

# PSYCHOLOGY IN MENTAL RETARDATION AND DEVELOPMENTAL DISABILITIES

## OFFICIAL PUBLICATION OF DIVISION 33

AMERICAN PSYCHOLOGICAL ASSOCIATION

Volume 31, Number 3

Spring 2006

## Issues in the Use of the "Flynn Effect" to Adjust IQ Scores When Diagnosing MR<sup>1</sup>

Stephen Greenspan<sup>2</sup>  
University of Colorado Health Sciences Center

The "Flynn effect" has begun to attract great interest among psychologists as well as in the popular press. It is generally accepted as a real phenomenon, although some respected scholars (e.g., Nettelbeck & Wilson, 2004; Rowe & Rodgers, 2002) have expressed doubts about the usual explanations for it. In the field of mental retardation (MR), especially in so-called "Atkins" death penalty exemption hearings, the phenomenon has been used typically to adjust individual IQ scores downward, although there has been at least one case in which it was argued that a so-called "reverse Flynn effect" should be used to adjust an individual's IQ score upward. In this paper, I shall describe the Flynn effect, discuss whether it occurs in people who have MR, and address the question of whether it is appropriate to use the Flynn effect to adjust IQ scores in order to support or dispute a diagnosis of MR, particularly in death penalty cases.

The Flynn effect is a phenomenon pointed out in a series of widely-cited papers by James R. Flynn (1984, 1987, 1998, 2006), a professor emeritus of political studies at New Zealand's University of Otago. It refers to the fact that data from virtually all nations in the developed world show an upward trend in performance on IQ tests from one generation to the next. The effect is most dramatic (around 0.7 point gain per year) in measures of "fluid" intelligence such as the Ravens Progressive Matrices, but is still impressive (around 0.3 point gain per year) on more "crystallized" (i.e., verbally- or content-loaded) measures such as the Wechsler intelligence scales. Flynn's initial position was that the shift in IQ

scores reflected improved test-taking ability rather than actual raised intellectual capacity (author of a book on Aristotle, he could not accept that any later generation is smarter than Aristotle's). His ideas have evolved over time, as stated by him in a recent clarifying e-mail, to: "a position that, while IQ gains cannot be equated with intelligence in any simple sense, they do signal piecemeal cognitive gains that have real-world significance. For example, people are better today at solving novel problems on the spot, more rational in their political discussions [maybe in New Zealand—SG], more likely to use their leisure in cognitively demanding ways, etc."

Flynn does not claim to have a full explanation as to why IQ scores have been trending upward over time. Neither, for that matter, does anyone else. He has explored a number of possible causes of this phenomenon, such as genetics, socioeconomic status, higher levels of education, and increased test sophistication. Flynn believes that most of the explanation for the Flynn effect can be attributed to environmental causes, including various social trends such as smaller family size, more professional work roles, use of leisure to play chess or certain video games, etc. Such environmental factors, according to Flynn, likely will prove to be more important in explaining the Flynn effect than formal schooling. Recently, Dickens and Flynn (2001) have attributed the sheer magnitude of IQ gains to a "hypothesized 'social multiplier' phenomenon, involving a feedback loop between environment and biology driven by rising societal expectations. An

updated description of this theory can be found in Flynn (2006).

To understand trends in IQ scores, it is necessary to look at what specific cognitive skills have risen or remained much the same. Traditional academic skills, as tapped by the General Information, Arithmetic and Vocabulary sub-tests of the WISC, Armed Forces entrance tests, and the SAT have not risen. The increases have been concentrated in tests such as the Similarities sub-test of the WISC and the Raven's Progressive Matrices, which ask you to solve problems without a previously learned method for doing so.

