Professional Psychology: Research and Practice
2008, Vol. 39, No. 6, 619–625

0735-7028/08/$12.00 DOI: 10.1037/a0012693

Adjusting IQ Scores for the Flynn Effect: Consistent With the Standard of Practice?

Leigh D. Hagan
Virginia Commonwealth University

Eric Y. Drogin
Harvard Medical School

Thomas J. Guilmette
Providence College

Should psychologists adjust obtained IQ scores to accommodate the *Flynn effect* (J. R. Flynn, 1985)? The authors surveyed directors of doctoral training programs approved by the American Psychological Association and board-certified school psychologists and completed a systematic review of IQ test manuals, contemporary textbooks on IQ testing, federally regulated IQ testing protocols, and various sources of legal and ethical guidance. They confirmed in each instance that such adjustments to IQ scores do not comport with prevailing standards of psychological practice. Results of IQ testing may be applied to a broad range of psycholegal issues, many of which cannot be anticipated. Psychologists assist examinees, courts, and other 3rd parties most effectively by administering and interpreting IQ tests in their intended fashion.

Keywords: practice standards, Flynn effect, IQ, intelligence testing

Each year psychologists assist in hundreds of thousands of legal determinations through evaluation reports and expert testimony based on scientific knowledge of measurement procedures, including intelligence testing. Psychologists' reports of IQ test data can have a major impact on access to services and even life-and-death decisions (*Atkins v. Virginia,* 2002). In addition to specific medicolegal evaluations, psychologists administering an IQ test for one purpose, such as treatment planning or special education, might find their work product used for a different purpose years later in a criminal proceeding, disability evaluation, or claim of damages in a lawsuit. Given the possible intended and unintended consequences of intelligence test records, understanding and comporting with practice standards is essential. A debatable but potentially emerging standard is whether psychologists should subtract points from an individual's obtained IQ score on the basis of the Flynn effect (FE; Flynn, 1985), a phenomenon in which IQ means have been shown to increase in the general population across time.

Why Standards Make a Difference

A standard is "a model accepted as correct by custom, consent, or authority" (Black, 2004, p. 1441). Standards establish parameters of practice and communicate the prevailing views of psychology to those outside of behavioral science. Psychological practice standards do not exist in a vacuum. Law, science, and ethical principles impact each other; none stand in isolation. In the psycholegal context, each guides the psychologist who, in turn, advises the court about prevailing standards.

The FE and Adjusting IQ Scores

The FE refers to the finding that the general population's average IQ test scores have increased over the past several decades (Flynn, 1985). Although some studies have reported an increase of about 0.30 IQ points per year (Flynn, 1999), the issues underpinning the changes in average scores over time are complex and exceed the scope of this article. The research-informed practitioner should note the differential impact of a host of variables, including gender and ethnicity ("Latest Thinking," 2007), age and culture (Flynn, 1987), level of industrial and technological development (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003; Flynn, 1987), the type of cognitive task being measured (fluid or crystallized), and where the score falls along the distribution curve (Zhou & Zhu, 2007).

LEIGH D. HAGAN received his PhD in counseling psychology from the University of Missouri—Columbia. He maintains an independent clinical and forensic practice in Chesterfield, Virginia, and is an affiliate assistant clinical professor of psychology at Virginia Commonwealth University. His areas of professional interest include capital defendant sentencing, custody evaluation, and practice standards for mental health professionals in the forensic arena.

ERIC Y. DROGIN received his PhD in clinical psychology from Hahnemann University and received his JD from the Villanova University School of Law. He maintains an independent forensic consulting practice and serves on the faculty of the Harvard Medical School, Beth Israel Deaconess Medical Center, as a clinical instructor and member of the Program in Psychiatry and the Law. His areas of professional interest include trial consultation and expert testimony.

THOMAS J. GUILMETTE received his PhD in counseling psychology from the University of Missouri—Columbia. He is a professor of psychology at Providence College and an adjunct associate professor of psychiatry and human behavior at the Warren Alpert Medical School of Brown University. His professional and research interests include neuropsychology and psychology and law.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Leigh D. Hagan, P.O. Box 350, Chesterfield, VA 23832. E-mail: lhagan@leighhagan.com

Flynn (2007) documented a wide range of score fluctuations, including a slight reverse of the FE, depending on which Wechsler scale was used. Some countries have actually shown a reverse FE in more recent years (Shayer, Ginsburg, & Coe, 2007).

