IAP  AP 101

APPLIED PSYCHOMETRICS 101: THE FLYNN EFFECT SERIES

#7: Is the Flynn Effect a Scientifically Accepted Fact?

This report is the second in a series of brief reports the will define, explain, and summarize the scholarly consensus regarding the validity of the Flynn Effect (FE). This brief report presents a summary of the majority of FE research (in tabular form of n=113 publications) which indicates (via a simple "vote tally" method) that despite no consensus regarding the possible causes of the FE, it is overwhelming recognized as a fact by the scientific community. 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), LLC

Is the Flynn Effect a Real and Scientifically Recognized and Accepted Fact?

Yes.

A review of the majority of the Flynn effect (FE) research literature (empirical research and commentaries) indicates that the Flynn effect (aka; norm obsolescence) is recognized as a scientific fact by the scientific community. This conclusion is based on the current author's review of the FE literature summarized in Table 1 (see Appendix A of this report). Table 1 is a comprehensive listing of the majority of the Flynn effect research studies and commentaries that have been presented in the professional literature.[1] As can be seen in Table 1, there are over 100 (*n* = 113) Flynn effect references spanning a period of 52 years (1958 to 2010).

As illustrated by Table 1, 113 Flynn effect manuscripts were reviewed as per their degree of support for the Flynn effect. References were either classified as:

- clearly presenting results and/or conclusions consistent with the Flynn effect ("yes" classification)
- clearly presenting results and/or conclusions against the validity of the Flynn effect ("no" classification)
- manuscripts that were somewhat mixed or neutral in their conclusions regarding the validity of the Flynn effect ("mixed" or "neutral")
- manuscripts deemed "not appropriate-NA" for inclusion in this classification analysis (i.e., reviews of books; research studies focused on other issues but which referenced the Flynn effect).[2]

According the classification of the 113 references, **86 (76.1%) were classified as affirmative ("yes")**, **9 (8.0%) classified as not supportive of the Flynn effect ("no"**), **4 (3.5%) were mixed or neutral, and 14 (12.4%) were not relevant or appropriate for the current analysis**. From this simple vote tally accounting method of the extant Flynn effect literature (which does not take into account the fact that some references should likely count more than others due to their larger sample sizes used and that some references report research summaries across multiple individual studies), it appears that one can conclude, with reasonable degree of scientific certainty, that the majority of the scientific and professional community that has written about the Flynn effect recognizes the validity of the Flynn effect in the context of norm-referenced intelligence testing.

The current author's conclusion is also summarized in an invited article published in a special issue on the Flynn effect in the Journal of Psychoeducational Assessment. My conclusion was:[3]

> *the consensus of most (but not all) intelligence scholars is that the Flynn Effect (FE), at the level of the global IQ score, is real. As a result, IQ test batteries are routinely re-standardized on a regular basis. Recently the FE focus has shifted. The publication of Flynn's What is intelligence? Beyond the Flynn Effect (2007) has directed attention towards developing a unified explanation of the causal mechanisms of the FE.*

K. McGrew, The Flynn Effect and Its Critics: Rusty Linchpins and "Lookin' for *g* and Gf in Some of the Wrong Places, 28(5) Journal of Psychoeducational Assessment (in press, 2010).

---

[1] Copies of almost all articles listed in Table 1 can be found on-line at the Atkins MR/ID Flynn Effect Archive Project originally created on January 11, 2010, and regularly updated by Dr. Kevin McGrew (http://www.iapsych.com/iqmr/fe/map.htm). The most recent revision is dated May 15, 2010.

[2] It is recognized that these classifications would benefit from inter-rater reliability analysis. Scholars interested in independently classifying all manuscripts in Table 1, as part of an inter-rater reliability classification study, should contact Kevin McGrew @ iap@earthlink.net

[3] This quote was from a final draft version of the manuscript and the quote may change in the final published versions of these article.

A review of the majority of comments presented in Table 1 by some of the leading scholars (and sometimes critics of the FE research—especially the interpretation of the causes of the FE) in intelligence testing (Alan Kaufman)[4], individuals involved in the standardization and publication in of the Wechsler batteries (Larry Weiss)[5], recognized intelligence theory and research methodologists (Joseph Rogers), and even the most vocal critics of the application of the Flynn effect in individual cases[6], indicates that **the scientific community recognizes the robustness of the Flynn effect in norm-referenced intelligence testing.** The following illustrative quotes capture the essence of the convergence of opinion regarding the validity of the Flynn effect.[7]

> *The Flynn effect (FE) is well known" and the "rate of increase in the United States has apparently remained a fairly constant 3 points per decade since the 1930s."* See *A. Kaufman, "In What Way Are Apples and Oranges Alike?*: A Critique of Flynn's Interpretation of the Flynn Effect, 28(5) Journal of Psychoeducational Assessment (in press, 2010).
>
> *The Flynn Effect (FE) is real" and "The FE has been shown to be near 3 points per decade on average across a large number of studies, countries, and tests.* See L.G. Weiss, Considerations on the Flynn Effect, 28(5) Journal of Psychoeducational Assessment (in press, 2010).
>
> *Even in the presence of a skeptical and critical scrutiny of the effect, it appears that there is more than just methodological artifact to be explained" and "Even with a healthy dose of skepticism, the effect rises above purely methodological interpretation, and appears to have substantive import*. J.L. Rodgers, A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both?, 26(4) Intelligence 354 (1999).
>
> *We do not seek to impugn or debunk the FE or its relevance to these cases; rather, our goal is to insist that those inclined to invoke this theory do so in a valid, responsible, and ethical manner*. L.D. Hagan, E.Y. Drogin, & T.J. Guilmette, IQ Scores Should Not Be Adjusted for the Flynn Effect in Capital Punishment Cases, 28(5) Journal of Psychoeducational Assessment (in press, 2010).

A complete reading of the information summarized in Table 1(and original manuscripts) 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, and methodological issues in various studies, the **inescapable conclusion is that the consensus of the scientific community is that the Flynn effect is real**. The robustness of this conclusion may best be represented by Joseph Rogers' statement where, after raising valid methodological issues regarding various statistical analysis and conclusions across Flynn effect studies, he concluded that even with a "*healthy dose of skepticism, the effect rises above purely methodological interpretation, and appears to have substantive import*." See J.L. Rodgers, A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both?, 26(4) Intelligence 354 (1999).

---

[4] See A. Kaufman, "In What Way Are Apples and Oranges Alike?": A Critique of Flynn's Interpretation of the Flynn Effect, 28(5) Journal of Psychoeducational Assessment (in press, 2010)).

[5] See L.G. Weiss, Considerations on the Flynn Effect, 28(5) Journal of Psychoeducational Assessment (in press, 2010).

[6] See L.D. Hagan, E.Y. Drogin, & T.J. Guilmette, Adjusting IQ scores for the Flynn effect: Consistent with the standard of practice?, 39 Professional Psychology: Research and Practice 619-625 (2008); L.D. Hagan, E.Y. Drogin, & T.J. Guilmette, IQ Scores Should Not Be Adjusted for the Flynn Effect in Capital Punishment Cases, 28(5) Journal of Psychoeducational Assessment (in press, 2010).

[7] These quotes are from final draft versions of these manuscripts and the quotes may change in the final published versions of these articles.

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

Appendix A

Table 1: Summary of Flynn effect (FE) research studies and commentaries: Classified by support for FE and relevant comments regarding the use of the FE adjustment to individual IQ scores

Reference[1]	Type of source	FE valid/real? (and countries of research)	Direct quotes illustrative of "yes/no" classification in preceding FE validity column [1]	Direct quotes of potential interest regarding the application of the FE adjustment to IQ test scores [2]
Fletcher, J., Stuebing, K., & Hughes, L. (2010). IQ scores should be corrected for the Flynn effect in high stakes decisions. *Journal of Psychoeducational Assesment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	These two approaches to estimating the mean and the precision of the effect support Flynn's aggregated estimate of the magnitude of norms obsolescence and are sufficiently precise to justify adjustments for high stakes decisions. There is variability across studies, and age/ability level, but this is true for any subject matter. The estimate of 3 +/- 1 is similar to the estimates for the conversion of WAIS III and WAIS IV scores for the middle of the distribution (where the sample size is larger) in Table 5.6 of the WAIS IV technical manual.	IQ test scores should be corrected for high stakes decisions in which a test with older norms is invoked as evidentiary support in the decision-making process. This could include not only Atkins cases involving capital offenses and the death penalty, but also intellectual disability (ID) decisions involving social security eligibility or special education where eligibility hinges on a specific score or range of scores. In all these contexts, the person may have previous IQ test scores that are higher than current scores, which may be reconciled by taking into account norms obsolescence. Adjusting an IQ score is not a violation of test administration. Rather, it is selecting an appropriate normative comparison (Gresham, 2009).

[1] Copies of almost all articles referenced can be found (and downloaded) at the following URL: http://www.iapsych.com/iqmr/fe/map.htm

[2] The quotes for the 2010 (in press) manuscripts in the *Journal of Psychoeducational Assessment* special issue on the Flynn effect are based on the final submitted manuscripts provided to Dr. Kevin McGrew by the editor of the journal, and not the final copy-edited versions. Thus, minor changes in wording (but not substantive changes) are possible in the final published articles. The exact quotes (with page #'s) will be inserted when the copy-edited "in press" versions are posted to the JPA website (then available for public access)

