JOURNAL OF CLINICAL AND EXPERIMENTAL NEUROPSYCHOLOGY
2007, 29 (5), 514–529

Psychology Press
Taylor & Francis Group

The Flynn effect and its relevance to neuropsychology

Merrill Hiscock

University of Houston, Houston, TX, USA

Evidence from several nations indicates that performance on mental ability tests is rising from one generation to the next, and that this "Flynn effect" has been operative for more than a century. No satisfactory explanation has been found. Nevertheless, the phenomenon has important implications for clinical utilization of IQ tests. This article summarizes the empirical basis of the Flynn effect, arguments about the nature of the skill that is increasing, and proposed explanations for the cause of the increase. Ramifications for clinical neuropsychology are discussed, and some of the broader implications for psychology and society are noted.

A sustained upward drift in mean cognitive ability now has been documented in several countries and is thought to have been occurring since the advent of the industrial revolution (Neisser, 1998b). This "rising curve" is often referred to as the Flynn effect, after the New Zealand political scientist James R. Flynn who discovered and characterized the phenomenon (e.g., Flynn, 1984, 1987, 1998a, 1999, 2006a). Implications of this phenomenon are numerous and profound, ranging from fundamental questions about the respective role of genes and environments in determining a person's intellectual capability to practical questions about the definition of mental retardation. The Flynn effect has the potential to change many aspects of neuropsychological theory and practice. This paper discusses implications of the Flynn effect for understanding and measuring cognitive ability before addressing some specific implications of the Flynn effect for clinical neuropsychology.

IQ tests in neuropsychology

IQ tests play a prominent role in clinical neuropsychology (Kaplan, Fein, Morris, & Delis, 1991). Diaz-Asper, Schretlen, and Pearlson (2004), for example, regard intelligence testing as "a cornerstone of neuropsychological assessment" (p. 90). Neuropsychologists' reliance on IQ tests is paradoxical, as the tests were not developed for the purpose of identifying deficits produced by brain damage, nor are they especially well suited to that objective (Boake, 2002; Lezak, 1988; Lezak, Howieson, & Loring, 2004). The substantial correlation between IQ and neuropsychological test performance in normal populations complicates the interpretation of low scores on neuropsychological tests in clinical settings (Diaz-Asper et al., 2004). Nonetheless, the results of IQ tests complement information about developmental, social, educational, and occupational history in providing a comprehensive portrayal of the patient. At the very least, the IQ test helps the clinician to estimate premorbid cognitive functioning, to formulate expectations of performance on other tests, and to determine the level of discourse at which to engage the patient. Multifaceted IQ tests such as the Wechsler and Stanford–Binet tests also provide a cognitive profile, which informs the clinician about relative strengths and weaknesses and may constitute a starting point from which to search for more specific deficits.

Normative data

Irrespective of how the IQ test is employed by the clinician, its usefulness depends on the adequacy of

---

The author thanks James R. Flynn for his helpful comments and updated information.

Address correspondence to Merrill Hiscock, Department of Psychology, University of Houston, Houston, TX 77204–5022, USA (E-mail: mhiscock@uh.edu).

http://www.psypress.com/jcen DOI: 10.1080/13803390600813841

normative data. Even profile analysis, which entails within-individual (ipsative) comparisons, will be corrupted if the norms for one or more subtests are inadequate or inappropriate for the particular patient population being served. Age-stratified and up-to-date norms are an important asset that enhances the usefulness of commercially available IQ tests. If norms are inaccurate or inappropriate, results will be difficult to interpret if not misleading.

With respect to IQ tests as well as other psychological tests, normative problems arise under three circumstances. One circumstance, relatively common in neuropsychology, involves normative data that may be inaccurate even for the population from which the data were obtained. This may occur when the sample size is inadequate, when the test materials are not standardized, or when the test is administered in a manner that deviates from the procedures used in clinical evaluation. Lezak et al. (2004) and Spreen and Strauss (1998) provide several examples of these sources of variability in normative neuropsychological test data. If the norms do not represent accurately the population on which the norms are based, they are even less likely to be appropriate for the clinical population to which the norms are being applied. A second variety of normative problems is observed when the normative data are representative of the normative population but the normative population is discrepant in some significant way from the population to which the norms are to be applied. The respective populations may differ in age, education, socioeconomic level, or general health status. In the United States, for example, norms based on patients at a Veterans Administration Medical Center may not reflect characteristics of the general population (e.g., Burke, 1985). The third category of normative problems pertains to norms that have become old. This is the problem to be considered in the present paper.

Currency of norms is generally regarded as a favorable attribute, but seldom does one find explicit justification for this belief. If a test's content is appropriate for use with a contemporary population, then why should it matter whether test norms were compiled 5 years ago or 50 years ago? It seems that there are two reasons for mistrusting older norms, one of which is rather obvious. The obvious reason is that tests sometimes are revised with respect to content, administration, or scoring rules. Periodic revisions of the Wechsler and Stanford–Binet IQ tests, for example, are accompanied by updated norms. The less obvious reason for preferring recent norms is that the distribution of the ability measured by the test may change over time.

THE FLYNN EFFECT

IQ gains, when considered across various tests given to children and adults in different countries, are roughly twice as large for "culture-reduced" tests as for tests with learned content (i.e., verbal tests). In studies of data from 20 different nations, Flynn (1987, 1994) found that scores on Raven's Progressive Matrices Test (RPM) increased about 0.6 IQ points per year between 1930 and 1990. This conclusion was supported by more intensive analyses of data from the United Kingdom and from four countries—the Netherlands, Israel, Norway, and Belgium—in which the Raven's test had been administered to huge military samples (Flynn, 1987, 1998b, 1999).

Although the rate at which Wechsler and Stanford–Binet IQ has increased is not as dramatic as the rate at which scores on Raven's test have increased, the Flynn effect does apply to the Wechsler and Stanford–Binet tests, as well as tests such as the Otis Intermediate Test of Mental Ability and the California Test of Mental Maturity (Flynn, 1987). In fact, Flynn's discovery of IQ gains over time began with his observation, from tables in the Wechsler Intelligence Scale for Children–Revised (WISC-R) manual, that children who completed both the WISC-R and the older WISC scored an average of 8 points higher on the WISC. Thus, the average IQ for the respective standardization samples must have increased during the interval between the norming of the WISC in 1947–1948 and the norming of the WISC-R in 1971–1973. In Flynn's (1999) words, "Clearly from one sample to the other, over a period of 25 years, Americans had gained 8 IQ points" (p. 6). Figure 1 shows typical mean increases over time for scores from the Wechsler and Stanford–Binet tests and for scores from the RPM.

The gain of 8 IQ points per 25 years, or about 0.3 points per year, was confirmed by examination of Wechsler and Stanford–Binet results for approximately 7,500 individuals between the ages of 2 and 48 years who had completed the original and revised versions of the same test (Flynn, 1984, 1987). Subsequently, after examining differences between WISC-R and Wechsler Intelligence Scale for Children–Third Edition (WISC-III) scores, and between Wechsler Adult Intelligence Scale–Revised (WAIS–R) and Wechsler Adult Intelligence Scale–Third Edition (WAIS-III) scores, Flynn (1998c) noted a discrepancy between the

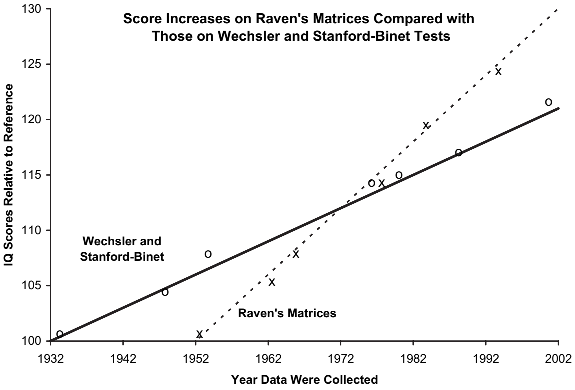

Figure 1. Smoothed curves depicting the differential increase in mean scores on Wechsler, Stanford–Binet, and Raven's Matrices tests during the middle and latter parts of the twentieth century. Adapted from Neisser (1998a). Copyright © 1998 by the American Psychological Association. Adapted with permission.

annual IQ gain as inferred from the children's tests (0.312 points) and that from the adult tests (0.171 IQ points). The source of the discrepancy has been identified by more recent analyses, in which WAIS-III normative data were compared with normative data from older as well as newer IQ tests (Flynn, 2006b). Those analyses indicate that IQs derived from the WAIS-III are inflated by 2.34 points. Consequently, the Flynn effect is underestimated when the WAIS-III is compared with older tests and overestimated when the WAIS-III is compared with more recent tests. Flynn (2006b) concludes that Wechsler and Stanford–Binet IQ continues to rise at the rate of 0.3 points per year. He recommends subtracting 2.34 points from IQs based on the WAIS-III to correct for its tendency to yield "inflated IQs even in the year in which it was normed" (p. 179).