Our research focuses on the straightforward question: Is it the standard of practice to adjust obtained IQ scores in light of the FE? To the extent that the empirical impact of the FE is blind to the purpose for which a test is administered, then practicing psychologists need to be cognizant of this issue, not just for criminal evaluations, but for special education, disability, employment, and any other purpose. Although mainstream recognition of the FE as an authentic psychometric consideration has increased, the question of how to accurately represent its impact for a particular individual's earned scores on IQ tests is a different question altogether.

Of particular importance to the evaluating psychologist is whether the observed changes in group mean scores over time apply reliably to a specific individual. The question here is whether the FE's broad construct applies to a specific evaluee's IQ test scores, particularly when the individual's obtained score is offered as evidence in support of a theory to prove a legal fact. Specifically, is it the generally accepted practice in the field of psychological testing to adjust a particular person's earned IQ scores or to recalculate norm means on the basis of the FE?

Flynn has advanced several different positions on this point. In 1987, he cautioned against placing unwarranted emphasis on individual IQ scores, asserting that "IQ tests do not measure intelligence but rather a correlate with a weak causal link to intelligence" (Flynn, 1987, p. 171). Later, he took the position that the Wechsler Adult Intelligence Scale (3rd ed.; Wechsler, 2002) might be reliable for scores below 70 and concluded that the FE was a factor of 0.25 rather than 0.30 (Flynn, 1998). Shortly thereafter, in 2000, he proposed abandoning the use of IQ scores for mental retardation determination rather than adjusting obtained scores, arguing that "the fact that people will get quite different scores on different IQ tests can be manipulated by psychologists to suit their clients' needs" (Flynn, 2000, p. 191).

In 2006, Flynn advocated adjusting individual IQ scores on the premise that doing so creates no greater error than failing to do so. He argued that resistance to the practice of subtracting points from an individual's obtained score was not particularly defensible. Yet, within the same article, he pointed out that the FE is not generally accepted in the clinical field. Most recently, with respect to deducting 0.30 IQ points per year, Flynn (2007) acknowledged that "recommending such a simple cure for obsolete norms assumes too much" (Flynn, 2007, p. 134).

Although Flynn's position about IQ scores varies in his scholarly articles, he steadfastly advocates subtracting obtained IQ points in criminal sentencings (e.g., *Berry v. Mississippi*, 2005; *Walker v. True,* 2005). To the extent that the FE is a function of IQ tests generally, and if adjusting an individual's obtained IQ scores is the accepted convention in clinical practice, then one would expect to find empirically based support for individual score adjustments across all IQ test purposes. One would not expect to find the discussion limited to a narrow range of purposes, such as capital case advocacy. Yet, the professional literature is almost silent on individual score adjustments outside of the criminal forensic arena.

Although the FE appears in hundreds of articles, most are of a technical nature or focus on social policy implications. Very few psychologists forward the position that an individual's obtained IQ scores should be reduced by a numerical factor based on the FE. Kanaya, Scullin, and Ceci (2003) argued for score adjustments on the basis of a large scale empirical study. Greenspan (2006), in a discussion article absent new empirical data, asserted that subtracting IQ points from an individual's obtained score is not only appropriate, but essential. Other psychologists have argued through their reports and testimony in the capital-sentencing context that adjusting scores is the normative practice (*Bowling v. Kentucky,* 2005; *Green v. Johnson,* 2008; *Howell v. Tennessee,* 2004; *People v. Superior Court [Vidal],* 2005; *Walker v. True,* 2005; *Walton v. Johnson,* 2006;), but they drew from work previously cited without adding to the empirical research base of knowledge.

Division 33 of the American Psychological Association (APA) called for an ad hoc committee to further study this issue and to find those areas in consensus on standards for psychologists (Olley, Greenspan, & Switzky, 2006). Beyond the works previously cited, we found no empirical studies advocating for FE-based score adjustments in special education, disability, parental rights termination, or any other purpose for which psychologists ordinarily administer IQ tests.