Hagan, L. D., Drogin, E. Y., & Guilmette, T. J. (2008*)*. IQ Scores Should Not Be Adjusted for the Flynn Effect in Capital Punishment Cases, *Journal of Psychoeducational Assessment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	We do not seek to impugn or debunk the FE or its relevance to these cases; rather, our goal is to insist that those inclined to invoke this theory do so in a valid, responsible, and ethical manner.	We conclude that the practice of altering an obtained IQ score based on the FE is insufficiently supported by scholarly literature or legal authority. We conclude that the practice of altering an obtained IQ score based on the FE is insufficiently supported by scholarly literature or legal authority. Altering obtained IQ scores based on the FE does not comport with the standard of forensic psychological practice, and there exists no legal mandate to make such adjustments. Psychologists serve an important function in capital punishment cases when they identify data limitations that may be attributable to the FE or any other error source. If an obtained score is considered to be invalid and if the "true" score is believed to be higher or lower within an estimated range, psychologists are justified in sharing this perspective in narrative form, but the current state of psychological science—particularly in light of the established variability of individual cases—does not support devising some other score based upon the FE and then substituting that score for the one obtained.
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),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	The aim of this paper is to challenge Flynn's (2007, 2009b) interpretation of the FE, not the existence of the effect itself, or even its magnitude. But the acceptance of the generational change in IQs applies only to global scores such as Wechsler Full Scale IQ. The Flynn effect is (FE) 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 (Flynn, 1984, 1987, 1998; The Psychological Corporation, 2003, 2008),	

			although that magnitude of steady gain pales in comparison to the gains observed in many other developed nations (e.g., 5-7 points in Belgium, The Netherlands, and Japan). The worldwide generational change has been observed most frequently using Wechsler's scales and Raven's (1938, 2000) matrices, but applies as well to data obtained on other tests such as the Stanford-Binet Intelligence Scale (Roid, 2003; Thorndike, Hagen, & Sattler, 1986) and the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983, 2004).	
Kaufman, A. (2010b). Looking through Flynn's rose-coloured scientific spectacles. *Journal of Psychoeducational Assessment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	But with the FE, in the U.S., a multitude of studies converge around 3 points per decade (Flynn, 2007, 2009b, this issue; Zhou et al., this issue) and the variability around 3 points is relatively minor (Fletcher et al., this issue).	Ultimately, I disagree with Flynn's explanation of the Flynn Effect, but I agree with his position that IQs should be adjusted for the effect in death-penalty cases. I respect the diversity of opinion on the topic of capital punishment, and the statistical complexity that surrounds the FE, but I am firmly in the camp with Reynolds et al. (this issue), Fletcher et al. (this issue), and Flynn (2006, 2007, 2009b) that IQs obtained on outdated norms should be adjusted for the FE in capital punishment cases. But for life-or-death decisions at the opposite end of the spectrum I believe the FE adjustment is warranted. Perhaps 3 points is a conservative estimate for individuals who score below 70, but what makes more sense if the FE is ultimately shown to be 4 or 5 points for low-functioning adults—to subtract 3 points or to subtract 0 points? Clinical judgments and court decisions will continue to be made based on the WAIS Full Scale IQ, with all its imperfections. And the FE—with all of *its* imperfections—cannot simply be dismissed.

Kaufman, A & Weiss , L. (2010). Guest Editors Introduction to the Special Issue of *JPA* on the Flynn Effect. *Journal of Psychoeducational Assessment*, *28 (5),* in press	Commentary/ synthesis: Research-based theoretical	Yes US	The Flynn Effect (FE) has apparently stopped, and even reversed, in some Scandinavian countries (Teasdale & Owen, 2008), but it has remained a virtual constant gain for children and adults in the U.S. for the past eight decades.	
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*), in press.	Empircal-based analysis	Yes US	The consensus of most (but not all) intelligence scholars is that the Flynn Effect (FE), at the level of the global IQ score, is real. As a result, IQ test batteries are routinely re-standardized on a regular basis. Recently the FE focus has shifted. The publication of Flynn's *What is intelligence? Beyond the Flynn Effect* (2007) has directed attention towards developing a unified explanation of the causal mechanisms of the FE. So, what do I believe about the FE? As a coauthor of a major IQ and achievement battery (WJ III) who was intimately involved in the statistical analysis and calculation of norms for the WJ-R and WJ III revisions, I have stared the reality of parts of the FE (on test norms) in the face on my computer screen. My most recent encounter was during the calculation of the WJ III NU (normative update), where the WJ III 2001 norms, originally calculated based on 1996 U.S. Census *projections* for year 2000, were recalculated with the *final* 2000 census *statistics* (see McGrew, Dailey & Schrank, 2007, for details), which resulted in a downward shift in global IQ scores of approximately 2.8 standard score points	
Reynolds, C., Niland, J., Wright, J., & Rosen, M. (2010). Failure to Apply the Flynn Correction in Death Penalty Litigation: Standard Practice of Today Maybe, but Certainly Malpractice of Tomorrow. *Journal*	Commentary/ synthesis: Research-based theoretical	Yes US	IQ tests are periodically revised and renormed to keep the content appropriate to current cultural contexts, ensure the representativeness of the normative or reference group (characteristics of the target population are constantly changing) and in order to maintain an average score of 100. The findings associated with these periodic revisions led researchers to observe that scores on standardized measures of intelligence have steadily risen over the last century, a phenomenon	Because of the central role IQ tests play in determining an individual's level of mental retardation, and because of the importance of mental retardation in determining a defendant's eligibility to be killed by the State, it is imperative that the FE, if it is real, be taken into account in capital cases. Since it is at this time a practical impossibility to renorm tests annually to

of Psychoeducational Assessment, 28 (5), in press.			termed the Flynn Effect (after James Flynn, the man who first documented these changes carefully and comprehensively, e. g., Flynn, 1984).	maintain a more appropriate reference group, to the extent corrections are available and valid, they should be applied to obtained scores so the most accurate estimate of standing possible is obtained. To do less is to do wrong—what possible justification could there be for issuing estimates of general intelligence in a death penalty case that are less than the most accurate estimates obtainable? The highest court in this country has made the determination that executing persons with mental retardation violates the Eighth Amendment's prohibition against cruel and unusual punishment. As a generally accepted scientific theory that could potentially make the difference between a constitutional and unconstitutional execution, the FE must be applied in the legal context.
Sternberg, R. (2010). The Flynn Effect: So What? *Journal of Psychoeducational Assessment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	In the end, it is not clear that we will find any one factor that causes the rise in IQs. Flynn's (2007) own interpretation of the rise as due to technological and educational enhancement seems plausible, but no explanation satisfactorily accounts for the uniformity of the rise, given that technology and education have not increased in quality and availability everywhere. Probably there are multiple interacting causes, including the increased complexity of the world and the increase in intelligence needed to adapt successfully to this world.	One might wonder why this matters, but in a capital case, where an IQ is being used as part of the evidence to determine whether a convicted killer has an IQ sufficient to understand the crime he or she has committed, misapplication of the FE to individual cases could have tragic results. The FE seems to apply in the aggregate, but it is extremely difficult to apply it in individual cases. With the stakes in such cases so high, can we really put enough faith into levels of IQ scores to draw sound conclusions? But the gist of my own essay is that the very use of IQ in such proceedings is ethically challenged, because such tests measure cognitive intelligence, not ethical intelligence, to be discussed later! Ethical intelligence—one's level of ethical reasoning and problem solving--may be more relevant than cognitive intelligence in such cases.

Weiss, L. G. (2010). Considerations on the Flynn Effect. *Journal of Psychoeducational Assessment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	The Flynn Effect (FE) is real. The FE has been shown to be near 3 IQ points per decade on average across a large number of studies, countries, and tests.	Still, the ethical position of an expert witness providing testimony is not to argue either for or against FE adjustments but to inform the court about the extant research on the topic. Adjustment of scores on obsolete tests for high stakes evaluations is controversial. The primary area of disagreement concerns the appropriateness of adjusting individual scores based on group data. There is no definition of when a test becomes obsolete. When asked privately, most FE researchers have ten years in mind.
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),* in press.	Empirical-based analysis	Yes US	The magnitude of the FE (as measured by change in PIQ scores) varies across ability groups, although the direction of the variation across ability groups is inconclusive. This finding is consistent for all Wechsler intelligence tests studied spanning the range from preschool to older adult. Overall, our findings suggest that the average IQ gain Flynn initially described may only be valid as an aggregated phenomenon.	
Flynn, J. R. (2010). Problems with IQ gains: The huge Vocabulary gap. *Journal of Psychoeducational Assessment, 28 (5),* in press.	Commentary/ synthesis: Research-based theoretical	Yes US	*[Note. This article is more correctly classified as an implicit "yes" for the FE as it did not investigate the FE per se, but consisted of Flynn countering arguments of Kaufman (2010a) at the level of the Wechsler subtests. Flynn's lengthy program of FE research clearly indicates his belief in the FE]*	A psychologist whose clinical judgment waivers because of a few IQ points is incompetent. But courts that use inflated IQs to kill people are morally remiss. Better the application of an approximate rule (Flynn, 2009a). Zhou et al's findings confirm that my rule (0.30 points per year) applies a conservative adjustment in the retardate range. The adjustment should also be made when law forbids benefits unless an IQ score of 70 or below is on record.
American Association on Intellectual and Developmental Disabilities (2010). Intellectual disability:	Commentary/ synthesis: Research-based	Yes General	As discussed in the *Users 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)	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

Definition, classification and systems of supports. Washington, DC: Author.	theoretical		are used in the assessment or interpretation of an IQ score. *[Note. The reference to the 10th edition above is correct, as this is a direct quote from the 11th edition that refers to the fact that the practice of adjusting for the Flynn effect was referenced in the 2007 Users Guide that accompanied the 10th edition]*	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.
Kanaya, T., Ceci, S. J (2010). The Flynn Effect in the WISC Subtests Among School Children Tested for Special Education Services. *Journal of Psychoeducational Assessment* OnlineFirst, published on May 19, 2010 as doi:10.1177/07342829 10370139	Empirical-based analysis	Yes US	The Flynn effect, a secular rise in IQ seen throughout the world, was examined on the WISC-R and WISC-III subtests in a longitudinal sample of more than 2,500 school children who were tested between 1974 and 2002. Multivariate analysis of variance and multiple regression analyses revealed that all the subtests experienced significant decreases in scores on the introduction of the WISC-III, as expected because of the Flynn effect…These findings add to the growing literature comparing the magnitude of the Flynn effect on crystallized versus fluid measures. Implications for special education testing and the current WISC-IV are discussed.	Although these findings may have the most immediate impact on school children, it is important to note that IQ tests play an important role in life-altering decisions for many people. For example, individuals who are diagnosed with mild MR because of their low IQ and concomitant behavioral measures do not qualify for the death penalty (*Atkins v. Virginia*, 2002), and can also be denied organ transplants (Shapiro, 2006). Full Scale IQ scores are used to determine Social Security Disability eligibility and occupational assignments, including decisions within the military, are also based on measures that are highly correlated with traditional IQ tests and other measures that have been found to experience the Flynn effect (Kanaya, Scullin, et al., 2003).
Lynn, R. & Meisenberg, G. (2010). The average IQ of sub-Saharan Africans: Comments on Wicherts, Dolan, and van der Maas. *Intelligence, 38*, 21–29	Empirical-based analysis	Yes US	In this section we summarize studies we consider sufficiently representative for inclusion. IQs are corrected for a Flynn effect of 3 IQ points a decade and calculated in relation to a British IQ of 100 (sd 15), and deduction of 2 IQ points from IQs calculated on American norms. These are described as FE/ British IQ adjusted. *[Note. This study is more correctly classified as an implicit "yes" for the FE as it did not investigate the FE per se, but in the data analysis made adjustment for the FE on the sample data (see prior column)]*	

