

UNITED STATES COURT OF APPEALS

FOR THE SIXTH CIRCUIT

BYRON LEWIS BLACK,

Petitioner-Appellant,

v.

WAYNE CARPENTER, Warden,

Respondent-Appellee.

No. 13-5224

Appeal from the United States District Court
for the Middle District of Tennessee at Nashville.
No. 3:00-cv-00764—Todd J. Campbell, District Judge.

Argued: December 8, 2016

Decided and Filed: August 10, 2017

Before: COLE, Chief Judge; BOGGS and GRIFFIN, Circuit Judges.

COUNSEL

ARGUED: Kelley J. Henry, OFFICE OF THE FEDERAL PUBLIC DEFENDER, Nashville, Tennessee, for Appellant. John H. Bledsoe, OFFICE OF THE TENNESSEE ATTORNEY GENERAL, Nashville, Tennessee, for Appellee. **ON BRIEF:** Kelley J. Henry, OFFICE OF THE FEDERAL PUBLIC DEFENDER, Nashville, Tennessee, for Appellant. Andrew H. Smith, OFFICE OF THE TENNESSEE ATTORNEY GENERAL, Nashville, Tennessee, for Appellee.

BOGGS, J., delivered the opinion of the court in which GRIFFIN, J., joined, and COLE, C.J., joined in part. COLE, C.J. (pg. 22), delivered a separate opinion concurring in the majority opinion except for Section II.E and concurring in the judgment.

OPINION

BOGGS, Circuit Judge. In 1986, Byron Black shot his girlfriend Angela's ex-husband, Bennie. Black pleaded guilty to malicious shooting and was sentenced to two years of imprisonment at a Davidson County, Tennessee, workhouse. In 1988, while on a weekend furlough from that workhouse, Black entered Angela's home, shot Angela in the head as she slept, and then shot nine-year-old Latoya and six-year-old Lakeisha (Angela's children by Bennie) once and twice, respectively, killing all three victims. Black returned to the workhouse at the end of his furlough before law-enforcement officers discovered the bodies.

Black's trial and post-conviction proceedings have spanned nearly thirty years. Seventeen years have elapsed since Black filed the federal habeas petition presently before us. The Supreme Court and the Tennessee courts have recently recognized limitations imposed by the Eighth Amendment on the power of states to execute mentally retarded persons. But, for the reasons that follow, these jurisprudential developments do not give Black a reprieve from his sentence of death. We affirm the district court's denial of post-conviction relief.

I

Black stood trial for the 1988 triple murder. A jury found Black guilty of murder and burglary and sentenced him to death for one murder and life imprisonment for the other two murders. The Tennessee Supreme Court affirmed on direct appeal. The Tennessee Court of Criminal Appeals denied post-conviction relief, and the Tennessee Supreme Court denied further post-conviction review. In 2000, Black filed a federal habeas petition in which he raised various claims including a claim that his mental retardation precluded the imposition of the death penalty. The petition was dismissed as meritless. *Black v. Bell*, 181 F. Supp. 2d 832, 883 (M.D. Tenn. 2001). Black appealed to our court, but the Supreme Court shortly thereafter decided *Atkins v. Virginia*, 536 U.S. 304, 321 (2002) (holding that the Eighth Amendment prohibits states from executing "mentally retarded criminals"), so we granted Black's motion to hold his appeal

in abeyance while Black exhausted an *Atkins* claim in the Tennessee courts. *Black v. Bell*, No. 02-5032 (6th Cir. July 26, 2002) (order).

The Tennessee trial court conducted an evidentiary hearing and denied Black's *Atkins* claim as meritless, the Tennessee Court of Criminal Appeals affirmed, and the Tennessee Supreme Court denied further review. *Black v. State*, No. M2004-01345-CCA-R3-PD, 2005 WL 2662577 (Tenn. Crim. App. Oct. 19, 2005), *perm. app. denied* (Tenn. Feb. 21, 2006). Our court then remanded Black's appeal to the district court so that it could consider Black's *federal* habeas claim in light of *Atkins*. *Black v. Bell*, No. 02-5032 (6th Cir. May 30, 2007) (order). The Supreme Court in *Atkins* had "[left] to the States the task of developing appropriate ways to enforce" its prohibition on executing mentally retarded criminals. *Atkins*, 536 U.S. at 317. The district court thus, quite understandably, looked to Tennessee law in analyzing Black's *Atkins* claim.

Tennessee had enacted a statute defining mental retardation as follows:

- (1) Significantly subaverage general intellectual functioning as evidenced by a functional intelligence quotient (I.Q.) of seventy (70) or below;
- (2) Deficits in adaptive behavior; and
- (3) The mental retardation must have been manifested during the developmental period, or by eighteen (18) years of age.

Tenn. Code Ann. § 39-13-203(a) (2003).

The United States Supreme Court recently referred to a definition of mental retardation substantially similar to this tripartite Tennessee definition as the "the generally accepted, uncontroversial intellectual-disability diagnostic definition." *Moore v. Texas*, 137 S. Ct. 1039, 1045 (2017).

For its part, the Tennessee Supreme Court held in 2004 that the first part of Tennessee's statutory definition of mental retardation imposed a "bright line rule" requiring an *Atkins* petitioner to demonstrate an IQ of seventy or below. *Howell v. State*, 151 S.W.3d 450, 456–59 (Tenn. 2004) (agreeing with the State that § 39-13-203(a)(1) "should not be interpreted to make allowance for *any standard error of measurement or other circumstances* whereby a person with an I.Q. above seventy could be considered mentally retarded" (emphasis added)).