### Is the Flynn Effect Found in People with MR?

Although the Flynn effect has begun to be used in the process of diagnosing MR, there have been relatively few studies exploring whether the effect actually occurs in people who have MR. Initially, Flynn (1987) addressed the topic of mental retardation (MR) mainly as a confounding variable, and removed subjects at the two ends of the IQ continuum. Writing about himself, Flynn (1987, p. 176) stated that he "...confined himself to studies with normal subjects, as gifted and retarded subjects give eccentric results..." In other words, Flynn doubted whether the effect could be found reliably in samples containing subjects who have MR. However, in a recent paper, Flynn (2005, p. 8) has revised that opinion, and now states that "the evidence that gains are at least as great at low levels (i.e., 55-70 IQ) as they are with subjects in the normal range is overwhelming". As example,

<sup>1</sup>The committee on Law and the Death Penalty has its own web-page, which can be accessed through the Division 33 web-site at: <http://www.apa.org/divisions/div33/lawanddeathpenalty.html>. Over the coming months, the committee will post various draft documents on this site, for the purpose of stimulating awareness and discussion among psychologists who may at some point become involved in death penalty or other forensic hearings involving persons who may have MR/DD."

<sup>2</sup>Appreciation is expressed to Keith Widaman and James Flynn for their detailed comments on an earlier draft. Professor Flynn also generously shared current papers. He can be contacted at jim.flynn@stonebow.otago.ac.nz. The author can be contacted at stephen.greenspan@gmail.com.



Flynn describes two analyses (N=1,607 and N=2,266) which show that the gains from both the WISC to the WISC-R and from the WISC-R to the WISC-III were very slightly greater for subjects with MR than for the overall norming samples.

Knowledge of the Flynn effect has likely contributed to the decision by the major IQ test publishers to revise and re-norm their tests every ten or fifteen years, although other considerations undoubtedly play a role. The fact that at certain points in time an IQ test is made more difficult in order to control for the upward population drift in test scores (or for any other reason), has obvious implications for the likelihood that a person will receive a test score which will qualify him or her for a diagnosis of MR. The impact of changing test norms on the diagnosis of MR can be addressed, independent of whether one attributes such changes to the Flynn effect or whether the Flynn effect can be found to occur in persons who have MR.

Such impact has been explored in a widely-cited study by Kanaya, Scullin and Ceci (2003). Specifically, they were interested in seeing what would happen to children whose WISC-R IQs fell right around the dividing line (IQ around 70) between borderline and mild MR when they were retested three or four years later on the new WISC-III. Kanaya, Scullin and Ceci hypothesized that more children scoring above 70 would be recommended for MR services when tested on the WISC-R and then retested on the WISC-III than would children tested or retested on the WISC-R or tested and retested on the WISC-III.

Support for this hypothesis was found in the fact that there was a mean drop of 4.5 IQ points (from 78.4 to 73.9 for children in the borderline range and 5.3 points for children in the mild MR range) when going from the WISC-R to the WISC-III, while there was essentially no change for the other two conditions. More dramatic support was found in the recommendations of school psychologists for MR classification of children scoring in the borderline (IQ between 71 and 85) category. While the percentage after retest remained around 9% for children in the WISC-R to WISC-R and WISC-III to WISC-III groups, it doubled to nearly 20% for children in the WISC-R to WISC-III condition.

Kanaya, Scullin and Ceci (2003) raise legitimate concerns about the implications for both individuals and society of changing IQ norms, given the importance placed on IQ score in determining eligibility for a variety of social programs, entitlements and protections. They indicate that the times "...to be particularly cautious are when a

test is either at the beginning or at the end of its norming cycle" (p. 789), with a test being least valid when administered at the end of its cycle. An obvious implication is that clinicians should exercise care when relying on an IQ score to diagnose MR, especially when the consequences, as in death penalty exemption hearings, can be very serious for the individual.

The finding that children with MR or borderline intelligence experienced an approximately five-point IQ drop when tested on the WISC-R and retested on the WISC-III shows clearly that the Flynn effect has important implications for the diagnosis of MR. The Flynn effect postulates that the general population is getting smarter, at least as measured by IQ tests. However, these studies (most of them involving adults) have all been cross-sectional in nature. The Kanaya, Scullin and Ceci (2003) study, involving lower-IQ children, may have been the only such investigation in which within-subject longitudinal methods were used, even if the time-frame between test 1 and test 2 was very short (about four years on average). While the children who took the WISC-R and then took the WISC-III experienced an average drop of about five points, those who took and retook the WISC-R or took and retook the WISC-III received IQ scores that were virtually identical. We do not know if longitudinal methods would produce similar results for the general population, given that the average normally-developing child has no need to have his or her intelligence tested once, let alone twice. Certainly, children and adults with (or without) MR can and do have IQ scores which go up, or down, over time, with these changes often being attributed to environmental factors, such as cognitive stimulation or neglect. Given the shifting nature of the MR definition (e.g., changes in IQ eligibility criteria and the increasing emphasis on adaptive