Rushton, J. P. & Jensen, A. R. (2010). The rise and fall of the Flynn Effect as a reason to expect a narrowing of the Black–White IQ gap? *Intelligence*, 38 (2), 213-219.	Commentary/ synthesis: Research-based theoretical	NA Commentary on group IQ differences		
Russell, E. (2010). Commentary: The "obsolescence" of assessment procedures. *Applied Neuropsychology, 17*, 60-67	Commentary/ synthesis: Research-based theoretical	Yes (but time qualified) US	However, in the last few years, it is becoming evident that the influence of the Flynn Effect may be attenuating. A recent commentary (Russell, 2007) describes how studies compiled from several countries indicated that the Flynn Effect appears to have plateaued for these advanced Western countries. The few studies that have directly examined the Flynn Effect since Russell's study have all found evidence of this plateauing effect	
Sundet, J. M., Eriksen, W., Borren, I. & Tambs, K. (2010). The Flynn effect in sibships: Investigating the role of age differences between siblings. *Intelligence, 38*, 38–44.	Empirical-based analysis	Yes Norway	The aim of the study was to investigate the relationship between the Flynn effect and the effects of age differences between siblings on the intelligence difference between them. In Norway, the secular trends in intelligence-test score means vary both in magnitude and direction…Regression analyses showed that the Flynn effect can be quite well predicted from the effects of the age differences between brothers on their intelligence-test scores. We conclude that the factors causing the Flynn effect also work within sibships. Hypotheses positing that the Flynn effect is solely caused by between-families factors (e.g. the heterosis hypothesis) are weakened. The present results also entail that the birth order effect observed in Norway is in part conditional on the Flynn effect.	
Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2010). Raven's test performance of sub-Saharan Africans:	Empirical-based analysis	Yes (but geographically qualified) Sub-Saharan	This paper presents a systematic review of published data on the performance of sub-Saharan Africans on Raven's Progressive Matrices…Raven's scores among African adults have shown secular increases over the years. It is concluded that the Flynn Effect has yet to take hold	

Average level, psychometric properties, and the Flynn Effect. *Learning and Individual Differences, 20 (3),* 135-151		Africa	in sub-Saharan Africa. Although the implications of our psychometric findings for the potential of the Flynn Effect in sub-Saharan Africa remain unclear, the Raven's tests and other IQ tests have shown robust increases in many populations	
Wicherts, J. M., Dolan, C. V., & Van der Maas, H. L. J. (2010a). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence, 38,* 1-20.	Empirical-based analysis	NA Purpose was to establish average IQ of sub-Saharan Africans		
Wicherts, J. M., Dolan, C. V., & Van der Maas, H. L. J. (2010b). The dangers of unsystematic selection methods and the representativeness of 46 samples of African test-takers [rejoinder to Lynn & Meisenberg, 2010] *Intelligence, 38,* 30-37.	Empirical-based analysis	NA Purpose was to establish average IQ of sub-Saharan Africans		

Brouwers, S. A., VandeVijver, F. J. R., & VanHemert, D. A. (2009). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences, 19(3),* 330-338.	Empirical-based analysis	Yes 45 Countries (US included)	The paper describes a cross-cultural and historical meta-analysis of Raven's Progressive Matrices. Data were analyzed of 798 samples from 45 countries (N=244,316), which were published between 1944 and 2003. Country-level indicators of educational permeation (which involves a broad set of interrelated educational input and output factors that are strongly related to economic development), the samples' educational age, and publication year were all independently related to performance on Raven's matrices. Our data suggest that the Flynn effect can be found in high as well as low GNP countries, although its size is moderated by education-related sample and country characteristics and seems to be smaller in developed than in emerging countries.	
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.	Empirical-based analysis	Yes US	*Daubert* motions oppose adjusting IQ scores. They argue that the rate of IQ gains over time (the Flynn Effect) cannot be set at 0.3 points per year with scientific exactitude; therefore, the adjustment formula that rate implies is inadmissible in capital cases. This ignores the fact that there is universal agreement in the scientific community that there have been substantial gains and that, therefore, the worst possible option is to simply leave inflated IQ scores unadjusted. That would undermine equity entirely.	Ideally, we would know exactly how to adjust IQ scores for obsolescence. It is quite possible that the rate of gain on Wechsler tests is 0.275 or 0.325 points per year. But one thing we know for certain: IQ gains have not been nil. Unadjusted IQs presume that fiction. All of the evidence suggests that a rate of 0.30 is about right, and varying it from case to case lacks any rationale. Ideally, a mechanical application of the usual rules of evidence would promote equity. But in reality, *Daubert* memos, if successful, are certain to perpetuate a monstrous injustice by making the death penalty a lottery. Justice Stewart (*Furman v. Georgia,* 1972) deserves the last word: ‘‘These death sentences are cruel and unusual in the same way that being stuck by lightning is cruel and unusual.’’ He did not want to see them wantonly and freakishly imposed.
Gresham, F. (2009). Interpretation of Intelligence Test Scores in Atkins Cases: Conceptual	Commentary/ synthesis: Research-based	Yes US	It is well established that there has been a substantial increase in measured intelligence test performance over time because IQ test norms become obsolete. As such, intelligence test norms have to periodically be recalibrated to maintain	However, states use IQ scores which are inextricably and directly dependant on norms for their meaning. These scores often are rigidly adhered to by many states (e.g., Virginia) to determine a person’s

and Psychometric Issues. *Applied Neuropsychology, 16*, 91–97	theoretical		their accuracy in reflecting an individual's level of intelligence. The general upward trend in IQ scores has become known as the Flynn Effect, named after James Flynn who first documented this phenomenon (Flynn, 1984). Based on his extensive review of the literature, Flynn established that Americans gain approximately 0.3 IQ points per year or 3 points per decade in measured intelligence. Thus, an IQ test normed in 1972 would reflect a 10.8 point gain in IQ today (36_0.3¼10.8 points).	eligibility for the death penalty. The view that the Flynn Effect does not reflect real changes in intelligence is moot because the courts often use an absolute level of intelligence (IQ<70) to determine whether an individual is eligible for capital punishment.
Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence,37(1),*16-24	Empirical-based analysis	Yes US	Results of five studies show that during the second half of the twentieth century there were increases in the Development Quotients (DQs) of infants in the first two years of life. These gains were obtained for the Bayley Scales in the United States and Australia, and for the Griffiths Test in Britain. The average of 19 data points is a DQ gain of approximately 3.7 DQ points per decade. Similar gains of approximately 3.9 IQ points per decade have been present among preschool children aged 4–6 years. These gains are about the same as the IQ gains of school age students and adults on the Wechsler and Binet tests.	
Ronnlund, M., & Nilsson, L. G. (2009). Flynn effects on sub-factors of episodic and semantic memory: Parallel gains over time and the same set of determining factors. *Neuropsychologia, 47(11),* 2174-2180.	Empirical-based analysis	Yes Sweden	The study examined the extent to which time-related gains in cognitive performance, so-called Flynn effects, generalize across sub-factors of episodic memory (recall and recognition) and semantic memory (knowledge and fluency).We conducted time-sequential analyses of data drawn from the Betula prospective cohort study, involving four age-matched samples (35–80 years; *N*= 2996) tested on the same battery of memory tasks on either of four occasions (1989, 1995, 1999, and 2004). The results demonstrate substantial time-related improvements on recall and recognition as well as on fluency and knowledge, with a trend of larger gains on semantic as compared with episodic memory.	
Shayer, M., & Ginsburg, D. (2009). Thirty years on - a large anti-Flynn effect?	Empirical-based analysis	NO UK	The negative Flynn-effect found on Volume & Heaviness for Y7 pupils is paralleled by a similar negative effect on attainment of formal operations by Y8 and Y9, compared with 1976. Yet at the	

(II): 13-and 14-year-olds. Piagetian tests of formal operations norms 1976-2006/7. *British Journal of Educational Psychology, 79,* 409-418.			same time the proportion of pupils using the top level of concrete operational thinking has increased on both tests. It seems that there has been a change either in general societal pressures on the individual or in the style of teaching in schools – or both – favouring a lower level of processing of reality. These data suggest that discussion of the Flynn effect may require a deeper analysis of test data than just examining standardized scores of psychometric tests…Yet the proportions on both tests showing the higher level thinking of formal operations are down radically from what they were in 1976.	
Teasdale, T. W. (2009). The Danish Draft Board's intelligence test, BØrge Priens PrØrve: properties and research applications through 50 years. *Scandinavian Journal of Psychology, 50(6),* 633-638.	Empirical-based analysis	Yes (but time qualified) Denmark	The availability of the BPP test scores made Denmark one of the best equipped countries for studying the Flynn Effect, together with Norway where draft-board intelligence data for decades since the 1950s are also available (Sundet, Barlaug & Torjussen,2004). In an early study, Teasdale and Owen (1987b) showed that the effect had manifested itself in BPP scores which had shown sharp rises between the late 1950s and the 1980s. They also noted that the gains were greater at the lower ends of the test scores, with little change at the highest percentiles (Teasdale & Owen, 1989). More recently, it has been found that the BPP scores have plateaued with very little change since the late 1990s up to the present time (Teasdale & Owen, 2005, 2008).	
Ashton, M. (2008). Review of What is intelligence? Beyond the Flynn effect. *Personality and Individual Differences, 44(5*), 1289-1291.	Book review	NA	NA-Review of Flynn's book	

