Psychology, Public Policy, and Law
2006, Vol. 12, No. 2, 170–189

1076-8971/06/$12.00 DOI: 10.1037/1076-8971.12.2.170

TETHERING THE ELEPHANT
Capital Cases, IQ, and the Flynn Effect

James R. Flynn
University of Otago

Capital offenders cannot be executed if they are mentally retarded. Therefore, the IQ scores of offenders are important, and the U.S. 4th Circuit Court of Appeals has held that the Flynn effect is relevant to interpreting their IQ scores. The Flynn effect (IQ gains over time) means that different IQ tests will give different scores purely as a result of when the tests were normed. Because execution must not be a random result of what test defendants take, a formula is provided to convert IQ scores to a common metric: the norms current at the time the test was taken. The formula also includes a correction based on evidence that the Wechsler Adult Intelligence Scale—Third Edition inflates IQs because of sampling error. Given the inevitability that opposing experts will offer conflicting diagnoses, IQ scores merit special attention in capital cases.

Keywords: death penalty, mental retardation, IQ gains

> These death sentences are cruel and unusual in the same way that being struck by lightning is cruel and unusual. . . . The Eighth and Fourteenth Amendments cannot tolerate the infliction of a sentence of death under legal systems that permit this unique penalty to be so wantonly and so freakishly imposed. (Justice Stewart, in *Furman v. Georgia,* 1972)

> It is the mark of an educated man . . . that in every subject he looks for only so much precision as its nature permits. (Aristotle, *Nichomachean Ethics*, 1, 3, 1094b, 24–26)

In *Furman v. Georgia* (1972), on behalf of the U.S. Supreme Court, Justice Stewart argued that the death penalty must be imposed with consistency and with due regard to the culpability of those who suffer its consequences. Thirty years later, in *Atkins v. Virginia* (2002), the Court held that the Eighth Amendment to the U.S. Constitution forbids the death penalty for those with mental retardation (MR). Subsequently, in *Walker v. True* (2005), the Fourth Circuit Court of Appeals held that in applying this standard, the "Flynn effect" had to be taken into account if it could be shown that it had affected the defendant's IQ score.

Because this effect bears my name, I have been approached by legal counsels in over a dozen postconviction cases and have sworn 11 affidavits in five states: Most cases were assessments of the IQ scores of defendants on death row; one was in support of a submission to the Florida Supreme Court advocating the relevance of the Flynn effect; another was a declaration requested by counsel for Walker (*Walker v. True,* 2005) when his case was returned to the court of appeals for decision.

Correspondence concerning this article should be addressed to James R. Flynn, Department of Political Studies, University of Otago, P.O. Box 56, Dunedin, New Zealand. E-mail: jim.flynn@stonebow.otago.ac.nz

Similar submissions by others have been welcomed by the courts. In California, the court in *People v. Superior Court* (2005) said that the Flynn effect must be considered in determining a defendant's IQ and noted that its relevance appeared to be generally accepted in the clinical field. Over the last year or so, two federal courts of appeal and eight courts in six states have discussed its relevance (*Black v. State,* 2005; *Bowling v. Commonwealth,* 2005; *Ex parte Murphy,* 2006; *In re Hicks,* 2004; *Myers v. State,* 2005; *People v. Superior Court,* 2005; *State v. Burke,* 2005; *State v. Murphy,* 2005; *Walker v. True,* 2005; *Walton v. Johnson,* 2005). In the most recent of these (*Ex parte Murphy,* 2006), despite a defendant with an IQ of 81 on the Wechsler Adult Intelligence Scale—Revised (WAIS–R), the appellate court granted a stay of execution directing the trial court to resolve the issue of MR. How a defendant with a measured IQ of 81 might be classified as mentally retarded will become apparent.

Although I have no great desire to discourage attorneys from paying for my expert opinion, it seems propitious to make public some of the considerations and data that must be assimilated if the Flynn effect is to clarify rather than confuse. Hereafter I shall refer to it by its descriptive name: IQ gains over time.

Relevance of IQ Gains Over Time

Most states that apply the death penalty have adopted a criterion for MR as both poor adaptive functioning and an IQ score of at least two standard deviations below the population mean, namely, an IQ of 70 or below. They have also recognized the possibility of measurement error in any particular test administration. This allows for the fact that a person's performance may vary from day to day because of his or her mood, general health, and so forth. Where there is only one test score on record, the effective criterion can be as high as 75.

Arguably, the court-administered criterion is more stringent than that set by the American Association on Mental Retardation (AAMR). First, the AAMR (2002, p. 58) recommends a score of approximately 70 to 75 irrespective of measurement error. This means that when there is a single testing, caution dictates a criterion of 75 to 80. Second, the AAMR (2002, p. 58) castigates states such as Connecticut for adopting fixed IQ scores as cutoff points for MR and stresses that IQ scores can be overridden by poor adaptive behavior and clinical judgment. The implication is that defendants who are classified as mentally retarded on behavioral criteria should not in addition have to produce a case based on IQ. However, defense attorneys find that the clinical judgments of prosecution and defense psychologists are almost always at variance and feel constrained to make a case in terms of IQ in order to be taken seriously. Both the courts and the AAMR specify that MR should be apparent during the defendant's developmental period, that is, prior to the age of 18.

The question of poor adaptive functioning usually turns on such sources of information as interviews and examinations of the offender by school or clinical psychologists, medical histories, failure to qualify for a driver's license or hold anything but an unskilled job, and testimony by family and friends as to degree of suggestibility. An IQ score at least two standard deviations below the mean carries weight because it isolates the bottom 2.27% of the defendant's age group. In practice, cases tend to fall into one of three categories:

1. During the defendant's school years, there was at least one clear diagnosis of MR consequent on inability to cope and a series of IQ scores at 70 or below.

2. The defendant was evaluated at school and was not diagnosed as mentally retarded. However, there is reason to believe that the diagnosis was affected by the fact that IQ scores were assumed to show that the defendant was less than two standard deviations below the mean when, in fact, they signaled the opposite.

3. The defendant never received a formal diagnosis prior to age 18, and therefore, prison diagnosis as an adult becomes crucial. Once again, IQ scores play a dual role: They must be interpreted properly to determine whether the IQ criterion of two standard deviations below the mean is met; and they must be interpreted properly so that clinical judgment of adaptive behavior is not influenced, perhaps subconsciously, by an inflated estimate of the defendant's standing relative to his or her peers.

The second and third categories pose the question of how a school psychologist or a court-appointed psychologist could be mistaken about whether an IQ score was two standard deviations below that of the average person. The answer is that all a score of 70 tells us is that someone was two standard deviations below the average person in a standardization sample, namely, the sample drawn to set the norms for the test that the person took. If the sample included only Americans with no more than a grade-school education, its norms would be too low. Someone who took that test and received a score of 85, a score one standard deviation below the average of those who set its norms, might actually be two standard deviations below the average American and might deserve a score of 70.

