intellectual disability assessments made every day in a wide range of clinical settings—from schools and universities to hospitals and military installations. As this Court has recognized, "the definition of intellectual disability by skilled professionals has implications far beyond the confines of the death penalty: for it is relevant to education, access to social

programs, and medical treatment plans." *Hall*, 572 U.S. at 710. So, while these diagnostic criteria inform *Atkins* claims, this use is relatively rare—and the suggestion that the criteria are somehow tailor-made for that context is flat wrong.

A. Diagnostic Criteria Are Developed In A Rigorous Process, Grounded In Scientific Evidence, Reflecting Clinical Experience

The DSM is a generally applicable text used by practitioners in a wide range of clinical areas. The DSM is the product of decades of research and empirical testing conducted by thousands of scientists and practitioners from across mental-health related fields. America First nonetheless attacks the DSM as the work of ideology, not science. And it claims the American Psychiatric Association's policy positions inform the development of the DSM. These charges, which are not limited to the DSM's diagnostic criteria for intellectual disability, are entirely unfounded.

1. The DSM classifies and defines mental disorders and provides associated criteria designed to facilitate reliable diagnoses and scientific research. These classifications, definitions, and criteria allow clinicians to identify the signs and symptoms of disorders, including affects, behaviors, cognitive functions, personality traits, physical symptoms, and durational markers that differentiate between disorder and normal variation. Recognizing the broad range of clinical domains in which these definitions and criteria are applied, this Court has noted that the DSM "offer[s] 'the best available description of how mental disorders are expressed and can be recognized by trained clinicians.'" *Moore*, 581 U.S. at 20 (citation omitted).

The DSM is developed through a robust process grounded in science and clinical experience. The first edition of the DSM, published in 1952, was the first manual of mental disorders to contain a glossary of diagnostic categories. The DSM has been revised several times to account for scientific developments. And in 2013, the American Psychiatric Association published the most recent full revision to the DSM, the DSM-5.

Development of the DSM-5 took 14 years.[32] The revision process began in 1999 with a series of conferences to develop a research agenda, which was published in 2002. Over the next five years, hundreds of researchers and clinicians reviewed the scientific literature and identified advances and gaps in knowledge. In 2007 and 2008, 28 DSM-5 Task Force members and 130 diagnostic work group members were selected to help draft revisions. These members included researchers and clinicians across a wide range of fields, including psychiatry, psychology, neuroscience, biology, genetics, statistics, epidemiology, and public health. Four hundred non-voting advisors also participated in the revisions.

A critical aspect of the development of the DSM-5 was clinical testing of proposed criteria to assess (among other things) the consistency of diagnoses. Starting in 2010, work groups conducted field trials to test how proposed criteria performed in real-world settings. They selected a sample of "target" disorders to test. A patient population at each site was screened for the target disorders using DSM-IV criteria or by

---

[32] *See* DSM-5-TR at 5-7 (detailing history of the Manual and revision process).

assessing qualifying symptoms likely to predict diagnoses under the DSM-5 disorders. Each patient was evaluated by a clinician blind to the initial screening diagnosis, and by a third clinician using the same process.[33] The results of these trials were published in the *American Journal of Psychiatry*.[34]

While the field trials were underway, the work groups also published proposed revised criteria for public comment. More than 8,000 comments were submitted, reviewed, and incorporated as appropriate. At the end of this process, the work groups submitted final proposed revisions to the Task Force. After several additional American Psychiatric Association committees reviewed the proposals, the Task Force provided a final recommendation to the Association Assembly's Committee on the DSM-5, the Assembly recommended approval by the Association's Board of Trustees, and the Board approved publication in December 2012.

After the DSM-5 was published, the Association established a process for additional substantive revisions to criteria and disorders, which allows professionals and members of the public to submit proposed revisions based on specific scientific advances. This process requires that proposed revisions to criteria are supported by scientific research documenting that the revision improves

---

[33] *Id.* at 7-8 (detailing field trials).

