No. 24-6058

IN THE

Supreme Court of the United States

BLAINE KEITH MILAM,

Petitioner,

v.

TEXAS,

Respondent.

ON PETITION FOR A WRIT OF CERTIORARI TO THE COURT OF CRIMINAL APPEALS OF TEXAS

BRIEF OF *AMICI CURIAE* RANDY W. KAMPHAUS, KEVIN S. MCGREW, CECIL R. REYNOLDS, W. JOEL SCHNEIDER AND MARC J. TASSÉ IN SUPPORT OF PETITIONER

JOHN H. BLUME
Counsel of Record
CORNELL LAW SCHOOL
159 Hughes Hall
Ithaca, NY 14853
(607) 255-1030
jb94@cornell.edu

EMILY C. PAAVOLA
900 Elmwood Avenue,
Suite 200
Columbia, SC 29201

120213

COUNSEL PRESS
(800) 274-3321 • (800) 359-6859

TABLE OF CONTENTS

Table of Contents

TABLE OF CITED AUTHORITIES

Page

Cases

Other Authorities

Cited Authorities

Page

Cited Authorities

Page

Cited Authorities

Page

BRIEF OF *AMICI CURIAE* RANDY W. KAMPHAUS, KEVIN S. MCGREW, CECIL R. REYNOLDS, W. JOEL SCHNEIDER AND MARC J. TASSÉ IN SUPPORT OF PETITIONER

INTEREST OF *AMICI*[1]

Dr. Randy Kamphaus is a research professor at the Ballmer Institute for Children's Behavioral Health at the University of Oregon. He previously served as Dean of the colleges of education at the University of Oregon and Georgia State University. Dr. Kamphaus' research is devoted to improving the measurement of psychological and educational constructs and advancing assessment practice. He has authored or co-authored books, scientific journals and book chapters on these topics, created psychological and educational tests, and holds one patent. His research has been funded by the U.S. Department of Education and its Institute of Education Sciences.

Dr. Kevin McGrew is the owner and Director of the Institute for Applied Psychometrics. He earned his doctorate in educational psychology, special education with emphasis on research methods and applied psychometrics, at the University of Minnesota. Dr. McGrew's teaching and research interests focus on theories of human intelligence, intelligence testing, adaptive behavior, non-cognitive variables important for learning, and applied psychometrics. Dr. McGrew is a co-author of the Woodcock-Johnson III and IV cognitive and achievement batteries, and the first author

1. No counsel for a party authored this brief in whole or in part, and no entity or person, other than *amici*, their members and counsel, made a monetary contribution to fund the preparation or submission of this brief. Counsel of record for both parties received notice and consented to the filing of this brief.

of the Woodcock-Johnson V battery. He has authored or co-authored over 100 peer-reviewed journal articles and book chapters, four books on intelligence test interpretation, and eight norm-referenced test batteries.

Dr. Cecil Reynolds is Emeritus Professor of Educational Psychology, Professor of Neuroscience, and Distinguished Research Scholar at Texas A&M University. His work focuses on psychological testing and assessment, and he has authored more than 45 books, including *The Handbook of School Psychology*, the *Encyclopedia of Special Education*, and the *Handbook of Psychological and Educational Assessment of Children*. Dr. Reynolds has authored or co-authored more than 40 commercially published tests and has published more than 300 scholarly publications. He is the former editor-in-chief of *Archives of Clinical Neuropsychology* and *Applied Neuropsychology*, as well as an editor of the *Journal of School Psychology*, *Psychological Assessment* and *Archives of Scientific Psychology*.

Dr. W. Joel Schneider is a professor at Temple University in the College of Education and Human Development. His research explores the validity of psychological assessment, the discovery of statistical procedures to increase diagnostic accuracy, and the creation of software to facilitate better clinical decision-making. Dr. Schneider completed his doctoral studies in clinical psychology at Texas A&M University. He teaches courses on psychological assessment, counseling, statistics and research methods. Dr. Schneider has authored or co-authored numerous scholarly publications focused on understanding and improving the validity of psychological assessment practices.

Dr. Marc Tassé is a co-author of *Intellectual Disability: Definition, Diagnosis, Classification, and Systems of*

Supports, published by the American Association on Intellectual and Developmental Disabilities. He is a professor at The Ohio State University and serves as the Director of the Ohio State Nisonger Center, a University Center for Excellence in Developmental Disabilities. Dr. Tassé has more than thirty years of experience in conducting research and providing clinical services in the field of intellectual disability. He has been involved in the development of several standardized tests and has published more than 175 scholarly articles.

Amici collectively have extensive expertise in the field of intellectual disability and the administration and interpretation of intelligence testing.