skills) and changes in the policies of schools and other MR-labeling institutions over the past few decades, it is impossible to say whether the class of people with MR—in line with the general Flynn effect—is getting smarter. However, there is no doubt that there has been a mean rise in adult IQ of individuals born with various congenital disorders such as Down, Prader-Willi and Williams syndromes, with this rise usually attributed to the advent of early special education programs and improved medical interventions (Hodapp & Dykens, 2003). At any event, the Kanaya, Scullin and Ceci (2003) study certainly demonstrates that the likelihood that someone whose true IQ falls around 70 will qualify for the MR label is heavily affected by what test the person happened to have been given.

### Use of the Flynn Effect in Atkins Hearings

One minor problem with relying on the Kanaya, Scullin and Ceci (2003) study to justify making changes in individual IQ scores in Atkins cases is that most of the (usually multiple) intelligence test scores in the files of capital defendants are versions of the WAIS, while the Kanaya, Scullin and Ceci (2003) study involved children and two versions of the WISC. It is fair, however, to assume that the same downward shift in IQ scores from the WISC-R to the WISC-III that was found in children at the upper boundary of mild MR will also be found in the shift from the WAIS-R to the WAIS-III for adults at the upper boundary of mild MR, where most defendants making an Atkins claim will be found. This is because it has been demonstrated that successive versions of the WAIS, as well as the Stanford-Binet and other adult IQ tests, show the same shift from easier to harder that was found

## PSYCHOLOGY IN MENTAL RETARDATION AND DEVELOPMENTAL DISABILITIES APA DIVISION 33 SPRIG 2006 VOLUME 31, NUMBER 3

### Editorial Policy

Psychology in Mental Retardation and Developmental Disabilities is an official publication of Division 33 of the American Psychological Association. It is devoted to keeping members informed about the activities of Division 33 and to news and comment concerning all aspects of service, research, dissemination, and teaching in psychology and mental retardation. Brief articles about policy issues in psychology and mental retardation, as well as descriptions of service programs and preliminary research summaries are invited. We are especially interested in articles inviting the reaction and comment of colleagues in future issues. Comments and letters will be published as

space allows. Manuscripts of between 500-1000 words must conform to APA style and should be submitted in triplicate (hard copy) and on diskette in Word format. Articles, comments, and announcements should be sent to Samuel J. Thios, Ph.D., Editor, Psychology in Mental Retardation and Developmental Disabilities, Department of Psychology, Denison University, Granville, OH 43023. Books, films, videotapes, and other material also may be submitted for possible review to the Editor.

Unless stated otherwise, opinions expressed are those of the author and do not necessarily represent official positions of Division 33.

Deadlines are September 15, Fall; March 1, Spring; June 1, Summer.



to be the case with successive versions of the WISC. Therefore, it is essential that psychologists rely on the general Flynn effect to consider old adult IQ scores in light of changed norms, and such a practice has been recognized by the courts (such as the fourth US Circuit Court, in *Walker v. True*, 2005) and by most psychologists.

An example of such a court-approved use of the Flynn effect occurred in a habeas (post-conviction) Atkins hearing for Jorge Vidal, Jr., held in Tulare County, California in 2004 (28 Cal. Rptr.3d 529). Defense expert Keith Widaman, a psychology professor at the University of California at Davis, cited the Flynn effect to justify changing Mr. Vidal's most recent IQ score downward. In 2003, Vidal received a full-scale IQ of 78, which was above the 70-75 IQ ceiling for diagnosing MR. The test—nearing the end of its lifecycle—used was the WAIS-R, which was published in 1981, over 21 years earlier. As the Flynn effect shows an upward population shift of .33 points per year, Widaman gave Vidal an adjusted IQ score of 71. His reasoning was that if Vidal were to be administered the same exact test items in 2004, he would score seven points lower on the more difficult norms likely to be used by the publisher now. Using the Flynn-generated estimate of a rise in population Wechsler IQ of .33 per year, Widaman argued that a downward adjustment of Vidal's IQ by seven points (.33 x 21) was justified, which would place his true IQ at (a MR-qualifying) 71 rather than (a MR-disqualifying) 78. An illustration of how the Flynn effect might be used to make similar adjustments in other, hypothetical, cases can be found in Table 1.