The district court considered Black's IQ scores as follows:

IQ Scores Before Age 18			
Date of Test	Name of Test	Score	Black's Approximate Age
1963	Lorge Thorndike	83	7
1964	Unknown	97	8
1966	Lorge Thorndike	92	10
1967	Otis	91	11
1969	Lorge Thorndike	83	13
IQ Scores After Age 18			
Date of Test	Name of Test	Score	Black's Approximate Age
1989	Shipley-Hartford	76	33
1993	WAIS-R	73	37
1997	WAIS-R	76	41
2001	WAIS-III	69	45
2001	Stanford-Binet-IV	57	45

Black argued to the district court that the Tennessee courts' denial of his *Atkins* claim was improper in part because those courts "refused to consider standard errors in test measurement [and] the 'Flynn Effect,'¹ permitted the State's experts to testify, and placed the burden of proof on the Petitioner." *Black v. Bell*, No. 3:00-0764, 2008 U.S. Dist. LEXIS 33908, at *15 (M.D. Tenn. Apr. 24, 2008). Black had argued in state court, and argued again to the

¹The Flynn Effect, named after intelligence expert James Flynn, is a "generally recognized phenomenon" in which the average IQ scores produced by any given IQ test tend to rise over time, often by approximately three points per ten years from the date the IQ test is initially standardized. *See Ledford v. Head*, No. 1:02-CV-1515-JEC, 2008 WL 754486, at *7 (N.D. Ga. Mar. 19, 2008); *see also* Am. Ass'n on Intellectual and Developmental Disabilities, *Intellectual Disability: Definition, Classification, and Systems of Supports* 36-41 (11th ed. 2010).

The WAIS-III test, for example, was published in 1997. When the WAIS-III was designed, it was administered to a "standardization sample" of 2,450 adults from the United States who were sorted into cohorts by age and other characteristics. D. Wechsler, The Psychological Corp., *WAIS-III Administration & Scoring Manual* (1997). IQ scores generated by the WAIS-III test essentially offer a measure of intelligence *relative to* the standardization sample of 2,450 people, all of whom took the test in 1995. The Flynn Effect would thus predict that average IQ scores generated by the WAIS-III in 2005 (ten years after it was normed) would be approximately three points higher, on average, than those generated in 1995, and would predict that scores generated by the same test in 2015 would be approximately six points higher, on average, than those generated in 1995.

But there is no legal or scientific consensus that *requires* an across-the-board downward adjustment of IQ scores to offset the Flynn Effect; rather, the Flynn Effect is one of many potential factors affecting the reliability and validity of any individual IQ score, and a professional who is assessing an individual's intelligence on the basis of an IQ score would take the Flynn Effect and other factors into consideration as part of that assessment.

district court, that his IQ scores should be reduced retroactively to account for both the standard error of measurement (SEM) and the Flynn Effect.²

The district court noted that the Tennessee Court of Criminal Appeals, in rejecting Black's argument to adjust his IQ scores downward to account for the SEM or the Flynn Effect, thoroughly considered the evidence provided by Black's experts and the State's experts. *Black v. Bell*, 2008 U.S. Dist. LEXIS 33908, at *15–20. The district court itself was “not persuaded” by Black's arguments. *Id.* at *21. Applying *Howell*, which had also guided the decision of the Tennessee Court of Criminal Appeals, the district court denied Black's *Atkins* claim on the basis that “the state court was not unreasonable in stating that the proof in the record did not support the conclusion, under a preponderance of the evidence standard, that [Black's] I.Q. was below

²The SEM is distinct from the Flynn Effect. The SEM allows for the possibility that an IQ score either overestimates or underestimates a subject's true IQ. Contrary to common understanding, a SEM of “five points” does not necessarily mean, for example, that a person with an IQ score of 75 *must* have a true IQ between 70 and 80. Rather, the SEM represents the *standard deviation* of true IQ scores from reported IQ scores. See, e.g., Leo M. Harvill, *An NCME Module on Standard Error of Measurement*, 10 *Educ. Measurement: Issues & Prac.* 33 (1991). Thus, a SEM of five points means that a person with a reported IQ of 75 is approximately 68% likely to have a true IQ within five points of 75 (i.e., between 70 and 80—one standard deviation on either side of 75), approximately 95% likely to have a true IQ within ten points (*two* standard deviations) of 75 (i.e., between 65 and 85), and approximately 99.7% likely to have a true IQ within fifteen points (*three* standard deviations) of 75 (i.e., between 60 and 90). It is therefore a gross oversimplification to attempt to account for error in measurement by retroactively reducing (or increasing) a reported IQ score by *one* SEM (or any number of SEMs).

Further, the SEM itself varies by test, subtest, and test-taker. The American Psychiatric Association states in its *Diagnostic and Statistical Manual of Mental Disorders* simply that “there is a measurement error of approximately 5 points in assessing IQ.” *Diagnostic & Statistical Manual of Mental Disorders* 41–42 (4th ed., text rev. 2000). But on the WAIS–III, for example, the SEM for an individual between the ages of 45 and 54, for the full-scale IQ score (as opposed, for example, to a verbal-only or performance-only scale score) is reported as only 2.23 points. See Am. Ass'n on Mental Retardation, *Mental Retardation: Definition, Classification & Systems of Supports* 51 (10th ed. 2002); see also *Hall v. Florida*, 134 S. Ct. 1986, 1995–96 (2014).

Thus, when experts acknowledge a SEM of “*up to* five points” on widely accepted IQ tests such as the Wechsler (WISC and WAIS series) tests, and a SEM of “*up to* eight points” on “group-administered” tests like the Lorge Thorndike, they are *not* saying that the *maximum gap* between reported score and true score is five (or eight) points, respectively. Nor are they saying that, other than probabilistically, any given reported IQ score should be viewed as being *up to* five (or eight) points higher or lower than the true IQ score. Rather, they are saying that the maximum *standard deviation* between reported score and true score is five (or eight) points—meaning there is *at least* a 68% likelihood that the individual's true score is within five (or eight) points of the reported score.