SUMMARY OF ARGUMENT

In accordance with *Atkins v. Virginia*,[2] a criminal defendant in Texas is ineligible for a death sentence if the defendant meets the three-pronged definition of intellectual disability: (1) deficits in intellectual functioning – indicated by an IQ score approximately two standard deviations below the population mean, adjusted for the standard error of measurement ("SEM"), *i.e.*, an IQ score of approximately 75 or less[3]; (2) adaptive deficits; and, (3) onset during the developmental period. For over a decade,

2. 536 U.S. 304 (2002).

3. *Id.* at 309, n.5 ("It is estimated 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 . . ."); *Hall v. Florida*, 572 U.S. 701, 713 (2014) ("The SEM reflects the reality that an individual's intellectual functioning cannot be reduced to a single numerical score. . . . Even when a person has taken multiple tests, each separate score must be assessed using the SEM . . .").

this Court's precedent has dictated that states' *Atkins* decisions must be "informed by the medical community's diagnostic framework."[4]

At a hearing before a state habeas court, it was undisputed that all four of Milam's full-scale IQ scores ("FSIQ"), obtained over a ten-year period, demonstrated significant deficits in intellectual functioning and fell within the range necessary to satisfy prong one, but the habeas court abandoned the clinical guidelines and adopted a state-drafted order finding the General Ability Index ("GAI") to be a "more reliable indicator of [Milam's] intellectual functioning than the FSIQ." App'x A at ¶ 165(a). Relying on the premise that the GAI could serve as a valid alternative measure, the habeas court determined that "the GAI score of 91 puts [Milam] in the 'average' range" and "does not come close to the cutoff for meeting prong one of an ID diagnosis." *Id.* at ¶ 165(d). The court further concluded the lingering effects of Milam's methamphetamine use could explain a purported discrepancy between the GAI and his prior IQ scores. *Id.* at ¶ 167(d), (e).[5]

The habeas court's findings depart significantly from the medical community's diagnostic framework for intellectual disability, especially regarding the use of the GAI. *Amici* are professionals with extensive knowledge and experience in the interpretation of IQ test results, diagnosis of intellectual disability, educational intervention and planning for supports. Intelligence

4. *Hall,* 572 U.S. at 721.

5. The Texas Court of Criminal Appeals adopted the habeas court's findings and conclusions in a summary order. *Ex parte Milam*, WR-79,322-04, 2024 WL 3595749 (Tex. Crim. App. July 31, 2024).

tests are not designed strictly, or even primarily, for the purpose of diagnosis. They are most frequently used in educational settings, where diagnosis is often a secondary concern. In certain circumstances, the GAI can be a useful tool for understanding an individual's specific strengths and weaknesses, allowing for more targeted educational or vocational support. But the GAI is not equivalent to a full-scale IQ score, and to use it instead of the FSIQ in the *Atkins* context creates a risk that a person with intellectual disability will be executed. In this brief, *Amici* will explain why the habeas court's conclusions reflect a fundamental misunderstanding of the medical and psychological community's knowledge of the appropriate use of the GAI.

STATEMENT OF FACTS

Prior to his trial in 2010, Milam was evaluated by four experts. Only one of those experts, Dr. Timothy Proctor, testified on behalf of the State that Milam did not meet the criteria for intellectual disability. However, following this Court's decisions in *Hall* and *Moore v. Texas*,[6] Dr. Proctor reevaluated his assessment and concluded, "[b]ased on the information currently available to me and the relevant diagnostic nomenclature and law at this time, it is my opinion that Mr. Milam meets criteria for intellectual disability." 2021 Ex. 1 at 6. Milam was granted an evidentiary hearing at which Dr. Proctor explained the basis for his changed opinion. The State offered testimony from Dr. Antoinette McGarrahan, whom the State retained only after being advised that Dr. Proctor had determined Milam was a person with intellectual disability.

6. 581 U.S. 1 (2017).

I. Summary of IQ Test Results.

Regarding prong one, deficits in intellectual functioning, the evidence established that Milam had obtained four FSIQ scores in the intellectual disability range, as follows:

Test Name & Administration date	Administered By	Obtained Score	Adjusted For Aged Norms	95% confidence interval
WAIS-IV November 2009	Dr. Paul Andrews	71	70	65-75
Stanford Binet Intelligence Scales, 5th ed. December 2009	Dr. Paul Andrews	78	76*	71-81
WAIS-IV March 2010	Dr. Proctor	68	67*	62-72
WAIS-IV September 2021	Dr. McGarrahan	80	76	71-81
*These scores are likely affected by practice effect.[7]				