Beaujean, A. A., & Osterlind, S. J. (2008). Using item response theory to assess the Flynn Effect in the National Longitudinal Study of Youth 79 Children and Young Adults Data. *Intelligence, 36(5),* 455-463.	Empirical-based analysis	No US	Thus, for the data used in this study, the Flynn Effect appears to be largely the result of changing item properties instead of changes in cognitive ability.	
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: Elsevier.	Commentary/ synthesis: book Research-based theoretical	Yes US	Gains in IQ over time render test norms obsolete within a decade of publication of an intelligence test. The Flynn Effect refers to the well-documented improvement in performance on tests of intelligence that takes place over decades.	Obsolete norms inflate IQs and drain the pool of those eligible to be classified as having mental retardation (MR). As a result, many are missing the services they need and capital offenders are being executed who should qualify as mentally incompetent. Better to err on the side of caution than to sentence convicted offenders to death. As the reader knows, we have offered a formula to be used for the time being. Some might complain about the lack of precision of our formula, but this formula is clearly better than having no adjustment of any kind. Proceeding without an adjustment formula during a time of enhanced normative performance ensures that persons, perhaps many persons, who deserve the label of MR will fail to receive it.
Hagan, L. D., Drogin, E. Y., & Guilmette, T. J. (2008*).* Adjusting IQ scores for the Flynn effect: Consistent with the standard of practice? *Professional Psychology: Research and Practice, 39,* 619-	Commentary/ synthesis: Research-based theoretical	No US	Should psychologists adjust obtained IQ scores to accommodate the *Flynn effect* (J. R. Flynn, 1985)? The authors surveyed directors of doctoral training programs approved by the American Psychological Association and board-certified school psychologists and completed a systematic review of IQ test manuals, contemporary textbooks on IQ testing, federally regulated IQ testing protocols, and various sources of legal and ethical guidance. They confirmed in each instance that such	Three conclusions emerge. First and foremost, adjusting obtained scores and recalculating norm means on the basis of the FE do not represent the convention and custom in psychology. Adjusting obtained IQ scores for this purpose is not the standard of practice. Second, recalculating an individual's actual data likely violates standardization procedures and departs from training

625.			adjustments to IQ scores do not comport with prevailing standards of psychological practice. Results of IQ testing may be applied to a broad range of psycholegal issues, many of which cannot be anticipated. Psychologists assist examinees, courts, and other 3rd parties most effectively by administering and interpreting IQ tests in their intended fashion.	practices, prevailing canons, guidelines, most treatises, and test instructional manuals. The current accepted convention does not support subtracting IQ points in a way that departs from the requirements of the test manual. "Evaluators must also be aware that there is no agreed upon method for how diagnostic conclusions should be influenced by the Flynn effect" (Young et al., 2007, p. 176). Psychologists cannot conclude that adjusting scores is the generally accepted practice in evaluations for special education, parental rights termination, disability, or any other purpose.
Khaleefa, O., Abdelwahid, S. B., Abdulradi, F., & Lynn, R. (2008). The increase of intelligence in Sudan 1964-2006. *Personality and Individual Differences, 45(5*), 412-413.	Empirical-based analysis	Yes Sudan	Results are reported for mean IQs on the Draw-a-Man Test for 4–10 year olds in Sudan in 1964 and 2006. There was a gain of 12.2 IQ points over the 42 years, representing 2.9 IQ points a decade. On the basis of the present limited evidence, it can reasonably be concluded that IQ increases in economically undeveloped nations have been increasing during the last half century or so at about the same rate as the IQ increases in the economically developed nations.	
Ronnlund, M., & Nilsson, L. G. (2008). The magnitude, generality, and determinants of Flynn effects on forms of declarative memory and visuospatial ability: Time-sequential analyses of data from a Swedish cohort study. *Intelligence, 36(3),* 192-209.	Empirical-based analysis	Yes Sweden	To estimate Flynn effects (FEs) on forms of declarative memory (episodic, semantic) and visuospatial ability (Block Design) time-sequential analyses of data for Swedish adult samples (35–80 years) assessed on either of four occasions (1989, 1994, 1999, 2004; n=2995) were conducted. The results demonstrated cognitive gains across occasions, regardless of age, with no evidence of narrowing gender gaps. Across the entire range of birth cohorts (1909–1969) the estimated gain approached 1 SD unit.	

Sundet, J. M., Borren, I., & Tambs, K. (2008). The Flynn effect is partly caused by changing fertility patterns. *Intelligence, 36(3),* 183-191.	Empirical-based analysis	Yes Norway	The overall impression from the present results quite clearly indicate that the increase of the relative frequency of small families at the expense of larger ones alone accounts for a distinct, but modest part of the Flynn effect both in General Ability (Fig. 3) and in the subtests (Table 2). The implication seems to be that changing fertility patterns should be taken into account when searching for the causes of the Flynn effect. It is a matter of preference whether the mean changes due to proportion changes should be considered a cause of the Flynn effect, or whether they should be considered as aggregation artifacts.	
Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence, 36(2)*, 121-126.	Empirical-based analysis	Mixed Denmark	Scores on cognitive tests have been very widely reported to have increased through the decades of the last century, a generational phenomenon termed the 'Flynn Effect' since it was most comprehensively documented by James Flynn in the 1980's. There has, however, been very little evidence concerning any continuity of the effect specifically into the present century. We here report data from a population, namely young adult males in Denmark, showing that whereas there were modest increases between 1988 and 1998 in scores on a battery of four cognitive tests–these constituting a diminishing continuation of a trend documented back to the late 1950's–scores on all four tests declined between 1998 and 2003/2004. For two of the tests, levels fell to below those of 1988. Across all tests, the decrease in the 5/6 year period corresponds to approximately 1.5 IQ points, very close to the net gain between 1988 and 1998. The declines between 1998 and 2003/4 appeared amongst both men pursuing higher academic education and those not doing so.	
Wicherts, J. M. (2008). Review of What is intelligence? Beyond the Flynn effect. *Netherlands Journal of Psychology, 64(1)*, 41-43.	Book review	NA	NA-Review of Flynn's book	

Fitzgerald, S., Gray, N. S., & Snowden, R. J. (2007). A comparison of WAIS-R and WAIS-III in the lower IQ range: Implications for learning disability diagnosis. *Journal of Applied Research in Intellectual Disabilities*, *20*, 323-330.	Empirical-based analysis	Yes UK	Our results are consistent with the Flynn effect affecting IQ scores (Flynn 2000). According to Flynn (1984), the populations' IQ is increasing at roughly 0.3 IQ points per year. Therefore, with the 16 years between the norming of the WAIS-R and the WAIS-III, the decrease in IQ score would be expected to be 4.8 Full Scale IQ points. We found a decrease of 4.1 IQ points.	Intelligence quotient score alone, and as part of a formal diagnosis of learning *disability [note: In this UK-based article, a learning disability refers to the WHO term that references mental retardation-not SLD as in the US*], is often used by professionals to make decisions about the provision of services (e.g. whether someone be admitted to mental health services or specialized learning disability services). Similarly, they are used to determine issues such as 'fitness to plead' and subsequent disposal into criminal justice or mental health services. The finding of a four-point shift in IQ scores between these tests must be taken into account when considering any such decision.
Flynn, J. R. (2007a). *What is intelligence? Beyond the Flynn Effect*. New York: Cambridge University Press.	Commentary/ synthesis: book Research-based theoretical	Yes US and international	*[Note - This is Flynn's synthesis book that summarizes most of what he has written and hypothesized. Thus, it clearly is a "yes" in the preceding column.]*	
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.	Commentary/ synthesis: Professional commentary response to Moore (2006)	Yes US	Correcting IQ scores for norms obsolete at the time the subject took the test is simply a *special* case of any other correction we might make when we discovered we had used the wrong norms.	On 28 February 2006, the US Navy-Marine Corps Court of Criminal Appeals stated: "In determining whether an offender meets this definition (author: of MR) standardized IQ scores scaled by the SEM and the Flynn effect will be considered." Not adjusting scores puts offenders at the mercy of chance: whether someone was lucky enough to take a relatively current IQ test or unlucky enough to take a test with obsolete norms.
Flynn, J. R. & Weiss, L. (2007). American IQ gains from 1932 to 2002: The WISC subtests and educational progress.	Empirical-based analysis	Yes US	Recent data from 12 pairs of tests representing eight standardization samples show that American IQ gains have occurred at a rate of 0.308 points per year from 1972 to 2002. Linked with earlier IQ gains, Americans have gained about 22 points over the 70 years between 1932 and 2002. Comparing	

*International Journal of Testing, 7(*2), 209-224.			the new WISC-IV (2002) and the old WISC-III (1989) shows a difference of only 2.5 points. However, they have only five subtests in common when full scale IQ is calculated. If one simulates a comparison of the WISC-III and WISC-IV standardization samples on the 10 subtests of the WISC-III, IQ gains over the intervening 12.75 years were no less than 3.83 points, yielding a minimum estimate of 0.300 points per year.	
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.	Commentary/ synthesis: Professional commentary response to Moore (2006)	Yes US	Moore, along with many other practicing psychologists, clearly has a problem with the idea that IQ scores may need adjusting. However, the science of psychology has fundamentally accepted the reality of the Flynn effect (Nesisser, 1998), with leading intelligence researchers putting their main efforts into understanding rather than disputing the phenomenon.	More to the point, however, is that IQ trends in Scandinavia have zero relevance to American norms, and the basic point of an Atkins hearing is to see if an individual qualifies for the diagnosis of MR when compared to the US population, and not the population of Norway, Denmark or any other country.
Hagen, J. (2007). The label mental retardation involves more than an IQ score: A commentary on Kanaya and Ceci (2007). *Child Development Perspectives,1(1)*,60-61.	Commentary/ synthesis: Research-based theoretical	NA Commentary on usefulness of IQ as a construct and other constructs in MR Dx		
Hiscock, M. (2007). The Flynn effect and its relevance to neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 29(5),* 514-529.	Empirical-based analysis	Yes US	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.	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 he test has not been revised, and new norms have not been disseminated

			The Flynn effect does not imply that the test has become inappropriate. It implies that the norms have become inappropriate.	
Kanaya, T., Ceci, S. J (2007a). Are all IQ scores created equal? The differential costs of IQ cutoff scores for at-risk children. *Child Development Perspectives,1(1)*,52-56.	Commentary/ synthesis: Research-based theoretical	Neutral/mixed US	The Flynn effect occurs because as IQ test norms get older, people perform better on them, raising the mean IQ by several points within a matter of years. The reason for this rise is unclear.	It may seem disconcerting that we cannot offer viable solutions to resolve this problem. However, the first step toward a solution is to raise awareness of the problem, specifically, the importance of exploring the role of the Flynn effect for children who are at risk for academic failure. Regardless of the outcome to this dilemma, it is clear that researchers, policymakers, and practitioners will need to work together to find viable solutions to the impact of the Flynn effect on children. To continue along the current path is to make critical life course decisions that will be detrimental to the lives of at-risk children.
Kanaya, T., Ceci, S. J (2007b). Mental retardation diagnosis and the Flynn Effect: General intelligence, adaptive behavior, and context. *Child Development Perspectives,1(1)*,62-63.	Commentary/ synthesis: Research-based theoretical	Yes US	We argued that because of the Flynn effect and changing IQ norms, the IQ cutoff score of 70 is resulting in MR misdiagnoses and subsequent misallocations of resources for children who are at risk for academic failure.	Widaman (2007) endorses the idea of subtracting points from an individual's IQ to compensate for the Flynn effect (approximately 0.3 points per year), even though there is some evidence to suggest that the Flynn effect may not be uniform across all IQ levels, thus resulting in some added error. He is correct, however, about the limitations of the recent research (e.g., Kanaya, Ceci, & Scullin, 2005), and we find his argument persuasive that ''it may be better and fairer . . . to employ an adjustment that is presumably a little too much forsome and a little too little for others than to employ no adjustment at all.'' Although a uniform adjustment will be a better and fairer than no adjustment at all, it should only be regarded as a temporary solution.
Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends.	Empirical statistical simulation study	Yes General	IQ test scores have risen steadily across the industrialized world ever since such tests were first widely administered, a phenomenon known as the *Flynn effect*. Although the effect was documented more than 2 decades ago, there is currently no	