Although the Flynn effect has been used mainly by psychologists testifying on

behalf of the defense, in theory it could also be used by psychologists testifying on behalf of the prosecution. This in fact occurred in a 2005 pre-trial Atkins hearing in Los Angeles for murder defendant Mauricio Silva. Neuro-psychologist Paul Satz, a professor emeritus at UCLA's Neuropsychiatric Institute, argued the existence of a novel "reverse Flynn effect" as a justification for changing one of Mr. Silva's IQ scores in a direction which contradicted the testimony of two defense experts, one of them me, to the effect that he qualified for a diagnosis of MR. Specifically, Dr. Satz argued that a 2004 full scale WAIS-III IQ of 74 (barely under the 75 MR ceiling) should be adjusted upwards by 5 points, placing it at a MR-disqualifying 79. Satz's reasoning was that because the predecessor test, the WAIS-R, had easier norms than the later WAIS-III, then if one went backwards 15 years, Mr. Silva would actually get retroactively smarter by 5 points. Satz came to this conclusion on the basis that if Silva had been administered the same exact WAIS-III test items 15 years earlier, in 1989 (something of course purely hypothetical as the WAIS-III did not exist then), he likely would have received a WAIS-III IQ score of 79, because of the likelihood that the test publishers would have used easier norms at the earlier date. The 5 point estimate comes from the backward trend over that 15 year period of .33 points per year (15 x .33= 4.95).

Satz's use of reverse re-scoring on a test that did not even exist 15 years earlier is, of course, more problematic from a logical standpoint than was Widaman's attempt to rescore an already existing test that was administered 21 years after it was initially normed. Nevertheless, psychologists and attorneys who argue that use of the Flynn

effect is necessary because it helps to neutralize the impact of a spuriously high IQ score should keep in mind that they are not the only ones who can attempt to make such a claim, even if most psychologists dispute the validity of the reverse Flynn effect. Professor Flynn himself considers the reverse Flynn effect to be unsupportable, as noted in a message to me that "by going back far enough, to the norms of, say, 1918, you could argue that virtually no person today is mentally retarded". Even Judge Johnson, while ruling against Silva's Atkins claim, found unpersuasive Dr. Satz's "unusual opinions on the 'reverse Flynn Effect.'" (Order Following Hearing on Mental Retardation in *People v. Silva*, Los Angeles Superior Court Case No. A753263, Nov. 14, 2005, slip op. at 11).

## Conclusion

The main reason why it is necessary for psychologists to adjust IQ scores down (or, more controversially, up) is, obviously, due to the primary importance given to IQ scores in diagnosing MR in various settings, including Atkins hearings. In spite of the emphasis given in recent AAMR diagnostic manuals to the importance of not using IQ cut-offs as literally as in the past, and on the importance of equally considering both IQ and adaptive functioning, the fact is that obtaining an IQ score below a specific ceiling (now typically more liberally interpreted as 75 rather than 70) still remains a requirement for receiving a diagnosis of MR in most settings, including the court system.

Typically, there are multiple IQ tests in the record of an Atkins applicant, especially when he has been on death row for 15 or more years, and these fluctuate around the upper IQ boundary of the mild MR range. An adjustment based on the Flynn effect can be very helpful to a defense counsel when one or two of these scores is a little bit too high. Before the Atkins ruling, it was not unusual for defendants, even in states which had their own Atkins-type statutes, to receive the death penalty because of a single IQ score (out of multiple ones) that was one or two points too high. Any practice which would reduce such an overly-literal reliance on an IQ number, especially when there is reason to question the validity of that number, is to be commended. However, a more direct approach would be to reduce drastically the importance given to IQ scores in making a diagnosis of MR.

