

---



---

**PSYCHOLOGY  
IN MENTAL RETARDATION  
AND DEVELOPMENTAL DISABILITIES**

---

**OFFICIAL PUBLICATION OF DIVISION 33**

**AMERICAN PSYCHOLOGICAL ASSOCIATION**

---

Volume 32, Number 3

Fall 2006

---

## Letter to the Editor:

### **Modification of Individual's IQ Scores is Not Accepted Professional Practice**

I am writing in response to the article by Dr. Stephen Greenspan that appeared in the Spring 2006 edition of the Division 33 newsletter (*Psychology in Mental Retardation and Developmental Disabilities*, Vol. 31, number 3). The article discussed adjusting individual IQ scores based on the "Flynn effect" primarily in regards to diagnosing mental retardation in capital litigation cases. It was noted that the phenomenon has received much attention post-Atkins and is typically used to adjust scores downwards. I have been involved in over a dozen capital cases where mental retardation was an issue, serving as an expert witness providing consultation, evaluation and testimony; I have worked both for the State and for the defense and I am well aware of the battleground in the legal arena related to the interpretation of IQ scores.

The "Flynn effect," an average rise in population performance on IQ measures, has been noted to have taken place over many decades and across cultures in industrialized nations (e.g. Flynn, 1984; Flynn, 1987; Sundet, Barlaug & Torjussen, 2004; Teasdale & Owen, 2000). The rate of rise has been found to vary across different tests (e.g. Flynn, 2006), subtests within tests (e.g. Sundet, Barlaug & Torjussen, 2004; Teasdale and Owen, 2000), ability level (e.g. Sundet, Barlaug & Torjussen, 2004; Teasdale and Owen, 2000), time (e.g. Sundet, Barlaug & Torjussen, 2004; Teasdale and Owen 2006), and countries (Flynn, 1987). Furthermore, it has long been recognized that it is unlikely that the rise would continue unabated; that it is likely to peak at some point. I have critically reviewed the literature regarding the "Flynn effect" and I do not believe that it is an appropriate basis upon which to modify individual scores. The reasons for my conclusions are outlined below.

First, recent research in Denmark (Teasdale & Owen, 2006) and Norway (Sundet, Barlaug & Torjussen, 2004) provides strong evidence that the phenomenon peaked in those two countries in the mid 1990's and has begun to reverse. These findings were based on updated longitudinal data sets that Dr. Flynn has described as representing "strong data" (Flynn, 1987) for assessing long-term IQ trends. Whether the effect has reached a peak or begun to decline in the U.S. on at least some types of tests is an open question. It has been demonstrated that performance on IQ tests and achievement tests are closely correlated and there has been a clear

decline in performance over the past twenty years by American students on some achievement tests and a failure to keep pace with other industrialized countries on others (Williams and Ceci, 1997).

Secondly, in the U.S. the rate of increase has been studied by using comparisons of test scores for individuals taking two different tests close in time to each other. The gains are not studied by comparing performance on the same test at different times, but rather by comparing performance on two different tests at the same time, including tests from different publishers. Hence these studies, which were actually designed and undertaken as a validity check to look at inter-test correlation between new test versions and established tests, contain a confound in that the tests themselves differ along a variety of domains.

The importance in this confound should not be underestimated. Wicherts, et al. (2004) explored changes in performance over time on numerous cognitive measures across several populations and consistently found that changes in performance levels differed across subtests. They found strong consensus that overall changes in performance between populations at different time periods were not due solely to changes in testing cohorts. In other words, there was not a steady, consistent rise over time across different types of measures, but rather notable differences in the degree of performance change depending on the nature of the particular cognitive task assessed. Three of the four highest rates of gain found in Flynn's 2006 article involved comparisons of tests from different publishers (i.e. Stanford-Binet editions versus Wechsler editions).

Thirdly, comparing different tests with each other then using scoring differences between the tests to establish a rate of rise in scores over time results in rates that fall across a notable range. The sixteen test-test comparisons provided in Flynn (2006) involved various combinations of comparisons between adult Wechsler, child Wechsler and Stanford-Binet tests. Only four comparisons result in yearly average rates of increase within plus or minus twenty percent of .3 points per year. In Dr. Flynn's 1984 study he provides eighteen similar Wechsler/Stanford-Binet comparisons with earlier versions of these tests. Only ten fall within plus or minus twenty percent of .3 points per year. The range becomes even greater when looking

at data across tests and countries such as in Flynn's 1987 paper in which only five of the twenty-nine rates reported fall within plus or minus twenty percent of .3 points per year. This figure that is cited with such precision is actually an average of wide-ranging estimates; it seems highly suspect to utilize it as a specific correction to an individual's scores.