[34] *See* Darrel A. Regier et al., *DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses*, 170:1 Am. J. Psychiatry 59, 61-64 (2013), https://psychiatryonline.org/doi/epdf/10.1176/appi.ajp.2012.12070999.

diagnostic validity across a range of criteria.[35] Submissions are then subject to an extensive multi-level review and validation process for inclusion in the revised manual.

The latest version of the DSM, DSM-5 Text Revision (DSM-5-TR) was published in 2022. The DSM-5-TR includes revisions approved through this process, along with updated descriptive text, based on the work of 200 mental-health professionals over several years.[36]

2. Notwithstanding this robust, decades-long process of scientific study and collaboration on the DSM, America First claims the DSM is "not science as generally understood" and lodges other scattershot attacks on the DSM and the professionals who developed it. These critiques are baseless.

Relying solely on quotations from a "Litigators Handbook," America First argues (at 15-16) that "'descriptions [of mental disorders] do not lend themselves to a single interpretation'" and that the DSM "lacks 'a unifying theory to guide diagnostic decision-making.'" That statement reflects a critical misunderstanding of psychiatric diagnosis. Many mental disorders present as a collection of symptoms across multiple domains and often lack a known pathophysiological basis.[37] This is not because mental disorders are not scientifically defined, but

---

[35] DSM-5-TR at 10.

[36] *Id.* at 10-11.

[37] *See* Daniel Morehead, *The DSM: Diagnostic Manual or Diabolical Manipulation?*, Psychiatric Times (Mar. 21, 2022), https://www.psychiatrictimes.com/view/the-dsm-diagnostic-manual-or-diabolical-manipulation.

because of the immense complexity of human cognition and responses to social and environmental stimuli.

Mental disorders often present on a spectrum and may have overlapping symptoms.[38] "Although some mental disorders may have well-defined boundaries . . . scientific evidence now places many, if not most, disorders on a spectrum, with closely related disorders that have shared symptoms, shared genetic and environmental risk factors, and possibly shared neural substates."[39] As a result, diagnosis necessarily requires clinical judgment—but that does not render psychiatric diagnoses unreliable or unscientific. On the contrary, they "frequently demonstrate predictive validity, genetic risk factors, neurological correlates, and other psychometric characteristics."[40] And more fundamentally, any imprecision in psychiatric diagnosis is consistent with diagnostic approaches in other medical specialties, where clinicians assess symptoms and match them to recognized disorders or diseases based on available

---

[38] *See* DSM-5-TR at 39-41 (describing severity levels of intellectual disability); *id*. at 45 (discussing co-occurrence of intellectual disability with other mental disorders).

[39] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* 8 (5th ed. 2013).

[40] *See* Mark L. Ruffalo, *What Is Meant by a Psychiatric Diagnosis?*, Psychology Today (June 7, 2020) https://www.psychologytoday.com/us/blog/freud-fluoxetine/202006/what-is-meant-psychiatric-diagnosis.

data and using professional judgment and experience.[41]

America First's assertion that the DSM-5 has created "diagnostic hyperinflation"—based on a single opinion piece which does not even mention intellectual disability—is likewise unfounded.[42] America First Amicus Br. 18 (citation omitted). A 2022 meta-analysis of diagnostic rates between two consecutive DSM editions shows that although evidence of increased rate of diagnosis was found for certain disorders including ADHD, autism, and eating disorders, there was no overall change in the rigor or stringency of diagnostic criteria from DSM-III to DSM-IV and DSM-5 and "no reliable tendencies for particular DSM revisions to be more inflationary or deflationary than others."[43] And most critically here, America First does not claim (much less cite evidence to show) that DSM-5 has resulted in any increased rate of diagnosis for intellectual disability.