A dichotomy sometimes emerges between scholarly empirical research and expert testimony in the courtroom. Cases abound in which expert witnesses have testified that adjusting an individual's obtained IQ score is the standard (*Commonwealth v. Prieto,* 2007; *Green v. Johnson,* 2008; *People v. Superior Court [Vidal],* 2005; *State v. Keel,* 2003; *Walker v. True,* 2005). In these same cases, however, other qualified experts have testified that adjusting IQ scores is not the accepted practice.

Other scholars and expert witnesses oppose adjusting IQ scores for several reasons. Moore (2006) challenged the proposition that adjusting individual IQ scores is the standard of practice. Lacritz and Cullum (2003) advised that "caution should be used in applying Flynn's philosophy to actual patients, as there are many sources of variance unaccounted for by his formulas that could impact an individual's score" (p. 529).

Young, Boccaccini, Conroy, and Lawson (2007) provided the closest analysis to date with respect to the standard of practice and IQ score adjustment in death penalty evaluations. They found that among experienced death penalty evaluators, most psychologists reported being aware of the FE either by name or the underlying construct, yet most (71%) of the psychiatrists surveyed had never heard of the concept underpinning the FE. Olley et al. (2006) also pointed out the lack of consensus about how to present IQ data for *Atkins* hearings (see *Atkins v. Virginia,* 2002) for the court to determine if the capital defendant meets the statutory criteria for mental retardation. We investigate whether there presently exists a standard for adjusting individually obtained IQ scores in a way that is accepted as correct in light of custom, consent, and authority.

Search for a Standard

Survey 1: Doctoral Program Directors

Participants were program directors of APA-approved clinical, counseling, and school psychology doctoral programs as identified by their respective APA Web sites. Of the surveys sent to each of 358 program directors, all respondents were program directors, IQ/intelligence instructors or supervisors, or a combination of both cat-

egories. The largest portion (43%) received their doctoral degree more than 20 years ago. Most (69%) taught or supervised doctoral students' IQ testing in the previous 3 years. We did not solicit information about the respondents' forensic experience specifically but did inquire about their knowledge of the FE in any arena.

The survey questions were not limited to any specific IQ testing purpose. Respondents were instructed to stop filling out the survey and return it if they were not at all familiar with the FE. The remaining items sought to determine whether graduate school faculty members were teaching their students to calculate, adjust, and list scores on the basis of the FE in ways that have previously been described in some cases as the accepted professional standard (*Commonwealth v. Prieto,* 2007) or as near universal (*Green v. Johnson,* 2008).

We found that of the 89 respondents, 36% indicated that their familiarity with the FE was slight or that they had no familiarity at all; 37% were moderately familiar, whereas 27% were very familiar.

Because our focal interest was in contemporary teaching practices, the balance of the data analysis was derived from the responses of those faculty who indicated that they had taught or supervised graduate student IQ testing and interpretation within the previous 3 years. Excluding those who had not taught or supervised students also eliminated respondents who were not at all familiar with the FE. Of the remaining 57 respondents, 93% reported that they had taught or supervised IQ testing in the past 3 years.

Table 1 reveals that, of this group, 82% indicated that it was only slightly important or not at all important for students to learn to calculate the FE when listing actual scores in the written report. In addition, although 61% believed that it was moderately or very important for students to learn to consider the FE when interpreting scores, only 18% indicated that it was very important, which is the same as the percentage who believed that it was not at all important.

Simply considering the FE is not the same enterprise as memorializing that thought process in the narrative of a written report. Thus, Table 2 reveals the frequency with which the participants taught their graduate students to comment on the FE in reports or to actually recalculate or adjust IQ scores based on the FE. As can be seen in Table 2, two thirds of the respondents never taught students to comment on the FE, and 9 out of 10 never taught their students to adjust or recalculate IQ scores.

The survey inquired about teaching students to adjust IQ scores depending on where in the distribution the score might fall. The vast majority (94%) reported that they never taught students to adjust obtained IQ scores, irrespective of their position in the distribution. Only 2% advocated adjusting IQ scores across the entire range.

Rather than adjusting obtained IQ scores, some psychologists have proposed compensating for the FE by adjusting the mean score from the published norms and then reporting the obtained score relative to the newly adjusted mean (*Green v. Johnson,* 2008). Teaching students to adjust obtained scores after recalculating the published means was even less likely, with 95% never instructing in this practice. None of the respondents indicated that they promoted this practice for all IQ testing referral questions.

