

# Journal of Psychoeducational Assessment

<http://jpa.sagepub.com/>

---

## **Failure to Apply the Flynn Correction in Death Penalty Litigation: Standard Practice of Today Maybe, but Certainly Malpractice of Tomorrow**

Cecil R. Reynolds, John Niland, John E. Wright and Michal Rosenn

*Journal of Psychoeducational Assessment* 2010 28: 477 originally published online 20 July 2010

DOI: 10.1177/0734282910373348

The online version of this article can be found at:

<http://jpa.sagepub.com/content/28/5/477>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Journal of Psychoeducational Assessment* can be found at:**

**Email Alerts:** <http://jpa.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jpa.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jpa.sagepub.com/content/28/5/477.refs.html>

# Failure to Apply the Flynn Correction in Death Penalty Litigation: Standard Practice of Today Maybe, but Certainly Malpractice of Tomorrow

Cecil R. Reynolds<sup>1</sup>, John Niland<sup>2</sup>,  
John E. Wright<sup>3</sup>, and Michal Rosenn<sup>4</sup>

## Abstract

The Flynn Effect is a well documented phenomenon demonstrating score increases on IQ measures over time that average about 0.3 points per year. Normative adjustments to scores derived from IQ measures normed more than a year or so prior to the time of testing an individual have become controversial in several settings but especially so in matters of death penalty litigation. Here we make the argument that if the Flynn Effect is real, then a Flynn Correction should be applied to obtained IQs in order to obtain the most accurate estimate of IQ possible. To fail to provide the most accurate estimate possible in matters that are truly life and death decisions seems wholly indefensible.

## Keywords

Intelligence, Flynn effect, death penalty, forensic psychology

Since the Supreme Court's decision in *Atkins v. Virginia* (536 US 304, 122 S. CT 2242, 2002) that the execution of the mentally retarded violates the Eighth Amendment's prohibition against cruel and unusual punishment, the importance of understanding and assessing mental retardation in criminal defendants has become critical, indeed a true matter of life and death, in capital felony cases. Determining whether a defendant's intellectual functioning is severely limited is essential to a judgment as to whether that individual is able to act with the level of moral culpability that merits particular forms of punishment. As the best measures of intellectual functioning, IQ tests are regarded as one of the primary indicia of mental retardation by both clinicians and courts. The consensus among mental health professionals is that an IQ of 70 to 75 or less on

---

<sup>1</sup>Texas A&M University, College Station, TX, USA

<sup>2</sup>Texas Defender Service, Austin, TX, USA

<sup>3</sup>Huntsville, TX, USA

<sup>4</sup>Harvard University, Cambridge, MA, USA

## Corresponding Author:

Cecil R. Reynolds, 101 Reynolds Court, Bastrop, TX 78602, USA

Email: [crrh@earthlink.net](mailto:crrh@earthlink.net)

a standardized, individually administered IQ test satisfies the IQ prong of the diagnostic criteria for mental retardation (e.g., see Flynn, 2006, 2007a; Reynolds, Price, & Niland, 2003, as well as various court cases cited in these articles).

IQ tests are periodically revised and renormed to keep the content appropriate to current cultural contexts, ensure the representativeness of the normative or reference group (characteristics of the target population are constantly changing), and to maintain an average score of 100. The findings associated with these periodic revisions led researchers to observe that scores on standardized measures of intelligence have steadily risen over the past century, a phenomenon termed the Flynn Effect (FE; after James Flynn, the man who first documented these changes carefully and comprehensively, e.g., Flynn, 1984). Among the various explanations offered for the effect, the predominant explanation—and the one adopted by Flynn—is that environmental changes relating to modernization have increased people's ability to manipulate abstract concepts, a skill that is heavily emphasized in IQ tests. However, the reason for the FE is controversial, as can easily be seen in other articles in this issue, but the existence of the effect has no significant scholarly challenges of which we are aware. The FE, whatever its cause, is as real as virtually any effect can be in the social sciences. Studies have observed an increase of 0.3 points per year in average IQs; thus, for a test score to reflect accurately the examinee's intelligence, 0.3 points must be subtracted for each year since the test was standardized (Flynn, 2006, 2007a, 2007b). Since the FE's increased scientific acceptance in the 1990s, it has become one of the reasons why IQ tests have been revised and renormed more frequently than in the past, typically occurring on a 10- to 11-year schedule now as opposed to a 20-year or more schedule in the past. Even so, the FE is observable in the years between revisions, and is certainly relevant where outdated test versions are used—especially where even 2 or 3 IQ points may determine whether a defendant is allowed to live or is killed.