The gains manifested on Wechsler IQ tests are not equally distributed across subtests. In keeping with the principle that increases for culture-reduced tests are greater than increases for tests of learned content, Performance IQ has risen more dramatically than Verbal IQ. Flynn (1998c) has used scores from about 1,000 children who took the WISC-R and WISC-III to calculate a 0.358 IQ points-per-year increase in Performance IQ and a 0.218 points-per-year increase in Verbal IQ. According to Flynn (1999), gains on the more culturally influenced Wechsler subtests such as Arithmetic, Information, and Vocabulary are "small or nil" in English-speaking countries. Likewise, gains on achievement tests are insignificant or nonexistent.

The Similarities subtest from the Wechsler IQ tests stands as an exception to the rule that culturally influenced tests are relatively invulnerable to the Flynn effect. Gains on the Similarities subtest are as large as those observed for Raven's RPM (Flynn, 1999). As pointed out by Flynn, the rapid increase in Similarities scores over time is informative because the test has a verbal format and does not appear to demand spatial reasoning. The ability that is increasing most dramatically thus appears to be broader than spatial ability or some other specific manifestation of nonverbal intelligence.

Characteristics of Raven's test

As noted previously, the most striking evidence for rising IQ comes not from the Wechsler or Stanford–Binet tests, but from Raven's Progressive Matrices Test (RPM), an untimed measure of intellectual ability that was first published by John C. Raven in 1938. The test requires no reading or speaking and only simple written responses. It is easily administered, it has satisfactory reliability, and the results correlate moderately well with scores from other tests that are purported to measure general intellectual ability (Burke, 1985; Llabre, 1984; Raven, 1960; Spreen & Strauss, 1998). The RPM is considered to be a culture-free test (Jensen, 1980), or at least a culture-reduced test, for which there are vast amounts of data from several Western nations (Flynn, 1999). Although the most widely used version of the RPM is the Standard Progressive Matrices, the easier Colored Progressive Matrices and the more difficult Advanced Progressive Matrices are available for use with low- and high-functioning groups, respectively.

Raven's Matrices test does not appear to be used frequently by American neuropsychologists. Nonetheless, it is an especially attractive test for theorists who posit the existence of a general intellectual factor such as Spearman's *g* (e.g., Jensen, 1998; Spearman, 1904). In the words of Neisser (1998a), "Whatever *g* may be, we at least know how to measure it. The accepted best measure, which has played a central role in analyses of the world-wide rise in test scores, is the Raven Progressive Matrices" (p. 9). Among those theorists who endorse Cattell's (1957) decomposition of *g* into fluid intelligence (ability to solve novel problems) and crystallized intelligence (domain-specific knowledge acquired over time), Raven's Matrices test is also regarded as the quintessential measure of fluid intelligence. Snow, Kyllonen, and Marshalek (1984), who conducted a multidimensional scaling analysis of scores from various cognitive ability tests, found that Raven's test occupied the territory at the very center of their model. In other words, the RPM was the best measure of the domain-independent abilities that are required for various kinds of figural, verbal, and numerical problem solving.

What is rising?

Three aspects of the Flynn effect provide potentially important clues as to the identity of the cognitive characteristic that is increasing over time. One clue is the constancy of the slope across time. Since 1932 and probably prior to that, test scores have been increasing at a rate of 3 to 6 IQ points per decade, depending on the IQ test used. The preponderance of evidence indicates that scores are continuing to rise at a constant rate (Flynn, 2006b). Any characterization of the cognitive ability that is increasing must account for the apparent inexorability of the increase. There is at least one exception, however. The Scandinavian countries currently are showing little or no rise in their test scores (Flynn, 2006a). As large IQ increases were seen in Norway prior to 1968, Flynn suggests that Scandinavia might have experienced early increases that have since abated. This raises the possibility that IQ increases in other industrialized nations will also end.

A second clue consists of evidence that IQ gains have been comparable for young children, older children, and adults (Flynn, 1984) and for various nations, ethnic and linguistic groups, and geographic regions (Flynn, 1999, 2006a). The IQ increase can be seen in children who have not yet been exposed to formal education and who, being young children, have had limited exposure to other aspects of the culture into which they were born. The increase has been documented in developed and developing nations. It has been reported in Europe, North American, South America, Asia, and Africa.

The third clue, which has been discussed above, consists of findings that the scores on culture-reduced tests, or tests of fluid intelligence, show an increase twice as large as that observed for tests of learned information, or tests of crystallized intelligence. The increase represents largely an enhancement of people's ability to solve certain kinds of problems rather than their acquisition of more information from the culture in which they live.

Attempts to describe the cognitive skill that is increasing over time fall into three categories. Some authors accept the rising scores at face value and assume that they reflect actual increases in general intelligence or in the ability to adapt to the new cognitive demands of a changing culture (Greenfield, 1998; Martorell, 1998; Sigman & Whaley, 1998; Williams, 1998). For example, Greenfield (1998) suggests that the increasing test scores indicate a rise in "culturally phenotypic intelligence," by which she means an adaptation to the increasing importance of visual electronic media and the decreasing emphasis on traditional print media. This characterization accords well with the greater rise of nonverbal IQ than verbal IQ, although it fails to explain why verbal IQ should rise at all. Greenfield's argument does not account for the ubiquity of the increase across individuals from different cultures, socioeconomic levels, and educational backgrounds unless one assumes that environmental shifts are uniform across diverse groups. It is also difficult to reconcile Greenfield's characterization of the cognitive skill with the absence of age differences in the IQ increase,and even more difficult to reconcile it with the long period over which scores have been rising. As Flynn (1999) has observed, "Ravens gains were large before the television era, much less before the computer-game era" (p. 9).

Lynn (1998) has proposed a modified version of the position that IQ gains reflect actual increases in intelligence. According to Lynn, only the 3-point-per-decade increase in IQ, as observed in the Wechsler data, reflects a genuine rise in intelligence. The larger increases on other tests, such as those shown on Raven's Matrices by northern European and Israeli military conscripts, are attributed to mathematics education. Lynn invokes the experimental study of Carpenter, Just, and Shell (1990) to argue that Raven's test requires the application of "the mathematical principles of addition, subtraction, progression, and the distribution of values" (Lynn, 1998, p. 212). If increasing proportions of adolescents from northern Europe and Israel remain in school and become more proficient with mathematics than were earlier cohorts, then, according to Lynn, this additional learning may account for the size of the increase in Raven's performance.

Insofar as Lynn's explanation concerns degrees of IQ increases, rather than the presence or absence of increases, the quality of the extant evidence is inadequate for evaluating the explanation. The magnitude of the Flynn effect varies too widely across studies. For example, Flynn (1987) found that the rise in Raven's scores in different studies of children between 8 and 16 years ranged from 0.19 to 1.25 IQ points per year. Given the marked variability across samples, one cannot conclude that the gain for children is smaller than the gain for older individuals, who are more likely to have benefited from improvements in mathematics education. Even if a reliable difference between children and adults were to be established, the

Colored Progressive Matrices test used with younger children may not be comparable in its cognitive demands to the Standard Progressive Matrices test used with adolescents and young adults. The Colored Matrices test is even less likely to be comparable to the Advanced Matrices test that was examined by Carpenter et al. (1990). The degree to which mathematical reasoning is required by the different forms of Raven's test is difficult to specify, but it does not appear to be a salient demand of either the Colored or Standard test (Carlson & Jensen, 1980; Costa, 1976; Villardita, 1985; Waltz et al., 1999). Also, Lynn's account fails to explain why the magnitude of the Flynn effect is so large for the Wechsler Similarities subtest.