Some researchers and testifying experts (Flynn, 2006; *Green v. Johnson,* 2008; Kanaya et al., 2003; *People v. Superior Court [Vidal],* 2005; *Walker v. True,* 2005;) have advocated adjusting the obtained IQ score, not just for each year after the publication of the test, but also for each year after the normative data were collected. This procedure accounts for the postulated lag between data collection and publication of the test manual. Flynn (2006) referred to this process as "the general rule" (p. 179).

The survey polled for this practice. Of the participants, 79% (45 out of 57) did not teach their students to make numerical adjustments to the obtained IQ, but of those who did, the majority (75% or 9 out of 12) relied on the year the norm group was collected when adjusting the IQ.

No consensus emerged about a scientific authority for adjusting scores. The much larger majority (86%) declined to identify any scientific, legal, regulatory, or ethical authority for adjusting obtained scores or recalculating means because they did not train students to use this practice.

Survey 2: Diplomates in School Psychology

The second survey queried clinicians who had achieved the advanced credential of board certification in school psychology from the American Board of Professional Psychology. We chose these psychologists because they frequently engage in intelligence testing and have considerable experience and expertise in interpreting archival test data.

Participants in Survey 2 were all of the 141 American Board of Professional Psychology school psychologists identified by the board's Web site. We received 28 usable returns, or 23% of the viable pool. The majority had over 20 years of experience. Most (93%) of the viable respondents had personally administered, scored, and interpreted more than 200 individual IQ tests.

The majority (68%) were moderately or very familiar with the FE. A large majority (94%) of the viable participants reported that they had never adjusted obtained IQ scores on the basis of the FE when reporting numerical IQ scores. Only one participant reported adjusting obtained scores in some cases (few but less than most). None reported doing so in most or all cases. Only one reported

Table 1
Percentage of Participants Who Considered It Important for Students to Learn to Calculate or Consider the FE in Written Reports

Item	Not important	Slightly important	Moderately important	Very important
Learning to calculate the FE when listing scores in written reports	46.4	35.7	14.3	3.6
Learning to consider the FE when interpreting scores in written reports	17.9	21.4	42.9	17.9

Note. Participants were program directors or instructors of IQ testing courses in clinical, counseling, or school psychology programs approved by the American Psychological Association ($n = 56$). FE = Flynn effect.

Table 2
Percentage of Participants Who Taught Students to Comment on the FE or Recalculate IQ Scores on the Basis of the FE in Written Reports

Item	Never	Yes, in all cases	In MR cases only	In certain other cases	In MR and certain other cases
Teach students to comment on the FE in reports	68.5	3.7	7.4	18.5	1.9
Teach students to recalculate IQ scores on the basis of the FE	91.9	0.0	3.6	3.6	0.0

Note. Participants were program directors or instructors of IQ testing courses in clinical, counseling, or school psychology programs approved by the American Psychological Association (n = 54 and 56, respectively, for Item 1 and Item 2). FE = Flynn effect; MR = mental retardation.

commenting on the FE in the written narrative. None of the respondents reported having adjusted archival scores retrospectively when reviewing previous IQ scores. These findings are consistent with the testimony in *Green v. Johnson* (2008) in which, out of 5,000 school-based IQ test reports between 1999–2001, only 6 mentioned the FE. None adjusted the obtained IQ scores.

Other Standards Authorities

The search for other IQ testing standards authorities led to the test manuals themselves because multiple authorities substantiate that the manual is the sine non qua for test administration and scoring.

We included current adult IQ tests fully meeting the criteria of the National Research Council (2002), instruments authorized by the Social Security Administration (SSA, 2006), measures identified from peer-reviewed published surveys of clinical practice patterns (Rabin, Barr, & Burton, 2005; Watkins, Campbell, Nieberding, & Hallmark, 1995), and those approved by the only two states that maintain lists of measures for capital mental retardation evaluations (Fla. Stat. § 921.137 [1], 2005; Virginia Department of Mental Health, Mental Retardation and Substance Abuse, 2005). Excluded were earlier versions of tests that psychologists might encounter in the evaluee's archives (e.g., the Wechsler Intelligence Scale for Children [3rd ed.] or the Stanford–Binet Intelligence Scales [4th ed.]) or tests constructed primarily for minors.