Flynn himself has made essentially that point in papers exploring the implications of his effect for diagnosing MR and related learning disorders (Flynn,

Table 1

### Illustrating Use of the Flynn Effect to Adjust Individual IQ Scores

Formula: (a) subtract the year the WAIS or WISC (or other IQ test) edition was published from the year it was administered<sup>1</sup>.

(b) multiply this number by 0.33

(c) subtract this number from the individual's Full-Scale IQ

Publication dates: WAIS=1955; WAIS-R=1981; (c) WAIS-III=1997  
WISC=1967; WISC-R=1974; WISC-III=1991; WISC-IV=2003

Example 1: In 1980, Andy H. received a WAIS Full-Scale IQ of 80.

(a) 1980-1955 = 25

(b) 25 x .33 = 8.25

(c) 80 - 8 = 72 (Andy H's Flynn-adjusted IQ)

Example 2: In 1983, Susan W. received a WISC-R Full-Scale IQ of 77

(a) 1983-1974=9

(b) 9 x .33 =2.97

(c) 77 - 3 = 74 (Susan W's Flynn-adjusted IQ)

<sup>1</sup>Flynn (2005) advocates pushing the test edition's date back another year or two, to reflect the year when norms were collected rather than year of a test's publication. He also presents convincing evidence that sampling problems with the WAIS-III justify subtracting an additional 2.34 IQ points from any WAIS-III score, even if obtained in the same year in which the test was published.



1985, 2000). Essentially he has argued the following: given that (a) IQ is an imperfect measure of "intelligence", (b) behavior in the world is a better indicator than IQ of how smart someone actually is, and (c) periodic changes in test norms create uncertainty about the extent to which one can rely on any particular test score, then it makes the most sense to base a diagnosis of MR on a combination of adaptive behavior and informed clinical judgment, with relatively little weight given to IQ scores. Flynn has decried the degree of importance which psychologists attach to individual IQ scores, and he is especially critical of the notion that an artificial invention—such as a number ending in zero, or a standard deviation unit—has any meaningful role to play in the process of determining whether or not someone has MR.

Given that Flynn's main point has been about the dubiousness of the proposition that IQ equals intelligence, one might think that he would suggest that an alternative to adjusting IQ scores would simply be to stop relying on them altogether when diagnosing MR. He seems, however, to have accommodated (as have other IQ critics, including myself) to the impossibility of loosening the magical hold that IQ plays on the general imagination. Flynn has come out strongly in favor of adjusting IQ scores in light of changing population norms, when diagnosing MR, especially in death penalty cases. His position on IQ adjustment is spelled out in an under-review paper (Flynn, 2005), in which he presents an IQ-adjustment formula that is essentially identical to the one presented by me in Table 1. Here he also points to the fact that given the adversarial nature of the American justice system, it is almost inevitable that the prosecution and defense will each have their own psychologists contradicting one another about whether the defendant has MR, leaving IQ as the only "neutral" criterion.

In addition to the purely statistical argument that IQ scores should reflect current norms, there is something else going on when a psychologist makes a Flynn-mediated adjustment. This has to do with the diagnostic process itself and the clinical judgments that are being made regarding whether an individual behaves in a manner congruent or incongruent with the MR behavioral taxon. In thoughtful clinicians, the process of diagnosing someone as MR or non-MR, as is also the case when making other psychiatric or educational diagnoses, is based on an intuitive judgment which integrates information from many different sources, including how an individual comes across in dealings with the examiner. This is not a bad thing; it is how it should be, assuming

that the clinician has sufficient knowledge and experience relating to MR (a big assumption, as many forensic psychologists testifying in Atkins cases lack relevant qualifications). In the case of MR, however, the clinician is so hemmed in by the necessity of obtaining IQ scores which fall within the MR range that the Flynn effect helps to explain away one or more scores that might otherwise prove problematic in justifying a diagnostic conclusion that he or she has (quite often, legitimately) already reached.