Lastly, it is important to note that the empirical literature has identified many factors that affect IQ scores which are not utilized to make individual score adjustments. For example, there is robust research of large differences in performance on IQ tests across ethnic groups likely in concert with socioeconomic status. For example, African Americans, on average, demonstrate a 10 to 15 point IQ difference compared to their Caucasian counterparts (e.g. Neisser, U., et al., 1995; Flynn, 1984; Williams and Ceci, 1997). African American students have long been at notably elevated risk of being identified as mentally retarded as compared to Caucasian students (Donovan & Cross, 2002). There is substantial debate on the reasons behind these ethnic group performance differences on measures of cognitive functioning, with environmental factors likely playing a large role. However, a clinician does not "correct" the IQ scores of African Americans by adding ten to fifteen points to their obtained scores. Similarly, persons from urban areas tend to score two to four points higher than those from rural areas, but again, clinicians do not "correct" for this by either adding to the scores of rural examinees or subtracting from the scores of urban examinees.

In sum, the "Flynn effect correction" lacks the empirical precision for altering individual IQ scores. Furthermore, it would be inappropriate to select particular effects to "correct" for while ignoring other more robust findings or larger magnitude effects in an effort to meet a particular social or legal agenda. In addition to looking at the scientific or clinical appropriateness of making score adjustments based on the "Flynn effect," it is important to also look at the professional and legal acceptance of the effect.

One question that is often put forth in court is whether modifying individual scores based on the "Flynn effect" is accepted professional practice. It does not appear to be. I have been in direct contact with representatives from PsychCorp/Harcourt Assessments, Inc. regarding their recommendation about such scoring modifications. The officially stated position



of PsychCorp/Harcourt Assessment, Inc. is that "Harcourt Assessment, Inc. stands behind the quality of the WAIS-III and accuracy of the IQ scores." (Personal communication Barbra-Ann Frazier, June, 2006). In additional communication the publisher states, "The Flynn effect has been in the literature for some years. We do not recommend a 'correction' because different studies find different results, and the effect occurs for all measures of cognitive ability. It is the primary reason that we re-norm tests periodically (so our norm-based scores reflect the contemporary population)." (Personal communication, Barbra-Ann Frazier, July, 2006).

Further evidence that modifying scoring or interpretation based on the Flynn Effect does not represent standard clinical practice can be found in the Amicus Brief filed by the American Psychological Association, American Psychiatric Association and American Academy of Psychiatry and the Law with the Supreme Court on behalf of Daryl Atkins in that seminal case. The amici speak extensively of the strong validity and reliability of the instruments used to measure intellectual functioning. There is no indication of a need or appropriateness to making any adjustment to the interpretation of the scores based on the "Flynn effect" or any other factors, other than noting the range of confidence associated with the standard error of measurement (APA Brief, LEXIS 2000 U.S. Briefs 8727).

The lack of support for scoring adjustments from the test publisher and the failure of the APA to mention such adjustments in the Amicus Brief are consonant with my clinical experience. In the course of my professional career I have had the opportunity to review over ten thousand psychological reports and I have never seen a single case where an individual score has been adjusted based on the "Flynn effect" outside of psychological reports submitted on behalf of the defense in post-Atkins capital cases where mental retardation was the central issue.

Legal precedents related to application of the "Flynn effect" is a final area of importance to discuss. Flynn, 2006, indicates that submissions by experts supporting the relevance of the "Flynn effect" have been "welcomed by the courts" and he cites numerous examples. While it is true that experts have submitted declarations and provided testimony in numerous court cases related to the matter, close review of the cases that he cites reflects a range of responses on behalf of the courts regarding the usefulness or relevance of the effect.

In the Vidal case in California,<sup>1</sup> application of a scoring modification based


on the "Flynn effect" was accepted, but the Superior Court's decision was vacated by the California Supreme Court,<sup>2</sup> and thus this case cannot be cited as approving or supporting application of the "Flynn effect." In McLaughlin v. Polk,<sup>3</sup> the federal district judge applied the "Flynn effect" when ordering an evidentiary hearing in which the petitioner was later judged to be mentally retarded.<sup>4</sup> Some courts have ruled that lower courts should consider the persuasiveness of the "Flynn effect" on a case-by-case basis, but do not approve or command application of it, e.g. State v. Burke,<sup>5</sup> and Walker v. True.<sup>6</sup> In Walton v. Johnson,<sup>7</sup> the Fourth Circuit Court of Appeals specifically stated that it made "no determination as to the validity of . . . [the "Flynn effect" and other arguments]; [it held] merely that Walton [was] entitled to be heard on them."<sup>8</sup> Some courts have heard evidence on the "Flynn effect" but have taken the position that the controlling state law has a bright line cut-off and therefore it does not apply, e.g. Black v. Tennessee,<sup>9</sup> and Bowling v. Kentucky.<sup>10</sup> Other courts appear to have ignored any mention of the "Flynn effect" in the majority opinion despite its having been an aspect of the arguments in the case, e.g. State v. Murphy,<sup>11</sup> and in Hicks.<sup>12</sup>