Six IQ tests met the inclusion criteria: the Wechsler Adult Intelligence Scale (3rd ed.; WAIS-III; Wechsler, 2002), the Stanford–Binet Intelligence Scales (5th ed.; Roid, 2003), the Kaufman Adolescent and Adult Intelligence Test (Kaufman & Kaufman, 1993), the Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003), the Multidimensional Aptitude Battery (2nd ed.; Jackson, 2003), and the Woodcock–Johnson Test (3rd ed.; Mather & Woodcock, 2001). We examined each test manual for citations of Flynn's publications, references to the FE, and any specific recommendation for dealing with the increase in scores over time.

The *WAIS-III Technical Manual–Revised* (Wechsler, 2002) acknowledges "IQ-score inflation over time" and thus recommends that "norms for a test of intellectual functioning should be updated regularly" (Wechsler, 2002, p. 9). The WAIS-III publisher specifically rejects the practice of adjusting obtained scores: "Still, there is no scientific justification for adjusting data to fit theory. As the publisher of the Wechsler series of tests, Harcourt Assessment does not endorse the recommendation made by Flynn to adjust WAIS-III scores" (Weiss, 2007, p. 1).

The Stanford–Binet Intelligence Scales and the Kaufman Adolescent and Adult Intelligence Test manuals cite Flynn (1987) but make no specific recommendation for dealing with this statistical observation beyond the general admonition to follow the scoring rules strictly. The Reynolds Intellectual Assessment Scales, the Multidimensional Aptitude Battery, and the Woodcock–Johnson Test do not reference the FE, either conceptually or by name.

Several other sources of authority illuminate whether adjusting individual obtained IQ scores is the model accepted as correct by custom or consent. The SSA eligibility determination process is one of the largest testing programs in the United States. More than 1 million individuals currently receive SSA benefits under the mental retardation criteria.

In an effort to assess the adequacy of disability determinations, the SSA engaged the National Research Council to "evaluate the existing determination process in the context of state-of-the-art scientific knowledge and clinical practice" (National Research Council, 2002, p. 1). The large-scale effort by the study group produced numerous recommendations but did not include a specific proposal to adjust individual obtained IQ scores either in current testing or for archival assessments. Instead, the study group recommended that "tests should undergo normative update, restandardization, or revision at intervals corresponding to the time expected to produce one *SEM* of change" (National Research Council, 2002, p. 125).

The SSA Program Operations Manual System articulates the disability evaluation protocol for mental retardation (SSA, 2006). The agency's policy specifically bars its reviewing staff psychologists from adjusting current and archival IQ tests scores provided by the examining psychologist (SSA, 2006). To date, no appellate court has reversed or remanded a denial of an SSA entitlement claim because of a failure to adjust IQ scores on the basis of the FE.

The use of IQ testing for special education is another substantial public policy issue impacting a large population. As many as 5 million children receive special education services under the Individuals With Disabilities Education Improvement Act of 2004. This regulation does not reference the FE and does not set a standard for adjusting an individual's obtained scores or recalculating the mean score against which the obtained score should be assessed (Individuals With Disabilities Education Improvement Act, 2004, § 300.532).

Next, our search for a standard of practice examined contemporary textbooks published for practicing clinicians and graduate students. We queried APA Online PsycNET book records for 1984 through 2007, using the keyword *IQ test*. A leading psychology

textbook publisher and current graduate school assessment faculty also contributed to a list of relevant titles. Other titles were found in the IQ testing section of the library of a university with APA-approved training programs in clinical and counseling psychology. Other titles surfaced in research publications cited earlier.

Because our interest focused on practice standards, the search included textbooks only of an applied nature. The search yielded 14 textbooks published between 1999 and 2007 (Flanagan & Harrison, 2005; Gleitman, Fridlund, & Reisberg, 2003; Groth-Marnat, 2003; Kaplan & Saccuzzo, 2005; Kaufman & Lichtenberger, 1999, 2006; Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005; Lichtenberger & Kaufman, 2004; Myers, 2007; Prifitera, Saklofske, & Weiss, 2005; Sattler & Hoge, 2006; Tulsky, Saklofske, & Ricker, 2003; Urbina, 2004; Weiss, Saklofske, Prifitera, & Holdnack, 2006).We examined each textbook for the presence of Flynn in the author index, FE in the subject index, and specific recommendations for dealing with the FE when reporting scores.