7. The practice effect refers to an artificial inflation that occurs when the same or a similar IQ test is given to an individual within a short period of time – approximately 12 months or less. *See* Intellectual Disability: Definition, Classification, and Systems of Supports 39 (AAIDD, 12th ed. 2021) [hereafter AAIDD-12]; Diagnostic and Statistical Manual of Mental Disorders 37 (APA, 5th ed. revised 2022) [hereafter DSM-5-TR]. Practice effect can occur when the second test, such as a new version of the WAIS, is "similar, but not identical to, the first test administered." James Ellis et al., *Evaluation of Intellectual*

Dr. Proctor and Dr. McGarrahan agreed on nearly all essential aspects of the clinical guidelines for assessment of intellectual functioning. For example, both experts testified that the two primary texts establishing diagnostic criteria for intellectual disability – AAIDD-12 and the DSM-5-TR – require the use of a comprehensive, norm-referenced IQ test (which includes the WAIS and the Stanford Binet),[8] application of the SEM to calculate a confidence interval at 95%,[9] and an adjustment to account for the age of the test's norms (*i.e.*, the Flynn effect).[10] Dr. McGarrahan testified that she regularly applies adjustments for the Flynn effect in her practice, and she agreed that the confidence interval reported on the chart above for her IQ test administration is correct. 2 SHRR 64.[11] Dr. McGarrahan acknowledged that when the clinical guidelines are properly considered, the FSIQ score that

---

Disability: Clinical Assessments in Atkins Cases, 46 Hofstra L. Rev. 1305, 1361 (2018).

8. 1 SHRR 16; 2 SHRR 61; *see* AAIDD-12 at 28-29; DSM-5-TR at 38. This brief refers to the 2023 evidentiary hearing record as [volume] SHRR [page number]. Exhibits filed in connection with the state habeas applications are cited as [year] Ex. [number]. The state habeas court's Findings of Fact and Conclusions of Law, dated November 1, 2023, is attached to the Petition for Writ of Certiorari as Appendix A and cited herein as App'x A [paragraph number].

9. 1 SHRR 18; 2 SHRR 63; *see* AAIDD-12 at 29; DSM-5-TR at 38.

10. 1 SHRR 22-24; 2 SHRR 63-65; *see* AAIDD-12 at 42; DSM-5-TR at 37.

11. Dr. McGarrahan testified that the appropriate interval is 70.8 to 80.8. The figures on the chart have been rounded up to the nearest whole number for simplicity.

Milam obtained on her testing is congruent with the three FSIQ scores he previously obtained. *Id.* at 65. Additionally, Dr. McGarrahan did not dispute any of Milam's prior IQ scores (*id.* at 68) and testified that the best practice is to consider all available FSIQ scores. *Id.* at 65. She further agreed that "looking solely" at the FSIQ scores, Milam's results satisfy the criterion for prong one. *Id.* at 78.

II. Dr. McGarrahan's Direct Testimony.

The expert opinions diverged, however, on the issue of the GAI. Dr. McGarrahan met with Milam for a total of eight hours and administered a neuropsychological test battery, including the WAIS-IV. *Id.* at 14.[12] Milam was cooperative and made good effort. *Id.* at 14, 25. He reported that he attended school until the fourth grade after which his parents withdrew him with the intent to homeschool. However, no additional schooling took place after Milam's father suffered a heart attack and his mother had to work to support their family. *Id.* at 15. Dr. McGarrahan agreed with Dr. Proctor that such a profound lack of education is a risk factor for intellectual disability.

---

12. Although the assessment of prongs two and three are beyond the scope of this amicus brief, it is worth noting that Dr. McGarrahan did not speak to any family members, friends, teachers, employers or any other collateral informants. She did not administer any adaptive behavior scales, nor did she seek to speak with anyone, other than Milam himself, about his developmental history and adaptive behavior. It is well-established that a clinically accurate assessment of intellectual disability relies on a rigorous collection of data, from as many collateral sources as possible, about an individual's typical functioning in a community setting. AAIDD-12 at 42. Reliance on self-reported information is strongly disfavored.

Id. at 17. Milam disclosed that he had a history of drug and alcohol use during the developmental period – yet another risk factor. *Id.* at 20. Specifically, Milam reported that prior to his arrest at age 18, he used "about a gram" of methamphetamine on the weekends. 2022 State's Ex. A at 4.

Although she agreed that the FSIQ Milam obtained on her testing was "not substantially different" from his prior test results, Dr. McGarrahan found it significant that Milam's Verbal Comprehension Index score ("VCI") was higher in September of 2021 compared to prior testing.[13] She considered this a "highly unlikely, very rare increase in his verbal skills." 2 SHRR 27. Milam's index scores were otherwise "consistently low" across testing (*id.* at 35), but because the VCI was significantly higher, Dr. McGarrahan believed the FSIQ was "not to be relied upon" and should be substituted with the GAI, which is calculated by removing two of the four index scores – working memory and processing speed (the two areas in which Milam performed most poorly). *Id.* at 35-36. Relying on the GAI score of 91, Dr. McGarrahan opined that Milam does not meet criteria for intellectual disability, although she believed "[h]e does have a substantial amount of cognitive difficulties," including severely impaired nonverbal problem solving and abstract reasoning skills, extremely slow processing speed, poor working memory and low average perceptual reasoning abilities. *Id.* at 25; 2022 State's Ex. A at 6-7.