Because of the central role IQ tests play in determining an individual's level of mental retardation, and because of the importance of mental retardation in determining a defendant's eligibility to be killed by the State, it is imperative that the FE, if it is real, be taken into account in capital cases. IQ ranges that indicate mental retardation are determined relative to the average score (which has been set by convention, albeit arbitrarily, at 100). The so-called average score is derived from a reference group, which is a snapshot of the population at one particular point in time. The determination of the intelligence component of the diagnosis of mental retardation (we do recognize that the actual diagnosis is far more complex than looking at an IQ—but the IQ is a crucial component, and we deal only with it here) should be based on the person's standing relative to the target population at the time the person was actually tested, not the target population when the test was normed. Because it is at this time a practical impossibility to renorm tests annually to maintain a more appropriate reference group, to the extent corrections are available and valid, they should be applied to obtained scores so the most accurate estimate of standing possible is obtained. To do less is to do wrong—what possible justification could there be for issuing estimates of general intelligence in a death penalty case that are less than the most accurate estimates obtainable?

The way in which the Flynn correction applies to an individual is illustrated by the following scenario: a person taking the same version of the same IQ test 10 years apart will, on average, experience a 3-point increase in his or her score over that time—not because of any actual increase in intellectual functioning, but because of latent social changes that manifest themselves in the test. Therefore, a 3-point correction downward of the obtained IQ is required to provide the most accurate estimate of intellectual functioning relevant to today's population. Without this correction, whether a criminal defendant is deemed mentally retarded and thus eligible for the death penalty can thus turn on when the IQ measure chosen was standardized (e. g., see Ceci, Scullin, & Kanaya, 2003, and Flynn, 2006). If there remains any doubt that we must provide the

most accurate IQ estimates we can, in all cases, but especially matters of death, we can take guidance from the U.S. Supreme Court, the ultimate arbiter of legal issues in the United States, whose members have repeatedly recognized, “the penalty of death is different in kind from any other punishment imposed under our system of criminal justice” (*Gregg v. Georgia*, 428 U.S. 153, 188, 1976). It thus requires “a greater degree of accuracy and factfinding than would be true in a non-capital case” (*Gilmore v. Taylor*, 508 U.S. 333, 342, 1993). This has led to specialized procedures for capital cases: attorney competency requirements; provisions for automatic appeal in many states; special requirements for jury sentencing (e.g., unanimous verdicts, consideration of mitigating evidence); and, in many states, review for proportionality. These measures and others are meant to address the heightened need for accuracy in capital cases to ensure that the exercise of the State’s ultimate authority to kill a defendant is meted out only to those deemed legally deserving of such a final and irrevocable punishment.

Though courts usually consider evidence regarding the FE relevant to their interpretation of defendants’ IQs, application of the effect is not mandatory. Judges are often particularly hesitant to conclude that, because a general effect exists, an individual’s IQ should be adjusted downward accordingly—apparently some psychologists also view generalizing a group effect to an individual as an undue leap of inference and therefore a reason not to make the Flynn correction. This is really a straw man argument.

First, nearly all effects in psychology are based on aggregated data and groups and subsequent probability estimations from groups to individuals. Any prediction formula, and these are used often by most all psychologists involved in forensic cases, in employment decisions, prediction of achievement levels, diagnosis of specific learning disabilities, college admissions, and so on to name a few, is based on groups and then the formulae are applied to individual cases. However, to argue the FE should not be applied to individuals belies the fact that all IQs, obtained or otherwise, are to a significant extent based on a group effect and derived from aggregated data. This is because we have only interval scaling, not ratio scaling, available to us in determining scores such as IQs. The determination of an individual’s IQ begins with defining the midpoint of the distribution of performance of a sample (a group of people) of a target population. With interval scaling, the only point we can initially locate accurately is the middle of the score distribution—we then measure outward toward the two ends of the distribution of scores based on the variance of the group we used to derive the scores. We then generalize this set of group effects to the performance of individuals and place them on the group distribution and demarcate their placement with the assignment of an IQ (for a more detailed explanation see Reynolds & Livingston, in press). As the group used to provide these statistics ages and becomes less like the current target population, applying any correction that can improve the accuracy of the placement of the individual on this continuum (e.g., the Flynn correction) improves IQ estimation for the individual. The admonishments of the U.S. Supreme Court, in multiple incarnations, that death penalty cases require special attention to accuracy apply an even more profound legal argument to applying this correction.

