The DEATH PENALTY and INTELLECTUAL DISABILITY

Edward A. Polloway, Editor

aaidd
American Association on Intellectual and Developmental Disabilities

In the 2002 landmark decision *Atkins v. Virginia*, the U.S. Supreme Court ruled that executing someone with intellectual disability (ID) is a violation of the Eighth Amendment of the U.S. Constitution, which prohibits cruel and unusual punishment. In its 2014 decision in *Hall v. Florida*, the Court ruled that, while states have the right to establish their own rules for handling *Atkins* cases, they are not free to ignore scientific and medical consensus regarding intelligence and the nature and diagnosis of ID. The Court rejected the use of an IQ test score of 70 as a bright-line cutoff for determining ID and ruled that all evidence pertinent to the claim, including adaptive behavior assessments, should be considered.

This AAIDD publication is the authoritative resource on the science that is the basis for the definition of intellectual disability and on the critical issues involved in its diagnosis. The book is a comprehensive and cogent analysis of what is involved in the determination of ID, particularly as it relates to defendants in death penalty cases. AAIDD assembled a group of notable scholars and clinicians to bring the best science to this discussion, and the work of this group is represented in the important resource published here.

aaidd
Knowledge | Support | Empowerment
American Association on Intellectual and Developmental Disabilities

39.95
ISBN 978-1-937604-13-4
53995>
9 781937 604134

10 Norm Obsolescence: The Flynn Effect

Kevin S. McGrew

Nature of the Problem

A person's IQ test score is based on the comparison of the person's tested performance to an age-appropriate norm reference group. The *norms* for an IQ test are developed to represent the snapshot of the general U.S. population (at each age level the test covers) at the time the norm or standardization data are collected (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, NCME], 1999). VandenBos (2007, defines a norm as "a standard or range of values that represents the typical performance of a group or of an individual [of a certain age, for example] against which comparisons can be made" [p. 631]). The person's test performance is compared to this standard reference group. For example, the WISC-R IQ test was published in 1974 and the WISC-R norm data was gathered on children ages 6 through 16 from 1971 through 1973 (Wechsler, 1974). (1972 is thus considered the official date of the WISC-R norm/standardization sample.) Thus, a child who is 7 years, 2 months old who was administered the WISC-R in 1974 would have the calculation of his or her IQ test score based on a comparison to the performance of children from ages 7 years, 0 months through 7 years, 3 months in the year 1972. (The WISC-R norm tables are provided in 3 month intervals within each year of age.) If the WISC-R was administered to a child of the same age (7 years, 2 months) in 1984, rather than being compared to other children of the same age in 1984, this child's performance would still be evaluated against similarly aged children from 1972. This second comparison results in a test-date/test-norm *mismatch* of 12 years (1984 – 1972 = 12). As explained next, comparing an individual's performance on an IQ test with outdated test norms results in *a comparison to a historical reference group from the past—not the person's contemporary peers.* This *norm obsolescence* problem is more commonly referred to as the *Flynn effect* (Flynn, 1984,

1985, 2000, 2006, 2007a, 2009). The Flynn effect produces inflated and inaccurate IQ test scores.

In simple terms, psychologists and psychological measurement experts typically describe the Flynn effect as the result of a "softening" of IQ tests norms with the passage of time. That is, individuals tested today on an IQ test normed many years earlier will obtain artificially inflated IQ test scores, because the older test norms reflect a level of overall performance that is lower than that of individuals in contemporary society. This is one of the primary reasons why authors and publishers of IQ tests make every effort to periodically provide "freshened" norms by collecting new nationally representative sample data for IQ test batteries. The professional consensus among test developers is that the "shelf life" of an IQ test's norms is approximately 10 years. According to Weiss (2010), Vice President of Pearson Clinical Assessment, the company and division that develops and publishes the various Wechsler IQ batteries, "there is no definition of when a test becomes obsolete. When asked privately, most Flynn effect researchers have 10 years in mind" (p. 492). If new norms are not provided, individuals tested using IQ tests with outdated norms will typically obtain inflated and inaccurate IQ test scores.

The Flynn effect recognizes that the normal curve distribution of intelligence shifts upward over time. Thus, the same raw score performance on an IQ test, when compared to outdated norms, will produce a markedly different IQ score when it is compared to updated norms based on a contemporary sample of abilities for a person of the same age. The person's tested performance (i.e., the number of correct responses across all parts of the IQ test) does *not* change, but the person's relative standing in the distribution of IQ scores across the population *does* change as a function of which norm reference group his or her performance is compared against. The same performance that is considered average in the contemporary norm sample, yielding an IQ test score of 100 in the distribution, will result in a higher IQ test score when using older norms (Schalock, 2012).

As a result of the Flynn effect, it is possible that one or more IQ test scores reported for an individual being considered for a diagnosis of intellectual disability (ID) may be inaccurate and inflated estimates. Given the high-stakes nature of *Atkins,* ID cases and their tendency to artificially focus on specific "bright line" cutoff scores, the impact of the Flynn effect must be recognized and an adjustment to the inflated scores is recommended.

