IAP  AP 101

APPLIED PSYCHOMETRICS 101: THE FLYNN EFFECT SERIES

#6: What is the Flynn Effect?

Norm obsolescence is recognized in the intelligence testing literature as a potential source of error in global IQ scores. Psychological standards and assessment books recommend that assessment professionals use tests with the most current norms to minimize the possibility of norm obsolescence spuriously raising an individual's measured IQ. This phenomenon is typically referred to as the Flynn Effect. This report is the first in a series of brief reports the will define, explain, and summarize the scholarly consensus regarding the validity of the Flynn Effect. The series will conclude with an evaluation of the question whether a professional consensus has emerged regarding the practice of adjusting dated IQ test scores for the Flynn Effect, an issue of increasing debate in Atkins MR/ID capital punishment hearings.

Kevin S. McGrew, Ph.D.

Educational Psychologist

Director

Institute for Applied Psychometrics (IAP)

What is the Flynn Effect?

The *APA Dictionary of Psychology* defines the Flynn effect as:

> *a gradual rise of IQ level that has been observed since the time when records of IQ first were kept. Although the average IQ remains 100 due to periodic renorming of IQ tests, raw scores have been rising. These increases have been roughly 9 points per generation (i.e., 30 years). The gains have been unequally distributed across the different abilities, with fluid abilities showing substantially greater gains than crystallized abilities*

APA Dictionary of Psychology, American Psychological Association 382 (Gary R. Vandenbos ed., 2007).

In simple terms, psychologists and psychological measurement experts typically describe the Flynn effect as resulting from 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 inflated IQ scores, as the older test norms[1] are too easy (are obsolete) for individuals in contemporary society. This is one of the primary reasons why authors and publishers of IQ tests must provide "freshened" norms via the collection of new nationally representative sample data for intelligence test batteries at least every 10 years (the generally accepted rule of thumb in the IQ testing industry). See L.G. Weiss, Consideration on the Flynn Effect, 28(5) Journal of Psychoeducational Assessment (in press, 2010). If the new norms are not provided, individuals tested on IQ tests with outdated norms will typically obtain inflated IQ scores. A visual explanation of the Flynn effect is represented in the figure below. See K.J. Sanborn, S.D. Truscott, L. Phelps, & J.L. McDougal, Does the Flynn Effect differ by IQ level in samples of students classified as learning disabled?, 21(2) Journal of Psychoeducational Assessment 145-159 (2003).

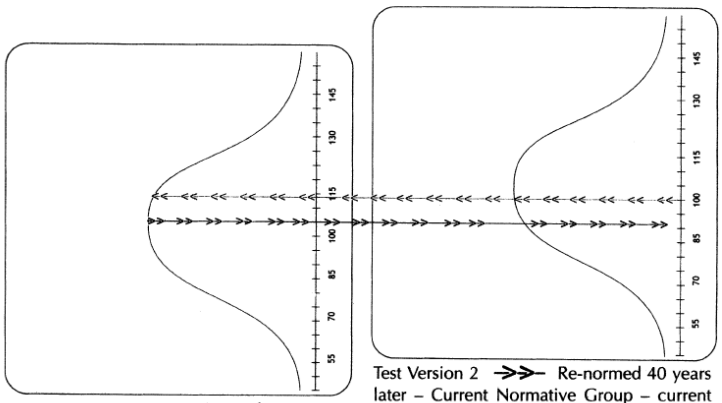


FIGURE 1. Visual representation of the Flynn Effect using two versions of a test, normed 40 years apart.

[1] As per the *APA Dictionary of Psychology*, a norm is "a standard 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" *APA Dictionary of Psychology*, American Psychological Association 631 (Gary R. Vandenbos ed., 2007).