Imagine that there are IQ gains over time that run at a rate of 0.3 IQ points per year and therefore, over 50 years, total 15 points. That means that even if a sample were perfectly representative of Americans in 1950, it is no longer representative circa 2000. Indeed, in that context, its norms would be far too low. Assume that a schoolchild aged 10 in 2000 is two standard deviations below contemporary 10-year-olds and therefore deserves an IQ of 70. If you give that child a test 50 years out of date, he or she will get a score of 85 and appear to be only one standard deviation below average. Naturally, judges want to know whether defendants were actually two standard deviations below their peers at the time they were tested and not how they rank against a group selected at some random date in the past. Therefore, the question of IQ gains over time cannot be ignored. Presumably, judges will also wish to know whether some IQ tests produce suspect scores in the sense that they rank people against a group that was unrepresentative even at the time it was selected. After all, as shown above, substandard samples are simply deceptive. They engender test scores that indicate that someone is less than two standard deviations below the population mean when, in fact, their performance ranks them at that level or below.

Recognition of the Relevance of IQ Gains

With the publication of the first systematic study of IQ gains in America (Flynn, 1984b), scholars began to adjust test results in the light of those gains. An early example was the famous Milwaukee Project. Garber (1988) had administered a variety of IQ tests to children who had been exposed to an enriched environment to estimate whether the intervention had reduced the risk of their being classified as mentally retarded. Since the tests had been normed at different dates, they produced radically higher or lower IQs until the scores were adjusted to take into account which tests had been normed recently and which had been normed as much as 30 years before. Since 1998, when the American Psychological Association issued *The Rising Curve: Long-Term Gains in IQ and Related Measures* (Neisser, 1998), no scholar published in a first-line journal has ignored the relevant data on IQ gains over time. Recent examples include *IQ and the Wealth of Nations* (Lynn & Vanhanen, 2002) and the study headed by the noted scholar Naomi Breslau, tracking the relationship between low birth weight, social disadvantage, and IQ during the years of school attendance (Breslau, Dickens, Flynn, Peterson, & Lucia, in press).

The problem of adjusting IQ scores across nations and across different kinds of tests is complex. The rate of gain varies by country—for example, from America to Israel to Norway—and by test—for example, from matrix tests to purely verbal IQ tests (Flynn, 2006). Those who deal with the scores of Americans on Wechsler and Stanford–Binet (SB) IQ tests are fortunate in that the rate of gain is roughly uniform. The standard practice is to deduct 0.3 IQ points per year (3 points per decade) to cover the period between the year the test was normed and the year in which the subject took the test. The evidence for that value is discussed in a subsequent section of this article. For now, take it for granted that the relevant rate of gain is 0.3 IQ points per year.

Why do scholars make these adjustments? Imagine you are running an educational project whose objective is to raise the IQs of young children in a depressed area. You give one IQ test before they enter the program at age 4 and another when they leave it at age 6. They appear to have gained 10 IQ points. Then you notice that the first test was normed 5 years before you administered it, the second 25 years before. Thanks to IQ gains over time, the entry test was normed on a better performing sample. This is clear because studies in which the same children took both tests reveal the following pattern: They scored 100 on the recent test, which shows that their performance was merely average; and yet, the very same children receive an IQ of 106 when scored on the norms of the older test. In other words, a sample of children drawn 20 years ago set a lower standard than children do today. The 6 points higher IQ is purely a matter of being scored against an earlier and lower performing group. Naturally, what concerns you is how much the experimental children really improved in their IQ performance. Therefore, you deduct 6 points from the IQ scores they received when they left the program: 0.3 points for each of the 20 years between the norming of the two tests. This shows that the experimental children really gained only 4 points from participating in your project, not the 10 points the unadjusted IQ scores suggested.

Failure to adjust the scores is to take flight from reality. Suppose you are coaching an athlete who aspires to qualify for the Olympic high jump. He jumps 6 ft 6 in., and you assure him that he will qualify. He replies: “But that was the standard in 1985. Since then, performances have improved, and today, I have to jump 7 feet to qualify. You are judging my performance in terms of the norms of yesterday rather than today.” He would do well to hire a new coach.

Clinical and school psychologists have been slow to recognize the problems posed when a new IQ test replaces an older one (for full names of all tests, see Table 1). Kanaya, Scullin, and Ceci (2003) found that after the Wechsler Intelligence Scale for Children—Third Edition (WISC–III) was published in 1991, there was an immediate rise in the number of children being classified as mentally retarded. This was because some school psychologists had begun to use the new test (normed in 1989) rather than the older WISC–R (normed in 1972). Thanks to IQ gains over time, children in the retardate range who took the new test were averaging 5.6 IQ points lower than those who were still taking the old test. Note, by the way, how close 5.6 points is to the 5.1 points entailed by a gain of 0.3 points per year over the 17 years between the selection of the two standardization samples. Failure to adjust for the difference between the two tests meant that beginning in 1992, whether a child was classified as mentally retarded was likely to depend on whether school psychologists had used up their stock of the old WISC–R test or had purchased the new WISC–III test.

The AAMR (2002) has cited the Flynn effect as evidence that it "is critically important to use standardized tests with the most updated norms" (p. 56) and notes score changes "between successive revisions" (p. 59) of IQ tests. It does not discuss the mechanics of translating scores on an old version of the WISC into scores on a current one. After all, if you are a practicing school psychologist and see a child with an obsolete IQ score, the obvious remedy is to administer the current test. However, when a court wants to establish whether defendants in their 40s showed signs of MR before the age of 18, they cannot regress them to childhood and administer a new test. If the tests taken had current norms at the time they were taken, all is well. But if the norms were obsolete, all that can be done is to reassess the IQ scores recorded in the light of IQ gains over time.

The California court cited in *People v. Superior Court* (2005) in California goes further than I would in asserting that the Flynn effect seems to be generally accepted in the clinical field. However, the comments of Trowbridge (2003) and Frumkin (2003) on how to deal with expert testimony about IQ scores show that the formula of deducting 0.3 points per year is making headway in capital cases.

IQ Gains and Capital Cases

The case for taking IQ gains into account in capital cases is best presented by noting the consequences of failing to do so. Consider a boy whose IQ was tested twice during his school years. In 1973, at age 6, he scored 75 on the WISC, which was normed in 1947–1948 and whose norms were 25.5 years out of date. In 1975, at age 8, he scored 68 on the new WISC–R, which was normed in 1972 and whose norms were only 3 years out of date. Both scores show that this boy was mentally retarded: The score at age 6 is deceptive because it was inflated by norms 25.5 years old, which is to say by over 7 IQ points. When those points are deducted, both scores become 68. The boy's true IQ did not alter between ages 6 and 8: All that changed were the norms.