Summary of Related Research

Origins of the Flynn Effect

Probably the first widely recognized scholarly report of IQ norm obsolescence was published by Lynn in 1983. Reflecting Lynn's early writings, some intelligence scholars refer to IQ norm obsolescence as the *Lynn-Flynn effect* (Woodley, 2012a). Recently, Lynn

(2013) provided evidence that 24 studies, the first being Runquist (1936), reported on the phenomenon of norm obsolescence before the "effect was rediscovered by Flynn" (1984). Lynn (2013) argued that the proper designation of IQ test norm obsolescence should be the "Runquist effect." Although Lynn (2013) provided a compelling argument (based on the customary practices in the history of science for naming phenomena), the term *Flynn effect* is used here given its prominent and frequent use in intelligence research and *Atkins* court cases.

Seventeen years prior to the 2002 *Atkins* decision, Flynn (1985) published an article in the *American Journal on Mental Deficiency* (now the *American Journal on Intellectual and Developmental Disabilities*). This article, titled "Wechsler Intelligence Tests: Do We Really Have Criterion of Mental Retardation?" first raised the issue of a possible "adjustment" in the context of an ID diagnosis. In hindsight, Flynn's 1985 article was the "canary in the coal mine" in that it first demonstrated that the Flynn effect may have a significant impact on the proportion of the population of individuals that would be identified as ID. At that time, Flynn proposed a form of adjusting for the softening of tests norms, although it was in a slightly different form than the current recommended Flynn effect adjustment procedure.

Flynn (1985) proposed that to account for the softening of test norms, an IQ test score of 70 on a "reference" IQ test (i.e., WAIS-R) would be set in as the *absolute criterion for mental retardation* (that is, on the intellectual functioning prong of the definition). Then, to account for norm obsolescence, each time a new IQ test was published there would be a lowering of the MR cutting line. Flynn's 1985 idea was that whenever a new IQ test was published, it would be given together with the established reference IQ test (e.g., WAIS-R) and the average mean IQ test score difference between the new test and the reference test would be used to "derive a new score equivalent to the old cutting line" (p. 243). Although different from what is now considered the standard Flynn effect adjustment approach (i.e., subtracting 3 IQ test score points from an individual's total IQ test score for every 10 years for which the test was administered to a person who was normed prior to the date of individual's testing), conceptually Flynn's 1985 proposal accomplished the same goal as the currently employed Flynn effect adjustment procedure.

Fifteen years later, and still 2 years prior to the *Atkins* decision, Flynn (2000) again sounded the alarm regarding the implication of norm obsolescence related to the diagnosis and classification of mental retardation:

> It is certain that over the past 50 years, literally millions of Americans evaded the label of mentally retarded designed for them by the test manuals. Whether this was good or bad depends on what one thinks of the label. Some will say millions avoided stigma. Others will say that millions missed out on needed assistance and classroom teachers were left unaided to cope with pupils for whom aid was needed (p. 197).

The potential impact of the Flynn effect on other diagnoses was also reported in 2001 and 2003. Truscott and colleagues (Sanborn, Truscott, Phelps, & McDougal, 2003; Truscott & Frank, 2001) reported on the impact of the Flynn effect on learning disability (LD) identification, not identification of individuals with ID. Although these authors did not offer or endorse any IQ test score adjustment procedure, these researchers concluded that

> A critical finding of this study is that the FE probably contributes to misdiagnosis of LD. If this research is combined with previous reports that academic achievement may be unaffected by the FE (Neisser, 1998) it strongly suggests that, over the life of a test version, IQ-achievement discrepancies, the most salient LD criterion, are exaggerated. One potential result of such an exaggeration of IQ-achievement discrepancies would be that, as test norms aged, fewer students would score in the mentally retarded range (Flynn, 2000) and more students would qualify for LD based on inflated severe discrepancies (p. 300).

In conclusion, the recognition of the impact of *norm obsolescence* (i.e., the Flynn effect) on IQ test scores, and more importantly, the potential for misdiagnosis of ID and other conditions (e.g., LD), has been recognized and documented as early as the 1980s. It continued to be discussed prior to and after the 2002 ID-related *Atkins* decision by researchers and professionals who did not anticipate nor were influenced by the 2002 *Atkins* decision. For obvious reasons (i.e., the life-or-death implications of the *Atkins* decision), there has been increased interest in the Flynn effect adjustment procedure since the *Atkins* decision. The facts indicate that the recognition of the impact of norm obsolescence on IQ test scores (and the idea of a norm obsolescence IQ test score adjustment) was established prior to the *Atkins v Virginia* (2002) U.S. Supreme Court decision.

Scientific Basis of the Flynn Effect

There is a scientific and professional consensus that the Flynn effect is a scientific fact. A complete reading of the extant Flynn effect research literature leads to the conclusion that, despite debates regarding the causes of the Flynn effect, differences in the rate of Flynn effect change in different countries. Whether the Flynn effect has started to plateau in Scandinavian countries or whether the Flynn effect differs by different levels of intelligence and different methodological issues in various studies, *the consensus of the relevant scientific community is that the Flynn effect is real* (Cunningham & Tassé, 2010; Fletcher, Stuebing & Hughes, 2010; Flynn, 2009; Greenspan, 2006, 2007; Gresham & Reschly, 2011; Kaufman, 2010a, 2010b; McGrew, 2010; Rodgers, 1999; Trahan, Stuebing, Fletcher, & Hiscock 2014; Weiss, 2010; Zhou, Zhu, & Weiss, 2010). The robustness of this conclusion may best be represented by Rogers' (1999) statement where, after raising valid methodological issues regarding various statistical analysis and conclusions across Flynn effect studies, that even with a "healthy dose of

skepticism, the effect rises above purely methodological interpretation, and appears to have substantive import" (p. 354).