Psychological Review, 114(3), 806-829.			generally agreed-on explanation for it. The author argues that the phenomenon *heterosis* represents the most likely cause.	
Nijenhuis, J. T., & vanderFlier, H. (2007). The secular rise in IQs in the Netherlands: Is the Flynn effect on g? *Personality and Individual Differences, 43(5),* 1259-1265.	Empirical-based analysis	Yes Netherlands	IQ scores have been increasing over the last half century, a phenomenon known as the Flynn effect. In this study, we focused on the question to what extent these secular gains are on the g factor. Two IQ batteries: the Interest-School achievement-Intelligence Test (ISI) and the Groningen Final Examination Primary Education (GALO) yielded small and modest negative correlations between standardized gains and g loadings. As these studies employ large samples this suggests that the combined literature now shows a modest negative relationship between d (the secular change in test score) and g Previous studies of the Flynn effect made extensive use of the Wechsler tests and the various versions of Raven's Progressive Matrices. This study has now also shown clear Flynn effects for the ISI and the GALO.	
Resing, W. C. & Tunteler, E. (2007). Children becoming more intelligent: Can the Flynn effect be generalized to other child intelligence tests? *International Journal of Testing, 7(2),* 191-208.	Empirical-based analysis	Yes Netherlands	In this article, time effects on intelligence test scores have been investigated. In particular, we examined whether the "Flynn effect" is manifest in children from the middle and higher IQ distribution range, measured with a child intelligence test based on information processing principles—the Leiden Diagnostic Test. The test was administered to two independent groups of children—one tested recently, another tested 25 years ago. Analyses of the test scores showed that over time, children's IQ scores increased at least 15 points for those in the higher SES group and 8 points for those in the middle SES group.	In conclusion, it is clear that intelligence tests should be restandardized every 10 years. This is the case for both more traditional Wechsler tests and for tests like the LDT designed from an information processing perspective. Using tests with old test norms will lead to an overestimation of the intelligence of tested individuals and has therefore consequences for individual diagnostic interpretations and decisions of the diagnostician using the tests. This not only applies to individuals with test scores within the normal range of the intelligence test score distribution, but also for children with scores at both the lower and higher end of this distribution.
Rodgers, J. L., & Wanstrom, L. (2007). Identification of a Flynn effect in the NLSY: Moving from the center	Empirical-based analysis	Yes US	Our study uses a national sample of U.S. children to test for the Flynn Effect in PIAT-Math, PIAT-Reading Recognition, PIAT-Reading Comprehension, Digit Span, and PPVT. An effect of the predicted magnitude was observed for PIAT-	

to the boundaries. *Intelligence, 35(2),* 187-196.			Math when maternal IQ was controlled. This finding in a large representative sample with thousands of variables supports more careful evaluation of the Flynn Effect, in demographic, geographic, environmental, and biological domains. In the NLSY, we identified a consistent and significant Flynn Effect of approximately the predicted magnitude in PIAT-M, the ability domain for which it was predicted.	
Russell, E. (2007). Commentary: The Flynn effect revisited. *Applied Neuropsychology, 14(4),* 262-266	Empirical-based analysis	Yes (but time qualified) US	The Flynn Effect postulates that intelligence is increasing over time. However, as an environment becomes optimal, a plateau occurs when general growth becomes largely determined by genetics. There is evidence that such a plateau is occurring for intelligence in countries with optimal social environments. In the United States, examination of adult Wechsler test scores between normings indicates a reduction of the FSIQ increase such that average FSIQ would plateau about 2024…a plateau in the Flynn Effect would have been reached in 2004. In summary this study presents evidence that the Flynn Effect may not now exist in United States as well as in a number of the other advanced countries. Consequently there is no longer any justification for renorming an intelligence test based on an indefinite increase in intelligence. Intelligence measures would now be invariant overtime and the social consequences of the Flynn Effect would be eliminated	
Senechal, C., Larivee, S., Audy, P. & Engelbert, R. (2007). The Flynn effect and mental retardation. *Canadian Psychology/Psychologie Canadienne, 48(4),* 256-270.	Empirical-based analysis	Yes Canada	In 1984, Flynn pointed out a 3 to 5 point increase per decade in major IQ tests in the United States. To respect the golden standard of the 100 point average and 15 point standard deviation, it became necessary to regularly restandardize the American IQ tests. In the first part of this paper we tackle three subjects. The observation of the "Flynn effect," its probable causes and its possible limit. In the second part we show that the "Flynn effect" might present a misdiagnosis risk for mental retardation when professionals use an obsolete version of an IQ test. This caveat especially applies	…we show that the "Flynn effect" might present a misdiagnosis risk for mental retardation when professionals use an obsolete version of an IQ test. This caveat especially applies given that the greatest risk of misdiagnosis is for borderline cases (with scores from 70 to 85).

			given that the greatest risk of misdiagnosis is for borderline cases (with scores from 70 to 85). We also raise the possibility that a more complex environment due to the evolution of technologies might necessitate adjustment of the functional adaptability criteria for mental retardation.	
Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on - a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975-2003. *British Journal of Educational Psychology, 77*, 25-41.	Empirical-based analysis	NO UK	The idea that children leaving primary school are getting more and more intelligent and competent – whether it is viewed in terms of the Flynn effect, or in terms of government statistics on performance in Key Stage 2 SATS in mathematics and science – is put into question by these findings. Thus the large drops in competence by 11- to 12-year-olds on Volume & Heaviness between 1975/76 and 2000/01, and the continuing fast drop in the 3 years following represent an important and objective finding – that is free from any process of adaptation to changing circumstances – that needs to addressed in assessing the overall impact of primary schooling on children.	
Weiss, L. G. (2007). Response to Flynn. *WAIS-III Technical Report*. San Antonio: Harcourt Assessments.	Commentary/ synthesis: Research-based theoretical	NA Response to Flynn's criticism of WAIS-III norms		
Widaman, K. (2007). Stalking the roving IQ score cutoff: A commentary on Kanaya and Ceci (2007). *Child Development Perspectives,1(1)*,57-59.	Commentary/ synthesis: Research-based theoretical	Yes US	Determining whether an individual meets diagnostic criteria for any particular condition can be fraught with difficulties. These difficulties are probably no more contentious or momentous than in the diagnosis of mental retardation (MR) in children during the elementary or middle school years, as Kanaya and Ceci (2007) outlined. As Kanaya and Ceci (2007) pointed out, this diagnostic task is made substantially more difficult by the well-known Flynn effect (Flynn, 1984, 1987) the steady increase in mean IQ performance of approximately 3 points per decade by populations	If Flynn-effect adjustments can dampen improper score fluctuations due to aging norms and thereby smooth out the proportions of students receiving IQs of 70 or below, I think such adjustments should be used; in fact, it would be inappropriate not to do so. This ''quick fix'' is admittedly imperfect and should be monitored by continuing research to ensure that it has no unintended negative consequences. Nonetheless, use of Flynn-effect adjustments, however imperfect, is likely to have fewer negative consequences than would the failure to use such adjustments. Support for this position in no way

			around the world.	undermines the call by Kanaya and Ceci for further concerted research efforts on the vexing problem of roving cutoff scores.
Young, B., Boccaccini, M. T., Conroy, M. A., & Lawson, K. (2007). Four practical and conceptual assessment issues that evaluators should address in capital case mental retardation evaluations. *Professional Psychology: Research and Practice, 38,* 169-178.	Empirical-based analysis	Yes US	The Flynn effect refers to the finding that the normal population's performance on intelligence tests has improved over the past 70 years (Flynn, 1985). Specifically, younger cohorts of individuals obtain higher scores on standardized measures of intelligence than older cohorts. Because of this trend, new revisions of existing intelligence tests are more difficult than older versions, meaning that a higher level of ability is needed to obtain a specific score on a newer test compared with an older test. Therefore, individuals who may have scored slightly above 70 on intelligence tests with older norms would be expected to score lower when tested using more recently normed versions of the same test or any other recently revised test.	Mental health professionals who conduct capital case mental retardation evaluations should be aware of the Flynn effect and be prepared to discuss their diagnostic conclusions in light of the effect, even if the Flynn effect had no influence on their conclusions. Evaluators must also be aware that there is no agreed-upon method for how diagnostic conclusions should be influenced by the Flynn effect. In our view, the Flynn effect is best used as a framework for explaining the variation in test scores across administrations of different versions of the same test, but explaining the effect does not require reporting adjusted scores.
Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29(2),* 109+.	Commentary/ Synthesis: Research-based theoretical	Yes US	Although brain-imaging and psychometric findings present a striking convergence of evidence seemingly in support of a fluid cognitive basis for general intelligence, number of sources provide contravening evidence indicating that fluid skills cannot be g. Most prominent, perhaps, is the rapid secular rise in IQ over the past century known as the Flynn effect (Flynn 1984; 1987; 1999).	
Bradmetz, J. & Mathy, F. (2006). An estimate of the Flynn Effect: Changes in IQ and subtest gains of 10-yr-old French children between 1965 and 1988. Psychological Reports, 99(3), 743-746.	Empirical-based analysis	Yes French	The increase in ECNI IQ scores between 1965 and 1988 for a sample of 8,640 French 10-yr.-old children conforms to the expectation of an increase in IQ of approximately 3 point per decade for an industrialised country on an intelligence test comparable with the WISC.	