Now imagine we split this example into two boys who are today subject to the death penalty. At school, they performed identically on IQ tests. But one is executed because he was unlucky enough to take the WISC and scored 75 when measured against its obsolete norms; the other is excused as mentally retarded because he was lucky enough to take the WISC–R and scored 68 when measured against its relatively current norms. Failure to adjust IQ scores in the light of IQ gains over time turns eligibility for execution into a lottery—a matter of luck about what test a school psychologist happened to administer.

IQ Scores and Criteria of MR

Prosecution experts sometimes claim that psychologists are still evaluating the significance of the Flynn effect. This is true. But it does not mean that we should wait for some distant day before applying our knowledge of the rate of IQ gains to adjust the IQ scores of defendants in current cases.

Sooner or later, we must decide whether retardate IQ gains represent enhanced ability to deal with everyday life and enhanced ability to make decisions for which one is culpable. If IQ gains mean that fewer and fewer Americans lack those capabilities, the AAMR criterion of MR should be altered. Perhaps it should be set at three standard deviations below the population mean (an IQ of 55 against current norms) rather than two standard deviations (an IQ of 70). I will argue in favor of the status quo.

First, we now have data that compare the performance of samples used to standardize the Vineland Adaptive Behavior Scales at two different times (Sparrow, 2006). For ages 7 to 18, subjects who took both tests actually found the 1984 norms more difficult to meet than current norms. That is, they received an overall Adaptive Behavior Composite of only 95.0 on the old test and one of 98.4 on the new test ($SD = 15$). This seems to indicate that American schoolchildren actually lost ground in terms of adaptive behavior since 1984.

However, their scores on the Communication and Socialization subtests were similar on the two versions. The lost ground was almost entirely on the Daily Living Skills subtest. The 1984 version of that subtest contains obsolete skills that would deflate the scores of contemporary children (items such as "sews or hems clothes," "makes own bed," and "uses a pay telephone"). The most judicious conclusion is that American children have marked time. In sum, they made no gains in terms of adaptive behavior during a period in which they made large IQ gains. This result suggests that IQ gains over time do not mean that fewer and fewer children find it difficult to cope with everyday life.

Second, IQ gains over time show a peculiar pattern (Flynn, 2006): They are huge on some WISC subtests, such as Verbal Similarities and Block Design, and virtually nil on others subtests, such as Vocabulary and General Information. I believe that IQ gains are measuring real skill gains but that these gains do not constitute global intelligence gains. The rise in Verbal Similarities scores signals the penetration of the scientific ethos into American minds. If asked what dogs and rabbits have in common, the prescientific children of 1900 tended to say, "you use dogs to hunt rabbits," a satisfactory concrete operational answer but one for which they would score zero. The postscientific children of today will say that they are both mammals, a formal taxonic answer for which they will score full marks. The visual culture of today has upgraded performance on Block Design. And it is important that America has more citizens who can see the world through scientific spectacles and can deal with the Internet. But once again, these gains do not mean that people of low intelligence are any better at concrete operational decisions in everyday life. They certainly have no greater fund of general information to inform those decisions and no greater core vocabulary to verbalize them.

Third, assume I am mistaken and that IQ gains in fact signal some sort of global intelligence gain. The main environment to which a low-IQ person must adapt is the expectations and behavior of other people. The more intelligent the average child becomes, the greater the cognitive demands we make on all schoolchildren. The more intelligent the average worker becomes, the less likely an employer will be to hire those who need a foreman to stand over them and tell them what to do. And what of the culpability of the low-IQ person? Cases often involve a ringleader of normal intelligence who has recruited someone of low IQ to participate in an armed robbery. The ringleader kills someone, but the low-IQ person present is an accessory. If everyone is getting brighter, low-IQ persons may be a bit more resistant to being manipulated, but the average person is also more persuasive in attempting to manipulate them.

For example, I was recently asked to comment on a case in which the companions of the defendant's youth had submitted affidavits. One testified that when the defendant and he were both 16 years of age, the companion pointed out a car and said: "That is Mrs. Smith's car. She and I are friends and she said I could borrow her tires. Would you go over and take them off and bring them to me?"

I believe that jurists should allow the AAMR to decide issues concerning what IQ score should serve as a criterion of MR. Whatever its decision, the salient point is this: No matter whether the criterion is set at an IQ of 55 or 70 or 85, the defendant must be assessed against current norms and not obsolete norms that inflate his or her score. Otherwise, one person will meet the criterion of mental retardation, and another person will be judged not to have done so, purely because one took a test with current norms and the other took a test with obsolete norms. No matter what the criterion, who meets it must not be a matter of chance. Someone who took the WISC and reaped an "unearned" bonus of 7 IQ points must not be executed, whereas someone who was lucky enough to take the WISC–R is exempt. The only way to avoid such inequity is to adjust IQ scores in the light of IQ gains over time.

IQ Gains and Evidence

Table 1 presents evidence showing that IQ gains, at least in the United States on Wechsler–Binet tests, have proceeded at the rate of 0.3 points per year or 3 points per decade. Prior to the SB–5, the SB standard deviation was set at 16 IQ points, which means that 68 (two standard deviations below the mean) was the cutting line for MR. In Table 1, all scores inclusive of older SB scores have been adjusted to a common metric of $SD = 15$, so that IQ gains can be calculated without ambiguity. All test comparisons are derived by analysis of data provided in the Wechsler and SB publishers' manuals. In every instance, they gave the same subjects two tests normed some years apart. The comparisons of WISC and WAIS are based on subjects ages 16–17, the age at which the norms of the two tests overlap. IQ gains from one test to the next have been divided by the number of years that separated the norming of the two tests. This gives us 12 estimates of the rate of IQ gains over time covering the period 1972 to 2002.