Flynn, in the 1980s and 1990s, articulated a third point of view regarding the nature of the ability that is being indexed by the rising test scores. Specifically, he asserted that the IQ gains have no relation to any form of real-world intelligence. In his 1987 article in *Psychological Bulletin*, Flynn used the term "abstract problem-solving ability (APSA)" to describe the characteristic that is increasing over time. According to Flynn, APSA is the ability to solve problems on an IQ test, and this ability can be highly divergent from "the real-world problem-solving ability called intelligence" (Flynn, 1987, p. 188). Flynn (1999) distinguished among (a) IQ test scores, (b) psychometric *g*, and (c) real-world intelligence. Real-world intelligence can be inferred from psychometric *g*, which in turn can be inferred from scores on IQ tests. The inferential chain can be represented as follows:

$$\begin{aligned} \text{IQ test scores} &\rightarrow \text{psychometric } g \\ &\rightarrow \text{real - world intelligence} \end{aligned}$$

Flynn accepted the validity of inferring psychometric *g* from IQ tests (which, he acknowledged, measure *g* and not much else) but he disputed the second inference. According to Flynn, neither academic achievement-test results nor "real-world achievements" support the conclusion that humans are becoming more intelligent.

If the cognitive ability that increases over time is as narrow as Flynn (1998a, 1999) initially construed it to be, then the relationship between test scores and their meaning is isomorphic. When divorced from real-world manifestations, intelligence becomes only "what the tests test" (Boring, 1923). The unknown ability that is rising is psychometric intelligence, and psychometric intelligence is defined as that which IQ tests measure. Given this tautology, any change in test scores reflects nothing more than a change in the ability to perform the tasks required by the test. The problem with this interpretation is its incompatibility with a wealth of evidence that connects psychometric *g* with real-world criteria. Intelligence theorists who agree on little else agree that, under the right circumstances, IQ test scores do predict educational and occupational outcomes (cf. Herrnstein & Murray, 1994; Sternberg, Grigorenko, & Bundy, 2001).

Flynn's new model of IQ rise

Flynn's original interpretation of the rising-curve phenomenon leaves us with two strikingly divergent characterizations of IQ disparities. On the one hand, population-wide increases in IQ are devoid of real-world significance and are environmentally determined. (There is no reason to believe that differential reproduction rates could account for the rapid rise of scores from one generation to the next; Flynn, 1994.) On the other hand, individual differences in IQ within a cohort are known to be associated with people's real-world attainments and seem to be influenced substantially by genetic inheritance (e.g., Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Herrnstein & Murray, 1994; Jensen, 1973; Plomin, 1990). How can IQ assume both guises? This paradox requires yet another interpretation of the Flynn effect, an interpretation that reconciles the intergenerational rise in IQ with the individual differences in IQ that are found within cohorts.

The new interpretation of the Flynn effect (Flynn, 2003, 2006a) rests on a quantitative model of the genesis of IQ differences that has been formulated by Dickens and Flynn (2001a, 2001b). The Dickens–Flynn argument begins with a reassessment of evidence that has been used to establish the heritability of IQ. While acknowledging that genes contribute to IQ patterns obtained from twin studies and adoption studies, Dickens and Flynn (2001b) assert that genetic differences among individuals are amplified by correlations between genes and environment. For instance, a young child who demonstrates an aptitude for a particular activity may be encouraged by parents and teachers to pursue that activity. Additional resources will be made available to the child, and the child's proficiency will increase. The increased proficiency will lead to successes that, in turn, will increase the child's interest in the activity and the motivation to improve his or her knowledge and skill. Dickens and Flynn refer to this kind of matching of genetic characteristic and environment, and the positive feedback loop that ensues, as a “multiplier effect.” Similar arguments have been made by others (e.g., Bell, 1968; Jensen, 1975; Plomin, 1986; Scarr, 1992). However, Dickens and Flynn use the principle of gene–environment correlation for a specific purpose—that is, to show that, in behavior genetics studies, heredity is credited for individual differences produced by this interactive process when, in fact, only a portion of the differences is the direct consequence of genetic endowment. Thus, the true heritability of IQ is not as great as that reflected in heritability indices, and the conflict between Flynn effect and findings from behavior genetics is reduced accordingly.

The Dickens and Flynn (2001b) model also incorporates a new means of determining the true impact of environmental variables that are not correlated with individuals’ intrinsic characteristics. These “exogenous environmental effects” can be divided into transient and persistent components. The authors argue that, at any specific time, the variance explained by a persistent exogenous effect may be weak relative to the variance attributable to transient factors. However, the same persistent factor will appear to be much stronger when it is compared with the average of all the exogenous variables that have been present at different times. In the words of Dickens and Flynn, “this collective averaging further diminishes the importance of random individual environment influences, whereas consistent factors acquire an impact beyond what we would expect viewing the individual in isolation” (p. 351).

The next task for the Dickens–Flynn model is to account for the large magnitude of IQ increases across generations, which appears to be much greater than the magnitude of the individual differences produced by gene–environment correlations. To explain the size of transgenerational IQ increases, Dickens and Flynn (2001b) and Flynn (2003, 2006a) emphasize the distinction between social multipliers (exogenous environmental effects) and individual multipliers (gene–environment correlations). Whereas individual multipliers influence an individual’s position in a societal pecking order, those environmental factors have little direct effect but instead exert their influence through interaction with the individual’s inherited abilities. Social multipliers, in contrast, have a pervasive effect on an entire cohort. To borrow one of Dickens and Flynn’s most effective examples, reinforcement for basketball accomplishments (an individual multiplier) accrues almost entirely to those individuals who have high levels of aptitude for the game. Unusually tall or talented individuals will capitalize on their own individual multipliers and rise to the top of the basketball hierarchy. However, the markedly increased popularity of basketball, presumably attributable to the widespread televising of professional games, constitutes a social multiplier. As an entire culture becomes progressively more interested in basketball, many more children play the game, and the general level of skill, coaching, and competition increases dramatically. The distribution of basketball performance is shifted upward, and, over a sufficient time span, the magnitude of that shift will be much greater than the effect that individual multipliers have on elite players. Gould (1996) has reached a similar conclusion from his analysis of major-league baseball statistics. As the overall skill level rises, the variability at the right tail of the distribution is reduced, and the gap between the most extraordinary batters and “ordinary” batters is reduced accordingly. To use Dickens and Flynn’s terminology, the baseball statistics seem to show that the influence of individual multipliers on the most elite players is constrained by biological limits that are less likely to constrain the influence of social multipliers on other players.

Athletics may provide the clearest examples of social multipliers, but Flynn (2003, 2006a) applies the same reasoning to intellectual, academic, occupational, and avocational realms. According to Flynn’s new theory, the IQ increase does reflect an actual increase in intellectual prowess, and the causal factors are environmental. Nevertheless, this increase in IQ can be reconciled with evidence that individual differences in IQ predict real-world outcomes as well as evidence that the individual differences are determined in part by genetic inheritance. Flynn’s specific account of the environmental factors that cause IQ to rise are summarized at the end of the following section.

Why are the scores rising?

Several writers have offered opinions and claims regarding variables that might be causing a sustained international rise in IQ. Often the putative cause depends on the writer’s perception of that which is changing. Thus Greenfield (1998), who interprets the Flynn effect as in increment in visuospatial ability, argues that the effect stems from an emerging cultural shift away from traditional verbal communication media and toward new visual and interactive media. Writers who equate the rise in test scores with an increase in general intelligence are more likely to favor causal explanations such as improved nutrition (Lynn, 1990, 1998; Sigman & Whaley, 1998) or greater exposure to formal education (Williams, 1998). As already noted, Lynn (1998) regards part of the increase in Raven's scores as a consequence of more extensive education in mathematics. He attributes the remainder of the Flynn effect—the "genuine" IQ increase—to nutritional improvements. However, as Flynn (2003) has pointed out, improved nutrition should benefit primarily those individuals who are most deprived. Consequently, if nutritional improvement were the primary cause of IQ increases, there should be disproportionate gains at low IQ levels, and this does not seem to be the case, at least in the United States.