Breslau, N., Dickens, W. T., Flynn, J. R., Peterson, E. L., & Lucia, V. C. (2006). Low birth weight and social disadvantage: Tracking their relationship with children's IQ during the period of school attendance. *Intelligence, 34(4),* 351-362.	Empirical-based analysis	Yes US	Fortunately, the Psychological Corporation gave 184 17-year olds both the WISC-III and WAIS-III in counterbalanced order; the results show that the latter gave full scale IQs 0.7 points higher than the former (Flynn, 1998, p. 1237). Moreover, the standardization of the WISC-IV in 2002 shows that the WISCIII's norms were becoming obsolescent at the familiar rate of 0.3 points per year (Flynn & Weiss, submitted for publication). Therefore, the WAIS-III scores must be reduced by 4.3 IQ points: 0.7 points to equate them with the WISC-III; and 3.6 points to allow for the WISC-III's obsolescence. *[Note. This study is more correctly classified as an implicit "yes" for the FE as it did not investigate the FE per se, but in the data analysis made adjustment for the FE on the sample data (see prior column)]*	
Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17(10),* 913-920.	Empirical-based analysis	Yes US	It is often asserted that Black Americans have made no IQ gains on White Americans. Until recently, there have been no adequate data to measure trends in Black IQ. We analyzed data from nine standardization samples for four major tests of cognitive ability. These data suggest that Blacks gained 4 to 7 IQ points on non-Hispanic Whites between 1972 and 2002. Gains have been fairly uniform across the entire range of Black cognitive ability.	
Dickens, W. T., & Flynn, J. R. (2006b). Common ground and differences. *Psychological Science, 17(10),* 923-924.	Commentary/ synthesis: Research-based theoretical	NA Commentary on race IQ differences		
Flynn, J. R. (2006a). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law, 12,* 170-189.	Empirical-based analysis	Yes US	The Flynn effect (IQ gains over time) means that different IQ tests will give different scores purely as a result of when the tests were normed.	Capital offenders cannot be executed if they are mentally retarded. Therefore, the IQ scores of offenders are important, and the U.S. 4th Circuit Court of Appeals has held that the Flynn effect is relevant to interpreting their IQ scores. The Flynn effect (IQ gains over time) means that different IQ tests will give different scores purely as a result of when the tests were

				normed. Because execution must not be a random result of what test defendants take, a formula is provided to convert IQ scores to a common metric: the norms current at the time the test was taken. The formula also includes a correction based on evidence that the Wechsler Adult Intelligence Scale—Third Edition inflates IQs because of sampling error. Given the inevitability that opposing experts will offer conflicting diagnoses, IQ scores merit special attention in capital cases.
Flynn, J. R. (2006b). Towards a theory of intelligence beyond g. *Behavioral and Brain Sciences, 29(2),* 132+.	Commentary/ synthesis: Research-based theoretical	NA Commentary on brain physiology and general intelligence (g)		
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.	Commentary/synthesis: Research-based theoretical	Yes US	The “Flynn effect” has begun to attract great interest among psychologists as well as in the popular press. It is generally accepted as a real phenomenon, although some respected scholars…have expressed doubts about the usual explanations.	There is a strong case to be made for the practice of adjusting a test score obtained on an outdated test to better reflect how an individual would look on the same test with the changed norms certain to be in place if the test were published closer to the time of the testing. Use of the Flynn effect is a useful, and valid, method for increasing the likelihood that a psychologist will correctly diagnose MR in someone deserving of that label.
Juan-Espinosa, M., Cuevas, L., Escorial, S., & Garcia, L. F. (2006). The differentiation hypothesis and the Flynn effect. *Psicothema, 18(2),* 284-287.	Empirical-based analysis	Yes US	The Flynn effect represents a difference on ability-level between groups of the same age but different cohort. The ability-level differentiation hypothesis represents a difference on the relevance of cognitive factors between groups of high and low ability. Hence, it should be possible to imitate the ability-level differentiation effect by comparing groups of he same age but different cohort. The indifferentiation hypothesis represents no differences on the relevance of cognitive abilities in all age groups within the same cohort. The aim of	

			the present study is to test the relationships between these phenomena. For this purpose we analyzed the American standardization samples of the WISC, WISC-R and WISC-III. Results support the link between the Flynn effect and the differentiation hypothesis	
Moore, R. B. (2006). Letter to the Editor: Modification of individual's IQ scores is not acceptable professional practice. *Psychology in Mental Retardation and Developmental Disabilities*, *32*(3), 11-12.	Commentary/synthesis: Professional commentary response to Greenspan 2006)	No US	In summary, the "Flynn effect" has received a great deal of attention in the post-Atkins era as a means for adjusting obtained IQ scores downwards. However, examination of the literature related to the phenomenon reveals that it lacks the empirical precision to be applied to individual scores. Furthermore, it is not accepted professional practice to adjust or modify individual scores based on the "Flynn effect" or any other phenomenon. If there are factors that lead the psychologist to believe that the scores do not represent an accurate or reliable measure of the individual's functioning, such issues are delineated in the discussion and interpretation of the scores; the scores themselves are not changed. Modification of individual scores is not accepted professional practice, for good reason, and should not be introduced into the court as such.	In sum, the "Flynn effect correction" lacks the empirical precision for altering individual IQ scores. Furthermore, it would be inappropriate to select particular effects to "correct" for while ignoring other more robust findings or larger magnitude effects in an effort to meet a particular social or legal agenda. In addition to looking at the scientific or clinical appropriateness of making score adjustments based on the "Flynn effect," it is important to also look at the professional and legal acceptance of the effect.
Rushton, J. P., & Jensen, A. R. (2006). The totality of available evidence shows the race IQ gap still remains. *Psychological Science, 17(10),* 921-922.	Commentary/ synthesis: Research-based theoretical	NA Commentary on group IQ differences		
Scullin, M. H. (2006). Large state-level fluctuations in mental retardation classifications related to introduction of renormed intelligence test. *American Journal of Mental Retardation, 111,* 322-335.	Empirical-based analysis	Yes US	It is now well-established that IQs on the Wechsler series of intelligence tests rose steadily throughout the 20th century at the rate of about .3 IQ points per year, although this rise has been masked by the periodic renorming of intelligence tests to reset the mean to 100 (Flynn, 1984, 1987, 1998). This IQ rise has been dubbed the *Flynn effect* after James Flynn, who has extensively documented this phenomenon.	However, the potential under classification of mental retardation due to dated test norms and other reasons discussed in the introduction raises an important question about the extent to which children and adults (especially urban dwellers and minorities) who should be receiving government SSI and disability insurance assistance for mental retardation are not benefiting from these services or are even aware that they might qualify for these

				services. In a recent decision, the United States Circuit Court of Appeals for the Fourth District held that a lower court must take into consideration the persuasiveness of evidence that the Flynn effect had an impact on the appellant's sole surviving developmental period WISC-R score, because under Virginia law a relevant issue was whether the appellant scored two *SD*s below the mean on an intelligence test (*Walker v. True,* 2005). Understanding the impact of the Flynn effect on IQs is especially relevant for death penalty cases in which the burden of providing the evidence for mental retardation falls on the defense.
Voracek, M. (2006). Phlogiston, fluid intelligence, and the Lynn-Flynn effect. *Behavioral and Brain Sciences, 29(2),* 142+.	Commentary/ synthesis: Research-based theoretical	NA Commentary on brain physiology and general intelligence (g)		
Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence, 33(1),* 93-106.	Commentary/ synthesis: Research-based theoretical	Yes US	This paper proposes one potential explanation for 100 years of rising population mean IQ in the United States associated with historical changes in access to schooling and educational practice	
Colom, R., Lluis-Font, J., & Andres-Pueyo, A (2005). The generational	Empirical-based analysis	Yes	Generational intelligence gains are one intriguing finding in science. Nutrition and cognitive stimulation are among the most remarkable causes of the upward trend in intelligence. The nutrition	

intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence, 33 (1),* 83-91		Spain	hypothesis predicts a primary impact on the most deprived, producing disproportionate gains at low intelligence levels. The cognitive stimulation hypothesis predicts gains along the intelligence distribution. In summary, the evidence reported in the present paper seems to be more consistent with the nutritional hypothesis than with the cognitive stimulation hypothesis.	
Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences, 39(4),* 837-843.	Empirical-based analysis	Yes (but time qualified) Denmark	In the 1980s reviewed evidence indicated that, through the preceding decades of the last century, population performance on intelligence tests had been rising substantially, typically about 3–5 IQ points per decade, in developed countries. The phenomenon, now termed the "Flynn Effect", has been variously attributed to biological and/or to social and educational factors. Although there is some evidence to suggest a slowing of the effect through the 1990s, only little evidence, to our knowledge, has yet been presented to show an arrest or reversal of the trend. Substantially replicating a recent report from Norway, we here report intelligence test results from over 500,000 young Danish men, tested between 1959 and 2004, showing that performance peaked in the late 1990s, and has since declined moderately to pre-1991 levels. [Note – emphasis added]	
Nettelbeck, T., & Wilson, C. (2004). The Flynn effect: Smarter not faster. *Intelligence, 32(1),* 85-93.	Empirical-based analysis	Yes Australia	Flynn (1999) has clearly documented widespread rises in mean IQs from substantial samples from some 20 nations representing western European/North American cultures or technologies. These increases in mean IQ, apparently without changes in variance, are presumed to be caused by environmental influences as yet unidentified. The Flynn effect has nothing to do with speed of processing as measured by IT, despite the effect being strongest for ability tests that earn bonus scores for quick performance.	