Most (79%) contemporary applied textbooks cite Flynn's research and mention the FE by name or as a concept. In contrast to the claim in *Walker v. True* (2005), none recommend adjusting scores or recalculating norm means as generally accepted practice. Some specifically recommend following the test manual directions and give detailed instructions toward that end. Others simply advise that the norms should be updated periodically.

Ethical canons and related guidelines serve as a source of authority for practice standards. APA's "Ethical Principles of Psychologists and Code of Conduct" (APA, 2002) do not comment specifically on score adjustment apart from asserting that "psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques" (9.02a).

Standards for Educational and Psychological Testing (American Educational Research Association, APA, & National Council on Measurement in Education, 1999) provides criteria for testing practices and the effects of test use. Standards 5.1 and 5.2 require the test administrator to carefully follow the standardized procedures and score the measure according to the test manual without departing from the publisher's instructions. These standards make no reference to the FE, adjusting individual scores, or recalculating norm means separate and apart from the test manual.

Neither the "Specialty Guidelines for Forensic Psychologists" (Committee on Ethical Guidelines for Forensic Psychologists, 1991) nor the latest draft revisions for these guidelines (Committee on Ethical Guidelines for Forensic Psychologists, 2008) advocate diverting from test scoring manual instructions.

The APA has promulgated policy statements regarding psychological testing (APA, 1996; Joint Committee on Testing Practices, 1998), general service guidelines (Committee on Professional Practice and Standards, 2007; Committee on Professional Standards, 1987), practice area guidelines (APA, 1998, 2004; Committee on Professional Practice and Standards, 1998), and related qualification guidelines (APA, 2001). All are pertinent, in part or whole, to professional responsibility when using IQ tests for a wide range of purposes. All are silent with respect to the FE. None establish a standard for adjusting obtained scores or for departing from test manual instructions.

Statutory and Case Law Authority

Duvall and Morris (2006) surveyed the statutes relevant to death penalty evaluations in the United States. Of the 38 death penalty states, none has a statute that mandates adjusting IQ scores on the basis of the FE. Case law in Tennessee (*Howell v. Tennessee,* 2004) and Kentucky (*Bowling v. Kentucky,* 2005) specifies that adjusting obtained scores on the basis of the FE is not sufficiently scientific. The latter court rejected factoring in the impact of the FE, finding that "the scientific community does not agree on the cause of this phenomenon" (*Bowling v. Kentucky,* 2005, p. 37). In *Green v. Johnson* (2008), the Fourth Circuit Court of Appeals observed for both the FE and the standard error of measurement that "neither *Atkins* nor Virginia law appears to require expressly that these theories be accounted for in determining mental retardation status" (p. 8).

Although appellate case law calls for consideration of the FE when not procedurally barred, there is no judicial consensus that adjusting obtained scores or recalculating norm means is generally accepted in the field. Some appellate courts have ruled that a trial court must consider evidence of the FE and determine the persuasiveness of the evidence (*Walker v. True,* 2005). However, this survey found no instance in which an appellate court ruled that the FE is compelling or controlling as a matter of law.

Conclusions and Implications for Practice

Three conclusions emerge. First and foremost, adjusting obtained scores and recalculating norm means on the basis of the FE do not represent the convention and custom in psychology. Adjusting obtained IQ scores for this purpose is not the standard of practice. Second, recalculating an individual's actual data likely violates standardization procedures and departs from training practices, prevailing canons, guidelines, most treatises, and test instructional manuals. In addition, the prevailing consensus calls for publishers to update norms periodically. Third, when choosing IQ tests or reviewing archival test data, psychologists should carefully consider potential compromises to validity and the differential impact of such compromises in light of race, culture, age, gender, and the weighting of cognitive demands of the instrument. Commenting on these issues in the report narrative is appropriate, but adjusting the numerical scores is not. The practitioner should heed the practice standard to use the most current version of a test.

The current accepted convention does not support subtracting IQ points in a way that departs from the requirements of the test manual. "Evaluators must also be aware that there is no agreed-upon method for how diagnostic conclusions should be influenced by the Flynn effect" (Young et al., 2007, p. 176). Psychologists cannot conclude that adjusting scores is the generally accepted practice in evaluations for special education, parental rights termination, disability, or any other purpose.