When the 12 estimates are averaged, they yield a rate of gain of 0.300 points per year plus or minus less than one hundredth of a point. The slight ambiguity

Table 1
Using Recent IQ Gains to Evaluate the WAIS–III

Tests compared	Gains	Period (yrs)	Rate
Comparisons without WAIS–III			
WISC–R (1972) & WAIS–R (1978)[a]	+0.90	6	+0.150
SB–LM (1972) & SB–4 (1985)[b]	+2.16[c]	13	+0.166
WISC–R (1972) & SB–4 (1985)[d]	+2.95[c]	13	+0.227
WISC–R (1972) & WISC–III (1989)[e]	+5.30	17	+0.312
WAIS–R (1978) & SB–4 (1985)[f]	+3.42[c]	7	+0.489
SB–4 (1885) & SB–5 (2001)[g]	+2.77[c]	16	+0.173
WISC–III (1989) & WISC–III/IV[h] (2001.75)[i]	+4.23	12.75	+0.332
WISC–III (1989) & SB–5 (2001)[j]	+5.00	12	+0.417
Comparisons with WAIS–III			
WAIS–R (1978) & WAIS–III (1995)[k]	+2.90	17	+0.171
WISC–III (1989) & WAIS–III (1995)[l]	−0.70	6	−0.117
WAIS–III (1995) & SB–5 (2001)[m]	+5.50	6	+0.917
WAIS–III (1995) & WISC–IV (2001.75)[n]	+3.10	6.75	+0.459
Comparisons with WAIS–III scores lowered by 2.34 IQ points			
WAIS–R (1978) & WAIS–III (1995)	+5.24	17	+0.308
WISC–III (1989) & WAIS–III (1995)	+1.64	6	+0.273
WAIS–III (1995) & SB–5 (2001)	+3.16	6	+0.527
WAIS–III (1995) & WISC–IV (2001.75)	+0.76	6.75	+0.113

Note. All dates assigned to tests refer to the date at which the test was normed. This is what is relevant, of course, not the date when the test was published. Another date that practitioners might like to have is that for the norming of the WISC: from 1947 to 1948. The data from all sources listed below have been analyzed and adapted to elicit comparisons.
[a]Wechsler Intelligence Scale for Children—Revised (WISC–R) & Wechsler Adult Intelligence Scale—Revised (WAIS–R): Wechsler (1981), Table 18. [b]Stanford–Binet (SB) LM & SB–4: Thorndike et al. (1986), Table 6.6. [c]Prior to the SB–5, the SB standard deviation was set at 16 IQ points, rather than the usual 15 points. The above estimates are all based on scores adjusted to a common metric of $SD = 15$. [d]WISC–R & SB–4: Thorndike et al. (1986), Table 6.7. [e]WISC–R & WISC–III: Flynn (1998), Table 1. [f]WAIS–R & SB–4: Thorndike et al. (1986), Table 6.9. [g]SB–4 & SB–5: Roid (2003), Table 4.1. [h]The alert reader will have noticed the peculiar label given in the comparison of the WISC–III with its successor test, namely, that the latter is called the WISC–III/IV. The WISC–IV dropped 5 of the 10 subtests of the WISC–III, and this rendered the two noncomparable in terms of estimating the rate at which the WISC–III had become obsolete. Fortunately, the Psychological Corporation had collected special data (see Note i below) that offered a solution. Flynn and Weiss (of the Psychological Corporation) used that data to simulate how the WISC–IV standardization sample would have performed on the unaltered WISC–III. They found that IQ scores would have been at least 1.33 points lower than the WISC–IV yielded—thus, the odd label WISC-III/IV, which refers to using a WISC–IV sample to assess norms for a test like the WISC–III. [i]WISC–III & WISC–III/IV: Flynn & Weiss (2006). The estimate given is the midpoint of the range of estimates for this pair of tests. [j]WISC–III & SB–5: Roid (2003), Table 4.6. [k]WAIS–R & WAIS–III: Wechsler (1997), Table 4. [l]WISC–III & WAIS–III: Wechsler (1997), Table 4.3. [m]WAIS–III & SB–5: Roid (2003), Table 4.7. [n]WAIS–III & WISC–IV: Psychological Corporation (2003), Table 5.12.

has to do with whether WAIS–III scores should be allowed to stand or whether they must be adjusted because its norms are suspect. When those who have committed capital crimes are tested postconviction, and they always are, a recognized test with norms for adults must be used. The only ones available are the SB–5 and the WAIS–III, with the latter the most popular choice. Therefore, any problem with its norms will have serious consequences.

The WAIS–III problem arises out of its effects within the total matrix of comparisons. Assume a sequence in which candidates average an IQ of 106 on Test A (normed 1980), 103 on Test B (normed 1990), and 100 on Test C (normed 2000). That would indicate a steady IQ gain of 0.3 points per year, as the norms get tougher to exceed over time. However, imagine that Test B disrupts this pattern wherever it appears—for example, candidates average 106 on Test A, 106 on Test B, and 100 on Test C. The effect is that wherever Test B occurs as the second test in a combination, it deflates the rate of gain to zero: The combination of A followed by B gives no gains over 10 years. And wherever Test B occurs as the first test in a combination, it inflates the rate of gain to improbable levels: The combination of B followed by C gives a huge gain of 6 points over 10 years. We would have a choice between concluding that the rate of gain was wildly eccentric and concluding that there was something wrong with the norms of Test B, namely, that its standardization sample was below average and thus inflated IQs by 3 points.

In accord with this logic, Table 1 is divided into three sections. The first section isolates the eight comparisons in which the WAIS–III does not appear. These give an average rate of gain of 0.283 points per year with a relatively narrow range of estimates (from 0.150 points per year to 0.489 points), particularly if one takes into account that no one estimate should be trusted. The second section isolates the four comparisons in which the WAIS–III does appear. These give an average rate of gain of 0.358 points per year with a huge range of estimates, running all the way from a "minus gain" (a loss of 0.117 points per year) to a gain of almost an IQ point per year—the latter far too large to be taken seriously. Looking at what is happening in those four comparisons, it is clear that the WAIS–III inflated IQ scores by about 2.34 points even at the time it was normed. That is, its normative sample was a bit substandard compared with other IQ tests, and therefore, its average performance was easier to beat. A performance that would be an IQ of 100 on the others becomes an IQ of 102.34 on the WAIS–III.

The third section of the table delivers the proof. When all WAIS–III scores are lowered by 2.34 points, the four comparisons in which the WAIS–III appears fall into line. They average rate of gain of falls from 0.358 points per year to 0.305 points per year, which is virtually identical to the rate yielded by the average of all 12 comparisons. And the wide range of estimates narrows to virtual parity with comparisons in which the WAIS–III does not appear.

In sum, when its scores are left unadjusted, the WAIS–III is behaving just like Test B. Whenever it is the later test, it produces tiny rates of gain or even losses (note the values of 0.171 and −0.117 in the second section of Table 1)—by inflating IQs. Whenever it is the earlier test, it produces huge rates of gain (note the values of 0.917 and 0.459 in the same section)—also by inflating IQs. When the WAIS–III was published in 1997, the results from that time to the present were of course unavailable. Therefore, it prompted speculation that the rate of IQ gains might have diminished (Flynn, 1998); that speculation is now seen to be amiss.