It is worth noting that “group-administered” tests like the Lorge Thorndike are not really “group tests” in the conventional sense: that is, the questions are not answered orally by groups of individuals. Rather, these tests are administered (much like the SAT or the LSAT) to individuals who each complete an *individual* written IQ test but may do so at the same time as others in a classroom-style setting under the guidance of a single administrator, instead of in a one-on-one setting as Wechsler-series tests (like the WAIS) are administered.

In short, SEM is complicated—and there is no authority that *requires* any adjustment, let alone a *downward* adjustment (when the true IQ score might just as well be *higher* than the reported score) to account for the SEM when analyzing IQ scores as part of an *Atkins* determination.

seventy before age 18.” *Id.* at *28–29. Nevertheless, the district court issued a certificate of appealability, and Black again appealed to our court.

In 2011, however, before we issued an opinion on that appeal, the Tennessee Supreme Court changed course and overruled *Howell*, holding that Tenn. Code Ann. § 39-13-203(a)(1) “does not *require* that raw scores on I.Q. tests be accepted at their face value and that the courts *may consider* competent expert testimony showing that a test score does not accurately reflect a person’s functional I.Q. or that the raw³ I.Q. test score is artificially inflated or deflated.” *Coleman v. State*, 341 S.W.3d 221, 224 (Tenn. 2011) (emphases added).

In light of *Coleman*, over a dissent, we again remanded Black’s *Atkins* claim to the district court. *Black v. Bell*, 664 F.3d 81, 84 (6th Cir. 2011). Even though the Tennessee Court of Criminal Appeals could not have known, at the time it denied Black’s state habeas relief, that the Tennessee Supreme Court would replace *Howell* with its opinion in *Coleman*, we held that the Tennessee Court of Criminal Appeals’ decision was “contrary to the latest Tennessee Supreme Court’s decision on this subject.” *Id.* at 96. And because *Atkins* allowed states to define the contours of *Atkins* itself (such that *Atkins* incorporated *Coleman*, so to speak, for purposes of Black’s claim), we held that the Tennessee Court of Criminal Appeals’ decision was “contrary to clearly established” *federal* “law under [the Antiterrorism and Effective Death Penalty Act (AEDPA)].” *Id.* at 100–01. Thus, because no court had yet evaluated Black’s *Atkins* claim under *Coleman*, we remanded Black’s *Atkins* claim for the district court to analyze it “according to the proper legal standard, which was set out by the Tennessee Supreme Court in *Coleman*.” *Id.* at 101. The district court denied Black’s claim, and for the reasons that follow, we affirm.

II

On remand, the district court conducted a de novo review of Black’s *Atkins* claim. The court accepted new briefing from Black and from the State. Black moved for an evidentiary hearing, and the court denied Black’s motion on the ground that our remand was a limited

³The *Coleman* court discussed “the validity and weight of *raw scores* of intelligence tests.” *Coleman*, 341 S.W.3d at 242 (emphasis added). The court was not referring to actual raw scores but rather to *reported* full-scale IQ scores unadjusted for Flynn Effect, SEM, or other factors.

remand directing the district court to review the record only, placing an evidentiary hearing “beyond the scope of the remand.” R.150. Nevertheless, on January 3, 2013, the district court held oral argument on the merits of Black’s *Atkins* claim, and the district court subsequently issued a 31-page opinion evaluating the record, analyzing the evidence provided by Black’s experts and the State’s experts, and concluding that Black had not “met *his burden* of proving intellectual disability by a preponderance of the evidence.” *Black v. Colson*, No. 3:00-0764, 2013 WL 230664, at *19 (M.D. Tenn. Jan. 22, 2013) (emphasis added).

On appeal, Black contends that the district court erred in perceiving our remand to be a limited remand; erred in denying Black an evidentiary hearing; erred in failing to apply a summary-judgment standard in ruling on Black’s *Atkins* claim; and erred in its merits determination that Black had not met his burden of establishing entitlement to *Atkins* relief. We address each issue in turn.

A. Our Remand Was a Limited Remand

We review the interpretation of our own mandate de novo. *United States v. Parks*, 700 F.3d 775, 777 (6th Cir. 2012). Under the mandate rule, a district court is bound by the scope of the remand issued by our court. *Mason v. Mitchell*, 729 F.3d 545, 550 (6th Cir. 2013); *Scott v. Churchill*, 377 F.3d 565, 570 (6th Cir. 2004). In concluding that we had issued a limited remand, the district court relied on this language from our prior opinion:

A complete review must apply the correct legal standard to all of the relevant evidence in the record. We therefore **VACATE** the district court’s denial of Black’s *Atkins* claim and **REMAND** the case for it to review the record based on the standard set out in *Coleman* and consistent with this opinion.

Black v. Bell, 664 F.3d at 101.

We agree that our remand was limited: the scope of the remand, as expressly stated in this quoted language, was a review of the record under *Coleman*.

Black contends that the district court “erroneously restricted its review to the state court record alone.” Appellant’s Br. 5. When AEDPA deference applies to an *Atkins* claim, the district court would indeed be limited to reviewing the record that was before the state courts.

Cullen v. Pinholster, 563 U.S. 170, 180–81 (2011). Here, however, because Black was entitled to a de novo review of his *Atkins* claim without AEDPA deference, the district court was free to consider the full record before it, including materials that were made part of the federal habeas record after the close of state habeas proceedings. Black argues that the district court “believed that it lacked authority . . . to consider record evidence presented in federal court.” Appellant’s Br. 7. But the record does not support Black’s argument: the district court, to be sure, stated that it was undertaking “a de novo review of the evidence admitted at the post conviction proceeding in state court,” *Black v. Colson*, 2013 WL 230664, at *6, and that it “fully considered the evidence in the state court record,” *id.* at *19, but nowhere in its memorandum opinion did the district court state that it was considering *only* the state-court record, or that it was declining to consider (or otherwise excluding) any of the exhibits that Black had provided to the district court in the course of the federal habeas proceedings.