13. The WAIS-IV is comprised of four indexes, which are factored together to produce the full-scale IQ score. They include verbal comprehension (VCI), perceptual reasoning (PRI), working memory (WMI) and processing speed (PSI). The Stanford Binet-5 measures similar elements of cognitive functioning but uses a different structure to produce the full-scale score.

Comparing the GAI score of 91 to Milam's previous FSIQ scores, Dr. McGarrahan concluded that such a "huge jump" must have occurred due to the "ongoing effects from methamphetamine use" that depressed Milam's previous scores. She stated that his prior testing produced a "flat profile" in which "[e]verything was down" and "negatively affected, from a number of things, but primarily the substantial drug abuse." 2 SHRR 34. Dr. McGarrahan testified that it can take up to a couple of years after use for the brain "to clear the fog, to clear the effects of the methamphetamine." *Id.* at 21. She believed that the effects of Milam's "meth fog" would have cleared by the time of her own testing, and Milam's subsequent improvement in verbal skills demonstrated "he had the capacity to learn" beyond what Dr. McGarrahan viewed as consistent with intellectual disability. *Id.* at 34.

III. Evidence Elicited on Cross-Examination and Rebuttal.

On cross-examination, Dr. McGarrahan admitted that the AAIDD instructs practitioners to use FSIQ for diagnostic purposes. *Id.* at 61; *see* AAIDD-12 at 28-29. She acknowledged that both AAIDD-12 and the DSM-5-TR require the use of a comprehensive IQ test whereas the GAI excludes two areas deemed by the medical community to be "critical components of intellectual functioning." 2 SHRR 85. In explaining why she chose to use the GAI, Dr. McGarrahan cited examples such as severe depression, poor motivation or motor difficulties as circumstances in which she believed the GAI might be a better measure of intelligence than FSIQ. *Id.* at 39-40. However, she testified that Milam was fully cooperative, made no effort to malinger, and did not exhibit any psychomotor

behaviors or physical disorders, nor were there any outside disturbances or other concerns that impacted her testing. *Id.* at 56-58.

In support of her opinion regarding the GAI, Dr. McGarrahan cited: (1) the *WAIS Technical Manual*; (2) *WAIS-IV Clinical Use and Interpretation*, by Diane Coalson and Susan Raiford [hereafter "Clinical Use"]; (3) *The Essentials of WAIS Assessment, 2nd edition*, by Elizabeth Lichtenberger and Alan Kaufman [hereafter "Essentials"]; and, (4) *WAIS-IV, WMS-IV and ACS Advanced Clinical Interpretation* [hereafter, "Advanced Interpretation"]. But none of these sources instructs practitioners to use the GAI as a substitute for FSIQ for diagnostic purposes.

Dr. McGarrahan conceded that the *WAIS Technical Manual* explicitly states, "the full scale IQ is the most reliable score" and "[t]he GAI does not replace the FSIQ." *Id.* at 81, 88. The textbook *Clinical Use* (edited by the same researchers who produced the WAIS Technical Manual), explains:

> [i]n our first WISC-IV book, we suggested that some practitioners may prefer the GAI as an alternative way of summarizing overall ability. This suggestion has led to an increasing number of psychological evaluations in which the GAI is described as a better estimate of overall ability than FSIQ whenever the WMI or PSI score are significantly lower than the VCI or PRI scores. As we subsequently stated, this is not what we intended a*nd can be a very problematic practice.*

Id. at 93-94. Consistent with this instruction, Dr. Kaufman's *Essentials* includes the following warning in both the text and in a separate caution box:

> [a]lways interpret a person's overall score on the WAIS-IV whenever a global score is essential for diagnosis (e.g., of intellectual disability) or placement (e.g., in a gifted program).

Id. at 101. Likewise, *Advanced Interpretation* provides the same cautionary language quoted above from *Clinical Use* and explains that "working memory and processing speed are essential components of a comprehensive assessment of intelligence"; excluding them is "poor practice" and leads to "unrealistically high estimates of intelligence" for patients with weaknesses in these areas. *Id*. at 160-162.[14]

Despite her acknowledgement of these guidelines, Dr. McGarrahan believed the texts authorized her to use "clinical judgment" to substitute the GAI in place of FSIQ. *Id*. at 89. She stated she relied most heavily on Dr. Kaufman's *Essentials*, but Dr. Kaufman himself testified in rebuttal that her reliance on his work was incorrect. 2 SHRR 184. Dr. Kaufman is a well-known and highly respected scholar in the field of intelligence testing who has worked in psychometrics, test development, test administration and interpretation for half a century. *Id*. at 167, 174. From 1971 to 1974, he worked directly with Dr. David Wechsler on revisions to the original Wechsler

14. The instructions from these authoritative texts are consistent with Dr. Proctor's testimony, as he offered a detailed explanation to the habeas court about why the use of the GAI in place of the FSIQ is contrary to clinical guidelines. 1 SHRR 16-17, 138-139.