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*, 349-362.	Empirical-based analysis	Yes (but time qualified) Norway	The present paper reports secular trends in the mean scores of a language, mathematics, and a Raven-like test together with a combined general ability (GA) score among Norwegian (male) conscripts tested from the mid 1950s to 2002 (birth cohorts c1935–1984). Secular gains in standing height (indicating improved nutrition and health care) were also investigated. Substantial gains in GA were apparent from the mid 1950s (test years) to the end 1960s–early 1970s, followed by a decreasing gain rate and a complete stop from the mid 1990s. The gains seemed to be mainly caused by decreasing prevalence of low scorers. From the early 1970s, the secular gains in GA were almost exclusively driven by gains on the Raven-like test. However, even the means on this particular test stopped to increase after the mid to late 1990s. It is concluded that the Flynn effect may have come to an end in Norway. It has long been known among intelligence test users that test performance improves from one generation to the next (the Flynn effect), necessitating new and stricter norms from time to time.	
Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., vanBaal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32(5),* 509-537.		Neutral/ mixed Netherlands	Stated otherwise, we wished to establish whether the Flynn effect is characterized by factorial invariance. To this end, we conducted five studies comprising a broad array of intelligence tests and samples. The results of the MGCFAs indicated that the present intelligence tests are not factorially invariant with respect to cohort. This implies that the gains in intelligence test scores are not simply manifestations of increases in the constructs that the tests purport to measure (i.e., the common factors). The overall conclusion of the present paper is that factorial invariance with respect to cohorts is not tenable. Clearly, this finding requires replication in other data sets. However, if this finding proves to be consistent, it should have implications for explanations of the Flynn effect. The fact that the	

			gains cannot be explained solely by increases at the level of the latent variables (common factors), which IQ tests purport to measure, should not sit well with explanations that appeal solely to changes at the level of the latent variables.	
Bocerean, C., Fischer, J. P., & Flieller, A. (2003). Long-term comparison (1921-2001) of numerical knowledge in three to five-and-a-half year-old children. *European Journal of Psychology of Education , 18(4)*, 405-424.	Empirical-based analysis	Yes French	This rise was expected, insofar as the Flynn effect is a general phenomenon that has been observed in a variety of cognitive tasks (the Flynn effect was found again recently on Piaget's formal reasoning). Everything points to the conclusion that the rise in performance observed here corresponds to a genuine increase in the numerical knowledge of children between the ages of three and five.	
Ceci, S.J., Scullin, M. & Kanaya, T. (2003). The difficulty of basing death penalty eligibility on IQ cutoff scores for Mental Retardation. *Ethics & Behavior, 13(1),* 11-17.	Empirical-based analysis	Yes US	As hypothesized, we discovered that there was a significantly greater decrement in IQ test scores when children were retested on a revised IQ test than when they were retested on an IQ test with the same norms. The median decrement between the WISC–R and WISC–III retestings was 5 IQ points, and the median decrement between the WISC and WISC–R retestings was 4.5 points. Ironically, these documented changes in IQ scores that are associated with changing norms have resulted in dramatic fluctuations in the numbers of persons deemed eligible for MR diagnosis	This form of IQ yo-yoing is the result of changing IQ test norms, and it occurs periodically, involving hundreds of thousands of Americans. As we argue here, it has profound implications for the entire society, particularly in the way the criminal justice system handles and classifies mentally retarded capital murder defendants.
Cocodia, E. A., Kim, J. S., Shin, H. S., Kim, J. W., Ee, J., Wee, M. S. W., & Howard, R. W. (2003). Evidence that rising population intelligence is impacting in formal education. *Personality and Individual Differences, 35(4),* 797-810.	Empirical-based analysis	Yes Australia Singapore Korea	Consensus is growing that rising IQ scores at least partly reflect rising population intelligence. Most Australian primary teachers did not perceive brighter children even over 30 years, but most in the two Asian nations did, particularly those in Singapore. General intelligence may have largely stopped rising in Western nations decades ago while visuospatial ability has been increasing, which with test sophistication has been pushing up IQ scores slightly. When the environmental improvements occur rapidly, teachers readily note brighter children.	

Flynn, J. R. (2003). Movies about intelligence: The limitations of g. *Current Directions in Psychological Science, 12(3),* 95-99.	Commentary/ synthesis: Research-based theoretical	Yes US	Even if I am mistaken in detail, sociological factors of some sort were the dominant cause of U.S. IQ gains.	
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.	Empirical-based analysis	Yes US	Using IQ records from 10 sites around the country, we examined longitudinal trends in IQ among mental retardation (MR) students during the Wechsler Intelligence Scale for Children (WISC) to WISC-Revised (R) transition in the 1970s and learning disability (LD) students during the WISC-R to WISC-3rd Revision (III) transition in the 1990s. Based on the work by Flynn [American Journal of Mental Deficiency 90 (1985) 236; Psychology, Public Policy, and Law 6 (2000) 191], we predicted a rise in IQ followed by a fall each time a newly normed IQ test is introduced into the schools and used to diagnose children as MR or LD. As expected, we found that mean IQs of MR and LD students followed a nearly identical parabolic trajectory, and differed depending on what year they were tested. IQs from the older norms were higher than IQs from the newly introduced norms. This systematic shift had a significant impact on the likelihood of a MR diagnosis. The broader implications of this pattern are discussed.	
Kanaya, T., Scullin, M. H., & Ceci, S. J. (2003). The Flynn effect and US policies: The impact of rising IQ scores on American Society via mental retardation diagnoses. *American Psychologist, 58(10),* 778-790.	Empirical-based analysis	Yes US	Over the last century, IQ scores have been steadily rising, a phenomenon dubbed the Flynn effect. Because of the Flynn effect, IQ tests are periodically renormed, making them harder. Given that eligibility for mental retardation (MR) services relies heavily on IQ scores, renormed tests could have a significant impact on MR placements. In longitudinal IQ records from 9 sites around the country, students in the borderline and mild MR range lost an average of 5.6 points when retested on a renormed test and were more likely to be classified MR compared with peers retested on the same test.	The magnitude of the effect is large and affects national policies on education, social security, the death penalty, and the military.

Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence, 31(5*), 461-471.	Empirical-based analysis	Yes US	This study investigated the secular rise in IQ scores over a 60-year period in 12- to 14-year-old Estonian schoolchildren. In 1934/1936, Juhan Tork adapted the U.S. National Intelligence Test for Estonia and administered it to more than 6000 schoolchildren. We administered the same test to 449 students in 1997/1998 and compared the results of 381 of these with a carefully matched sample of 307 from the testing in the 1930s. We found a rise of nearly 1 S.D. on subtests using basic information processing algorithms such as Comparison and Symbol–Number, but only 0.50 S.D.s on verbal subtests such as Sentence Completion and Concept Comprehension…As such, these results supported Rushton's [Pers. Individ. Differ. 26 (1999) 381] finding that the secular rise over time is not occurring on the g factor. In Estonia, the Flynn effect is not a Jensen effect	
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.	Empirical-based analysis	Yes US	The study investigated whether the FE was evident in two LD samples using two different measures of cognitive ability, and, whether the FE differentially affected scores of students at different IQ levels. Even though results for Sample 1 were not significant, this research presents clear evidence that the FE affects at least some IQ scores for students in the LD population.	Regardless of whether IQ gains as reported by Flynn (e.g., 1999) and the current study reflect actual increases in the population's cognitive abilities, increases in the normative performance in IQ tests over time mean that IQ tests must be standardized frequently. Otherwise, individuals whose test performances are scored against outdated, obsolete norms will obtain inflated IQ scores. Use of such obsolete norms will interfere with the accurate diagnosis of children thought to have a learning disability by making it more likely that the child will qualify as LD and less likely that the child will score in the mentally retarded ranges.

Gobet, F., Campitelli, G., & Waters, A. J. (2002). Rise of human intelligence: Comments on Howard (1999). *Intelligence, 30(4),* 303-311.	Empirical-based analysis	No US & International	Based upon the evidence that the best chess players in the world are becoming increasingly represented by relatively young individuals, Howard [Intelligence 27 (1999) 235–250.] claimed that human intelligence is rising over generations. We suggest that this explanation has several difficulties and show that alternative explanations relating to changes in the chess environment, including increased access to chess knowledge, offer better explanations for the increased presence of young players at top-level chess.	
Colom, R., Juan-Espinosa, M., & Garcia, L. F. (2001). The secular increase in test scores is a "Jensen effect". *Personality and Individual Differences, 30(4),* 553-559.	Empirical-based analysis	Yes Spain	The results in the present study are indicative of the fact that secular changes in fluid test scores are within the domain of the so-called ``Jensen effects'' (Rushton, 1998). Clearly, there is a positive correlation between a fluid g and the secular changes in several cognitive tests. More specifically, a fluid g is positively correlated with the secular change in the observed test scores at the successive Spanish standardizations of the DAT battery.	
Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108(2),* 346-369.	Empirical statistical simulation study	Yes General	Some argue that the high heritability of IQ renders purely environmental explanations for large IQ differences between groups implausible. Yet, large environmentally induced IQ gains between generations suggest an important role for environment in shaping IQ. The authors present a formal model of the process determining IQ in which people's IQs are affected by both environment and genes, but in which their environments are matched to their IQs. The authors show how such a model allows very large effects for environment, even incorporating the highest estimates of heritability.	

Howard, R. W. (2001). Searching the real world for signs of rising population intelligence. *Personality and Individual Differences, 30(6),* 1039-1058.	Empirical-based analysis	Yes US & international	The present study looked at several real world indicators for evidence of a rise….All these trends have various possible causes other than rising intelligence. However, together, and with other recent empirical evidence, most indicators suggest population intelligence really could be rising.	
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.	Empirical-based analysis	Yes US	This research examined two samples of students classified as learning disabled (LD) for evidence of the phenomenon known as the Flynn effect (FE; Flynn, 1999)…. A secondary analysis examined potential differences in the FE by ethnicity and/or gender. Results indicate that the FE does affect Wechsler IQ and component scores of students classified as LD. Further, the effect varies by task. There were no significant differences in the FE by race and/or gender. The study suggests that LD classification may be substantially impacted by the FE over the life of an IQ test version.	This research clearly supports Flynn's (1999) contention that the FE has significant ramifications for the identification of students for special education, particularly regarding the LD classification. 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.

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.	Commentary/ synthesis: Research-based theoretical	Yes US & international	Evidence has put the phenomenon of massive IQ gains over time beyond doubt; data from 20 nations show not a single exception.	If IQ tests are to play a role of minimal respectability, researchers will have to renorm them every 7 years. Nonetheless, 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.
Kane, H. D. (2000). A secular decline in Spearman's g: evidence from the WAIS, WAIS-R and WAIS-III. *Personality and Individual Differences, 29(3),* 561-566.	Empirical-based analysis	Yes (but time qualified) US	It has also been well documented (e.g. Flynn, 1987) that the mean IQ of the populations of a number of nations has increased by approximately 15 IQ points over the last few decades. This increase has occurred conjointly with economic development. A reasonable inference from the `Flynn effect' is that the influence of Spearman's g should be falling concomitant with the increases in IQ. In other words, Spearman's law of diminishing returns should also occur on a secular level. From the post hoc analysis, a further conclusion may be drawn that the Flynn effect exerted its greatest influence between 1955 and 1981, the standardization dates of the WAIS and WAIS-R, respectively.	
Kane, H., & Oakland, T. D. (2000). Secular declines in Spearman's g: Some evidence from the United States. *Journal of Genetic Psychology, 161(3),* 337-345.	Empirical-based analysis	Yes US	On a related topic, the mean level of intelligence in a number of countries seemingly has increased significantly over the past decades, with the total accumulated gains usually on the order of 15 intelligence quotient (IQ) points.	