At oral argument before our court, Black’s counsel stressed that the district court erred by failing to consider certain exhibits, namely the declaration of Dr. Marc J. Tassé, R.120-1, and the declaration of Dr. Stephen Greenspan, R.120-2. But nothing in the record indicates that the district court *didn’t* consider these exhibits—which were made part of the federal habeas record in 2008—when it issued its opinion in 2013. Indeed, at the oral argument before the district court in January 2013, Black’s counsel brought *both* declarations to the attention of the district court, including record citations to each, and the district court in no way indicated that it would decline to examine those items. R.160 at 22 (“I would be remiss to not point out another objective measure of Mr. Black’s adaptive functioning in affidavit of Dr. Ste[ph]en Greenspan. And that’s at Docket Entry 120-2.”); *id.* at 60 (“The Court: Is that what you called the screening test? Ms. Henry: Yes, sir. And you will see in Docket Entry 120-1, there is testimony there from Dr. Mar[c] Tass[é], who is the nation’s leading expert on assessing intelligence.”).

We therefore hold that the district court did not err in apprehending the scope of its remand. The district court understood that its task was to conduct a de novo review of the record before it—including, *at a minimum*, a de novo review of the state-court record applying *Coleman* in the same way that the Tennessee Supreme Court would have done if the *Atkins* claim were instead before that court. And while the district court was not *prohibited* under *Pinholster* from

considering additional evidence beyond the state-court record (because the district court was not subject to AEDPA's constraints), it was not error for the district court not to state whether and to what extent it was considering materials such as Dr. Tassé's and Dr. Greenspan's declarations that were part of the federal habeas record only. Indeed, as noted above, when the district court heard oral argument, it did—without cavil—engage with aspects of the declarations of both Dr. Tassé and Dr. Greenspan.

B. The District Court Did Not Abuse Its Discretion in Denying an Evidentiary Hearing

Relatedly, Black argues that the district court erred in denying him an evidentiary hearing. We review the district court's denial of an evidentiary hearing for abuse of discretion. *Cornwell v. Bradshaw*, 559 F.3d 398, 410 (6th Cir. 2009); *Getsy v. Mitchell*, 495 F.3d 295, 310 (6th Cir. 2007) (en banc). The fact that Black was “not disqualified from receiving an evidentiary hearing under [AEDPA] does not entitle him to one.” *Bowling v. Parker*, 344 F.3d 487, 512 (6th Cir. 2003). Rather, when a court is able to resolve a habeas claim on the record before it, it may do so without holding an evidentiary hearing. *See Sawyer v. Hofbauer*, 299 F.3d 605, 612 (6th Cir. 2002).

Here, the district court did not abuse its discretion in denying Black's motion for an evidentiary hearing. Notably, even if we had authorized the district court to entertain new evidence in evaluating Black's *Atkins* claim, Black has not identified any evidence that he would introduce other than exhibits already made part of the state or federal habeas record. And while Black has cited authorities that support *allowing* an evidentiary hearing, Appellant's Br. 11, 15–16, 26, Black fails to support the contention that an evidentiary hearing was *required* in order for the district court properly to evaluate the voluminous record before it under *Coleman*. At oral argument, Black's counsel argued that an evidentiary hearing would have provided Black an opportunity to direct the court's attention to the findings and conclusions, for example, of post-conviction expert Dr. Tassé. But, as we have stated, Black *was* able to bring Dr. Tassé's declaration to the district court's attention at the oral argument before that court, and, in any event, the district court's task was to review the record in the same way the Tennessee Supreme Court would have reviewed it under *Coleman*—and the district court's thorough 31-page opinion

reflects that it was able to do that within the scope of our limited remand and without conducting an evidentiary hearing.

***C. Principles of Summary Judgment
Do Not Apply to a Merits Ruling on a Federal Habeas Claim***

Black’s brief on appeal makes various assertions that the district court should have applied a summary-judgment standard in conducting its review, but Black cites no authority for this supposed rule—a rule that would mean, it is worth noting, that Black would prevail so long as *any reasonable juror* would grant him relief, giving Black the benefit of all reasonable factual inferences. Appellant’s Br. 5 (“On remand, Black’s request for an evidentiary hearing was denied. The district court erroneously . . . resolved factual disputes in favor of Respondent.”); *id.* 8 (“The district court compounded its error by failing to follow well-settled principles of summary judgment in its memorandum opinion. The district court credited the testimony of the State’s witnesses in the face of the expert opinions of Black’s witnesses. The district court refused to draw inferences in favor of Black. Rather, it did just the opposite.”); *id.* 28–29 (apparently treating the *Atkins* proceeding as a summary-judgment proceeding at which Fed. R. Civ. P. 56 governs because it was “a summary proceeding” without an evidentiary hearing).

Summary-judgment procedures simply do not apply to a federal habeas court’s final adjudication of an *Atkins* claim. Rather, it is Black who had the burden of proving, by a preponderance of the evidence, that he was entitled to relief. See *Parke v. Raley*, 506 U.S. 20, 34 (1992) (discussing “the preponderance of the evidence standard applicable to constitutional claims raised on federal habeas”); Tenn. Code Ann. § 39-12-203(c) (“The burden of production and persuasion to demonstrate intellectual disability by a preponderance of the evidence is on the defendant.”). Part of the confusion in Black’s briefing appears to arise from the fact that the State had filed a “Motion to Dismiss and for Summary Judgment” in the pre-*Coleman* federal habeas proceedings—and indeed, when Black originally filed his petition in 2002, before *Atkins* was decided, the district court granted “summary judgment” to the State on Black’s claims.