An accurate score on an IQ test can make a meaningful difference, and the descriptive label the psychologist applies to it can also make a difference (Guilmette, Hagan, & Giuliano, 2008). Highly skilled and conscientiously committed psychologists may find that these critical medico-legal evaluations stir significant personal and ethical dilemmas. Those who thoughtfully reflect on the clinical and forensic issues as well as their qualifications and experience and elect to decline or accept these referrals are to be

commended for their professional posture. Those who decide to undertake these forensic evaluations should proceed cautiously and continuously educate themselves about developments in the law, ethics, practice standards, and science.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychological Association. (1996). Statement on the disclosure of test data. *American Psychologist, 51,* 644–648.

American Psychological Association. (1998). Guidelines for the evaluation of dementia and age-related cognitive decline. *American Psychologist, 53,* 1298–1303.

American Psychological Association. (2001). APA's guidelines for test user qualifications: An executive summary. *American Psychologist, 56,* 1099–1113.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060–1073.

American Psychological Association. (2004). Guidelines for psychological practice with older adults. *American Psychologist, 59,* 236–260.

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

Berry v. Mississippi, 544 U.S. 950 (2005) [Flynn affidavit].

Black, H. C. (2004). *Black's law dictionary* (8th ed.). St. Paul, MN: Thomson-West.

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

Committee on Ethical Guidelines for Forensic Psychologists. (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior, 15,* 655–665.

Committee on Ethical Guidelines for Forensic Psychologists. (2008). *Specialty guidelines for forensic psychology: Third official draft.* Retrieved March 19, 2008, from http://ap-ls.org/links/22808sgfp.pdf

Committee on Professional Practice and Standards. (1998). *Guidelines for psychological evaluations in child protection matters.* Washington, DC: American Psychological Association.

Committee on Professional Practice and Standards. (2007). *Record keeping guidelines.* Washington, DC: American Psychological Association.

Committee on Professional Standards. (1987). *General guidelines for providers of psychological services.* Washington, DC: American Psychological Association.

Commonwealth v. Prieto, Fairfax (VA) Cir. Ct., FE 2005-1764 (March 5, 2007).

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise—The Flynn effect in rural Kenyan children. *Psychological Science, 14,* 215–219.

Duvall, J. C., & Morris, R. J. (2006). Assessing mental retardation in death penalty cases: Critical issues for psychology and psychological practice. *Professional Psychology: Research and Practice, 37,* 658–665.

Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.). New York: Guilford.

Fla. Stat. (2005). § 921.137[1].

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 1972–1985: How to compensate for obsolete norms. *Perceptual and Motor Skills, 86,* 1231–1239.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54,* 5–20.

Flynn, J. R. (2000). The hidden history of IQ and special education: Can the problems be solved? *Psychology, Public Policy, and Law, 6,* 191–198.

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. (2007). *What is intelligence? Beyond the Flynn effect.* New York: Cambridge University Press.

Gleitman, H., Fridlund, A. J., & Reisberg, D. (2003). *Psychology* (6th ed). New York: Norton.

Greenspan, S. (2006, spring). Issues in the use of the "Flynn effect" to adjust IQ scores when diagnosing MR. *Psychology in Mental Retardation and Developmental Disabilities Newsletter, 31,* 3–7.

Green v. Johnson, WL 352028 (4th Cir. 2008).

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). New York: Wiley.

Guilmette, T., Hagan, L. D., & Giuliano, A. J. (2008). Assigning qualitative descriptions to test scores in neuropsychology: Forensic implications. *Clinical Neuropsychologist, 22,* 122–139.

Howell v. Tennessee, 151 S.W.3d 450 (Tenn. 2004).

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 *et seq.*

Jackson, D. N. (2003). *Multidimensional Aptitude Battery: Manual* (2nd ed.). Port Huron, MI: Sigma Assessment Systems.

Joint Committee on Testing Practices. (1998). *Rights and responsibilities of test takers: Guidelines and expectations.* Washington, DC: American Psychological Association.

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.

Kaplan, R., & Saccuzzo, D. (2005). *Psychological testing: Principles, applications, and issues* (6th ed.). Belmont, CA: Thomson-Wadsworth.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence Test: Manual.* Circle Pines, MN: American Guidance Services.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment.* New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). New York: Wiley.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment.* New York: Wiley.