The research literature regarding the Flynn effect is extensive. Trahan et al. (2014) found over 4,000 articles in their comprehensive literature review. (Most all norm obsolescence references and articles can be found at the regularly updated *Flynn Effect Archive Project* [http://www.atkinsmrdeathpenalty.com/2010/01/atkins-mrid-capital-punishment-flynn_11.html]. As of 2014, this archive includes approximately 190 publications.) A thorough treatment of all this research is beyond the scope of the current chapter. Fortunately, key contemporary Flynn effect issues bearing on an ID diagnosis in the *Atkins* context were covered in a special 2010 issue of the *Journal of Psychoeducational Assessment* (*JPA*). A variety of invited scholars confirmed the scientific consensus regarding the validity of the Flynn effect. For example, Dr. Alan Kaufman (2010a), arguably the most prominent scholar on intelligence testing and interpretation of the various Wechsler IQ tests, stated that

> The Flynn effect (FE) is well known: Children and adults score higher on IQ tests now than they did in previous generations (Flynn, 1984, 2007, 2009b). The rate of increase in the United States has apparently remained a fairly constant 3 points per decade since the 1930s (p. 382).

The consensus of almost all authors who contributed to the *JPA* Flynn effect issue (Fletcher et al., 2010; Flynn, 2010; Hagan, Drogin, & Guilmette, 2010a; Kaufman, 2010a, 2010b; Kaufman & Weiss, 2010; McGrew, 2010; Reynolds, Niland, Wright, & Rosenn, 2010; Sternberg, 2010; Weiss, 2010; Zhou et al. 2010) was that IQ test norm obsolescence (i.e., the Flynn effect) is an established scientific fact. The following select quotes from recent peer-reviewed articles capture the essence of the convergence of opinion regarding the validity of the Flynn effect.

> The Flynn effect (FE) is real. The FE has been shown to be nearly 3 points per decade on average across a large number of studies, countries, and tests (Weiss, 2010, p. 491).

> The point is that a person tested on an outdated test will earn spuriously high scores as each year goes by, and the amount of the spuriousness amounts to about 3 points per decade for Americans (Kaufman, 2010b, p. 503).

> 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 (Reynolds et al., 2010, p. 478).

> The Flynn effect is a well-established psychometric fact documenting substantial increases in measured intelligence test performance over time (Gresham & Reschly, 2011, p. 131).

Since the publication of the 2010 special *JPA* Flynn effect issue, many additional Flynn effect research and commentary articles have appeared (e.g., Battarjee, Khaleefa, Ali, & Lynn, 2013; Baxendale, 2010; Cunningham & Tassé, 2010; Hagan, Drogin, & Guilmette, 2010b; Kanaya & Ceci, 2011, 2012; Lynn, 2013; Nijenhuis, 2013; Nijenhuis, Cho, Murphy, & Lee, 2012; Nijenhuis, Murphy, & van Eeden, 2011; Nijenhuis & van der Flier, 2013; Pietschnig, Voracek, & Formann, 2011; Nijman, Scheirs, Prinsen, Abbink, & Blok, 2010; Rindermann, Schott, & Baumeister, 2013; Rönnlund, Carlstedt, Blomstedt, Nilsson, & Weinehall, 2013; Skirbekk, Stonawski, Bonsang, & Staudinger, 2013; Trahan et al., 2014; Wai & Putallaz, 2011; Woodley, 2011, 2012a, 2012b; Young, 2012). The continued flow of the Flynn effect related to peer-reviewed articles confirms the consensus that the Flynn effect is a scientifically important and studied phenomenon among intelligence scholars.

Adjusting IQ Test Scores for the Flynn Effect in Atkins *Cases Is Best Practice*

Not only is there a scientific consensus that the Flynn effect is a valid and real phenomenon, there is also a consensus that individually obtained IQ test scores derived from tests with outdated norms must be adjusted to account for the Flynn effect, particularly in *Atkins* cases. (The use of a Flynn effect correction in clinical settings is less of an issue given that psychologists in such settings typically have more leeway to interpret scores as ranges, invoke clinical judgment, and incorporate information regarding measurement error in interpretation of the scores when making a diagnosis. In contrast, certain high stakes settings [e.g., *Atkins* cases; eligibility for Social Security Disability benefits] may have strict point-specific cut-scores [i.e., "bright line" criteria] where examiners, or the recipients of the scores [e.g., the courts], do not allow for such clinical interpretation. Thus, the Flynn effect adjustment is more relevant, appropriate, and primarily discussed in literature and law dealing with this type of high stakes IQ testing.) The most prominent and relevant professional consensus-based guidelines for ID diagnosis (Schalock et al., 2007, 2010, and 2012) support a Flynn effect adjustment for scores based on obsolete IQ test norms. *Intellectual Disability: Definition, Classification, and Systems of Supports* (11th ed.; Schalock et al., 2010), based on an expert-consensus process, provides a written guideline that endorses the appropriateness of the Flynn effect adjustment in the diagnosis of ID. (The 11th edition was created using a group-based consensus process conducted by the AAIDD Ad Hoc Committee on Terminology and Classification [Schalock et al., 2010]). AAIDD recommends that psychologists use the most recent versions of IQ tests and, if scores are reported from an IQ test with outdated norms, a correction for the age of norms is warranted (Schalock et al., 2007). The 11th edition states