Silverstein, M. L., & Nelson, L. D. (2000). Clinical and research implications of revising psychological tests. *Psychological Assessment, 12(3),* 298-303.	Commentary/ synthesis: Research-based theoretical	Yes US	The Flynn effect refers to regularly observed increases in IQ scores over time. This cohort effect is usually observed in relation to intelligence, and it corresponds to higher levels of educational attainment across successive generations. It also has important bearing on virtually all tests of cognitive function, although it has not received much attention in the fields of cognitive or neuropsychological assessment as a concept, despite the periodic renorming of select tests that typically corrects for this type of cohort effect.	
Teasdale, T. W., & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence, 28 (2),* 115-120.	Empirical-based analysis	Yes (but time qualified) Denmark	Changes are shown in the distribution of scores on a set of tests used by the Danish draft board since the late 1950s until the present day. Whereas there were marked gains in earlier decades, especially in the lower end of the distribution, the last 10 years have only seen very modest gains. Such gains as have occurred appear primarily to manifest themselves in a test of visuo-spatial abilities.	
Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54(1*), 5-20.	Commentary/ synthesis: Research-based theoretical	NA Commentary on group IQ differences		
Howard, R. W. (1999). Preliminary real-world evidence that average human intelligence really is rising. *Intelligence, 27(3)*, 235-250.	Empirical-based analysis	Yes US & International	The overall results are consistent with the view that average human intelligence really is rising. The chess player age drop and its pattern are what would be expected from a rise and the real-world data cited by Flynn against the idea of a rise do not hold up well on close examination	
Rodgers, J. L. (1999). A critique of the Flynn Effect: Massive IQ gains, methodological artifacts, or both? *Intelligence, 26(4),* 337-356.	Commentary/ synthesis: Research-based theoretical	Yes US & international	The Flynn Effect proposed by Flynn (1984;1987) is reviewed and evaluated. Even in the presence of a skeptical and critical scrutiny of the effect, it appears that there is more than just methodological artifact to be explained. But the acceptance of the effect has been too quick. The proper explanations for the effect will not be meaningful until the nature of the effect is much better understood than it is now. Six questions	

			are raised that have not been adequately answered. Even with a healthy dose of skepticism, the effect rises above purely methodological interpretation, and appears to have substantive import. But its nature, meaning and causes are still far from being well understood.	
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.	Empirical-based analysis	Yes US	Flynn used data covering 1932 to 1972 to put U.S. gains at about 0.300 IQ points per year. For post-1972, comparison of WISC-R versus WISC-III particularly the data of Zimrnerrnan and Woo-Sam, gives a rate of 0.312. However, comparison of WAIS-R versus WAIS-III shows that the current rate may be as low as 0.171. It is hypothesized that the discrepancy may be due to sampling error and it is suggested that post-1972 US gains be put at about 0.25 points per year.	
Flynn, J. R. (1998a). IQ gains over time: Toward finding the cause. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 25-66). Washington, DC: American Psychological Association.	Commentary/ synthesis: Research-based theoretical	Yes US & 19 other nations	Data now are available for 20 nations, and there is not a single exception to the finding of massive IQ gains over time.	
Flynn, J. R. (1998b). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science*, *30*, 541-553.	Empirical-based analysis	Yes Israel	The 1976 to 1984 data allow a comparison of male gains with female gains on both tests. They are virtually identical, allowing for the fact that the adjusted percentage of females tested was higher in 1976 to 1978 than it was in 1982 to 1984. Females gained 0·2 points more on the Matrices, and 0·1 IQ points less on the Instructions test. Matrices gains for both sexes run well above Verbal IQ gains. The male gains on the Matrices (4·89 IQ points) and on the Instructions test (2·92) give a ratio of 1·675 to one; the female gains (5·09 and 2·82) give a ratio of 1·805 to one. This pattern is also in line with international data, although the ratios are not as high as some.	

Neisser, U (1998). *The rising curve: Long-term gains in IQ and related measures*. Washington, DC: American Psychological Association.	Commentary/ synthesis: book Research-based theoretical	Yes US and international	Scores on intelligence tests are *rising*, not falling; indeed, they have been going up steeply for years. This rapid rise is not confined to the United States; comparable gains have occurred all over the industrialized world. The gains are steady and systematic: Performance on IQ tests has been going up 3 points per decade ever since the first Stanford-Binet was normed in 1932.	
Loehlin, J. C. (1997). Dysgenesis and IQ: What evidence is relevant? *American Psychologist, 52(11),* 1236-1239.	Commentary/ synthesis: Research-based theoretical	NA Commentary on group IQ differences		
Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. American Psychologist, 51(2), 77-101	Commentary/synthesis: Research-based theoretical	Yes US & international	Perhaps the most striking of all environmental effects is the steady worldwide rise in intelligence test performance. Although many psychometricians had noted these gains, it was James Flynn (1984, 1987) who first described them systematically. His analysis shows that performance has been going up ever since testing began. The "Flynn effect" is now very well documented, not only in the United States but in many other technologically advanced countries. The average gain is about 3 points per decade—more than a full standard deviation since, say, 1940.	
Williams, W. M., & Ceci, S. J. (1997). Are Americans becoming more or less alike? Trends in race, class, and ability differences in intelligence. *American Psychologist, 52(11)*, 1226-1235.	Commentary/ synthesis: Research-based theoretical	NA Commentary on group IQ differences		
Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American*	Commentary/ synthesis: Research-	Mixed US	There is no doubt that the average scores have been rising somewhat. We cannot trust the Flynn effect to bails us out, given	

Psychologist, 52(7), 685-699.	based theoretical		that extrapolation from the observed trends proves to be dangerous.	
Spitz, H. H. (1989). Variations in Wechsler interscale IQ disparities at different levels of IQ. *Intelligence*, *13*, 157-167.	Empirical-based analysis.	Yes (but ability level qualified) US	The present survey reports results based on groups that had taken the WAIS and WAIS-R (nromed 25 years after the WAIS) and that had mean IQs ranging from 50 to 130. Results in the average IQ range are consistent with Flynn's hypothesis, but test differences gradually diminish above and below the average range, and in the retarded range are also reversed (the newer test giving a higher score) and show increasing disparity with decreasing IQ.	
Spruill, J., & Beck, B. L. (1988). Comparison of the WAIS and WAIS-R: Different results for different IQ groups. *Professional Psychology: Research and Practice*, *19*, 31-34	Commentary/ synthesis: Research-based theoretical	Yes (but ability level qualified) US	When comparing the WAIS and WAIS-R, most researchers have found WAIS scores to be higher, although a few have found WAIS-R scores to be higher. We also found the typical results (WAIS score > WAIS-R score) when the entire subject group was compared. However, when subjects were divided into different groups on the basis of their intelligence classification, the differences between the WAIS and WAIS-R IQ scores varied as a function of the subject's intellectual level. This may explain the contradictory findings in the literature. Indeed, those researchers who found WAIS-R scores to be equal to or higher than WAIS scores studied mentally retarded subjects.	
Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.	Empirical-based analysis	Yes US & 13 other countries	The rates of gain are expressed in IQ points per year, so multiplied by 30 years, they give an estimated gain over one generation. Setting aside the most extreme cases, and discounting France a bit, the generational gains range from 5 to 25 points and the median is 15 points, or a full standard devia The current generation has made massive gains on all kinds of IQ tests.	

Spitz, H. H. (1986). Disparities in mentally retarded persons' IQs derived from different intelligence tests. *American Journal of Mental Deficiency*, *90*, 588-591.	Empirical-based analysis	No US	For mildly and moderately mentally retarded individuals, the Wechsler Adult Intelligence Scale—Revised measures at about the same level as the Wechsler Adult Intelligence Scale.	
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.	Empirical-based analysis	Yes US	In this paper I have attempted to show that Wechsler scales can be grossly unrepresentative of the general population and that psychologists who use them to classify subjects as 2 SDs below the mean may err by anything from .27 to a full SD.	Rather than revising our criterion of mental retardation with each new test, psychologists should fix it at a particular point in time, perhaps at the standards set by the WISC sample of 1947-1948 and the WAIS sample of 1953-1954. In practice, this would mean keeping a WAIS-R score of 70, because of its equivalence to the old WAIS, *but lowering the cutting line* for the WISC-R down to 65 or even 60 for older subjects, so as to establish equivalence with WISC-WAIS. Whenever a new test is published, it should be administered to a large sample of retarded subjects along with the appropriate reference test from the past, so as *to derive a new score equivalent to the old cutting line.* The soundness of these reflections depends on future research, studies of the WISC-R and WAIS-R and, far more important, studies that determine whether or not we can freeze our criterion of mental retardation at a given point in time. [Note. *Emphasis added*]
Flynn, J. R. (1984a). IQ gains and the Binet decrements. *Journal of Educational Measurement*, *21*, 283-290.	Empirical-based analysis	Yes US	This data also suggest that Americans have gained about 12 IQ points from 1932 to 1972 with verbal gains being a point lower and performance gains a point higher.	

Flynn, J. R. (1984b). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29-51.	Empirical-based analysis	Yes US	This study shows that every Stanford-Binet and Wechsler standardization sample from 1932 to 1978 established norms of a higher standard than its predecessor. The obvious interpretation of this pattern is that representative samples of Americans did better and better on IQ tests over a period of 46 years, the total gain amounting to a rise in mean IQ of 13.8 points. The implications of this finding are developed.	
Simon, C. L. & Clopton, J. R. (1984). Comparison of WAIS and WAIS-R scores of mildly and moderately mentally retarded adults. *American Journal of Mental Deficiency, 89*, 301-303.	Empirical-based analysis	No US	The Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Adult Intelligence Scale-Revised (WAIS-R) were administered to 29 mildly and moderately mentally retard adults. The WAIS Verbal and Full Scale IQs were significantly lower than were corresponding WAIS-R IQs. This difference is an exception to the general pattern of IQss being higher for tests that were standardized earlier.	
Spitz, H. H. (1983). Intratest and intertest reliability and stability of the WISC, WISC-R, and WAIS Full Scale IQs in a mentally retarded population. *The Journal of Special Education*, *17*, 69-80.	Empirical-based analysis	Yes US	The reliability and stability of the Full Scale IQ scores of retarded adolescents and young adults on 3 Wechsler Scales were measured…The mean WISC-R IQ was significantly lower than the mean WISC-IQ and both were dramatically lower than the mean WAIS IQ.	
Light, M. L., & Chambers, W. R. (1958). A comparison of the Wechsler Adult Intelligence Scale and Wechsler-Bellevue II with mental defectives. Am*erican Journal of Mental Deficiency*, 62, 878-881.	Empirical-based analysis	No US	The W.A.I.S. tends to yield IQ's 3 to 4 IQ points higher than the W. B. II IQ's on the Verbal and Full Scale tests.	