Intelligence Scales for Children. *Id.* at 173. He also served as a consultant for the development of the WAIS-IV. *Id.* Dr. Kaufman has published hundreds of articles, books and test manuals on intelligence testing, and his work was cited by this Court in *Hall*. 572 U.S. at 713.

Dr. Kaufman explained that *Essentials* was intended to serve as a "nuts-and-bolts book" about the WAIS-IV, and the goal was "to make it simple and straightforward." 2 SHRR 176. There are approximately 70 books in the Essentials series (covering various tests and subjects) and all include "callout boxes or cautions" to highlight "aspects of interpretation that are nonintuitive or . . . important." *Id.* In this case, *Essentials of WAIS-IV Assessment* includes a caution box to indicate that, apart from occurrence of "spoiled or invalid" subtests, "the full-scale IQ is the score of choice for diagnosis." *Id.* at 184. Given that Dr. McGarrahan expressed no concern that any index scores were invalid or spoiled, there was no reason for her not to accept the FSIQ as reliable. Dr. Kaufman reiterated that variability in index scores does not make the FSIQ invalid (*id.* at 179) and described the GAI as a short-form measure lacking the comprehensive nature of tests required by AAIDD-12 and the DSM-5-TR. *Id.* at 181 ("the GAI fits into the short form, not comprehensive category"). He stated that the GAI is "helpful to inform" assessment, but it was "not meant to sabotage diagnosis." *Id.* at 183. In sum, Dr. Kaufman's testimony established that the GAI should not be used for diagnostic purposes without a valid clinical reason, such as lack of effort, disruption of testing, or other circumstances that were not present in this case by Dr. McGarrahan's own admission.

IV. The Habeas Court's Findings.

Although the habeas court had the information it needed to make a legal decision adequately "informed by the medical community's diagnostic framework,"[15] the court instead sanctioned a line of reasoning that departs significantly from the clear consensus of the medical community. The court adopted the State's proposed findings of fact and conclusions of law, which determined that the GAI is a more reliable indicator of Milam's intellectual functioning than the FSIQ and its use "is supported by the professional literature and evidence." App'x A at ¶ 165(a). The court concluded Dr. McGarrahan's view on this topic was further supported because "the manuals provide for the exercise of clinical judgment in reaching the ultimate decision." *Id.* at ¶ 165(e). The court's findings also adopted Dr. McGarrahan's suggestion that "meth fog" likely "suppressed [Milam's] original IQ scores, while explaining the later jump in scores." *Id.* at ¶ 167(d).

ARGUMENT

I. The GAI is Not a Substitute for FSIQ.

The purpose of intelligence testing spans far beyond the narrow confines of diagnoses for *Atkins* purposes. Indeed, the fundamental principles that comprise the clinical consensus were conceived long before *Atkins* was even contemplated, and none were developed for the specific task of litigating the question of intellectual disability in an adversarial setting. Although diagnosis is one of the tasks performed by mental health professionals, the central purpose of the medical community (and the body of literature and tools it produces) focuses largely

15. *Hall*, 572 U.S. at 721; *Moore*, 581 U.S. at 5.

on the goal of providing appropriate supports to people with intellectual deficits.

The GAI was developed in the early 1990s for use with the Wechsler Intelligence Scales for Children, Third Edition ("WISC-III") in ability-achievement discrepancy analyses used to identify children with learning disabilities.[16] A discrepancy between IQ and academic performance is one of the hallmark indicators of a potential learning disability. A classic example is a child who obtains an average full-scale IQ of 100 but performs several years below her grade-level peers in one or more specific subjects. In many school districts nationwide, a discrepancy of a certain magnitude is often required for a student to qualify for educational intervention services. A complicating factor, however, is that many students with learning disabilities "exhibit cognitive processing deficits in working memory and processing speed concomitant with their learning disabilities." *Clinical Use* at 80. In such cases, their depressed performance on working memory and processing speed tasks lowers the FSIQ, "which decreases the magnitude of the discrepancy between ability and achievement and may result in denial of needed special education services." *Id.* In this situation, the GAI (which removes working memory and processing speed from the composite) can be used as the comparison point to refine academic placement decisions.