An example of this can be found in the Vidal case where Widaman noted that Mr. Vidal fit the behavioral profile for mild MR, as reflected in noted instances of his extreme gullibility. I admit that this warmed my heart, as I have been advocating the importance of gullibility as a key defining characteristic of the mild MR category for the past few years (Greenspan, 1998, 2003; Greenspan, Loughlin & Black, 2001). Widaman actually used evidence of gullibility as a basis for another interesting argument, which was that Vidal's very low verbal IQ was a more valid indicator than full scale IQ because of the likelihood (not established empirically, to my knowledge) that gullibility is a logical outcome of low verbal IQ. The trial judge accepted Widaman's view of Vidal as having MR based on the use of both verbal IQ and Flynn-adjusted full scale IQ. A California appellate court overturned the trial judge's ruling, specifically rejecting the attempted substitution of verbal IQ for full scale IQ. However, Widaman's use of the Flynn effect to adjust one of Mr. Vidal's full scale IQ scores was explicitly noted as an acceptable practice by the appellate court. Unfortunately, we are not yet at a point where evidence that someone fits the MR behavioral taxon (such as by presenting as a very gullible and socially naïve individual) is sufficient basis for the diagnosis, in the face of one or more IQ scores in the high 70's. Hence the need, universal and hardly limited to Professor Widaman, to use the Flynn effect to bring one or more IQ outliers more into line with one's diagnostic conclusion. Interestingly, Flynn (2005) also cited gullibility—in an illustration involving a low-IQ person whose lack of "moral autonomy" allowed him to be tricked into committing a crime—to justify making adjustments which bring an individual's IQ more into line with his actual social functioning.

Although initially skeptical (mainly because of a strong contrarian tendency), I have become firmly convinced that use of the Flynn effect to adjust individual IQ scores is an appropriate, indeed essential, practice.

#### AMERICAN PSYCHOLOGICAL ASSOCIATION

##### Mental Retardation and Developmental Disabilities Division 33

###### President

Bruce L. Baker  
UCLA  
1285 Franz Hall  
Box 951563  
Los Angeles, CA 90095-1563

###### President-Elect

Johannes Rojahn  
Ctr. for Cognitive Development  
George Mason University  
4400 University Dr MSN 206  
Fairfax, VA 22030-4444

###### President-Elect-Designate

Stephen F. Warren  
The Schiefelbusch Institute  
for Life Span Studies  
1052 Dale  
1000 Sunnyside Avenue  
University of Kansas  
Lawrence, KS 66045

###### Past President

Sara Sparrow  
Yale University  
333 Cedar St  
New Haven, CT 06510

###### Secretary-Treasurer

Robert Hodapp

###### APA Council Representative

James Mulick

###### Members-at-Large

Patricia Meinhold  
Laura Lee McIntyre

###### Newsletter Editor

Samuel J. Thios  
Department of Psychology  
Denison University  
Granville, OH 43023

###### Newsletter Editorial Assistant

Laura Bracken

###### Division 33 Committees

###### Constitution and Bylaws

Stanley E. Lunde, Chair  
Albert J. Berkowitz  
Olivia J. Hooker  
Marie S. Cisseay

###### Edgar A. Doll Award

William MacLean, Chair  
Philip W. Davidson  
Laraine Glidden

###### Ethics

Patricia Meinhold, Chair  
Richard Fox  
Lawrence Hartlege  
James Mulick  
Barbara McCasland

###### Fellows

Wayne Silverman, Chair  
John Belmont  
John Borkowski  
Laraine M. Glidden  
Michael Guralnick  
Carl Haywood

###### Finance

Michael Cataldo, Chair  
Susan Heimlich  
Steven Reiss

###### Legislation and Public Affairs

Joel M. Lavy, Chair  
Gershon Berkson  
Linda Bilsky  
Alan A. Baumeister  
Jack Stark

###### Membership

Warren B. Zigman, Chair

###### Nominations and Elections

Sara Sparrow, Chair  
Warren Zigman  
Gail O'Connor  
Stephen Schroeder

###### Professional Standards

Shelia H. Richard, Chair

###### Program

Johannes Rojahn, Co-Chair

###### Publications

Samuel J. Thios, Chair

###### Quality of Services

Allan G. Barclay, Chair



Changes in test norms, in a harder direction, from one version of an IQ test to another, is certainly a major consideration to be taken into account, along with other factors—such as practice effect, test reliability, aspects of test administration, etc.—when considering how much weight to attach to a particular test score.