Williams (1998) has mentioned a large number of other environmental changes that might have had a favorable effect on the intelligence of children in various countries. Her list of possible causes includes: (a) a greater emphasis on procedural knowledge than on declarative knowledge; (b) teaching to the test; (c) improved health; (d) smaller families; (e) better educated parents; (f) changes in parenting style; (g) a trend toward urbanization; and (h) changing patterns of stress on pregnant women and their babies. When we add these eight variables to the previously mentioned factors—better nutrition, more formal education, and a cultural shift from verbal to nonverbal media—we have a large choice of explanations for the Flynn effect. Some explanations are more plausible than others. Some are better able to account for the long history of the rising curve, whereas others are better able to account for its pervasiveness across nations in recent decades. The explanations are not mutually exclusive; two or more may be partially correct. Yet it is possible that none of the explanations is correct. This is the position originally taken by Flynn: "I believe it is fair to say that up to now, efforts to identify the environmental factors that have caused IQ gains have not come to much" (Flynn, 1998a, p. 49).

Building on the work of Dickens and Flynn (2001a, 2001b), Flynn recently has constructed a framework for explaining the rising test scores. Flynn (2006a) attributes the rise in IQ (especially in the United States during the first half of the twentieth century) primarily to increases in the number of years of formal education. He notes that the average number of years of public education in the United States rose from 8 to 10 between World War I and World War II (Tuddenham, 1948). Formal education exemplifies the power of a social multiplier. As students spend more years in school, "each student is surrounded by fellow students who are more competent, better students make better teachers for the next generation of students, parents become more serious about schooling and homework, the lengths of the school day and school year tend to increase" (p. 15). The transgenerational impact of increased education presumably explains why the Flynn effect is apparent in young children.

IQ gains since World War II, according to Flynn (2006a), can be attributed to a shift of emphasis from reading, writing, arithmetic, and other "disciplined" learning to "on-the-spot problem-solving skills." This educational shift seems to be associated with several demographic trends, such as greater urbanization and affluence, decreasing family size, changes in the kinds of work that people do, and the increasing importance of leisure activities. Perhaps many contemporary workplaces require more on-the-spot problem solving (fluid intelligence) than did workplaces of the past, and perhaps some contemporary leisure activities (e.g., playing video games, playing poker via the Internet, and using computer software) require more fluid intelligence than did leisure activities of the past. As occupational and avocational activities evolve, and as educational systems adjust to those societal changes, successive generations of children will continue to perform better on tests of abstract problem solving while demonstrating little or no improvement in traditional academic skills. Alternatively, societal emphasis may turn in a different direction. In either case, Flynn would invoke the concept of the social multiplier to explain a rapid increase in the standard of performance within selected intellectual realms.

IMPLICATIONS FOR CLINICAL NEUROPSYCHOLOGY

Certain practical implications of the Flynn effect are especially relevant to clinical neuropsychology, and those are the implications that are emphasized here. It should be noted, however, that Flynn's concept of social multipliers resembles the venerable neuropsychological principle of dissociation (e.g., Luria, 1973). Flynn (2006) has described an analogy between cultural variables, which may cause certain cognitive skills to be developed preferentially, and brain damage, which may selectively impair some cognitive functions without affecting other functions. Both kinds of dissociation—societal and pathological—are means of fractionating the global pattern of general intelligence. Thus, in principle, findings from clinical neuropsychology and cognitive neuroscience can be used to guide hypotheses about how normal brains might be expected to change as social multipliers reinforce particular cognitive skills while neglecting other skills.

Old tests and old norms

The most obvious implication of rising test norms for the clinical neuropsychologist is the risk of being misled by old tests and old norms. Yet, in the past, when a new test became available, clinicians typically were more worried about being misled by the new test and the new norms. When the WAIS–R was introduced in 1981, clinical neuropsychologists recognized that this test, with its new norms, yielded lower scores than did the 1955 WAIS (Bornstein, 1987; Chelune, Eversole, Kane, & Talbott, 1987; Reitan & Wolfson, 1990). Some of these authors acknowledged that the discrepancy was partly attributable to a population-wide rise in IQ, but they were more concerned about differences between the WAIS and WAIS–R that might alter subtest profiles. As neuropsychologists, their primary focus was on the neuropsychological validity of pattern analysis. Readers were cautioned that decision rules based on the older test might not be applicable to the newer test.

Transitional problems notwithstanding, the Flynn effect constitutes a compelling reason to adopt new tests and new norms. Nearly 60 years have passed since norms for the original WISC were obtained in 1947 and 1948. The WAIS was published more than 50 years ago, and the Wechsler–Bellevue tests were published more than 60 years ago. If Full Scale IQ has been increasing by 3 points per decade, then the results from those old tests and norms would be inflated by 15 to 18 points. If the earlier tests continue to be used routinely for testing, and if the norms are not adjusted for the Flynn effect, the average IQ for the general population will be 115 to 118 instead of 100. A patient with a tested IQ of 100, obtained currently from any of the early tests, would be expected to have an IQ of 82–85 if tested on a current Wechsler IQ test. This would be the case despite the appropriate age adjustments having been made in the calculation of IQ.

This is an extreme example, but the same problem will be experienced to some degree whenever a test with old norms is used to measure IQ or to estimate premorbid IQ. Hiscock, Inch, and Gleason (2002) describe an instance in which awareness of the Flynn effect led to a dramatic change in the way Raven's Matrices scores were interpreted. In the 1980s, before the Flynn effect was widely known, RPM raw scores were converted to IQs or *z*-scores on the basis of Raven's norms from 1960 or from norms published by Peck in 1970 and reprinted in Lezak (1983). Irrespective of which norms were used, Hiscock et al. found that patients with traumatic brain injuries tended to score above the population average on the RPM. It appeared at the time that Raven's Matrices test was singularly insensitive to the effects of head trauma and that RPM scores might even be useful for estimating premorbid intellectual ability. Subsequently, the availability of more recent norms (J. Raven, Raven, & Court, 1995) allowed the investigators to make retrograde as well as anterograde adjustments for the Flynn effect. Using an adjustment of 0.6 IQ points per year, Hiscock et al. found almost perfect congruity between the effects of retrograde and anterograde corrections. After adjustment for the Flynn effect, patients' RPM scores were commensurate with their WAIS–R IQ. It was concluded that the RPM is neither more nor less sensitive than the WAIS–R to the effects of traumatic brain injury.

Researchers who track their participants psychometrically in long-term longitudinal studies often face a difficult dilemma. Does one choose a test and stay with that test for the duration of the study even though the test and its norms have been superseded? Or does one switch to the newer test and thereby jeopardize the comparability of early and late assessments? Similar obsolescence problems arise in clinical practice when an IQ test with superseded norms is retained in a battery of tests in order to preserve the validity of the interpretive rules (Reitan & Wolfson, 1985) or when an old version of an IQ test is administered to a patient who has taken the same test in the past. In both instances, the IQ derived from the old test will overestimate the patient's actual IQ. On the other hand, switching to a newer version of the test may necessitate comparing scores based on different sets of items.

It may be counterintuitive to avoid retesting a patient with an IQ test that had been administered several years ago. The clinician may reason that, if the test had been appropriate for the patient 20 years ago, it should still be appropriate so long as the necessary age adjustments are made. The Flynn effect does not imply that the test has become inappropriate. It implies that the norms have become inappropriate. The WAIS norms for 60-year-olds, for example, apply to individuals who were approximately 60 years old when the test was normed and not to individuals who since have become 60 years old. Members of that earlier cohort of 60-year-olds, if alive today, would be more than 110 years old. Unless norms for the old IQ test can be collected from contemporary 60-year-olds, the IQ obtained from the WAIS will not be comparable to the IQ obtained from the WAIS in the years immediately following its publication in 1955. This argument, of course, is based on psychometric considerations and does not preclude readministering the old test for the purpose of making item-by-item comparisons. It may be instructive, for instance, to note differences between present and past performance on Block Design or Vocabulary subtests as indicated by the respective raw scores and by the specific items that were passed and failed.