But the district court’s decision that Black now appeals was not *summary* judgment—it was judgment. Indeed, nothing in the 2011-13 habeas proceedings leading up to the district court’s January 2013 memorandum opinion was styled “summary judgment” at all: the State

filed a “Brief Opposing [Black’s] *Atkins* Claim,” and Black filed a “Brief In Support Of His *Atkins* Claim,” but nothing in the record appears to justify (and Black does not direct us to anything in the record that would justify) Black’s contention that the district court’s oral argument and opinion constituted a summary-judgment proceeding. Nor is there any support for the proposition that the district court’s *Atkins* determination was transformed into a summary-judgment ruling because the district court declined to hold an evidentiary hearing, as Black’s brief seems to imply. Appellant’s Br. 5. The district court’s *Atkins* determination was a final judgment on the merits of Black’s *Atkins* claim, in which the district court properly weighed the evidence, made credibility determinations, and declared one party the victor.

At such a proceeding, under *Atkins* (as it incorporates state law), Black had to prove every element of his mental-retardation claim “by a preponderance of the evidence,” without receiving the benefit of having any inferences drawn in his favor. Tenn. Code Ann. § 39-12-203(c); *see Coleman*, 341 S.W.3d at 233 (“The statute places the burden on the criminal defendant to prove by a preponderance of the evidence that he or she had an intellectual disability at the time of the offense and requires the trial court rather than the jury to make the decision.”).

We therefore hold that the district court did not err when it resolved the factual disputes before it rather than employing a summary-judgment standard.

D. The District Court’s Merits Ruling Was Correct

We review the district court’s denial of habeas relief de novo. *Bigelow v. Williams*, 367 F.3d 562, 569 (6th Cir. 2004). But we review underlying factual findings for clear error, and we bear in mind that, contrary to the assertions in Black’s brief, Black carries the burden of persuasion:

Our review of the district court’s factual findings is highly deferential. We start from the premise that a district court’s factual findings in a habeas proceeding are reviewed for clear error. *Lucas v. O’Dea*, 179 F.3d 412, 416 (6th Cir. 1999). “‘Clear error’ occurs only when [the panel is] left with the definite and firm conviction that a mistake has been committed. If there are two permissible views of the evidence, the factfinder’s choice between them cannot be clearly erroneous.” *United States v. Kellams*, 26 F.3d 646, 648 (6th Cir. 1994). We are

also mindful that in a habeas proceeding the petitioner “has the burden of establishing his right to federal habeas relief and of proving all facts necessary to show a constitutional violation.” *Romine v. Head*, 253 F.3d 1349, 1357 (11th Cir. 2001).

Caver v. Straub, 349 F.3d 340, 351 (6th Cir. 2003).

The Supreme Court “[e]ft] to the States the task of developing appropriate ways to enforce” its decision in *Atkins*, 536 U.S. at 317, but the Court has invalidated state procedures for evaluating *Atkins* claims when those procedures are “[n]ot aligned with the medical community’s information,” *Moore*, 137 S. Ct. at 1044 (2017) (invalidating Texas scheme where “indicators of intellectual disability [were] an invention of the [Texas Court of Criminal Appeals] untied to any acknowledged source”), and thereby “creat[e] an unacceptable risk that persons with intellectual disability will be executed.” *Ibid.* (quoting *Hall*, 134 S. Ct. at 1990; *see also id.* at 1992 (invalidating Florida scheme that foreclosed “all further exploration of intellectual disability” where prisoner’s seven IQ scores in the evidentiary record were all above 70 (ranging from 71 to 80) and two IQ scores that had been excluded from the record were under 70)).

To prevail on his *Atkins* claim under *Coleman*, Black would need to “prove by a preponderance of the evidence”:

- (1) Significantly subaverage general intellectual functioning as evidenced by a functional intelligence quotient (I.Q.) of seventy (70) or below;
- (2) Deficits in adaptive behavior; and
- (3) The intellectual disability must have been manifested during the developmental period, or by eighteen (18) years of age.

Coleman, 341 S.W.3d at 233 (quoting Tenn. Code Ann. § 39-13-203(a) (2010)).⁴

⁴The only difference between this statute and the 2003 version quoted in Part I, *supra*, is that the term “intellectual disability” replaced the term “mental retardation” in the 2010 version of the statute. In 2014, the Supreme Court in *Hall* used the term “intellectual disability” and acknowledged that previous opinions of the Court had used the term “mental retardation” to describe the same phenomenon. *Hall*, 134 S. Ct. at 1990. But the next year, in *Brumfield v. Cain*, 135 S. Ct. 2269, 2277, 2291 (2015), the Court used both terms in the same decision. Because the vast majority of Black’s legal proceedings transpired before the term “mental retardation” began to fall out of favor, and because *Atkins* itself used “mental retardation,” we have also used that term throughout this opinion, but we use “intellectual disability” in this section because it is the predominant term used by *Coleman*.

Black argues that the district court wrongly concluded that he did not have significantly subaverage general intellectual functioning as evidenced by a functional IQ score of seventy or lower before he turned eighteen. The district court's conclusion largely rested on its analysis of the series of IQ tests that Black has taken over the course of his life, *see Black v. Colson*, 2013 WL 230664, at *6–7, and the crux of Black's argument is that the court wrongly analyzed those IQ scores.

As set forth in Part I, *supra*, Black's school records reveal IQ scores ranging from 83 to 97 when Black was age seven to thirteen. After those tests, the next IQ test on record was administered to Black in 1989 (at age 33) before he stood trial for the triple murder: he scored 76. During Black's first post-conviction proceeding in state court, he was twice administered the WAIS–R (once in 1993 at age 37, once in 1997 at age 41) and scored 73 and 76, respectively. And during federal habeas proceedings (*after* his death sentence had been upheld by the Tennessee courts), Black scored 69 on the WAIS–III and 57 on the Stanford-Binet-IV, both administered in 2001 when Black was 45.