There is a strong case to be made for the practice of adjusting a test score obtained on an outdated test to better reflect how an individual would look on that same test with the changed norms certain to be in place if the test were published closer to the time of the testing. Such an adjustment in past IQ scores, to bring them in line with the gradual toughening of norms, is clearly justified, especially if it contributes to a loosening of the concreteness with which clinicians and agencies tend to interpret IQ scores and cut-offs.

Use of the so-called reverse Flynn effect to make scores go up, on the other hand, strikes me as more problematic, from both a logical and an ethical standpoint. Going back in time to see how an individual would have scored at a particular point in the past when the test was not even in existence is a stretch. Flynn denounces this practice as well as the opposite one, whereby some defense attorneys have attempted to show how an individual's IQ could be lowered even further at some future time, as unjustified and illogical (by treating some future event as if it had already happened or, conversely, by altering history to invent a test at a time when it did not yet exist). Adjusting an actual test score to take into account changes in norms during the period since its publication is another matter entirely, however. As Flynn (2005, p.8) has stated, "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 IQ test you happened to take [and when you take it]". Flynn, an American living in New Zealand, further states that "the U.S. Constitution cannot tolerate a death penalty that is the random result of whatever IQ test a psychologist happened to have in stock" (p.17).

Given that mild MR is still a somewhat inadequately-defined category, it is important to err in very close cases on the side of being overly inclusive, especially given the potentially fatal consequence of a false negative diagnostic conclusion. Use of the Flynn effect is a useful, and valid, method for increasing the likelihood that a psychologist will correctly diagnose MR in someone deserving of that label.

## References

- Dickens, W.T. & Flynn, J.R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, 108, 346-369.
- Flynn, J.R. (1984). 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. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.
- 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. (2000). The hidden history of IQ and special education: Can the problem be solved? *Psychology: Public policy and law*, 6, 191-198.
- Flynn, J.R. (2005). [Tethering the elephant: Capital cases, IQ and the Flynn effect.](#) Unpublished manuscript. Dunedin, New Zealand: University of Otago.
- Flynn, J.R. (2006). The Flynn effect: rethinking intelligence and what affects it. In C. Flores-Mendoza & R. Colom (eds), [Introduction to the psychology of individual differences.](#) Porto Alegre: ArtMed. (English translation from the Portuguese is available from jim.flynn@stonebow.otago.ac.nz).
- Greenspan, S. (1998). A contextualist perspective on adaptive behavior. In R. Schalock (Ed.), [Measuring adaptive behavior](#) (pp. 61-80). Washington, DC: American Association on Mental Retardation.
- Greenspan, S. (2003). Vulnerability as an overlooked aspect of the mental retardation phenotype. In H.N. Switzky & S. Greenspan (Eds.), [What is mental retardation?](#) Washington, DC: AAMR/disabilitybooksonline.com.
- Greenspan, S., Loughlin, G. & Black, R. (2001). Credulity and gullibility in persons with mental retardation. In L. M. Glidden (Ed.), [International Review of Research in Mental Retardation](#) (vol. 24, pp. 101-135). New York: Academic Press.
- Hodapp, R.M. & Dykens, E.M. (2003). Looking to the 21<sup>st</sup> century: Toward an etiology-based classification system. In H.N. Switzky & S. Greenspan (Eds.), [What is mental retardation?](#) Washington, DC: AAMR/disabilitybooksonline.com.
- 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.
- Nettelbeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence*, 32, 85-93.
- Order Following Hearing on Mental Retardation in People v. Silva, Los Angeles Superior Court Case No. A753263 (Nov. 14, 2005)
- Rowe, D. C., & Rodgers, J. L. (2002). Expanding variance and the case of historical changes in IQ means: A critique of Dickens and Flynn (2001). *Psychological Review*, 109, 759-763.
- Walker v. True, No. 04-016 (4<sup>th</sup> Cir., February 17, 2005).