Lacritz, L. H., & Cullum, C. M. (2003). The WAIS-III and WMS-III: Practical issues and frequently asked questions. In D. S. Tulsky, D. H. Saklofske, G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. Bornstein, et al. (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 491–532). Boston: Elsevier.

The latest thinking on intelligence: Interview with James Flynn. (2007, June). *The Psychologist, 20,* 356–357.

Lichtenberger, E. O., & Kaufman, A. S. (2004). *Essentials of WPPSI-III assessment.* New York: Wiley.

Mather, N., & Woodcock, R. W. (2001). *Woodcock–Johnson 3rd Ed.: Examiner's manual.* Itasca, IL: Riverside.

Moore, R. B. (2006, fall). Modification of individual's IQ scores is not accepted professional practice. *Psychology in Mental Retardation and Developmental Disabilities, 31,* 11–12.

Myers, D. G. (2007). *Psychology* (8th ed.). New York: Worth.

National Research Council. (2002). *Mental retardation: Determining eligibility for social security benefits.* Washington, DC: National Academy Press.

Olley, J. G., Greenspan, S., & Switzky, H. (2006, winter). Division 33 Ad Hoc Committee on Mental Retardation and the Death Penalty. *Psychology in Mental Retardation and Developmental Disabilities, 31,* 11–13.

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).

Prifitera, A., Saklofske, D. H., & Weiss, L. D. (2005). *WISC-IV clinical use and interpretation.* New York: Academic Press.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20,* 33–65.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test: Professional manual.* Lutz, FL: Psychological Assessment Resources.

Roid, G. H. (2003). *Stanford–Binet Intelligence Scales, 5th Edition: Technical manual.* Itasca, IL: Riverside.

Sattler, J. M., & Hoge, R. D. (2006). *Assessment of children: Behavioral, social and clinical foundations* (5th ed.). LaMesa, CA: Sattler.

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on—A large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. *British Journal of Educational Psychology, 77,* 25–41.

Social Security Administration. (2006). *Disability evaluation under Social Security: DI 22510.021. Consultative examination report content guidelines—Mental disorders.* Washington, DC: Author.

State v. Keel, 90 CRS 8033 (Sup. Ct. Edgecombe Co., N.C.) (unpub.), *cert. denied,* 357 N.C. 465, 586 S.E.2d 462 (N.C. 2003).

Tulsky, D. S., Saklofske, D. H., & Ricker, J. H. (2003). *Clinical interpretation of the WAIS-III and WMS-III: Practical resources for the mental health professional.* Boston: Elsevier.

Urbina, S. (2004). *Essentials of psychological testing.* New York: Wiley.

Virginia Department of Mental Health, Mental Retardation and Substance Abuse. (2005). *List of standardized measures of intellectual functioning.* Retrieved July 25, 2007, from http://www.dmhmrsas.virginia.gov/documents/ofo-standardizedmeasures.pdf

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

Walton v. Johnson, 440 F.3d 160 (4th Cir.) (*en banc*), *cert. denied.* 126 S.Ct. 2377 (2006).

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26,* 54–60.

Wechsler, D. (2002). *Technical manual (updated) for the Wechsler Adult Intelligence Scale, 3rd ed. and Wechsler Memory Scale, 3rd ed.* San Antonio: Psychological Corporation.

Weiss, L. G. (2007). *WAIS-III technical report: Response to Flynn.* Retrieved March 14, 2007, from http://harcourtassessment.com/NR/rdonlyres/98BBF5D2–F0E8–4DF6–87E2–51D0CD6EE98C/0/WAISIII_TR_lr.pdf

Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV: Advanced clinical interpretation.* New York: Academic Press.

Young, B., Boccaccini, M. T., Conroy, M. A., & Lawson, K. (2007). Four practical and conceptual assessment issues that evaluators should address in capital case mental retardation evaluations. *Professional Psychology: Research and Practice, 38,* 169–178.

Zhou, X., & Zhu, J. (2007, August). *Peeking inside the "blackbox" of Flynn effect: Evidence from three Wechsler instruments.* Poster session presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Received October 26, 2007
Revision received April 28, 2008
Accepted May 6, 2008 ■