Table 1 shows only that the WAIS–III's norms are atypical. However, the alternative to considering that its norms are suspect is to doubt the norms of all four of the other tests that occur in comparison with it: the WAIS–R, the WISC–III, the SB–5, and the WISC–IV. It is logically possible that rather than the WAIS–III team selecting a substandard sample, the architects of all these other tests selected elite samples. If that is so, all tests but the WAIS–III are deflating IQs and many are being labeled as mentally retarded who should not qualify. But the odds are 15 to 1 against it. Table 1 suggests a general rule: Deduct 0.3 IQ points per year from the scores of defendants for every year that passed between when the test was normed and when the test was taken. It also suggests an addendum: Deduct an extra 2.34 points from WAIS–III scores on the grounds that it gave inflated IQs even in the year in which it was normed.

IQ Gains and Low-IQ Subjects

Before the general rule is applied, the question naturally arises as to whether IQ gains at low-IQ levels are similar in magnitude to those at the mean. If not, the rule of 0.3 points of obsolescence per year might have to be modified for those on the borderline of retardation—for example, those who score 75 on an obsolete test and need a deduction of 5 points to qualify as mentally retarded. Flynn (1985, 1998) and Spitz (1989) have collected large bodies of data at low-IQ levels. These have been supplemented by studies done by the testers themselves and by Kanaya et al. (2003).

Table 2 is not designed to allow conversion of IQs from one test to another but to use combinations of tests and their standardization samples to estimate the rate at which the norms of preceding tests became obsolete. The most important column is that headed "Predicted versus actual" which compares the predicted obsolescence (given by the general rule) with the actual value yielded by the data. The predicted obsolescence is based on taking the years between the standardization of each pair of tests and multiplying by 0.3 IQ points. Using the WISC tests as an example yields the following results: The WISC and WISC–R were normed 24.5 years apart, so multiplying by 0.3 gives 7.35 points as the predicted obsolescence; the WISC–R and WISC–III were normed 17 years apart, which gives 5.10 points; the WISC–III and WISC–IV were normed 12.75 years apart, which gives 3.825 points; and so forth for all of the other test included in the table.

Table 2 shows that the WISC tests behave with remarkable consistency. The data sets are large, and at every IQ level from 55 to 100, the predicted and actual values are very close. If anything, at low-IQ levels, the original WISC norms have become obsolete a bit faster (at about 0.4 points per year) than the rate set by the general rule (0.3 points).

But why are the WAIS and WAIS–R data different? Particularly at very low-IQ levels, we find what appear to be IQ losses rather than IQ gains. Throughout the world, in the relevant period (1953–1954 to 1978), IQ gains have occurred on all manner of tests, and the only controversy has been whether the obvious gains in the bottom half of the curve were matched by gains in the top half

Table 2
IQ Differences by IQ Level: Predicted Versus Actual Score Differences

WISC & WISC–R[a]		WISC–R & WISC–III		WISC–III & WISC–III/IV[b]	
IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual
90–100 = 81–91	7.35/8.52	90–110 = 86–104	5.10/5.02	100 = 96–98	3.825/4.33
80–90 = 71–81	7.35/8.89	80–90 = 75–85	5.10/5.33	85 = 81–84	3.825/3.83
70–80 = 61–71	7.35/8.52	75–80 = 70–75	5.10/4.78	70 = 65–70	3.825/3.83
55–70 = 45–60	7.35/9.70	60–75 = 53–68	5.10/6.77	55 = 49–56	3.825/3.83

WAIS & WAIS–R		WAIS-R & WAIS–III	
IQ differences by level	Predicted vs. actual	IQ differences by level	Predicted vs. actual[c]
80–100 = 73–93	7.35/6.90	100 = 96–98	5.10/5.34 (3.00 + 2.34)
[70–80 = 67–77][d]	nil/3.47	85 = 81–83	5.10/5.34 (3.00 + 2.34)
65–70 = —	—	70 = 65–69	5.10/5.34 (3.00 + 2.34)
[55–65 = 56–66][d]	nil/−1.23	55 = 49–54	5.10/5.84 (3.50 + 2.34)
[40–55 = 46–61][d]	nil/−6.28	—	—

Note. Dashes indicate that no data were available. Numbers for each comparison (running from higher to lower IQs) are as follows: WISC & WISC–R, 326, 415, 259, 170; WISC–R & WISC–III, 652, 1055, 215, 108; WISC–III & WISC–III/IV, 244 (only total for all subjects at all IQ levels is given); WAIS & WAIS–R, 2 studies, 1 study, —, 3 studies, 1 study. The average study had 52 subjects and none had under 20; WAIS–R & WAIS–III, 192 (only total for all subjects at all IQ levels is given). The data from the following sources have been analyzed and adapted: WISC & WISC–R: Flynn, 1985, Table 2; WISC–R & WISC–III: Zimmerman & Woo-Sam, 1997, Table 1, and Kanaya et al., 2003; WISC–III & WISC–III/IV: Psychological Corporation, 2003, Table 5.9; WAIS & WAIS–R: Spitz, 1989, Table 1; WAIS–R & WAIS–III: Wechsler, 1997, Table 4.1. WISC = Wechsler Intelligence Scale for Children; WAIS = Wechsler Adult Intelligence Scale.
[a]WISC–R scores adjusted to simulate scoring against the White members of the normative sample, thus allowing comparison with the WISC whose sample was White. [b]See Table 1 Note h: Flynn and Weiss found that (at least) 1.33 points had to be deducted from WISC–IV IQs to equate them with WISC–III IQs, so that was added to the WISC–III & WISC–IV differences at all levels—to produce the "WISC–III & WISC–III/IV" differences. [c] WAIS–III actual scores have been treated as being 2.34 points too high, so that amount was added to the "actual" differences between it and the earlier test. [d]The bracketed scores yield no real comparisons due to differences in scoring conventions.

(Flynn, 2006). Can low-IQ Americans above the age of 16 (the ages covered by the WAIS) be the only group in the world that differs?

Table 3 provides the answer by comparing the WISC–R and WAIS–R. Because these tests were normed only 6 years apart, one would expect WISC–R scores (1972) to be only slightly higher than WAIS–R scores (1978)—and they are higher at the mean. But low-IQ levels are wildly out of step. At a WISC IQ of 46, we get a reverse IQ gain (a loss) of fully 15.65 points, which gradually erodes as we go up through IQs of 64 and 73 and 81 and finally turns into a gain at the mean. This peculiar pattern suggests that the score differences at low-IQ levels measure not norms becoming obsolescent thanks to IQ gains but disparate scoring conventions.