> As discussed in the *User's Guide* (Schalock et al., 2007) that accompanies the 10th edition of this *Manual,* best practices require recognition of a potential Flynn effect when older editions of an intelligence test (with corresponding older norms) are used in the assessment or interpretation of an IQ score. (p. 37)

As suggested in the *User's Guide to Mental Retardation: Definition, Classification, and Systems of Supports* (Schalock, 2007, pp. 20–21),

> The main recommendation resulting from this work [regarding the Flynn effect] is that all intellectual assessment must use a reliable and appropriate individually administered intelligence test. In cases of tests with multiple versions, the most recent version with the most current norms should be used at all times. In cases where a test with aging norms is used, a correction for the age of the norms is warranted. (p. 37)

The AAIDD's more recent *User's Guide to Intellectual Disability: Definition, Classification, and Systems of Supports* (Schalock et al., 2012) states

> The *Flynn effect* refers to the increase in IQ scores over time (i.e., about 0.30 points per year). The Flynn effect affects any interpretation of IQ scores based on outdated norms. Both the 11th edition of the manual and this *User's Guide* recommend that in cases in which a test with aging norms is used as part of a diagnosis of ID, a corrected Full Scale IQ upward of 3 points per decade for age of norms is warranted. (p. 23)

A consensus among the professional and scientific community of intelligence and ID scholars has emerged. This consensus is that given the high-stakes nature of *Atkins* ID cases and their tendency to artificially focus on specific "bright line" cutoff scores, *a Flynn effect correction to a person's scores in this setting is now considered best or standard practice.* This conclusion is supported by a significant number of scholars and researchers in the areas of intelligence and ID (Cunningham & Tassé, 2010; Fletcher et al., 2010; Flynn, 2006, 2007b; Flynn & Widaman, 2008; Greenspan, 2006, 2007; Gresham & Reschly, 2011; Kaufman, 2010b; McVaugh & Cunningham, 2009; Reynolds et al., 2010; Schalock , 2007; Schalock, 2012). One example of this support is the statement of Reynolds et al. (2010) that "as a generally accepted scientific theory that could potentially make the difference between a constitutional and unconstitutional execution, the Flynn effect must be applied in the legal context" (p. 480). Reynolds et al. (2010) go as far as to state that "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" (p. 480).

A minority of scholars have offered a different approach to the issue of correcting IQ test scores due to the Flynn effect. Weiss (2010), while acknowledging the scientific validity of the Flynn effect, advocates that experts should simply inform the fact finder of what the research shows and the trier-of-fact should evaluate and decide if and how to apply it when interpreting individual scores. Hagan et al. (2010b) also agree with the need to consider the Flynn effect in capital cases but their disagreement "lies in how psychologists should convey IQ scores in light of the observation that mean scores drift over time" (p. 420). It is important to note that the more conservative positions of Weiss (2010) and Hagan et al. (2010a, 2010b) represent a minority position in the professional literature. More importantly, they do not argue against the scientific validity of the Flynn

effect or even the need to consider the effect in *Atkins* cases. Rather, their difference of opinion with the majority is only as to whether a specified score adjustment should be made to the original score or whether testifying experts should instead address the Flynn effect in narrative form.

Recently, legal scholars have also supported the application of the Flynn effect correction in *Atkins* cases. Young's (2012) recent law review article ("A More Intelligent and Just *Atkins:* Adjusting for the Flynn Effect in Capital Determinations of Mental Retardation or Intellectual Disability") concluded that

> adjusting for the Flynn effect reflects a practice consistent with both *Atkins* and the known world of IQ measurements. While a freakish strike of lightning is difficult to avoid, the potentially deadly and unconstitutional consequences of refusing to account for the Flynn effect are wholly preventable. Thus, for the intelligent and just enforcement of *Atkins*, courts and juries should adjust IQ score from outdated tests for the Flynn effect. (p. 663)

What Is the Correct Flynn Effect Adjustment for Norm Obsolescence?

The AAIDDs' *User's Guide* (Schalock, 2012) recommends a Flynn effect correction of 3 points per decade (0.3 points per year). The 3 points per decade rule-of-thumb is consistent with the previously cited comments of Kaufman (2010a, 2010b) and Weiss (2010), and is also consistent with the recommendation of most scholars in the areas of intelligence and ID (e.g., Fletcher et al., 2010; Gresham & Reschly, 2011; Trahan, et al., 2014; Widaman, 2007).

The 3 points per decade rule-of-thumb is based primarily on Flynn's (2009) seminal article where he synthesized the results of 14 estimates of IQ test score gains over time. Flynn reported an average IQ test score change, across the 14 studies, of 0.311 points per year. An average mean score of 0.299 points was reported for the Wechsler comparisons only. Flynn concluded that "the evidence suggests that a rate of 0.30 is about right, and varying it from case to case lacks any rationale" (p. 104).

More recently Fletcher et al. (2010) applied more precise quantitative meta-analytic procedures to Flynn's (2009) data and reported a weighted mean of 2.80 points per decade. After removing two outlier studies, the weighted mean per decade was 2.96. These researchers concluded that "the level of precision we reported of a mean of about 3 and a *standard error of the mean* (SEM) of about 1 supports the correction and is consistent with the Flynn correction of 3 points per decade" (p. 472). In the most comprehensive meta-analysis research synthesis of 285 studies, Trahan et al. (2014) found that for modern intelligence tests the Flynn effect size was a similar 2.93 points per decade. These researchers concluded that their "findings are consistent with previous research and with the argument that it is feasible and advisable to correct IQ scores for the Flynn effect in high-stakes decisions" (p. 22).