Admittedly, a revised IQ test may have been altered in ways that compromise the clinical interpretation of patterns of scores across subtests. This is precisely the problem that had caused consternation after release of the WAIS–R (Bornstein, 1987; Chelune et al., 1987; Reitan & Wolfson, 1990). On the other hand, analysis of patterns from a previous version of the test will also be compromised if the Flynn effect has produced differential increases in the various subtest scores. For instance, the so-called ACID profile from the children's Wechsler IQ test (relatively low scaled scores on Arithmetic, Coding, Information, and Digit Span) is commonly interpreted as an indicator of poor attention, impaired auditory short-term memory, and learning difficulties (e.g., Rourke, Bakker, Fisk, & Strang, 1983). Although that interpretation might have been valid with respect to the 1949 WISC during the 1950s, the same pattern would require a different interpretation today because some of the subtest scaled scores would have been affected more than others by nearly 60 years of exposure to the Flynn effect (Flynn & Weiss, 2006). Continuing to use an old test and its old norms is not an effective means of preserving the validity of clinical decision rules.

Age-related change

Another important implication of the Flynn effect, which may be less obvious than the first, concerns age norms. Clinical neuropsychologists who work with adult clients are accustomed to norms that indicate age-related change in performance, and clinicians know that some tests are especially sensitive to the effects of aging. The Wechsler Digit Symbol subtest is one example. On the WAIS–R, the average raw score for Digit Symbol decreases substantially between the ages of 25 and 65 years. A raw score of 58 is equivalent to a scaled score of 10 (average) if the examinee is 25 years old. However, if the examinee is 65 years old, the same raw score is equivalent to a scaled score of 15, which is 1.67 standard deviations above average.

The norms for different age categories are obtained using a cross-sectional (rather than longitudinal) methodology. Different cohorts are selected to represent 25-year-olds and 65-year-olds. Consequently, the performance differences between age groups on Digit Symbol may be confounded by cohort effects, or differences between the groups other than age per se (Cozby, 2003). The Flynn effect is based on birth date, not age, and thus is a potential confounding variable. Although there remains some ambiguity about the strength of the Flynn effect in different cohorts, the preponderance of evidence indicates that scores are rising by similar amounts in all age groups (Flynn, 1984, 1998a; Hanson, Smith, & Hume, 1985; Lynn, 1998). Therefore, for the purpose of discussing age norms, the Flynn effect is assumed to be independent of age. In other words, the following examples are based on the assumption that age effects and the Flynn effect are additive.

Given that age and cohort effects are additive, one effect can be disentangled from the other using logic similar to that applied by Schaie (1994) to cross-sectional data collected as part of the Seattle Longitudinal Study. Having found divergent cohort gradients for different indices of mental abilities, Schaie recognized that his cross-sectional patterns either underestimated or overestimated the actual age-related changes in those abilities. Similarly, if one knows the magnitude of the Flynn effect with respect to a specific Wechsler subtest, then one can subtract the Flynn-effect component from the age-related decrement observed in the cross-sectional norms. The best estimate of true age-related deterioration of performance is the tabulated age-related change minus the change attributed to the Flynn effect. These calculations, in other words, provide an estimate of the degree to which normative age-related changes in test performance are contaminated by cohort effects.

WAIS–R Digit Symbol norms can be used to illustrate the procedure. As mentioned previously, comparison of norms for 25-year-olds and 65-year-olds indicates that performance falling at the average ($z=0$) for the younger cohort would fall well above average ($z=+1.67$) for the older cohort. In other words, mean performance of the two age groups differs by 1.67 standard deviations, which reflects a decline of about 0.4 standard deviations per decade. But a difference of 0.20 standard deviations (3 IQ points) per decade would be expected on the basis of the Flynn effect alone. That is, about half of the apparent age-related deterioration in normative Digit Symbol performance can be attributed to the Flynn effect. If Digit Symbol performance is more susceptible to the Flynn effect than are most other WAIS–R subtests—which seems plausible, given that the Flynn effect is strongest for culture-reduced tests—then more than half of the apparent age-related decline could be attributed to the Flynn effect. Dickinson and Hiscock (2005) have undertaken additional analyses in which the magnitude of the Flynn effect for each subtest is calculated individually from data published in the WAIS-III administration manual (Wechsler, 1997). These calculations indicate that 60% of the normative age-related decline in Digit Symbol performance is attributable to the Flynn effect.

The previous example relies on an indirect method of decomposing normative differences between age groups into age-related and birth-date-related sources. A more direct method can be applied to a set of Raven's Matrices data in which raw scores for 10 cohorts are available (Flynn, 1998a). A graph of the data is shown in Figure 2. The curve depicts the median raw scores for 10 British cohorts with birth dates ranging from 1877 to 1967. The five points on the left represent five age groups (25, 35, 45, 55, and 65 years) that were tested in 1942, and the five points on the right represent five groups with the same mean ages that were tested in 1992. These data allow two kinds of analysis. Comparisons among different age groups tested during the same year reflect the additive effects of age and birth date. The data also allow comparisons between groups of the same age that were tested 50 years apart. The latter comparisons isolate the effects of birth date in the absence of age differences. The data and calculations are shown in Table 1.

In Table 1, the 10 data points from Figure 2 (from left to right) are denoted as Groups A to J. Groups A to E were tested in 1942, and Groups F to J were tested in 1992. The average annual difference between age groups tested in the same year can be used to estimate the combined effects of birth date and age. For the 1942 data, the average difference between groups (Average 1 in the table) was 0.51 points/year. Average 3, shown at the bottom of Table 1, represents the mean difference between each age group in 1942 and the corresponding age group in 1992. The average increase is 0.36 points/year. This constitutes an estimate of the magnitude of the Flynn effect alone for the 50-year interval from 1942 to 1992. Consequently, the proportion of the age-related decline that can be attributed to the Flynn effect is 0.36 divided by 0.51, or 71%. Consequently, if the 1942 medians are used as age-group norms, no more than 29% of the differences between any two groups could be