Let us look at the test manuals. The WISC–R makes it difficult to get even a low IQ (Wechsler, 1974, pp. 150, 152). An IQ of 48 requires a standard score of 26, which entails a raw score of 147, or at least 74 items correct (some items give 2 points for a correct answer, but others give only 1). The WAIS–R is notorious for what came to be called the "tree-stump" phenomenon. An IQ of 48 requires a standard score of 13, but that can be "earned" with zero raw score points (Wechsler, 1981, pp. 93, 142), which is to say that an inert subject who answered not one item correctly qualifies. The WISC–R items in question may appear easy, but they are not easy for those three standard deviations below average (the bottom one eighth of 1% of the population), and getting them right is not nearly as easy as getting no items right at all. In Table 3, the lowest IQ group certainly did better on the WAIS–R than the 45.90 they scored on the WISC–R. But they had to score higher: No items correct by anyone would have given them 48. That they scored 61.55 simply illustrates the potency of the tree-stump phenomenon.

There is only one hypothesis that could justify using WAIS–R data at low-IQ levels to calculate obsolescence. The WISC to WISC–R data combine scores for all ages from 5 to 15 years. Treating them collectively, a rate of gain of 0.4 points per year (see Table 2) would give them a 2.4-point gain over 6 years. But imagine

Table 3
Evidence That WAIS–R Scores Imply a Huge Variation of IQ Gains at Low-IQ Levels (the "Tree-Stump" Phenomenon)

WISC–R[a]	WAIS–R[a]	Difference	% difference intact
45.90	61.55	−15.65	100
63.67	71.05	−7.38	47
72.73	77.23	−4.60	29
81.13	82.70	−1.57	10
(91)	(91)	nil	nil
101.90	101.00	+0.90	−6

Note. The data from all sources have been analyzed and adapted to elicit comparisons. Sources: Spitz, 1989, Table 2, and Wechsler, 1981, p. 48. WAIS–R = Wechsler Adult Intelligence Scale—Revised; WISC–R = Wechsler Intelligence Scale for Children—Revised.
[a]WISC–R & WAIS–R comparison studies (numbers from low to high IQs): 2 studies, 2 studies, 3 studies, 1 study, —, 1 study. The study of those with average IQs (WISC–R = 101.90) had 80 subjects. The remaining 8 had a total of 242 for an average of 30, and none had fewer than 20.

there were large age differences. If gains by 5-year-olds turned into losses by age 15, then the WISC values and the WAIS–R values for ages 16–17 could be reconciled. To test this, we must do three things: (a) On the assumption that the tree-stump effect might quickly fade away, pick a level above the bottom IQ level—I will choose the IQ range of 55–70; (b) posit that the 7.38-point score difference between the WISC–R and WAIS–R at that level really does represent an IQ loss over 6 years—by 16- to 17-year-olds; and (c) calculate a progression from ages 5 to 15 in which IQ gains turn into IQ losses, such that the gain for all ages collectively remains at 2.4 points.

As Table 4 shows, we must posit a progression in which a 12.20-point gain (over 6 years) for 5-year-olds gradually turns into a 7.38-point loss by age 16. Such values are simply too huge to be credible: No existent data yield a rate of gain or loss of even half their magnitude. Table 4 also includes WISC gains by age from actual data: A reasonable gain of about 2 points (over 6 years) shows no tendency to vary with age. The contrast reinforces the point: WAIS–R data at low-IQ levels tell us something about different scoring conventions but nothing about IQ trends over time.

This leads to a piece of unfinished business: Reverting to Table 2 and its WAIS to WAIS–R data, because this comparison includes the tainted WAIS–R, there now is reason to be suspicious of the values at very low-IQ levels. A look at the WAIS manual shows that this suspicion is justified. An IQ of 48 requires a standard score of 16, which entails a raw score of 43, or at least 22 items correct (Wechsler, 1955, pp. 79, 101)—which means that one has to do better than a tree stump to score even a modest IQ. It is hardly surprising that these data show IQ losses up to the 70–80 score range.

In sum, one must keep in mind the point of the test comparisons. They help to estimate how much test norms have become obsolete (over time) at low IQ levels. There are no scoring convention problems with the low-IQ WISC data, and it vindicates a rate of 0.3 points per year. The tree-stump phenomenon plays havoc with the WAIS data. But a look at the bottom of Table 2 shows that it does so primarily at very low IQ levels. The WAIS to WAIS–R data start to make sense at an IQ of 70, and that, after all, is where adjusting IQ scores makes a difference as to whether someone is classified as mentally retarded. The most recent data, the WAIS–R to WAIS–III, are little affected by the tree-stump phenomenon: When the WAIS–III scores are adjusted by the usual 2.34 points, those data vindicate 0.3 points at all levels. An IQ of 45 on the WAIS–III requires only one right answer (Wechsler, 1997, pp. 181, 197). However, an IQ of 48 requires 19 correct, and I suspect that the fact that there is no discrepancy at the level of 55 is due to small numbers.

Where does this leave the defendants on death row? The IQ scores most relevant to their fate are those taken before the age of 18, which almost always means some version of the WISC rather than the WAIS. Because WISC scores do not have to be converted into WAIS scores, the problems that attend such conversions are irrelevant. All that has to be done is to update obsolete WISC IQs in terms of current WISC norms, and when that is done, we stand on firm ground.

As for testing postconviction defendants, sometimes we inherit WAIS–R scores from the past: Today's defendant may have been assessed on the WAIS–R when convicted of a previous offense as an adult. Those scores should be adjusted

Table 4
Evidence That WAIS–R Scores Imply a Huge Variation in WISC IQ Gains by Age

Age in years	5	6	7	8	9	10	11	12	13	14	15	16
Pattern implied[a]	12.20	10.42	8.64	6.86	5.08	3.30	1.52	−0.26	−2.04	−3.82	−5.60	−7.38
Actual pattern[b]	—	—	1.97	2.02	2.02	2.05	2.06	1.99	2.03	2.07	2.13	—
Number	—	—	55	114	128	132	156	114	112	92	75	—

Note. Dashes indicate that no data were available. WAIS–R = Wechsler Adult Intelligence Scale—Revised; WISC = Wechsler Intelligence Scale for Children.
[a]"Pattern implied" shows the trend from WISC IQ gains to losses, over ages 5–16, that would have to exist to match WAIS–R losses at ages 16–17 (IQ level 55–70).
[b]"Actual pattern" shows the actual trend of WISC IQ gains by age, which evidences no tendency for gains to diminish from ages 7–15 (all IQ levels). The source of these data (Flynn, 1984a, Table 2), which have been analyzed and adapted for comparison, gives gains over a 24.5-year period, so they are reduced to cover a period of 6 years.

according to the general rule. Anyone tested today should, of course, take the WAIS–III rather than the WAIS–R. The Psychological Corporation has made a determined effort to collect fuller low-IQ samples, so that the WAIS–III does not duplicate the WAIS–R's problems (Wechsler, 1997). To prevent any misunderstanding, it should be noted that there is nothing wrong with using the WAIS–III to assess whether someone is mentally retarded. It is not the test content but the test norms that pose problems. And these pose a problem only if the IQ scores they generate are taken at face value rather than adjusting them.