The best available research syntheses consistently converge on a Flynn effect rule-of-thumb of 3 IQ test score points per decade (of IQ test norm obsolescence). Although

scientific journals may report Flynn effect results to the second decimal place (e.g., 3.11 per decade or 0.311 per year), the psychometrics of IQ testing and research cannot partition human behavior with such precision. As noted by Widaman (2007), much of the variation between scores from different Flynn effect studies is due to sampling and measurement error. Using Flynn effect adjustment formulae that use numbers to the second decimal place would be akin to slicing butter with a laser beam. Consequently, the current best estimate of IQ norm obsolescence, and the recommended Flynn effect adjustment, is 3 IQ points per decade, or 0.3 points per year.

Researching the Flynn Effect "Black Box": Implications for Practice

Recently a significant portion of Flynn effect research has shifted from a focus on the secular changes in the global IQ test scores over time to changes on more specific intellectual abilities, possible differential effects by level of intelligence, and a search for the cause of the Flynn effect (Kaufman, 2010a). Zhou et al. (2010) characterized this shift to a focus on the "black box" of the Flynn effect.

The cause of the Flynn effect. In the context of the special articles in the 2010 *JPA* Flynn effect issue, Weiss (2010) stated that "Except for Flynn, there is general agreement . . . that we know precious little about the causes of the effect" (p. 487). Explanations and theories have touched on such causative variables as genetics, environmental factors (e.g., nutrition, education, improved public health, increased use of computer games), ethnicity, and different societal risks and benefits associated with different generations (Kaufman & Weiss, 2010; Weiss, 2010). Flynn (2007a), in his book *What Is Intelligence? Beyond the Flynn Effect*, suggests that the effect that bears his name is due to systematic shift in societies from concrete to abstract scientific thinking. Confounding the search for the cause(s) of the Flynn effect has been idiosyncratic and armchair-based speculations (Weiss, 2010).

In the current context, knowing that the Flynn effect exists trumps a lack of consensus regarding causation. The impact of norm obsolescence on IQ test scores is real and the professional consensus is that it should be accounted for in *Atkins* ID determination. Understanding the "why" of the Flynn effect is beyond the scope of the current chapter and is not necessary for recognizing the scientifically and professionally based consensus that IQ test scores suffering from norm obsolescence need to be adjusted in *Atkins* cases. As stated by Kaufman (2010b), "The Flynn effect is a fact, even if its cause is elusive, and it must be considered carefully when making high stakes decisions such as the death penalty" (p. 503).

Differential Flynn effects by specific intellectual abilities. The foundation of Flynn's (2007a) theoretical explanation of the Flynn effect is based primarily on the interpretation of differential rates of score changes as a function of different specific intellectual abilities (e.g., smaller gains on verbal and crystallized ability tasks and larger changes on visual-spatial and abstract fluid reasoning tasks—not a singular focus on the global IQ test score). If differential specific ability Flynn effects are eventually found to be valid, the potential implication is that different Flynn effect adjustments

may be recommended for different composite or cluster "part" scores in IQ tests, and not just the global IQ score. This would introduce a new layer of complexity in the interpretation of IQ test scores (and part scores) in *Atkins* cases.

Although the recent methodologically sophisticated attempt by Zhou et al. (2010) to examine differential ability Flynn effects within the Wechsler tests represents an important step forward in this area of inquiry, their research produced inconsistent and contradictory findings. Although differential specific ability Flynn effect findings may eventually be identified, currently the supporting research results are sparse, mixed in results, and suffer from significant measurement and methodological flaws (McGrew, 2010). The foundation of Flynn effect causal theory, which hinges on the presence of differential specific ability Flynn effects, has been questioned on logical, theoretical, measurement and methodological grounds (Kaufman, 2010a, 2010b; McGrew, 2010; Weiss, 2010). Currently the extant research is not mature enough to support differential specific-ability Flynn effect adjustments in clinical or forensic contexts.

Differential Flynn effects by level of intelligence. The use of the 3 IQ test score points per decade Flynn effect adjustment rule-of-thumb has been questioned by research suggesting that the Flynn effect may not be uniform across all levels of general intelligence (Kanaya & Ceci, 2007; Kanaya, Ceci, & Scullin, 2003; Sanborn et al. 2003; Zhou et al., 2010). More important has been the suggestion that the Flynn effect may be larger at the IQ score range at the threshold for ID diagnosis. Cunningham and Tassé (2010) have referred to this research as the investigation of the Flynn effect in the "zone of ambiguity" (IQ test scores from 71–80). Studies reviewed by Cunningham and Tassé (2010) report IQ per decade changes ranging from roughly 4 to 5 points in the zone of ambiguity. Zhou et al. (2010) also reported differential Flynn effects by level of intelligence, but the results were inconsistent in the directions of the variation and may differ for different tests or age groups.