The district court relied strongly on the IQ testing done during Black's school-age years as most probative of Black's mental condition prior to age eighteen. *Id.* at *10. Not surprisingly, Black maintains that this reliance is misplaced. First, Black argues that these test scores are invalid because the tests were “group-administered.”⁵ In the state post-conviction proceedings, Dr. Daniel H. Grant, a neuropsychologist and forensic psychologist, testified that the appropriate mental-health testing models establish that group-administered tests are unreliable and should not be used to determine intellectual disability. Dr. Greenspan's declaration avers that group-administered tests are not acceptable for intellectual-disability determinations because they have much weaker reliability and validity and there is a lack of information about the circumstances under which the tests were administered. And Dr. Tassé's declaration avers that group-administered tests “are not well normed nor possess the psychometric properties necessary to be used in diagnostic decision-making.” Dr. Tassé states that these tests “serve a screening purpose” but that he would not rely upon results from these

⁵As noted in Part I, *supra*, “group-administered” tests are written tests completed by individuals on their own; they are simply administered in a classroom setting as is the case with the SAT or other paper-based standardized tests.

tests “when making or refuting a diagnosis of mental retardation.” Of course, these declarations do not, without more, provide much help for Black: even if Black had persuaded the district court to reject his childhood IQ scores as useful for “making or refuting a diagnosis of mental retardation,” he would still have fallen short of carrying *his burden* to prove that he *was* intellectually disabled by age eighteen.

Moreover, a state expert and psychologist, Dr. Eric Engum, testified during state post-conviction proceedings that group-administered tests *are* relevant when considering whether an individual is intellectually disabled. While agreeing with Dr. Grant that these tests are not as accurate as individually administered tests, Dr. Engum believes that they are properly used as indicators of how well a child is functioning; if the test raised a concern about a child’s intellectual capacity, the child would have been referred for more testing. Although the SEM for group-administered tests is higher (up to eight points) than the SEM for individually administered tests (up to five points),⁶ Black was not referred for more testing (and indeed, Black graduated high school with a standard diploma), and all his childhood test scores would still be well above the numerical threshold for intellectual disability even if they were retroactively adjusted downward by one SEM.

Black next argues that even his adulthood IQ tests administered between 1989 and 1997, the scores from which fall in the low-to-mid 70s, overstate his level of intellectual functioning and that his results should be construed as below 70 when adjusted for the Flynn Effect. At oral argument, Black’s counsel argued that the Supreme Court’s decision in *Brumfield v. Cain*, 135 S. Ct. 2269 (2015), “require[s]” us to look at the “Flynn-adjusted scores” as reported in Dr. Tassé’s report. R.120-2; Oral Argument 25:10-26:00 (discussing *Brumfield* and *Hall*). But neither *Brumfield* nor *Hall* imposes any such requirement—indeed, neither case even mentions the Flynn Effect.

What they do mention is the SEM. *Brumfield*, 135 S. Ct. at 2278 (rejecting the argument “that *Brumfield*’s reported IQ score of 75 somehow demonstrated that he could not possess subaverage intelligence,” where Louisiana law categorically prohibited consideration of factors

⁶See n.2, *supra*.

such as the SEM when a defendant's reported IQ score was above 70); *Hall*, 134 S. Ct. at 1995–96 (“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.”). But as noted above, the SEM accounts for the possibility that an individual’s true IQ score is *either higher or lower* than the reported score. And while the Supreme Court has rejected rigid rules that *prevent* a court from considering evidence of the SEM altogether, *see, e.g., id.* at 1999–2001, the Court’s decisions in no way *require* a reviewing court to *make a downward* variation based on the SEM in *every* IQ score, let alone to do the same with the Flynn Effect.

Further, while the Tennessee Supreme Court in *Coleman* held that “an expert should be *permitted* to base his or her assessment of the defendant’s ‘functional intelligence quotient’ on a consideration of” “a particular test’s standard error of measurement, the Flynn Effect, the practice effect, or other factors affecting the accuracy, reliability, or fairness of the instrument or instruments used to assess or measure the defendant’s I.Q.,” *Coleman* only *requires* a downward adjustment to counteract the Flynn Effect when the IQ test administered to a given individual is an “older version” than the then-current version of the test on the market. *Coleman*, 341 S.W.3d at 242 n.55. Black has not raised any argument that any of his specific IQ scores is *required* to be corrected for the Flynn Effect under *Coleman* because an earlier-normed version of the test was administered.

Rather, Black’s argument is that we should retroactively lower his IQ scores because *his* experts say that we should. Black submitted evidence from various experts about the impact of the Flynn Effect. Dr. Grant testified, for instance (in the state post-conviction hearing), that the Flynn Effect should result in a four-point reduction in his IQ score from the 1993 testing, lowering the score from 73 to 69. Dr. Grant also said that the Flynn Effect should lower the 1997 score by five points from 76 to 71. Dr. Grant also opined that the WAIS–III, administered in 2001, which produced a score of 69, was a more accurate instrument than the WAIS–R and thus produced more accurate results. Dr. Greenspan’s declaration avers that the Flynn Effect would reduce the 1993 test by four points to 69 and the 1997 test by *six* points to 70. Dr. Greenspan also agreed that the 2001 test (with a score of 69) used a more current instrument than previous assessments had. Similarly, Dr. Tassé opined that the Flynn Effect would reduce

Black's 1993 results by four points to 69 and his 1997 results by five points to 71. Dr. Tassé further maintained that the 2001 WAIS–III results should be lowered to a score of 67 due to the Flynn Effect.