The numbers in the WAIS–R to WAIS–III data are scant. As for the rate of obsolescence since the WAIS–III was normed in 1995, we cannot be certain of the rate until the WAIS–IV appears. It is to be hoped that the Psychological Corporation will offer a careful analysis of trends, one based on, say, 500 subjects. In the meantime, we should use a rate of 0.3 points per year.

Aristotle's Opinion

The prefatory quotation from Aristotle is not intended as an appeal to authority but to encapsulate his wise advice: Do not ignore decent data just because it would be preferable to have perfect data. Given the problems involved in selecting representative standardization samples and given the changes in scoring conventions at low-IQ levels, the emergence of the general rule is impressive. The rate of 0.3 points per year holds true, both at the mean and low-IQ levels, where the data are best—namely, the WISC data. Why are the WAIS data less perfect? It is because they represent adult data, and it is much more difficult to select representative samples of adults. With schoolchildren, as long as you locate a good sample of schools and test everyone, you are home free, whereas adults are not gathered together in one institution and can be located only if you visit all of their workplaces or homes. It is likely that adult samples can be accurate only to within a few points. It is particularly difficult to set adult norms at a level three standard deviations below the mean, where, typically, only 2 or 3 members of a standardization sample provide scores.

Confronting the total body of data, one can conclude either that the rate of gain is like the epicycles of ancient astronomy, with heavenly bodies moving forward to one point, then reversing themselves, then moving forward again; or that a coherent pattern shines through. The options that confront us are sobering: Adjust scores for obsolescence or accept scores likely to be misleading. Despite the difficulties with the WAIS data, I believe the choice is clear. This is an appeal to prudence but no worse because of that.

Table 1 supplies all of the information needed to adjust scores. It specifies when all Wechsler–Binet tests were normed, including the old WISC of 1947–1948 (as distinct from when they were published: usually a year or two later). It recommends deducting 0.3 IQ points per year from the scores of defendants for every year that passed between the date when the test was normed and the date when the test was taken. In addition, an extra 2.34 points should be deducted from WAIS–III scores on the grounds that that test gave inflated IQs even in the year in which it was normed.

John Doe and Mary Smith

It may be helpful to refer to an actual case (with alterations only to avoid identification of the parties) to illustrate what these recommendations mean in practice. John Doe was convicted of murder and sentenced to death contingent on a determination of whether he is mentally retarded. He was born in 1967. During his developmental years, he was assessed only once: in 1975, at the age of 8, by Mary Smith, a school psychologist.

Fortunately, we have her report, which states that she gave him the WISC (normed 1947–1948) and not the WISC–R (1972). This may seem odd given that the new (at that time) WISC–R was published in 1974. However, IQ tests are costly, and schools tend to exhaust their old supply before purchasing the latest edition of a test. Sometimes, knowledge of when a test was published will tell you what version was used in a case where it is not specified. For example, if it was administered in 1973, it had to be the WISC rather than the WISC–R; if administered in 1990, it had to be the WISC–R rather than the WISC–III; and so forth. Note that the reverse is not true: The date a test is administered will not tell you whether the latest version was used.

Mary Smith was taught to assess adaptive functioning independently of IQ scores. In fact, school psychologists sometimes find it difficult to compartmentalize the two. You may find telltale signs in the report. In this case, Mary noted John's poor performance in reading and arithmetic despite extra tutoring, but then rejected a diagnosis of MR on the basis of a WISC IQ score of 75. Today, we know something that she could not have known. Fully 27.5 years had passed between the norming of the WISC and the day the defendant took it (1975 minus 1947.5). Therefore, its norms were 8.25 IQ points out of date (27.5×0.3 points per year). So John's score of 75 should have been lowered to 67, easily in the MR range. John is likely to be executed simply because his school's budget did not extend to purchasing the new WISC–R. On its newer norms, he would have scored 68 (only 3 years had passed) and probably would have been classified as mentally retarded.

Given the paucity of assessment during his developmental years, the defendant was interviewed by both a defense and a prosecution psychologist. Inevitably, the former noted hesitancy and conceptual vagueness typical of MR, and the latter found him far too alert and fluent to be mentally retarded. Therefore, much hinges on the IQ scores. The defense psychologist administered the WAIS–III, recorded an IQ of 70, and cited this as confirmation of the assessment of MR. One week later, the prosecution psychologist administered the WAIS–III and scored his IQ at 72, which left him in no doubt that an assessment of MR was inappropriate.

In fact, the two scores both indicate MR. The WAIS–III gives a bonus of 2.34 points. This in itself lowers the two scores to 67.66 and 69.66. The fact that its norms were 10 years out of date when it was administered has to count for something. Allowing 3 points reduces the two scores to 64.66 and 66.66, clearly in the retardate range. The second score is 2 points higher, but because the second testing was only 1 week after the first, the subject probably gained those points from practice effects.

I have not mentioned measurement error, that is, the fact that a single testing of a person may give an IQ score accurate only between ±5 points. That is because this defendant has three scores, all of which (when properly interpreted) put him at 67 or below—and that means the possibility of measurement error is much reduced. But imagine that this defendant had only one score, which when properly interpreted was 71. The odds that his true IQ justified a classification of MR would be almost 50%. A score of 76 would be needed to be confident that a diagnosis of MR was inappropriate.

Playing Games With Time

No prosecutor should be allowed to argue that because IQ scores are rising, a person tested 20 years ago (and who scored 70 against the norms of that time) would probably do better today and score 76. No defense attorney should be allowed a similar gambit: to argue that a person who today scores 71 (against current norms) would probably score 65 against the norms of 20 years hence.

The United States Supreme Court (*Atkins v. Virginia,* 2002) rightly undermined such arguments by setting as a criterion of MR that a person should be two or more standard deviations below the mean—if we add the corollary that the mean refers to the average performance of Americans at the time the person was tested. The prosecution should not make the unwarranted assumption that the defendant will gain ground on his or her cohort in terms of IQ as the defendant ages: Otherwise, a performance 20 years after the defendant's initial test will still put him or her two standard deviations below the mean and at an IQ of 70. The defense should not make the unwarranted assumption that the defendant will lose ground on his or her cohort as the defendant ages: Otherwise he or she will still score less than two standard deviations below the mean and at an IQ of 71. Such arguments could become valid only if the AAMR reached a firm conclusion about the significance of IQ gains over time and altered its criterion of MR. As noted above, that decision should be left to them rather than lawyers.