Similar to the differential Flynn effect by specific ability research, the ability-specific research has not been fully vetted through a sufficiently large number of studies and has been questioned on methodological grounds (McGrew, 2010; Widaman, 2007; Zhou et al., 2010). As summarized by Weiss (2010), "a small number of studies have suggested differential Flynn effect by ability level, but not enough is known about this at present" (p. 492). Reynolds et al. (2010) reinforce this conclusion, when after commenting on the Zhou et al. (2010 differential Flynn effects by levels of intelligence findings, that the results were inconsistent and "for now, best practice is the application of the Flynn correction as a constant by year across the distribution" (p. 480). Until more studies replicate the possibility of larger Flynn effects near the ID diagnostic threshold, the 3 points per decade Flynn effect rule-of-thumb should be employed across all levels of general intelligence.

Implications for Practice

The following implications are based on the integration of the content of the current chapter as well as the recommendations from the *User's Guide to the 10th edition,* the *11th edition,* and the *User's Guide to the 11th edition* (Schalock et al., 2007, 2010, 2012):

First, the potential problem of norm obsolescence can be minimized, but not always eliminated, by assessment professionals using IQ tests with the most up-to-date norms. When a new version of an IQ battery is published (e.g., WAIS-IV replaces WAS-III), assessment professionals should use the newest version (WAIS-IV) in *Atkins* cases. Assessment professionals have an ethical responsibility to stay abreast with the publication of new versions of IQ batteries and when the option exists to select among different IQ tests to administer to an individual. The relative degree of norm obsolescence of each possible IQ test should be one important factor incorporated into the IQ test selection decision.

Second, in cases where current or historical IQ test scores are impacted by norm obsolescence (i.e., Flynn effect), and the scores are to be used as part of the diagnosis of ID in *Atkins* or other high stakes decisions, the global scores impacted by outdated norms should be adjusted downward by 3 points per decade (0.3 points per year) of norm obsolescence.

Third, the recommended formula for the Flynn effect adjustment is: *FE adjustment = (Date test administered – date test was normed) × 0.3.* Stated simply, subtract the date the IQ test was normed (see point seven below) from the date the test was administered to the individual, multiply the obtained difference by 0.3. The obtained Flynn effect adjustment value should then be subtracted from the inflated obtained IQ score. The final Flynn effect adjustment value should be an integer value. Thus, the treatment of decimals in the final value should adhere to standard mathematical rules of "rounding to the nearest integer." The rationale for the particular rounding strategy employed should be described in the report. Current research does not support the application of different Flynn effect adjustment values for different part scores on IQ tests or at different levels of general intelligence. The best scientific evidence and professional consensus is that until sufficient research evidence produces evidence to the contrary, the 3 points per decade (0.3 points per year) adjustment rule-of-thumb should be used only on the global IQ test score and should be employed uniformly across all levels of general intelligence.

Fourth, both the original obtained (unadjusted) and Flynn effect adjusted scores should be included in all reports or court related statements or declarations provided by assessment professionals.

Fifth, the rationale for employing a Flynn effect correction should be described with supporting references. This chapter is intended to serve this function and can be cited as an authoritative source for the use of the Flynn effect adjustment in reports.

Sixth, when writing and discussing the Flynn effect, such as in psychological reports, legal declarations, or expert testimony, professionals should make frequent use of the term *norm obsolescence* when explaining the Flynn effect. Norm obsolescence is a much more descriptive and understandable means for conveying the essence of the Flynn effect.

Seventh, the calculation of the years of norm obsolescence should be based on the difference between the year the test was administered to an individual and the best

estimate of the year the IQ test was *normed* (see also Chapters 7 and 8). The data of publication of an IQ test does not accurately capture the time period when the test norm data were gathered. For example, the WISC-R IQ test was published in 1974 and the WISC-R norm data was gathered on children from 6 through 16 years of age from 1971 through 1973 (Wechsler, 1974). Thus, the middle most year of the actual norm data collection period is 1972. For the WISC-R, the year 1972 should be subtracted from the date of testing to determine the number of years of norm obsolescence. The test norm years reported for the different IQ tests by Flynn (2009) are recommended for uniformity purposes. For tests not reported in Flynn (2009), professionals need to consult the technical manuals for the IQ test in question and establish the best year estimate that is at the middle of the norm data collection period. If not readily available, professionals should seek the expertise of the test authors, publisher, or other intelligence test experts who may possess this information.

This chapter concludes with an example from an *Atkins* case. In 1998 an individual was administered the WAIS-R and obtained a Full Scale IQ of 80. Despite knowing that the WAIS-R had been revised and published as the WAIS-III in 1997, the psychologist administered the WAIS-R despite 20 years of norm obsolescence. The WAIS-R was published in 1981 and the best estimate of the date the actual test norms were gathered, as per the recommended procedures above, is 1978. Thus, the difference between the date of WAIS-R testing (1998) and date of test norming (1978) was 20 years, Using the 0.3/year Flynn effect adjustment, the best estimate of the magnitude of IQ test score inflation due to norm obsolescence is 6 IQ test score points ($0.3 \times 20 = 6.0$). Thus, this individual's Flynn effect adjusted WAIS-R score is 74 ($80 - 6 = 74$). This example represents one of the most dramatic instances of norm obsolescence (20 years) and also reflects the fact that the examiner did not engage in proper practice by administering the WAIS-III which was available at the time the individual was assessed.

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.