Flynn, 2006, cites one additional case, Myers v. State,<sup>13</sup> but that case has to do with the removal of county commissioners from their positions, not mental retardation. There is another Oklahoma case, Myers v. State,<sup>14</sup> which does deal with the issue of mental retardation but there is no mention of the "Flynn effect" in the decision.

In summary, the "Flynn effect" has received a great deal of attention in the post-Atkins era as a means for adjusting obtained IQ scores downwards. However, examination of the literature related to the phenomenon reveals that it lacks the

empirical precision to be applied to individual scores. Furthermore, it is not accepted professional practice to adjust or modify individual scores based on the "Flynn effect" or any other phenomenon. If there are factors that lead the psychologist to believe that the scores do not represent an accurate or reliable measure of the individual's functioning, such issues are delineated in the discussion and interpretation of the scores; the scores themselves are not changed. Modification of individual scores is not accepted professional practice, for good reason, and should not be introduced into the court as such.

## References

- APA Brief, LEXIS 2000 U.S. Briefs 8727  
 Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.  
 Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.  
 Flynn, J.R. (2006). Tethering the elephant – capital cases, IQ and the Flynn effect. *Psychology, Public Policy and Law*, 12, 170-189.  
 Kanaya, T., Scullin, M.H. & Ceci, S.J. (2003). The Flynn effect and U.S. policies: The impact of rising IQ scores on America society via mental retardation. *American Psychologist*, 58, 778-790.  
 Neisser, U., et al. (1995) Stalking the Wild Taboo – Intelligence: Knowns and unknowns. *Report of a Task Force established by the Board of Scientific Affairs of the American Psychological Association*.  
 Sundet, J.M., Barlaug, D.G., & Torjussen, T.M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362.  
 Teasdale, T.W. and Owen, D.R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence*, 28, 115-120.  
 Teasdale, T.W. and Owen, D.R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, 39, 837-843.  
 Wicherts, J.M., Dolan, C.V., Hessen, D.J., Oosterveld, P., van Baal, G.C.M., Boomsma, D.I., and Span, M.M. (2004). Are intelligence tests measurements invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509-537.  
 Williams, W.M. and Ceci, S.J. (1997). Are Americans becoming more or less alike? Trends in race, class, and ability differences in intelligence. *American Psychologist*, 52, 1226-1235. 

Roger B. Moore, Jr., Ph.D.  
 The Center for Psychological Wellness, PA  
 Clinical and Forensic Psychology

<sup>1</sup>People v. Superior Court (Vidal), 129 Cal. App. 4th 434 (2005).

<sup>2</sup>People v. Superior Court (Vidal), 32 Cal. Rptr. 3d 4, 116 P.3d 478 (2005).

<sup>3</sup>\_\_\_\_ (E.D. N.C. \_\_\_\_ ) (unpublished order).

<sup>4</sup>McLaughlin v. Polk, \_\_\_\_ (E.D. N.C. 2005) (unpublished).

<sup>5</sup>2005 Ohio 7020 (2005).

<sup>6</sup>399 F.3d 315 (4th Cir. 2005). An evidentiary hearing was held on remand to the district court, at which the Flynn effect was argued in concert with the SEM and the practice effect: the latter two factors were determined to be speculative but the "Flynn effect" was not specifically addressed. Walker v. True, No. 03-0764 (Memorandum Opinion, 30 August 2006). 7407 F.3d 285 (4th Cir. 2005).

<sup>8</sup>Id. at 297.

<sup>9</sup>2005 Tenn. Crim. App. LEXIS 1129 (2005).

<sup>10</sup>163 S.E.3d 367 (2005).

<sup>11</sup>2005 Ohio 423 (2005).

<sup>12</sup>375 F.3d 1237 (11th Cir. 2004).

<sup>13</sup>278 P. 1106 (Okla. 1929).

<sup>14</sup>133 P.3d 312 (Okla. 2006).