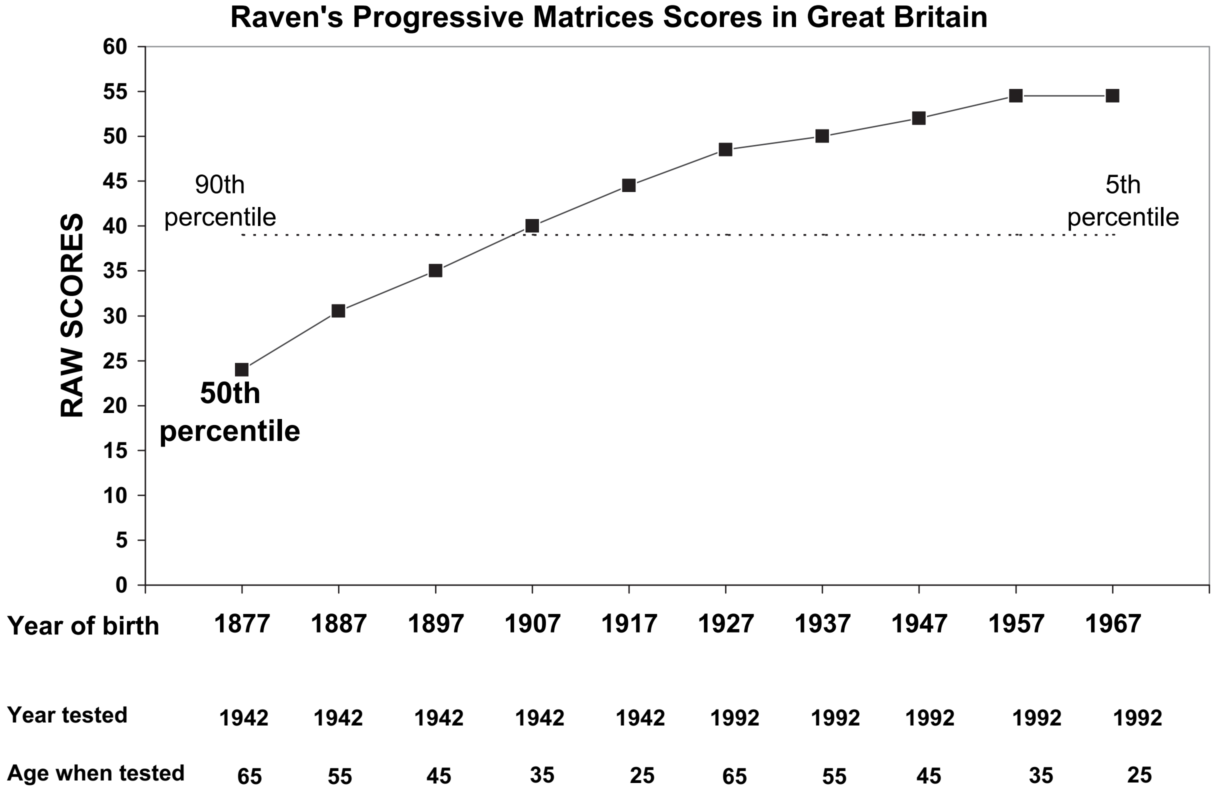

Figure 2. Median raw scores for 10 samples of British adults with birth dates ranging from 1877 to 1967. The five data points on the left represent groups tested in 1942 at ages 65 to 25 years (from left to right). The five data points on the right represent groups tested in 1992 at ages 65 to 25 years (from left to right). The dashed line indicates that a raw score falling at the 90th percentile in 1942 fell at the 5th percentile in 1992. Data from Raven et al. (1995). Figure adapted from Flynn (1998a). Copyright © 1998 by the American Psychological Association. Adapted with permission.

TABLE 1
Age and year-of-testing effects on Raven's Progressive Matrices raw scores for British adults tested in 1942 or 1992

Group	*Year of birth*	*Year tested*	*Age*[a]	*Raw score*
A	1877	1942	65	24
B	1887	1942	55	30
C	1897	1942	45	35
D	1907	1942	35	41
E	1917	1942	25	44
F	1927	1992	65	49
G	1937	1992	55	51
H	1947	1992	45	53
I	1957	1992	35	55
J	1967	1992	25	55

Average 1 (average of annual birth date and age differences combined for 1942 data):
Average [(E – A) + (E – B) + (E – C) + (E – D) + (D – A) + (D – B) + (D – C) + (C – A) + (C – B) + (B – A)] = 0.51
Average 2 (average of annual birth date and age differences combined for 1992 data):
Average [(J – F) + (J – G) + (J – H) + (J – I) + (I – F) + (I – G) + (I – H) + (H – F) + (H – G) + (G – F)] = 0.16
Average 3 (average of annual difference between comparable age groups in 1942 and 1992):
Average [(F – A) + (G – B) + (H – C) + (I – D) + (J – E)] = 0.36
[a] In years.

attributed to an actual decline in performance associated with aging per se. The majority of the decline in performance from one age group to the next could be accounted for by the Flynn effect.

The same method yields surprising results when applied to the 1992 data. It is evident from Figure 2 that the slope of the curve is much smaller for the 1992 data than for the 1942 data. As indicated by Average 2 in Table 1, the combined effect of birth date and age for groups tested in 1992 is only 0.16 points/year, which is less than the estimated magnitude of the Flynn effect alone. The flattening of the curve for the 1992 data seems to reflect a ceiling effect, as the median score for 25- and 35-year-olds indicates that half of the individuals in those cohorts had raw scores above the 86% level. Other data indicate that Raven's Matrices scores are continuing to rise as rapidly as in the past (Neisser, 1998b). In fact, the unremitting upward shift in the distribution has prompted the development of a more difficult version of the Standard Progressive Matrices (J. Raven, Raven, & Court, 1998).

Even though the refined cohort norms that are available for Raven's Matrices and the Wechsler IQ tests are not available for most neuropsychological instruments, it sometimes is possible to estimate the degree to which the Flynn effect has influenced performance on specific neuropsychological tests. For instance, Dickinson and Hiscock (2006) found six comparable sets of norms for the Trail Making Test that were published between 1968 and 2004. Analysis of those norms showed that overall performance on Part B of the test has improved significantly over the interval of 36 years. Mean completion time has decreased from 85 s to 60 s. The magnitude of this cohort effect is only slightly less than the difference between median times for 30- and 65-year-olds in normative data (e.g., Spreen & Strauss, 1998). This implies that a substantial portion of the age-related slowing of performance is attributable to the Flynn effect (at least in the middle and upper ranges of the distribution). The lesser degree of improvement on Part A of the Trail Making Test between 1968 and 2004 was not statistically significant. Results for other tests were either negative, or indeterminate because of a paucity of data. This kind of analysis obviously is limited by the availability, quality, and comparability of archival data.

An alternative strategy for estimating the magnitude of the Flynn effect for a neuropsychological test entails comparing age-related declines as estimated from both cross-sectional and longitudinal data. If the age-related change observed in cross-sectional data is commensurate with the change observed in longitudinal data, then the cross-sectional results can be presumed to index a true age-related deterioration. If, however, cross-sectional data show greater decline than do longitudinal data, the difference probably reflects a cohort effect. When this strategy was applied to scores from the Boston Naming Test by Connor, Spiro, Obler, and Albert (2004), the cross-sectional data showed a linear decline that was more than 50% greater than the decline seen in the longitudinal data, and the quadratic component of the decline was twice as great in the cross-sectional data as in the longitudinal data. The authors ruled out practice effects and attrition as factors that might have caused the cross-sectional results to overestimate the actual amount of age-related decline. Connor et al.'s results seem to indicate that one third to one half of the age-related change observed in the cross-sectional data is a cohort (Flynn) effect.

These attempts to disentangle age-related changes from birth-date effects are only as accurate as the data on which they are based and only as valid as the explicit and implicit assumptions on which the analyses depend. The resulting estimates of age-related and birth-date components cannot be considered to be definitive, even for the WAIS–R Digit Symbol subtest and Raven's Matrices test. Nonetheless, it seems clear that year of birth accounts for a substantial proportion of what appears to be an age-related decline in normative performance. Quite possibly, the differential strength of the Flynn effect across tests accounts for the apparent differences in sensitivity to the effects of aging. What appears to be differential sensitivity to aging may be, in fact, differential sensitivity to the Flynn effect (Dickinson & Hiscock, 2005). Alternatively, the characteristics of a test that make it sensitive to aging may be the very characteristics that make a test sensitive to the Flynn effect as well. A statistical technique known as age-period-cohort (APC) modeling may prove useful in future attempts to identify the relative contribution of age and cohort effects to changes in mental ability and neuropsychological functioning (Holford, 1991).

A better understanding of how the Flynn effect influences age-group norms will not only benefit clinical neuropsychology and the psychology of aging, but will also help to elucidate the Flynn effect. If researchers can specify which neuropsychological tests are "contaminated" by the Flynn effect and the degree to which they are contaminated, then it may be possible to identify the characteristics of a test that render it susceptible to the Flynn effect. For Digit Symbol and the other Performance subtests from the Wechsler IQ tests, a substantial proportion of the apparent age-related performance decline can be attributed to the Flynn effect; 70-year-olds perform worse on Digit Symbol partly because their skills have actually decreased over the previous 50 years of their lives, but the drop in performance is due largely to the fact that they were born 50 years earlier than the 20-year-olds in the same normative sample. This may also apply to some neuropsychological tests.