On the other hand, the State presented testimony that the impact of the Flynn Effect was overstated by Black's experts. While Dr. Engum was aware of the Flynn Effect and the need to revise and restandardize IQ tests, he questioned the appropriateness of relying on the Flynn Effect to lower IQ scores retroactively based on the passage of time. Dr. Susan Vaught, a neuropsychologist, testified that it was not standard practice to correct scores due to the Flynn Effect nor was it routinely considered by practitioners as a basis for lowering an IQ score. Upon consideration of the parties' evidence (including specific mention of Dr. Grant's, Dr. Engum's, and Dr. Vaught's testimony), the district court concluded that the Flynn Effect provided "weak support for the statutory requirement that [Black] have scores at or below 70 before he turned age 18." *Black v. Colson*, 2013 WL 230664, at *10. The court accepted the existence of the Flynn Effect but concluded that the 1993 and 1997 tests were not as probative of Flynn's mental ability before age eighteen as the earlier tests, and declined to accept Black's argument that retroactively reducing IQ scores was a "scientifically valid remedy" to account for the Flynn Effect. *Ibid.*

Black further argues that the district court should have credited the 2001 IQ tests that placed Black's IQ score at 57 and 69. The district court noted, however, that Black was 45 years old when these tests were administered (and, incidentally, Black was 45 years old before he was ever "diagnosed as having mental retardation," *id.* at *13). The 2001 IQ scores were also generated after Black had been under a sentence of death for more than a decade. Unlike in a competency hearing under *Ford v. Wainwright*, 477 U.S. 399 (1986), where these scores might be probative of a prisoner's insanity *at the time of execution*, these recent scores have far less probative value, if any, in showing Black's mental capacity before he turned eighteen. Black has argued that his mental retardation at age 45 was (unless rebutted by the State) evidence of *lifelong* mental retardation sufficient to satisfy the requirement that mental retardation manifest itself before age 18; indeed, Black presented expert witnesses' findings that Black had a brain

disorder, perhaps caused by fetal alcohol spectrum disorder, but the district court found those experts were “not persuasive.” *Id.* at *14.

Specifically, Dr. Albert Globus, a neuropsychiatrist, examined Black and conducted an extensive review of his past medical records and social history. While he did not conduct any IQ testing, Dr. Globus reviewed recent positron emission tomography (PET) scans of Black’s brain, which revealed “definite abnormalities,” including “changes in the cerebral cortex, the brain ventricles, and the white matter indicating organic damage to the structure of the brain.” Dr. Globus also observed “[h]ypometabolism of glucose in the orbito-frontal cortex, the medial and polar temporal cortex, and the caudate and/or the putamen.” Based on Black’s life history, Dr. Globus opined that Black had an organic brain disorder with an onset well before his current offense. Dr. Globus concluded that these findings were “consistent” with Black’s having an IQ of 70 or lower, which rendered him intellectually disabled—but while Dr. Globus stated that “evidence of early onset brain damage secondary to alcohol ingestion by [Black’s] mother” was “sufficient to produce an IQ lower than all but two or three per cent of the population,” Dr. Globus’s evaluation of Black’s mental ability centered around Black’s *current* ability (in 2001, when Dr. Globus wrote his report). Dr. Globus did *not* affirmatively state that Black’s IQ was 70 or lower before age eighteen.

The district court made several specific page citations to Dr. Globus’s testimony. *See, e.g., id.* at *11. But the district court did not assign great weight to Dr. Globus’s findings because Dr. Globus had not substantiated the facts concerning alcohol use by Black’s mother that Dr. Globus relied upon in his report, and because Dr. Globus admitted that the brain scans that he analyzed did not actually reveal whether Black’s brain abnormalities were caused by fetal alcohol spectrum disorder or instead by an adulthood injury. *Ibid.*

Dr. Ruben Gur, a neuropsychologist, also concluded that Black suffered from a brain disorder. Dr. Gur noted damage in Black’s frontal- and temporal-lobe functions and commented that Black’s “deficits are particularly pronounced in executive functions, memory and emotion processing.” Dr. Gur opined that these limitations potentially resulted from certain exposures during Black’s childhood. These exposures may have included his mother’s alcohol consumption while pregnant with him, or lead poisoning arising from his childhood living

conditions. Black also suffered several head injuries while playing football, although no formal diagnosis of concussion was ever made. At the time of Dr. Gur's report, Dr. Gur noted that Black demonstrated symptoms associated with serious psychiatric disorders, including paranoid and delusional beliefs—but these disorders are *not* necessarily concomitants of mental retardation.

The district court thoroughly evaluated all these reports, and the district court elected to disregard this most recent evidence of Black's mental ability because the district court was not persuaded that any injury that might have caused mental retardation had occurred before Black turned eighteen. *Id.* at *14.

In short, Black's argument requires three steps: (1) reject Black's childhood "group-administered" IQ scores (83, 97, 92, 91, 83); (2) either rely exclusively on the 2001 IQ scores (69, 57), or else apply a downward adjustment to the pre-2001 adulthood IQ scores (76, 73, 76) to account for the Flynn Effect and the SEM, so as to reduce those scores to below 70; and (3) presume that the adulthood scores, in the absence of contradictory childhood IQ scores (and by disregarding evidence put on by the State to rebut Black's contention that his mother's alcohol consumption caused Black to suffer any brain damage that caused any level of mental retardation), are evidence of lifelong mental retardation that must have manifested itself before age eighteen. Each of these three steps is a necessary condition for Black to prevail on his *Atkins* claim as we see it. And Black has not shown us any authority that would support taking *any* of these steps.

At the end of the day, without stronger evidence that Black's childhood IQ scores did not accurately reflect his intellectual functioning before he turned eighteen, the district court held that Black could not carry his burden of showing, by a preponderance of the evidence, that he had significantly subaverage general intellectual functioning before he turned eighteen.