Zhou, Zhu, and Weiss (2010) point to another controversy surrounding the FE that may be applicable to the size of the correction needed in an individual case. The FE may not be constant across the full distribution of IQs. Depending on the method and assumptions, they show the FE may vary in nonsignificant magnitudes across the full range of IQ. Using the largest and most stable samples, which were collapsed across scales after a statistical demonstration of constant effects by instrument, and assessed using analysis of variance (ANOVA) and analysis of covariance (ANCOVA), a larger FE was observed at the lower end of the IQ range, that is, obtained IQs less than or equal to 79. However, using an equipercentile approach to equating, disparate results were seen where on some tests the prior pattern was upheld but not on others—a reversal of the pattern occurring in some instances. However, the larger sample size and the use of the verbal composite to block and thus lessen any regression effects in the first analyses, appears more

reliable in its results, and perhaps a stronger fit to theories of cognitive development as well. The changes in magnitude of the FE reported in Zhou et al. may also be related to chronological age at the time of testing as it varied, albeit inconsistently, by age appropriate version of the Wechsler Scales. This could be addressed if tests that examine a large age range with a common set of core tasks could be subjected to similar analyses (e.g., the Reynolds Intellectual Assessment Scales; Reynolds & Kamphaus, 2003); however, the necessary data on these instruments are not available. Taken as a whole, and noting some of the inconsistencies in the results, the Zhou et al. analyses support the idea that an even larger correction may need to be applied to low IQs, especially given the paucity of individuals scoring at the extremes in the samples evaluated, but this remains for future research with much larger samples. For now, best practice is the application of the Flynn correction as a constant by year across the distribution.

Where courts strictly adhere to score cutoffs in determining mental retardation, a single point can mean the difference between a constitutional and unconstitutional execution—even if courts were to drop such a rigid adherence, the most accurate estimation of the defendant's IQ is still required. The FE, though certainly not without its detractors, is nevertheless a generally accepted scientific theory. Flynn has demonstrated its clear applicability to individual cases quite clearly as well.<sup>1</sup> (Flynn, 2006, has dealt eloquently with a number of other objections to applying the Flynn correction to individuals, which space limitations do not allow us to address.) The United States has decided to allow states to determine mental retardation almost exclusively by reference to a ranking system that quantifies an individual's standing relative to a reference sample. In doing so, psychologists who work in this system and the courts themselves must ensure that this system is applied as accurately as possible so that no overinclusion into the category of death-eligible individuals occurs.

## Conclusion

In criminal proceedings, the law's primary concern is that justice is meted according to the procedures and guarantees contained in the federal and state constitutions. These constitutional concerns, as well as the need for accuracy, are at their highest when the death penalty is at issue. The highest court in this country has made the determination that executing persons with mental retardation violates the Eighth Amendment's prohibition against cruel and unusual punishment. As a generally accepted scientific theory that could potentially make the difference between a constitutional and unconstitutional execution, the FE must be applied in the legal context. Those who oppose the Flynn correction must either dispute the scientific validity of the FE (and we see no such serious challenges—the remaining issue seems to be over why it occurs, a debate that is irrelevant to whether it should be applied), have a poor understanding of the death penalty and the writings of the Supreme Court on the matter, or perhaps simply do not understand interval scaling and its implications for how test scores are derived, the purpose and application of reference samples (i.e., norm groups), or how predictions are made in psychology (unfortunately, a too common state of affairs in professional psychology, e.g., see Reynolds, 2010). If the FE is real, the failure to apply the Flynn correction as we have described it is tantamount to malpractice. No one's life should depend on when an IQ test was normed.

## Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

## Funding

The author(s) received no financial support for the research and/or authorship of this article.

## Note

1. See, for example, affidavit of James R. Flynn in the case of Earl Wesley Berry, August 8, 2004, which noted a 7-point difference in two of the defendant's IQ scores, obtained 1 month apart but from different versions of the Wechsler Adult Intelligence Scale (WAIS) test. Adjusting the scores according to the FE yielded an identical score across the two tests.

## References

- Ceci, S., Scullin, M., & Kanaya, T. (2003). The difficulty of basing death penalty eligibility on IQ cutoff scores for mental retardation. *Ethics and Behavior, 13*, 11-17.
- Flynn, J. R. (1984). The mean IQ of Americans: Massive gains from 1932 to 1978. *Psychological Bulletin, 95*, 29-51.
- Flynn, J. R. (2006). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law, 12*, 170-189.
- Flynn, J. R. (2007a). *What is intelligence? Beyond the Flynn effect*. New York, NY: Cambridge University Press.
- Flynn, J. R. (2007b). Capital offenders and the death sentence: A scandal that must be addressed. *Psychology in Mental Retardation and Developmental Disabilities, 32*, 3-7.
- Reynolds, C. R. (2010). Measurement and assessment: An editorial view. *Psychological Assessment, 22*, 1-4. doi:10.1037/a0018811
- Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources.
- Reynolds, C. R., & Livingston, R. A. (in press). *Mastering psychological testing*. Boston, MA: Allyn & Bacon.
- Reynolds, C. R., Price, R., & Niland, J. (2003). The role of neuropsychology in capital felony (death penalty) defense. *Journal of Forensic Neuropsychology, 3*, 89-123.
- Zhou, X., Zhu, J., & Weiss, L. A. (2010). Peeking inside the "black box" of the Flynn effect: Evidence from three Wechsler instruments. *Journal of Psychoeducational Assessment, 28*, 399-411.