America First also argues (at 21-22) that supposed disagreements between the American Psychiatric Association and the American Association on Intellectual and Developmental Disabilities (AAIDD)

---

[41] *See* Morehead, *supra* (explaining that across the medical field, clinicians often "treat syndromes nonspecifically—as syndromes [or collection of symptoms] rather than the discrete diseases behind the syndrome"); Kenneth S. Kendler, *Potential Lessons for* DSM *From Contemporary Philosophy of Science*, 79(a) JAMA Psychiatry 99, 99-100 (2022).

[42] *See* Fabian Fabiano & Nick Haslam, *Diagnostic inflation in the DSM: A meta-analysis of changes in the stringency of psychiatric diagnosis from DSM-III to DSM-5*, 80 Clinical Psychology Rev. (2020).

[43] *Id.*

regarding the adaptive functioning criterion shows the DSM is not informed by science. But this critique misunderstands how the DSM works.

America First seizes (at 24-25) on a revision in the DSM-5-TR that removed *explanatory* text found in DSM-5 discussing "relatedness" between the adaptive functioning and intellectual functioning criteria for intellectual disability diagnosis. The DSM-5 diagnostic criteria—which are the operative diagnostic elements—never required a direct relationship between the adaptive functioning and intellectual functioning criteria. Recognizing that this explanatory text was misunderstood by some readers "to change the diagnostic criteria for Intellectual Disability to add a fourth criterion"—a fourth criterion that never existed—the DSM-5-TR revised the explanatory text to remove the confusion.[44] Far from resolving a "disagreement" between two medical organizations, this textual revision merely clarified what the DSM-5's diagnostic criteria had always required.

B. Diagnostic Criteria For Assessing Intellectual Disability Are Applied In A Wide Range Of Clinical Settings

The diagnostic criteria for assessing intellectual disability are used in a wide range of clinical settings. America First's suggestion that they are tailor-made

---

[44] *See* American Psychiatric Association, *Intellectual Developmental Disorder (Intellectual Disability)*, https://www.psychiatry.org/getmedia/497935b7-4543-4343-af1e-797063c22191/APA-DSM5TR-IntellectualDisability.pdf (last visited Sept. 23, 2025).

for death penalty cases or driven by ideology in opposition to the death penalty is unsupported.

1. Clinicians undertake hundreds of thousands of intellectual disability diagnostic assessments every year for a wide range of purposes. School psychologists rely on diagnostic criteria to determine whether students require additional academic support or are eligible for benefits and protections under federal and state law. Clinicians rely on diagnostic criteria when evaluating individuals to determine if they qualify for other federal and state benefits, such as Social Security or Medicaid support. Psychiatrists and psychologists employed by the Armed Forces rely on diagnostic criteria when evaluating service members for fitness and providing support for service members with combat-related illnesses. And courts rely on expert testimony applying diagnostic criteria to evaluate competency.

While diagnostic criteria are also used in the death penalty context, this use is relatively infrequent in comparison. *See Hall*, 572 U.S. at 710. For example, between 2002 and 2013, approximately 371 prisoners brought a claim under *Atkins*, representing only 7.7% of prisoners on death row.[45] In that same ten-year period, psychologists and psychiatrists would have assessed likely over one million other individuals for intellectual disability applying the same criteria.

2. America First nonetheless suggests that the scientific research and medical diagnostic materials of "the APAs"—principally amici American

---

[45] *See* John H. Blume et al., *A Tale of Tale of Two (and Possibly Three)* Atkins*: Intellectual Disability and Capital Punishment Twelve Years After the Supreme Court's Creation of a Categorical Bar*, 23 Wm. & Mary Bill of Rts. J. 393, 396 (2014).

Psychiatric Association and American Psychological Association—are informed by ideological opposition to the death penalty. In essence, America First asserts that medical professionals have altered diagnostic criteria applied in hundreds of thousands of intellectual disability assessments—assessments that can have life-altering consequences—because of an ideological mission aimed to put a thumb on the scales in favor of *Atkins* claimants. That is absurd.