Having reviewed the entire record, we cannot find fault with the district court's conclusion; after all, even if Black's childhood IQ scores were reduced by *both* eight points to account for the SEM (using the higher SEM applicable to group-administered tests, rather than five points for individually administered tests) *and* up to four points to counteract the Flynn

Effect,⁷ they *all* would still exceed seventy. To be sure, there is almost always a *possibility* that a reported IQ score significantly higher than 70 is an inaccurate reflection of a true IQ score of 70 or below—indeed, there is approximately a one-in-300 chance that a reported IQ of 92 on a group-administered test (like Black’s 1966 Lorge Thorndike score) reflects a true score lower than 70. But that possibility does not satisfy Black’s burden to prove his intellectual disability by a preponderance of the evidence.

E. Implications of the Flynn Effect

There is good reason to have pause before retroactively adjusting IQ scores downward to offset the Flynn Effect. As we noted above, see n.1, *supra*, the Flynn Effect describes the apparent rise in IQ scores generated by a given IQ test as time elapses *from the date of that specific test’s standardization*. The reported increase is an average of approximately three points per decade, meaning that for an IQ test normed in 1995, an individual who took that test in 1995 and scored 100 would be expected to score 103 on that same test if taken in 2005, and would be expected to score 106 on that same test in 2015. This does not imply that the individual is “gaining intelligence”: after all, if the same individual, in 2015, took an IQ test that was *normed in 2015*, we would expect him to score 100, and we would consider him to be of the same “average” intelligence that he demonstrated when he scored 100 on the 1995-normed test in 1995. Rather, the Flynn Effect implies that the longer a test has been on the market after initially being normed, the higher (on average) an individual should perform, as compared with how that individual would perform on a more recently normed IQ test.

At first glance, of course, the Flynn Effect is troubling: if scoring 70 on an IQ test in 1995 would have been sufficient to avoid execution, then why shouldn’t a score of 76 on that same test administered in 2015 (which would produce a “Flynn-adjusted” score of 70) likewise suffice to avoid execution? Further, even if IQ tests were routinely restandardized every year or two to reset the mean score to 100, and even if old IQ tests were taken off the market so as to avoid the Flynn Effect “inflation” of scores that is visible when an IQ test continues to be administered

⁷Of Black’s five childhood IQ scores, the 1969 Lorge Thorndike test is the most susceptible to Flynn Effect inflation. The Lorge Thorndike test was published in 1957, so a reduction of the 1969 score by approximately four points would offset the maximum expected inflation of that score that would be attributable to the Flynn Effect.

long after its initial standardization, that would only mask, but not change, the fact that IQ scores are said to be rising.

Indeed, perhaps the most puzzling aspect of the Flynn Effect is that it is true. As Dr. Tassé states in his declaration, “[t]he so-called ‘Flynn Effect’ is NOT a theory. It is a well-established scientific fact that the US population is gaining an average of 3 full-scale IQ points per decade.” The implications of the Flynn Effect over a longer period of time are jarring: consider a cohort of individuals who, in 1917, took an IQ test that was normed in 1917 and received “normal” scores (say, 100, on average). If we could transport that same cohort of individuals to the present day, we would expect their average score today on an IQ test normed in 2017—a century later—to be thirty points lower: 70, making them mentally retarded, on average.

Alternatively, consider a cohort of individuals who, in 2017, took an IQ test that was normed in 2017 and received “normal” scores (of 100, on average). If we could transport that same cohort of individuals to a century ago, we would expect that their average score on a test normed in 1917 would be thirty points higher: 130, making them geniuses, on average.

It thus makes little sense to use Flynn-adjusted IQ scores to determine whether a criminal is sufficiently intellectually disabled to be exempt from the death penalty. After all, if *Atkins* stands for the proposition that someone with an IQ score of 70 or lower in 2002 (when *Atkins* was decided) is exempt from the death penalty, then the use of Flynn-adjusted IQ scores would conceivably lead to the conclusion that, within the next few decades, almost no one with borderline or merely below-average IQ scores should be executed, because their scores when adjusted downward to 2002 levels would be below 70. Indeed, the Supreme Court did not amplify just what moral or medical theory led to the highly general language that it used in *Atkins* when it prohibited the imposition of a death sentence for criminals who are “so impaired as to fall within the range of mentally retarded offenders about whom there is a national consensus,” 536 U.S. at 317. If *Atkins* had been a 1917 case, the majority of the population now living—if we were to apply downward adjustments to their IQ scores to offset the Flynn Effect from 1917 until now—would be too mentally retarded to be executed; and until the Supreme Court tells us that it is committed to making such downward adjustments, we decline to do so.

III

Because Black cannot show that he has significantly subaverage general intellectual functioning that manifested before Black turned eighteen, we need not analyze whether Black has the requisite deficits in adaptive behavior, which he would *also* be required to demonstrate in order to be entitled to *Atkins* relief.

IV

In sum, the district court did not err in denying Black's *Atkins* claim under the applicable standard set forth by the Tennessee Supreme Court in *Coleman*.

AFFIRMED.

CONCURRENCE

COLE, Chief Judge, concurring in the opinion except for Section II.E. I concur with the majority opinion except as to the section discussing the implications of the Flynn Effect. In holding that Black did not prove that he had significantly subaverage general intellectual functioning, we concluded that Black's childhood IQ scores would be above 70 even if we adjusted those scores to account for both the SEM and the Flynn Effect. Accordingly, I would not address the question of whether we should apply a Flynn Effect adjustment in cases generally because it is unnecessary to the resolution of Black's appeal. Regardless, courts, including our own in *Black I*, have regarded the Flynn Effect as an important consideration in determining who qualifies as intellectually disabled. *See, e.g., Black v. Bell*, 664 F.3d 81, 95–96 (6th Cir. 2011); *Walker v. True*, 399 F.3d 315, 322–23 (4th Cir. 2005).