The important Flynn effect concept illustrated in Figure 1 is that the normal distribution of intelligence has shifted upward over time (normal curve distribution on the right-hand side of the figure is 40 years newer than that on the right-hand side of the figure). As illustrated, the same raw score performance on an IQ test, when compared to the old (left-hand side of figure) norms will produce a markedly different IQ score when compared to the new (right-hand side of figure) contemporary sample of abilities for a person of a specific age. The person's tested performance (number of correct responses across all parts of the IQ test) did not change, but the reference or normative comparison group used to derive an estimate of the persons relative standing in the distribution of IQ scores does change as a function of which norm reference group the individual's performance is compared against. As illustrated by the top arrowed line, the same performance that is considered average in the contemporary norm sample (IQ score of 100 in the distribution on the right-hand side) will result in a much higher IQ score (in this scenario, a score of 112) when using the older, softer, easier norms. The 12 point difference is based on the accepted research-based Flynn effect of a change of approximately 3 IQ points for each decade that a tests' norms are older than the date when an individual is tested (3 x 40 years [4 decades] = 12 IQ points in this example).[2]

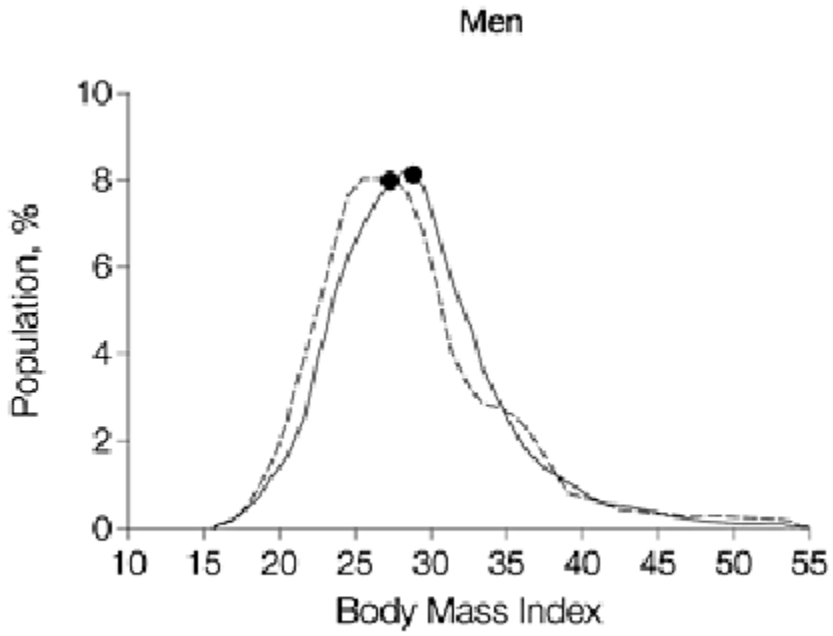
A non-IQ example

It is sometimes easier to grasp the nature of Flynn effect concept by considering a non-intelligence test example. Similar to the gathering of regular new IQ norm data, the U. S. *Center for Disease Control's (CDC) National Center on Health Statistics* (NCHS; http://www.cdc.gov/nchs/) conducts regular national health surveys to evaluate changes and trends in the health of the general population. One of these surveys is the *National Health and Nutrition Examination Survey (NHANES).* A recent empirical article published in the *Journal of the American Medical Association* compared nationally representative body mass index (BMI) data across an eight year period (normative data from the NHANES obtained in 2007-2008 was compared with normative data obtained from 1999 through 2006). See K.M. Flegal, M.D. Carroll, C.L. Ogden, & L.R. Curtin, Prevalence and trends in obesity among US adults, 1999-2008, 303(3) Journal of the American Medical Association 235-241 (2010).

A comparison of the national norm results for 40-59 year old males can be seen in Figure 2 below.[3] Similar to the hypothetical Flynn effect figure presented above, one can visually see a clear shift in what is considered average or normative body mass index (BMI) for males over this eight year period. The black circles at the top of the respective 1990-2000 (dashed line) and 2007-2008 (solid line) BMI distributions represents the average (median) body mass index for U.S. males at two different points in time in the U.S population. These average BMI points are analogous to the average IQ of 100 in different norms samples gathered at different times. As summarized by the studies' authors, there are clear changes in what is considered normative BMI (and thus, obesity) in 40-59 year old males over a period of eight years. More importantly in the current context, the shift in what is considered normal or normative has changed in the population, a shift that is conceptually identical to that represented by a shift in what is considered normative or the population average performance on intelligence tests (i.e., the Flynn effect).