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

Batterjee, A. A., Khaleefa, O., Ali, K., & Lynn, R. (2013). An increase in intelligence in Saudi Arabia, 1977–2010. *Intelligence, 41*(2), 91–93. doi: 10.1016/j.intell.2012.10.011

Baxendale, S. (2010). The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology, 32*(7), 699–703. doi: 10.1080/13803390903493515

Cunningham, M. D., & Tasśe, M. J. (2010). Looking to science rather than convention in adjusting IQ scores when death is at issue. *Professional Psychology: Research and Practice, 45*(5), 413–419. doi: 10.1037/a0020226

Fletcher, J., Stuebing, K., & Hughes, L. (2010). IQ scores should be corrected for the Flynn effect in high stakes decisions. *Journal of Psychoeducational Assessment, 28*(5), 469–473.

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*(3), 236–244.

Flynn, J. R. (2000). The hidden history of IQ and special education: Can the problems be solved? *Psychology Public Policy and Law, 6*(1), 191–198.

Flynn, J. R. (2006). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law, 12*, 170–189. doi:10.1037/1076-8971.12.2.170

Flynn, J. R. (2007a). *What is intelligence? Beyond the Flynn effect*. New York: 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), 3–7.

Flynn, J. R. (2009). The WAIS-III and WAIS-IV: *Daubert* motions favor the certainly false over the approximately true. *Applied Neuropsychology, 16*, 98–104. doi: 10.1080/09084280902864360

Flynn, J. R., & Widaman, K. F. (2008). The Flynn effect and the shadow of the past: Mental retardation and the indefensible and indispensible role of IQ. In L. M. Glidden (Ed.), *International review of mental retardation* (Vol. 35, pp. 121–149). Boston, MA: Elsevier.

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

Greenspan, S. (2007). Flynn-adjustment is a matter of basic fairness: Response to Roger B. Moore, Jr. *Psychology in Mental Retardation and Developmental Disabilities, 32*(3), 7–8.

Gresham, F., & Reschly, D. J. (2011). Standard of practice and Flynn effect testimony in death penalty cases. *Intellectual and Developmental Disabilities, 49*(3), 131–140. doi: 10.1352/1934-9556-49.3.131

Hagan, L. D., Drogin, E. Y., & Guilmette, T. J. (2010a). IQ Scores should not be adjusted for the Flynn effect in capital punishment cases. *Journal of Psychoeducational Assessment, 28*(5), 474–476. doi: 10.1177/0734282910373343

Hagan, L. D., Drogin, E. Y., & Guilmette, T. J. (2010b). Science rather than advocacy when reporting IQ scores. *Professional Psychology Research and Practice, 41*(5), 420–423.

Kanaya, T., & Ceci, S. J (2007). Mental retardation diagnosis and the Flynn effect: General intelligence, adaptive behavior, and context. *Child Development Perspectives, 1*(1), 62–63. doi: 10.1111/j.1750-8606.2007.00013.x

Kanaya, T., & Ceci, S. J (2011). The Flynn effect in the WISC subtests among school children tested for special education services. *Journal of Psychoeducational Assessment, 29*(2), 125–136. doi:10.1177/0734282910370139

Kanaya, T., & Ceci, S. (2012). The impact of the Flynn effect on LD diagnoses in special education. *Journal of Learning Disabilities, 45*(4), 319–326. doi: 10.1177/0022219410392044

Kanaya, T., Ceci, S. J., & Scullin, M. H. (2003). The rise and fall of IQ in special ed: Historical trends and their implications. *Journal of School Psychology, 41*(6), 453–465. doi:10.1016/j.jsp.2003.08.003

Kaufman, A. (2010a). "In what way are apples and oranges alike?": A Critique of Flynn's interpretation of the Flynn effect. *Journal of Psychoeducational Assessment, 28*(5), 382–398. doi: 10.1177/0734282910373346

Kaufman, A. (2010b). Looking through Flynn's rose-coloured scientific spectacles. *Journal of Psychoeducational Assessment, 28*(5), 494–505. doi: 10.1177/0734282910373573

Kaufman, K., & Weiss, L. (2010). Guest editor's Introduction to the special issue of *JPA* on the Flynn effect. *Journal of Psychoeducational Assessment, 28*(5), 379–381. doi:10.1177/0734282910373344

Lynn, R. (1983). IQ in Japan and the United States shows a growing disparity. *Nature, 306*, 291–292.

Lynn, R. (2013). Who discovered the Flynn effect? A review of early studies of the secular increase in intelligence. *Intelligence,41*(6), 765–769. doi: 10.1016/j.intell.2013.03.004

McGrew, K. (2010). The Flynn effect and its critics: Rusty linchpins and "lookin' for *g* and Gf in some of the wrong places." *Journal of Psychoeducational Assessment, 28*(5), 448–468. doi:10.1177/0734282910373347

McVaugh, G. S., & Cunningham, M. D. (2009). Atkins v. Virginia: Implications and recommendations for forensic practice. *The Journal of Psychiatry and Law, 37*, 131–187.

Nijenhuis, J. T. (2013). The Flynn effect, group differences, and *g* loadings. *Personality and Individual Differences*, 55, 224–228. doi:10.1016/j.paid.2011.12.023

Nijenhuis, J. T., Cho, S. H., Murphy, R., & Lee., K. H. (2012). The Flynn effect in Korea: Large gains. *Personality and Individual Differences, 53*(2), 147–151. doi: 10.1016/j.paid.2011.03.022

Nijenhuis, J. T., Murphy, R., & van Eeden, R. (2011). The Flynn effect in South Africa. *Intelligence, 39*(6), 456–467.