BROADER IMPLICATIONS OF THE FLYNN EFFECT

The problems associated with rapidly changing normative test performance are hardly trivial matters for the clinical neuropsychologist. Nevertheless, as implied by the title of Flynn's (1999) article, "Searching for Justice: The Discovery of IQ Gains over Time," the rising-curve phenomenon has broader ramifications for psychology and society. The impetus behind much of Flynn's effort to document and explain the rising curve is his interest in understanding the average IQ difference between racial or ethnic groups (e.g., Herrnstein & Murray, 1994; Jencks & Phillips, 1998; Jensen, 1969). Flynn points out that the mean IQ difference between generations born 30 years apart, as measured by Raven's Matrices, is larger than the mean IQ difference between American whites and blacks. Moreover, because children and their grandparents share the same genes, we can be sure that the intergenerational IQ difference is not genetic irrespective of the extent to which within-group variability is genetic. Thus, the Flynn effect constitutes a compelling example of large between-group IQ differences that are completely environmental. Flynn argues that IQ differences between ethnic or racial groups are analogous to intergenerational differences. The differences between ethnic groups are also likely to be environmental in origin notwithstanding data showing within-group differences to be genetically influenced. The reader is referred to pages 12–16 of Flynn (1999) for a detailed discussion of this explanation for the IQ gap between white and black Americans.

Flynn (2000) also has noted the dilemma that the rise in IQ has created for psychologists who use IQ tests to classify people as mentally retarded. After questioning the justification for using an IQ of 70 as a fixed value that separates the retarded from the nonretarded, Flynn uses IQ norms to show that the percentage of Americans who have met this criterion has ranged at different times during the past 50 years from 0.5 to 4.3%. As scores rise throughout the population, fewer and fewer individuals meet the IQ criterion for mental retardation. Then, when a new IQ test is published, along with new norms, there is a dramatic increase in the proportion of the population with scores falling below 70. Flynn concludes that awareness of the rising scores, and of the adjustments provided by newer tests and newer norms, gives psychologists a choice. Either they can select the version of the IQ test that is more likely to yield the desired classification, or they can disregard IQ testing and classify individuals solely on the basis of adaptive functioning. In a more recent publication, Flynn (2006b) offers a third option—that is, calculating the interval between the time at which a test was normed and the time at which it was administered and then adjusting the IQ downward to counteract the Flynn effect. Flynn's article provides all the information necessary to make the adjustments, including a list of the years during which the various Wechsler and Stanford–Binet IQ tests were normed.

Whereas Flynn's analysis of the mental-retardation problem is based on IQ test norms, Kanaya, Scullin, and Ceci (2003) have used data from several American school districts to show how changes in IQ test norms actually affect the diagnosis of mental retardation (MR). From their analysis of longitudinal data, the authors found that the magnitude of the Flynn effect for children in the borderline and mild MR ranges of IQ was similar to that for children in the middle of the IQ distribution. Moreover, children in the borderline range who were tested initially with the WISC–R and then retested with the WISC-III were more than twice as likely to be classified as MR as children who were evaluated twice with the same test. As expected, the Flynn effect was most likely to affect the classification of children who scored close to 70, the cutoff for MR. Even though a WISC-III IQ that falls a few points below 70 is equivalent to a WISC–R score slightly above 70, children were nearly three times as likely to be placed into the MR category if they scored 66–70 on the WISC-III than if they obtained a comparable score of 71–75 on the WISC–R.

Kanaya et al.'s (2003) analyses confirm Flynn's (2000) portrayal of the Flynn effect as a force that, over time, progressively inflates IQ scores until a new test, with new norms, causes the scores to "plummet back toward baseline" (p. 787). The authors caution that "there is reason to believe that many students are diagnosed as MR based on the year in which they are tested and test norms used rather than on their cognitive ability" (pp. 786–787). In addition to educational consequences of IQ classifications (e.g., eligibility for special-education services), Kanaya et al. also mention several financial, legal, and occupational consequences. The most dramatic consequence in the United States is eligibility for the death penalty. The U.S. Supreme Court ruled in 2002 that a mentally retarded individual cannot be executed after conviction for a capital crime. Although the intellectual classification of adult defendants is not determined solely by IQ tests, IQ is often an important factor. Thus, a defendant's intellectual classification literally may be a matter of life or death. Accordingly, the decision to execute or not execute a person may hinge on whether the person's IQ has been properly corrected for the Flynn effect. With increasing frequency, courts in the United States are recognizing the need for such corrections (see Flynn, 2006b).

CONCLUSIONS

A pervasive increase over time in performance on IQ tests is well established. The magnitude of the increase is especially marked when culture-reduced tests of general intelligence, such as Raven's Matrices, are used. The Flynn effect also raises scores from Wechsler and Stanford–Binet IQ tests, and the increases are sufficiently large as to present an interpretive problem for practitioners who administer IQ tests to their clients. Not only does the Flynn effect cause published norms for Full Scale IQ to become progressively less appropriate over time, but it also causes different subtest norms to change at different rates. Clinical neuropsychologists who use old versions of IQ tests not only will overestimate IQ but also will risk misinterpreting ipsative indicators such as Verbal–Performance disparities and subtest profiles.

Rapidly changing norms are especially problematic in situations that require reevaluation of a client after a new version of a test has been introduced. If the psychologist chooses the previous version of a test—the version used in the initial assessment—changes in performance can be measured directly but only at the cost of sacrificing the currency of the norms. To the degree that the subtests have been influenced unequally by the Flynn effect, the decision rules that are commonly applied to patterns of scores obtained from the test will be undermined. If the psychologist opts for the revised version of the test, the benefit of up-to-date norms is offset by interpretative ambiguity associated with altered content. The psychologist at first will not know whether clinical decision rules remain valid for the new test and will have to await the availability of empirical evidence to support the continued use of those rules. The clinician's dilemma becomes particularly acute if the test results will influence classification (e.g., mental retardation) or intervention (e.g., special education).

Norms, after being compiled and published, are static but the Flynn effect continues. Consequently, potential Flynn-effect problems do not end with the acceptance of a new test and current norms. The norms will gradually grow old and fail to reflect the recent rise in ability. As time passes, retesting an individual with the same test requires that the Flynn effect be taken into account even if the test has not been revised, and new norms have not been disseminated.

Neuropsychologists who evaluate elderly patients face an additional Flynn-effect problem—namely, the confounding of cross-sectional age-group norms with year of birth. This is the classical cohort effect, to which cross-sectional data are always susceptible. Normal decrements in subtest scores with increasing age reflect an unknown mixture of actual age-related deterioration of performance and birth-date-related decline (i.e., the Flynn effect). Preliminary analyses suggest that, for two of the tests most sensitive to age-related change—Raven's Matrices and Wechsler Digit Symbol—the Flynn effect may account for more than 50% of the decline in raw scores between the ages of 20 and 70 years. If this finding is confirmed in future research, it will necessitate modifying the interpretation of what previously has been regarded as biologically based decline in mental ability. Disentangling age-related decline from the Flynn effect is made more difficult by the likelihood that the tests most sensitive to age-related biological deterioration are the very tests that are most vulnerable to the Flynn effect (Dickinson & Hiscock, 2005).

IQ may be tip of the Flynn-effect iceberg. The worst case scenario for clinical neuropsychology involves widespread Flynn-effect contamination of neuropsychological test norms. This is a plausible concern because neuropsychological instruments tend to be relatively culture free, and many of the test norms are old. The problem is compounded to the extent that the various neuropsychological tests and norms have been published at different times. Not only are the tests likely to be differentially vulnerable to the Flynn effect, but their respective norms are out of date in varying degrees.

The uneven quality of normative data for neuropsychological tests is a problem that has been recognized and, to some extent, redressed in recent years (e.g., Heaton, Grant, & Matthews, 1991). However, if normative neuropsychological test scores tend to rise substantially over time, either norms will have to be updated frequently or else old norms will have to be adjusted mathematically for the passage of time between collection of the normative data and collection of the clinical data. It should also be noted that new norms may underestimate the actual magnitude of the Flynn effect if those norms reflect the pooling of current and previous normative data, or if they have been accumulated over a protracted period.