The GAI may shed greater light on a student's specific strengths, which is an important element of developing

16. *The WISC-III in Context*, in WISC-III Clinical Use and Interpretation: Scientist-Practitioner Perspectives, 1-38 (A. Prifitera, et al. eds) (1998).

a student's individual education plan or other academic support. Additionally, the GAI may be helpful "when physical or sensory disorders invalidate performance on the working memory or processing speed tasks, or both." *Id.* at 81. A student with attention deficit hyperactivity disorder, for example, may be so inattentive during testing as to invalidate these portions of the IQ test. In such circumstances, it would be better for the clinician to rely on the GAI than to simply deny services to an individual in need. As explained by some of the same researchers who produced the *WAIS Technical Manual*:

> [w]e intended the GAI to be used only where there are sound clinical reasons to exclude WMI and PSI, such as invalid administration due to lack of effort; sensory or physical impairments; disturbance of the testing session; etc. In some of these situations it may be possible to prorate a single subtest, which would be a better practice.

Id.

Unlike specific learning disabilities, intellectual disability is not diagnosed based on a discrepancy analysis. If an individual has a full-scale IQ that is significantly subaverage (approximately two standard deviations below the population mean), concurrent with adaptive deficits manifested prior to adulthood, they meet the criteria for intellectual disability. Although a person may have *both* intellectual disability and a learning disability, a discrepancy analysis would only be used to assess the comorbid learning disability – not to diagnose intellectual

disability.[17] Moreover, for practical purposes, a student who meets the criteria for intellectual disability will typically receive special education services because of that diagnosis alone. Rarely would an additional diagnosis of specific learning disorder be necessary for such a student to obtain appropriate intervention.

For diagnostic purposes, it is well-established that FSIQ should be used to assess prong one. *See* AAIDD-12 at 28 ("In reference to determining significant limitations in intellectual functioning, a full-scale IQ score should be used."). FSIQ is the best means of accurately and reliably determining global intelligence. David Wechsler, who developed the dominantly used Wechsler series of IQ tests, described intelligence as:

> the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable.[18]

---

17. DSM-5-TR at 45 (explaining specific learning disorders "may co-occur with intellectual developmental disorder. Both diagnoses are made if full criteria are met for intellectual developmental disorder and a communication disorder or specific learning disorder.").

18. Lisa Whipple Drozdick et. al., *The Wechsler Adult Intelligence Scale – Fourth Edition and the Wechsler Memory Scale – Fourth Edition*, *in Contemporary Intellectual Assessment: Theories, Tests, and Issues* 197, 198 (Dawn P. Flanagan & Patti L. Harrison eds., 3d ed. 2012) [hereafter, *Contemporary Assessment*]

The subtests of the Wechsler scales thus measure "many different mental abilities," including "abstract reasoning . . . perceptual skills, verbal skills, and processing speed."[19] No sub-test (or partial set of sub-tests) alone can assess the entire range of cognitive abilities; instead, they are aggregated to produce a FSIQ which is "the score most representative of . . . global intellectual functioning."[20]

The GAI is not equivalent to FSIQ. It is a "part score"[21] examining only the subject's verbal and perceptual reasoning abilities. It therefore provides a much narrower view and is neither a more valid nor a more complete measure of cognitive ability than FSIQ. Working memory and processing speed are essential components of an individual's general intelligence, and excluding these areas from consideration by relying on a part-score like the GAI creates a substantial risk of overlooking significant

---

(citing David Wechsler, *The Measurement of Adult Intelligence* 3 (1939)).

19. *WAIS III WMS III Technical Manual* 2-3 (David Tulsky, Jianjun Zhu & Mark Ledbetter eds., 1997).

20. *Contemporary Assessment* at 200 (describing the FSIQ as "a robust predictor of an array of important life outcomes"). The Stanford-Binet similarly produces a FSIQ score from the aggregate of multiple sub-tests. *Id.* at 249-52. Both tests "are well established, cover multiple areas that provide a reasonably comprehensive profile, and are carefully researched IQ tests." Denis Keyes et al., *Mitigating Mental Retardation in Capital Cases: Finding the "Invisible" Defendant*, 22 Mental & Physical Disability L. Rep. 529, 536 (1998).

21. A part score is the aggregate of the results of some, but not all, of the subtests on the WAIS.

cognitive deficits.[22] The medical community's clinical consensus establishes that "part score interpretation should not be standard practice during the consideration of ID diagnosis and eligibility."[23] As the AAIDD's diagnostic manual clearly explains, "[t]here is no reason to question the validity of the full scale IQ, even in individual cases where there is significant factor/part score variability." AAIDD-12 at 28.