Nijenhuis, J. T., & van der Flier, H. (2013). Is the Flynn effect on g? A meta-analysis. *Intelligence*, 41, 802–807.

Nijman, E. E., Scheirs, J. G., Prinsen, M. J., Abbink, C. D., & Blok, J. B. (2010). Exploring the Flynn effect in mentally retarded adults using a nonverbal intelligence test for children. *Research in Developmental Disabilities, 31*, 1404–1411. doi: 10.1016/j.ridd.2010.06.018

Reynolds, C., Niland, J., Wright, J., & Rosenn, M. (2010). Failure to apply the Flynn correction in death penalty litigation: Standard practice of today maybe, but certainly malpractice of tomorrow. *Journal of Psychoeducational Assessment, 28*(5), 477–481.

Rindermann, H., Schott, T., & Baumeister, A, (2013). Flynn effect in Turkey: A comment on Kagitcibasi and Biricik (2011). *Intelligence, 41*, 178–180. doi: 10.1016/j.intell.2013.02.003

Rodgers, J. L. (1999). A critique of the Flynn effect: Massive IQ gains, methodological artifacts, or both? *Intelligence, 26*(4), 337–356.

Rönnlund, M., Carlstedt, B., Blomstedt, Y., Nilsson, L. G., & Weinehall, L. (2013). Secular trends in cognitive test performance: Swedish conscript data 1970–1993. *Intelligence, 41*(1), 19–24.

Runquist, E. A. (1936). Intelligence test scores and school marks in 1928 and 1933. *School and Society, 43*, 301–304.

Sanborn, K. J., Truscott, S. D., Phelps, L., & McDougal, J. L. (2003). Does the Flynn effect differ by IQ level in samples of students classified as learning disabled? *Journal of Psychoeducational Assessment, 21*(2), 145–159.

Schalock, R. L., Buntinx, W. H. E., Borthwick-Duffy, S. A., Luckasson, R., Snell, M. E., Tassé, M. J., & Wehmeyer, M. L. (2007). *User's guide to mental retardation: Definition, classification, and systems of supports (10th ed.).* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Gomez, S. C., Lachapelle, Y., Luckasson, R., Reeve, A., Shogren, K. A., Snell, M. E., Spreat, S., Tassé, M. J., Thompson, J. R., Verdugo-Alonso, M. A., Wehmeyer, M. L., & Yeager, M. H. (2010). *Intellectual disability: Definition, classification, and systems of supports (11th ed.).* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Luckasson, R., Bradley, V., Buntinx, W. H. E., Lachapelle, Y., Shogren, K. A., Snell, M. E., Thompson, J. R., Tassé, M. J., Verdugo-Alonso, M. A., and Wehmeyer, M. L. (2012). *User's guide to intellectual disability: Definition, classification, and systems of supports.* Washington, DC: American Association on Intellectual and Developmental Disabilities.

Skirbekk, V., Stonawski, Bonsang, E., & Staudinger, U. M. (2013). The Flynn effect and population aging, *Intelligence, 41*(3), 169–177. doi: 10.1016/j.intell.2013.02.001

Sternberg, R. (2010). The Flynn effect: So what? *Journal of Psychoeducational Assessment, 28*(5), 434–440. doi: 10.1177/0734282910373349

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014, June 30). The Flynn effect: A meta-analysis. *Psychological Bulletin*. Advance online publication. http://dx.doi.org/10.1037/a0037173

Truscott, S. D., & Frank, A. J. (2001). Does the Flynn effect affect IQ scores of students classified as LD? *Journal of School Psychology, 39*(4), 319–334.

Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence, 39*(6), 443–455. doi:10.1016/j.intell.2011.07.006

Wechsler, D. (1974). *The Wechsler Intelligence Scale for Children—Revised (WISC-R).* San Antonio,TX: Psychological Corporation.

Weiss, L. G. (2010). Considerations on the Flynn effect. *Journal of Psychoeducational Assessment, 28*(5), 482–493. doi: 10.1177/0734282910373572

Widaman, K. (2007). Stalking the roving IQ score cutoff: A commentary on Kanaya and Ceci (2007). *Child Development Perspectives,1(1),*57–59. doi: 10.1111/j.1750-8606.2007.00011..x

Woodley, M. A. (2011). Heterosis doesn't cause the Flynn effect: A critical examination of Mingroni (2007). *Psychological Review, 118*(4), 689–693. doi: 10.1037/a0024759

Woodley, M. A. (2012a). A life history model of the Lynn-Flynn effect. *Personality and Individual Differences, 53*(2), 152–156. doi: 10.1016/j.paid.2011.03.028

Woodley, M. A. (2012b). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence, 40*(2), 189–204. doi:10.1016/j.intell.2011.12.002

Young, G. W. (2012). A more intelligent and just *Atkins:* Adjusting for the Flynn effect in capital determinations of mental retardation or intellectual disability. *Vanderbilt Law Review*, 615–755.

Zhou, X., Zhu, J., & Weiss, L. (2010). Peeking inside the "blackbox" of the Flynn effect: Evidence from three Wechsler instruments. *Journal of Psychoeducational Assessment, 28*(5), 399–411.

Vandenbos, G. (2007). *APA dictionary of psychology.* Washington, DC: American Psychological Association.