These arguments can be discredited by extension. Why stop at 20 years? Why not appeal to a posited future in which scores have risen so high that no one qualifies as mentally retarded? Why not appeal to a distant day whose norms are so tough that everyone now living would qualify as mentally retarded? The current state of our knowledge dictates three propositions: Defendants must be assessed against the criterion of MR that exists at present; their test scores must be interpreted in terms of the norms of the year in which they are tested; and their performance must be one that has actually occurred, not an event in a nonexistent future.

I also wish to call attention to another argument put forward by prosecutors, namely, that the Flynn effect is a "group phenomenon" and cannot be applied to individuals. As the reader now knows, this is just a senseless mantra. When the group making the IQ gains is composed of Americans, those gains render test norms obsolete and inflate the IQ of every individual being scored against obsolete norms.

Common Ground

The purpose of this article is not to undermine confidence in IQ scores. Quite the contrary. America's adversarial system of justice makes it almost inevitable that when prosecution and defense experts interview defendants, they will reach opposite conclusions about whether MR is present. IQ scores provide a less subjective weight to put into the scales. But IQ scores are not like a loaf of bread that simply is what it is. They are messages whose meaning is not intrinsically ambiguous but does require interpretation. Once the mode of interpretation is applied, adjusted IQ scores usually tell a coherent story.

There is nothing odd about this. Thermometers send us messages about temperature. But in order to interpret them, we have to know whether the norms are Fahrenheit or Centigrade. And compare the formulae for conversion:

$$\text{Temperature: } (C \times 9/5) + 32 = F, \quad (1)$$

$$\text{Intelligence: Test Score} - (I \times 0.3) = IQ, \quad (2)$$

where I is the interval between when the test was normed and when the subject sat; and the test is a Wechsler–Binet test administered in America. In addition, if the Wechsler test happens to be the WAIS–III, an extra 2.34 points should be deducted from the test score. Certainly, the IQ formula is simpler than the temperature formula.

I am not a jurist but a moral philosopher with expertise in the theory and measurement of intelligence. However, as the prefatory quotation from Justice Stewart makes clear, law must be consistent with at least the most basic rules of morality, so that grievous penalties are not imposed "wantonly." To paraphrase his words, the U.S. Constitution cannot tolerate a death penalty that is the random result of whatever IQ test a school psychologist happened to have in stock. The legitimacy of capital punishment is a matter of debate. Where psychologists should set the criterion for MR will always have to be reexamined in the light of new evidence. But no jurist, moral thinker, psychologist, or person of ordinary decency will believe that the Supreme Court intended to let loose an elephant in a crowd.

References

American Association on Mental Retardation. (2002). *Mental retardation: Definition, classification, and systems of support* (10th ed.). Washington, DC: Author.

Atkins v. Virginia, 536 U.S. 304, 122 S. CT 2242 (2002).

Black v. State, No. M2004-01345-CCA-R3-PD, 2005 Tenn. Crim. App. LEXIS 1129 (Tenn. 2005).

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

Breslau, N., Dickens, W. T., Flynn, J. R., Peterson, E. L., & Lucia, V. C. (in press). Low birthweight and social disadvantage: Tracking their relationship with children's IQ during the period of school attendance. *Intelligence*.

Ex parte Murphy, No. WR-38,198-03, 2006 Tex. Crim. App. (Tex. Jan. 18, 2006).

Flynn, J. R. (1984a). IQ gains and the Binet decrements. *Journal of Educational Measurement, 21,* 283–290.

Flynn, J. R. (1984b). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency, 90,* 236–244.

Flynn, J. R. (1998). WAIS–III and WISC–III: IQ gains in the United States from 1972 to 1995; how to compensate for obsolete norms. *Perceptual and Motor Skills, 86,* 1231–1239.

Flynn, J. R. (2006). Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn Effect: Rethinking intelligence and what affects it]. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à Psicologia das Diferenças Individuais* (pp. 387–411). Porto Alegre, Brazil: ArtMed. (English translation available from James R. Flynn: jim.flynn@stonebow.otago.ac.nz)

Flynn, J. R., & Weiss, L. G. (2006). *American IQ gains from 1932 to 2002: The significance of the WISC subtests*. Manuscript submitted for publication.

Frumkin, I. B. (2003). *Mental retardation: A primer to cope with expert testimony* (p. 3). Retrieved January 6, 2006, from http://www.nlada.org/DMS/Documents/1066919805.15/Mental%20Retardation.pdf

Furman v. Georgia, 408 U.S. 238 (1972).

Garber, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk*. Washington DC: American Association on Mental Retardation.

In re Hicks, 375 F.3d 1237 (11th Cir. 2004).

Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on American society via mental retardation diagnoses. *American Psychologist, 58,* 778–790.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.

Myers v. State, 278 P.1106 (Okla. 2005).

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.

People v. Superior Court (Vidal), 129 Cal. App. 4th 434, 28 Cal Rptr. 3d 529 (5th Ct. App. 2005), *vacated and later proceedings at* People v. S.C., 2005 Cal. LEXIS 13290 (Cal. Nov. 17, 2005).

Psychological Corporation. (2003). *The WISC–IV technical manual*. San Antonio, TX: Author.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, fifth edition: Technical manual*. Itasca, IL: Riverside.

Sparrow, S. (2006). [Table 8.2: Correlations between the Vineland–II and Vineland ABS domains and adaptive behavior composite]. Unpublished raw data.

Spitz, H. H. (1989). Variations in Wechsler Interscale IQ disparities at different levels of IQ. *Intelligence, 13,* 157–167.

State v. Burke, 2005 Ohio 7020 (2005).

State v. Murphy, 2005 Ohio 423 (2005).

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford–Binet Intelligence Scale, fourth edition: Technical manual*. Chicago: Riverside.

Trowbridge, B. (2003). *US Supreme Court finds execution of the mentally retarded "cruel and unusual"; you have to pass a test before you can be put to death?* (pp. 7–8). Retrieved January 6, 2006, from http://www.trowbridgefoundation.org/docs/execution.htm

Walker v. True, 399 F.3d 315 (4th Cir. 2005), *after remand,* 401 F.3d 574 (4th Cir. 2005).

Walton v. Johnson, 407 F.3d 285, 295–97 (4th Cir., 2005).

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale: Manual*. New York: Psychological Corporation.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised: Manual*. New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised: Manual*. New York: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition: Manual*. San Antonio, TX: Psychological Corporation.

Zimmerman, I. R., & Woo-Sam, J. M. (1997). Review of the criterion-related validity of the WISC–III: The first five years. *Perceptual and Motor Skills*, *85*, 531–535.

Received August 1, 2005
Revision received March 20, 2006
Accepted March 24, 2006 ■