America First points to the fact that amici have taken positions opposing the death penalty. But neither the American Psychiatric Association, which publishes the DSM, nor the American Psychological Association, which does not publish diagnostic criteria, have taken a position on the propriety of the death penalty. Rather, the two organizations have called on jurisdictions that impose the death penalty to adopt policies that protect criminal defendants' constitutional rights, particularly in matters related to competency.

These positions flow directly from, and are consistent with, this Court's holdings in *Atkins* and its progeny. For example, the American Psychiatric Association has taken the position that jurisdictions should not impose capital punishment unless it is "administered fairly and impartially in accord with the basic requirements of due process," and addresses ethical issues related psychiatrists' role in evaluating the competency of prisoners.[46] And the American

---

[46] American Psychiatric Association, *Position Statement on Capital Punishment* (July 2025), https://www.psychiatry.org/getattachment/7f9572e6-2143-4c3b-83c0-6f380363ad26/Position-Capital-Punishment.pdf. The APA's current Position Statement on Capital Punishment does not call for a moratorium.

Psychological Association passed a resolution in 2001—nearly a year before this Court decided *Atkins*—calling on jurisdictions "not to carry out the death penalty until the jurisdiction[s] implement[] policies and procedures" that ameliorate common deficiencies in the administration of the penalty.[47] And these formal position statements, of course, do not necessarily represent the views of the thousands of scientists and clinicians who independently research, study, and publish on issues related to diagnostic criteria, scholarship that informs development of the criteria.

Every field of science and medicine develops over time. The purpose of the DSM and its diagnostic criteria is to facilitate reliable diagnoses by clinicians that, in turn, facilitate effective treatment. Refinement of the diagnostic criteria, including the criteria for intellectual disability, is evidence of the scientific method at work—not politics or ideology.[48]

---

[47] American Psychological Association, *The Death Penalty in the United States* (Aug. 2001), https://www.apa.org/about/policy/death-penalty.

[48] America First also wrongly claims the position taken by amici in *Roper v. Simmons*, 543 U.S. 551 (2005) was inconsistent with the position that amici took 20 years later in *United States v. Skrmetti*, 145 S. Ct. 1816 (2025). In *Roper*, which involved juvenile crime, amici explained that scientific evidence supported the conclusion that adolescents are more impulsive than adults, a finding supported by anatomical studies of the brain. *See Roper* Am. Psych. Ass'n et al. Amici Curiae Br. 4-15, (July 19, 2004) (No. 03-633), 2004 WL 1636447; *Roper* Am. Med. Ass'n, Am. Psychiatric Ass'n et al. Amici Curiae Br. 3-20 (July 16, 2004), 2004 WL 1633549. That has nothing to do with whether minors "ha[ve] the emotional and cognitive maturity

CONCLUSION

The Court should affirm the judgment below.

Respectfully submitted,

AARON M. PANNER
BRENNA L. DARLING
KELLOGG, HANSEN, TODD, FIGEL & FREDERICK, P.L.L.C.
1615 M Street, NW
Suite 400
Washington, DC 20036

Counsel for Amicus Curiae American Psychiatric Association

MELISSA ARBUS SHERRY
Counsel of Record
PETER A. PRINDIVILLE
LATHAM & WATKINS LLP
555 11th Street, NW
Suite 1000
Washington, DC 20004
(202) 637-3386
melissa.sherry@lw.com

DEANNE M. OTTAVIANO
General Counsel
AMERICAN PSYCHOLOGICAL ASSOCIATION
750 First Street, NE
Washington, DC 20002

Counsel for Amicus Curiae American Psychological Association

September 24, 2025

---

required to provide informed consent/assent," to medical decisions made in collaboration with parents or guardians and experienced clinicians. *Skrmetti* Am. Psych. Ass'n et al. Amici Curiae Br. 14-15 (Sept. 3, 2024) (No. 23-477), 2024 WL 4101400.