Part score discrepancy is not unusual among individuals with intellectual disability, as they tend to have areas of relative strength and weakness just as other people do. "The typical person with an IQ of 70 has at least one part score in the average range of 90 or higher."[24] As demonstrated by the results of numerous studies, variability in part scores does not weaken the relationship between FSIQ and its predictive value for long-term achievement outcomes.[25] Thus, there is

---

22. Patrick C. Kyllonen and Raymond E. Christal, *Reasoning Ability is (Little More Than) Working Memory Capacity?!*, Intelligence, 14(4), 389-433 at 426 (1990) (finding "a consistent and remarkably high correction" between working memory and general reasoning ability).

23. Randy G. Floyd, et al., *Theories and Measurement of Intelligence* in 1 American Psychological Association Handbook of Intellectual and Developmental Disabilities 385, 412-413 (Laraine Masters-Glidden ed. in chief, Leonard Abbeduto, Laura Lee McIntyre, Marc J. Tasse eds. 2021) [hereafter, *Theories*].

24. *Theories* at 413.

25. *See* Kotz, K.M., et al., *Validity of the General Conceptual Ability Score From the Differential Ability Scales as a Function of Significant and Rare Interfactor Variability*, in School Psychology Review, 37(2), 261-278 (2008); Watkins, M., et al.,

"no scientific foundation" for the idea that part score variability invalidates the FSIQ.[26]

II. The Role of Clinical Judgment.

The habeas court's reliance on Dr. McGarrahan's clinical judgment was misplaced. Clinical judgment is not mere opinion, nor is it a license to disregard established clinical consensus. On the contrary, the proper exercise of clinical judgment is based on "familiarity with – and use of – best practices in the field."[27] Clinical judgment is a special type of judgment rooted in a high level of clinical experience, training, and knowledge. It is based on an extensive collection of data. Its purpose is to "enhance the quality, validity, and precision" of the clinician's decision, resulting in "high-quality" and "valid" assessment results.[28]

Clinical judgment is not "a facile excuse that is used when a person's opinion is challenged (e.g., 'I don't have

---

Validity of the Full Scale IQ When There is Significant Variability Among WISC-III and WISC-IV Factor Scores, in Applied Neuropsychology, 14(1), 13-20 (2007); Freberg, M.E., et al., *Significant Factor Score Variability and the Validity of the WISC-III Full Scale IQ in Predicting Later Academic Achievement*, in Applied Neuropsychology, 15, 13-139 (2008); Daniel, M.H., *"Scatter" and the Construct Validity of FSIQ: Comment on Fiorello et al.*, in Applied Neuropsychology, 14, 291-295 (2007); McGill, R.J., *Invalidating the Full Scale IQ Score in the Presence of Significant Factor Score Variability: Clinical Acumen or Clinical Illusion?*, in Archives of Assessment Psychology, 6(1), 49-79 (2016).

26. *Theories* at 413.

27. Robert L. Schalock and Ruth Luckasson, Clinical Judgment 11 (2d ed. 2014).

28. *Id*. at 7, 11.

to explain it – that is my clinical judgment!')."[29] Nor is it "a justification for abbreviated evaluations," "a substitute for insufficiently explored questions,"[30] or "a shield when one draws conclusions that are not supported by the assessment results, observations, and/or case records."[31] As she herself acknowledged, none of the sources cited by Dr. McGarrahan supported her decision to substitute the GAI for the full-scale IQ in this case. When confronted with the actual text of the literature, her only response was an invocation of her "clinical judgment." This is precisely what clinical judgment is *not*, and the court's heavy reliance on Dr. McGarrahan's clinical judgment was contrary to clinical guidelines.

III. There is no Scientific Support for the Habeas Court's "Meth Fog" Conclusions.

a. Milam's FSIQ scores are not divergent.

The habeas court's findings about the purported role of "meth fog" to explain a recent "jump" in scores are built on a faulty premise. The court concluded that Milam's three prior FSIQ scores were significantly divergent from Dr. McGarrahan's test results because the court accepted her view that the GAI score of 91 should be considered the more reliable measure of Milam's general intelligence. As discussed above, this conclusion was erroneous and is not supported by the scientific evidence. When all four FSIQ scores are examined, as they should be, Milam's test results are clearly congruent.