The Flynn effect may be regarded by the clinical neuropsychologist as an unwelcome source of complexity that can make the interpretation of test scores—already a challenging undertaking—even more difficult. On the other hand, neuropsychological tests may hold clues that will lead to a better understanding of the Flynn effect. The search for rising scores on neuropsychological tests will reveal, at the very least, the scope of the Flynn effect. Is it restricted to intellectual tests, or does it affect various tests of perception, attention, processing speed, learning, memory, working memory, language, motor skill, and executive functions? If the Flynn effect does extend beyond the realm of IQ tests, as indicated by the recent findings of Connor et al. (2004) and Dickinson and Hiscock (2006), the next question would concern its differential impact on different kinds of tests. Does it influence some categories of tests (e.g., working memory) more than others (e.g., language)? Does it affect speeded performance more than performance on untimed tests? Does it affect broad-band tests more than tests of specific functions? Does it affect tests according to the strength of the association between test performance and IQ? Whereas a detailed analysis of the Flynn effect in neuropsychological testing might not lead to an immediate and complete explanation of the phenomenon, it will answer some questions of practical and theoretical importance.

Original manuscript received 11 May 2005
Revised manuscript accepted 15 May 2006

REFERENCES

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review, 75,* 81–95.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 363–405.

Boring, E. G. (1923, June 6). Intelligence as the tests test it. *The New Republic,* 35–37.

Bornstein, R. A. (1987). The WAIS–R in neuropsychological practice: Boon or bust? *The Clinical Neuropsychologist, 1,* 185–190.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250,* 223–250.

Burke, H. R. (1985). Raven's Progressive Matrices: More on norms reliability, and validity. *Journal of Clinical Psychology, 41,* 231–235.

Carlson, J. S., & Jensen, C. M. (1980). The factorial structure of the Raven Coloured Progressive Matrices Test: A reanalysis. *Educational and Psychological Measurement, 40,* 1111–1116.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431.

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: Harcourt, Brace & World.

Chelune, G. J., Eversole, C., Kane, M., & Talbott, R. (1987). WAIS versus WAIS–R subtest patterns: A problem of generalization. *The Clinical Neuropsychologist, 1,* 235–242.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: Belknap Press.

Costa, L. D. (1976). Interset variability on the Raven Colored Progressive Matrices as an indicator of specific ability deficit in brain-lesioned patients. *Cortex, 12,* 3–40.

Cozby, P. C. (2003). *Methods in behavioral research* (8th ed.). Boston: McGraw-Hill.

Diaz-Asper, C. M., Schretlen, D. J., & Pearlson, G. D. (2004). How well does IQ predict neuropsychological test performance in normal adults? *Journal of the International Neuropsychological Society, 10,* 82–90.

Dickens, W. T., & Flynn, J. R. (2001a). Great leap forward: A new theory of intelligence. *New Scientist, 21,* 44–47.

Dickens, W. T., & Flynn, J. R. (2001b). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108,* 346–369.

Dickinson, M. D., & Hiscock, M. (2005, February). *The influence of the Flynn effect on age norms for the Wechsler IQ test* [Abstract]. Poster session presented at the annual meeting of the International Neuropsychological Society, St. Louis, MO.

Dickinson, M. D., & Hiscock, M. (2006, February). *The Flynn effect in neuropsychological assessment* [Abstract]. Poster session presented at the annual meeting of the International Neuropsychological Society, Boston.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29–51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Flynn, J. R. (1994). IQ gains over time. In R. J. Sternberg (Ed.), *The encyclopedia of human intelligence* (pp. 617–623). New York: Macmillan.

Flynn, J. R. (1998a). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 25–66). Washington, DC: American Psychological Association.

Flynn, J. R. (1998b). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science, 30,* 541–553.

Flynn, J. R. (1998c). 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. (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 problem be solved? *Psychology, Public Policy, and Law, 6,* 191–198.

Flynn, J. R. (2003). Movies about intelligence: The limitations of *g*. *Current Directions in Psychological Science, 12,* 95–99.

Flynn, J. R. (2006a). Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn effect: Rethinking intelligence and what affects it]. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à psicologia das diferenças individuais* [Introduction to the psychology of individual differences] (pp. 387–411). Pôrto Alegre, Brasil: ArtMed. (English trans. available from jim.flynn@stonebow.otago.ac.nz).

Flynn, J. R. (2006b). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law, 12,* 170–189.

Flynn, J. R. (2006c). Towards a theory of intelligence beyond *g*. *Behavioral and Brain Sciences, 29,* 132–134.

Flynn, J. R., & Weiss, L. G. (2006). *American IQ gains from 1932 to 2002: The significance of the WISC subtests. Manuscript submitted for publication.*

Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin.* New York: Harmony Books.

Greenfield, P. M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81–123). Washington, DC: American Psychological Association.

Hanson, R., Smith, J. A., & Hume, W. (1985). Achievements of infants on items of the Griffiths scales: 1980 compared to 1950. *Child: Care, Health and Development, 11,* 91–104.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead–Reitan battery: Demographic corrections, research findings, and clinical applications.* Odessa, FL: Psychological Assessment Resources.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hiscock, M., Inch, R., & Gleason, A. (2002). Raven's Progressive Matrices performance in adults with traumatic brain injury. *Applied Neuropsychology, 9,* 129–138.

Holford, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual Review of Public Health, 12,* 425–457.

Jencks, C., & Phillips, M. (1998). *The Black–White test score gap.* Washington, DC: Brookings Institution Press.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.

Jensen, A. R. (1973). *Educability and group differences.* New York: Harper & Row.

Jensen, A. R. (1975). The meaning of heritability in the behavioral sciences. *Educational Psychologist, 11,* 171–183.

Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kanaya, T., Scullin, M. H., & Ceci, S. (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, E., Fein, D., Morris, R., & Delis, D. C. (1991). *WAIS–R as a neuropsychological instrument.* San Antonio, TX: The Psychological Corporation.

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Lezak, M. D. (1988). IQ: R.I.P. *Journal of Clinical and Experimental Neuropsychology, 10,* 351–361.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Llabre, M. M. (1984). Standard Progressive Matrices. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1). Kansas City, MO: Test Corporation of America.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology* (B. Haigh, Trans.) New York: Basic Books.

Lynn, R. (1990). the role of nutrition in the secular increases of intelligence. *Personality and Individual Differences, 11,* 273–286.

Lynn, R. (1998). In support of the nutrition theory. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 207–215). Washington, DC: American Psychological Association.

Martorell, R. (1998). Nutrition and the worldwide rise in IQ scores. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 183–206). Washington, DC: American Psychological Association.

Neisser, U. (1998a). Introduction: Rising test scores and what they mean. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 3–22). Washington, DC: American Psychological Association.

Neisser, U. (Ed.). (1998b). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Plomin, R. (1986). *Development, genetics, and psychology.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Plomin, R. (1990). The role of inheritance in behavior. *Science, 248,* 183–188.

Raven, J. C. (1938). *Progressive Matrices: A perceptual test of intelligence: Individual form.* London: H.K. Lewis.

Raven, J. C. (1960). *Guide to the Standard Progressive Matrices*. London: H.K. Lewis.

Raven, J., Raven, J. C., & Court, J. H. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section J. General overview.* Oxford, UK: Oxford Psychologists Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 1. General overview. Introducing parallel versions of the CPM and SPM together with a more powerful version of the SPM (SPM-Plus).* Oxford: Oxford Psychologists Press.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and clinical interpretation.* Tucson, AZ: Neuropsychology Press.

Reitan, R. M., & Wolfson, D. (1990). A consideration of the comparability of the WAIS and WAIS–R. *The Clinical Neuropsychologist, 4,* 80–85.

Rourke, B. P., Bakker, D. J., Fisk, J. L., & Strang, J. D. (1983). *Child neuropsychology: An introduction to theory, research, and clinical practice.* New York: Guilford.

Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development, 63,* 1–19.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49,* 304–313.

Sigman, M., & Whaley, S.E. (1998). The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 155–182). Washington, DC: American Psychological Association.

Snow, R. E., Kyllonen, C. P., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 47–103). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Spearman, C. S. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–209.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47,* 1–41.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist, 3,* 54–56.

Villardita, C. (1985). Raven's Colored Progressive Matrices and intellectual impairment in patients with focal brain damage. *Cortex, 21,* 627–634.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santos, M. de M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–123.

Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale–Third Edition.* San Antonio, TX: Psychological Corporation.

Williams, W. M. (1998). Are we raising smarter children today? School- and home- related influences on IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 125–154). Washington, DC: American Psychological Association.