[2] The basis and acceptance of the 3 pts/decade rule-of-thumb will be discussed in future issues in this series.

[3] National health statistics are reported separately by male and females given the clear differences in male and female normative growth patterns. The same is not the case for measuring intelligence. Thus, the use of national physical measurement data based on gender should be considered analogous to comparing a person's IQ test performance to the norms of intelligence across genders.



Flegal, K. M. et al. JAMA 2010;303:235-241.

Figure 2. Smoothed frequency distributions of body mass index (BMI) for men aged 40 to 59: Comparision of 1999-2000 and 2007-2008 national normative data from NHANES survey

Thus, when a doctor (today) evaluates a male between the ages 40 to 59 to ascertain whether the individual is obese or not, the most recent normative information should be used when making a decision and any subsequent medical recommendations for possible treatments. If a doctor (today) used the year 1990-2000 normative male (ages 40-59) BMI information for the basis of his or her conclusions and diagnosis, as well as any treatment or intervention prescriptions, such decisions would be based on incorrect and out-of- date normative comparisons to the person's current peers. The nationally representative BMI information summarized in Figure 2 is a clear example of how changes and shifts in a population must be taken into account when evaluating a person's obesity with regard to appropriate norm standards.

Comparing a 45 year old male's BMI index to 1999-2000 male BMI norms, and in turn deciding whether to prescribe or not prescribe a particular set of treatments, would be considered (by some) medical malpractice. If a doctor only had old BMI statistical graphs and charts (based on the 1999-2000 data), one would expect that a prudent and reasonable medical professional would recognize that his or her normative reference information is dated, and knowing the average BMI for 2007-2008, would either: (a) add a constant (the average difference in BMI for 40-59 year old males in 1999-2000 and 2007-2008) to the 1999-2000 BMI charts/figures at his or her disposal or, (b) subtract the same constant value from the patients current BMI and then evaluate that "adjusted" BMI against the BMI normative data from 1999-2000. This type of BMI effect adjustment is conceptually identical to the recommended Flynn effect adjustment when evaluating a person's IQ test performance that has dated norms.

For additional information

This report, and future reports in this series, draws from publications available at the *ICDP (Intellectual Competence and Death Penalty blog* – www.atkinsmrdeathpenalty.com) Atkins *MR/ID Flynn Effect Archive Project*.

Author information and conflict of interest disclosure

Dr. Kevin S. McGrew, Ph.D., is an Educational Psychologist with expertise and interests in applied psychometrics, intelligence theories and testing, human cognition, cognitive and non-cognitive individual difference variables impacting school learning, models of personal competence, conceptualization and measurement of adaptive behavior, measurement issues surrounding the assessment of individuals with disabilities, brain rhythm and mental timing research, and improving the use and understanding of psychological measurement and statistical information by professionals and the public. Prior to establishing IAP, Dr. McGrew was a practicing school psychologist for 12 years. McGrew received his Ph.D. in Educational Psychology (Special Education) from the University of Minnesota in 1989.

Dr. McGrew is currently Director of the *Institute for Applied Psychometrics* (IAP), a privately owned applied research organization established by McGrew. He is also the *Research Director for the Woodcock-Munoz Foundation* (WMF), Associate Director for *Measurement Learning Consultants* (MLC), and a *Visiting Professor in Educational Psychology* (School Psychology) at the University of Minnesota.

Dr. McGrew authored the current document in his role as the Director of IAP. The opinions and statements included in this report do not reflect or represent the opinions of WMF, MLC, or the University of Minnesota.

More complete professional information, including his professional resume, bio, and conflict of interest disclosures can be found at each of his three professional blogs and web page:

- www.iqscorner.com
- www.atkinsmrdeathpenalty.com
- www.ticktockbraintalk.blogspot.com
- www.iapsych.com