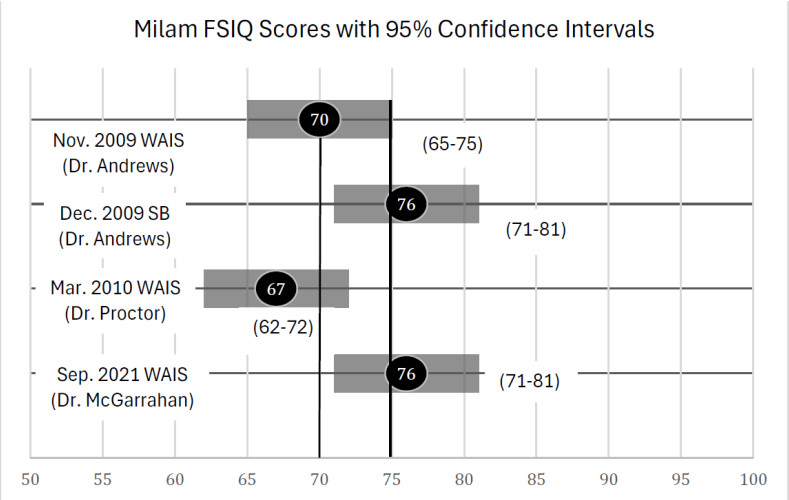
---

29. *Id*. at 15.

30. *Id*.

31. Marc J. Tassé, *Adaptive Behavior Assessment and the Diagnosis of Mental Retardation in Capital Cases*, 16 Applied Neuropsychology 114, 121 (2009).

As this Court has recognized, "the analysis of multiple IQ scores jointly is a complicated endeavor."[32] It is essential that clinicians evaluate the totality of the evidence, making sure to consider the SEM in calculating a 95% confidence interval for each separate score.[33] As depicted on the chart below, there is substantial overlap across the confidence intervals for each of Milam's four FSIQ scores, demonstrating strong consistency over time.



32. *Hall*, 572 U.S. at 714 (citing Schneider, *Principles of Assessment of Aptitude and Achievement*, in The Oxford Handbook of Childhood Psychological Assessment 286, 289-291, 318 (D. Saklofske, C. Reynolds, V. Schwean, eds. 2013)).

33. *Id.*; *see also*, *Theories* at 415 ("Given that it is statistically inappropriate to arithmetically average IQs, clinicians may benefit from evaluating the 95% confidence intervals for each score and collectively interpreting the complete set of scores using clinical judgment.").

It is true that Milam's performance on the Verbal Comprehension Index increased on Dr. McGarrahan's testing after he had been incarcerated for over a decade. However, this is the only index score that showed any appreciable increase. Milam's most significant gain occurred on the sub-test entitled "information," which measures a subject's general fund of knowledge.[34] In a 1976 study examining incarcerated people, Bolton and colleagues found that after an average of 19 months, inmates scored higher on IQ tests, particularly on verbal intelligence sub-tests, than they did upon initial confinement.[35] The researchers attributed this to the inmates' increased opportunity to practice verbal skills in the prison setting.

It is not surprising that Milam made gains in this area, particularly because he entered prison with an exceptionally low level of formal education. People with intellectual disability can and do learn new skills, and contrary to Dr. McGarrahan's testimony, there is no specific "ceiling" for improvement above which one would definitively rule out intellectual disability. Of critical importance here is that despite his increased verbal skills, Milam's overall FSIQ did not rise out of the intellectual disability range. There is therefore no "big jump" to be explained.

b. Even if "meth fog" exists, it would not explain the evidence in this case.

Even if Milam's FSIQ scores had been discrepant, which they were not, we found no scientific support for

34. 2022 State's Ex. A at 11.

35. Bolton, N., et al., *Psychological Correlates of Long-Term Imprisonment: A Longitudinal Analysis*, in The British Journal of Criminology, 16(1), 38-47 (January 1976).

the proposition that his previous use of methamphetamine would offer an adequate explanation of the available data. Prior to his arrest, Milam reported using approximately one gram of meth on the weekends. His first three IQ administrations occurred approximately one year after his arrest. To be sure, competent mental health professionals should ensure that the subject is not actively under the influence of drugs or alcohol *at the time of testing*. But as far as *Amici* are aware, nothing in the clinical literature suggests that a persistent "meth fog" might influence the testing up to one year after intermittent use.

Moreover, even if such a phenomenon exists, there is no reason to believe that the long-term effects of methamphetamine would depress only a subject's Verbal Comprehension Index score, as Dr. McGarrahan claimed. As Dr. McGarrahan testified, verbal comprehension is particularly *resilient* to external effects. 2 SHRR 29-30. Her suggestion that Milam's previous drug use might have strictly suppressed only his VCI score in 2009 and 2010 is unfounded, and she herself testified she was not claiming "it's definitively an issue." *Id.* at 153. The habeas court deviated from established clinical consensus in finding that "meth fog" explained the data in this case.

CONCLUSION

The Court should grant the Petition and reverse the judgment below.

Respectfully submitted,

JOHN H. BLUME
Counsel of Record
CORNELL LAW SCHOOL
159 Hughes Hall
Ithaca, NY 14853
(607) 255-1030
jb94@cornell.edu

EMILY C. PAAVOLA
900 Elmwood Avenue,
Suite 200
Columbia, SC 29201