The Lynn-Flynn Effect and School Psychology:
A Call for Research

A. Alexander Beaujean, University of Missouri & Applewood Centers, Inc.
& Shawn F. Guiling, University of Missouri & Columbia Public Schools

Defined, the Lynn-Flynn Effect (LFE) is the continued rise of psychometric IQ test scores (approximately .3 IQ points/year), an effect seen in many parts of the world, both in developed nations and undeveloped countries (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003; Rushton & Jensen, 2003; Sundet, Barlang, & Torjussen, 2004). The LFE is named after British differential psychologist Richard Lynn and New Zealand political scientist James R. Flynn, who “re-discovered” the effect in the early 1980’s—Lynn (Lynn, 1982; Lynn & Hampson, 1986) publishing data about the effect in Great Britain and Japan, with Flynn (1983, 1984, 1999) focusing more on the United States (but also see Flynn, 1987).

In the 20+ years research has been done in this field, the findings have been enigmatic. While multiple sources have found that psychometric IQ has been rising, general intelligence (g; Spearman, 1904) has not increased (Jensen, 1998; Kane & Oakland, 2000; Must, Must, & Raudik, 2003), and IQ endophenotypes have shown a mixed reaction, with chronometric measures (i.e., reaction times) showing no decrease (Nettelbeck & Wilson, 2004), but head size showing a marked increase (Storfer, 1999). In addition, although LFE appears to affect the entire range of the IQ distribution, there does appear to be a definite concentration among those at the lower end (Colom, Luís-Font, & Andres-Pueyo, 2005; Teasdale & Owen, 1980).

Another aspect of the LFE that has puzzled researchers is that although there are mean increases in average psychometric IQ scores, ethnic group differences on the same IQ tests have not diminished (Murray, 1999; Jensen, 1998; Rushton, 1999, 2003). While some have posited that the LFE, ipso facto, implies IQ malleability and, hence, the inevitability of the distributional convergence of Black and White IQ scores (Flynn, 1987), the one standard deviation difference between Black and White test takers is as pervasive today as it ever was (Rushton & Jensen, 2003, 2005; but also see Ceci, Rosenblum, & Kumpf, 1998). This is likely due to the fact that variance involved in the LFE is not made up of the same factors as those involved in the Black-White IQ gap (Wicherts et al., 2004).

To date, the LFE has mostly been a topic of research for differential psychologists, with various parties giving their explanation as to why the effect exists (e.g., Blair, Gamsonb, Thornec, & Bakerd, 2005; Brand, 1996; Burt, 1952; Eysenck & Schoenthaler, 1997; Lynn, 1980, 1990; Mingroni, 2004; Rodgers, 1999) or, perhaps, why it does not (Beaujean, 2005; Sundet et al., 2004; Teasdale & Owen, in press). Within this scholarship though, there has been some applied research that has tested to see how ubiquitous the effect is, with the majority of the findings showing the effect is present in a multitude of subpopulations, including those with various learning exceptionalities (Bolen, Aichinger, Hall, & Webster, 1995; Kanaya, Scullin, & Ceci, 2003; Sanborn, Truscott, Phelps, & McDougall, 2003; Truscott & Frank, 2001). Unfortunately, there has been little serious discussion within the field of school psychology, per se, as to the field's response; moreover, what little text is given over toward it seems rather haphazard and undeveloped. For example, in the fourth edition of NASP's Best Practices, Reschly and Grimes (2002) write:

The newest revisions and most recent norms for a test should be used because recent studies show that the stringency of norms changes over time and more recent norms typically are tougher than older norms. The now well-known Flynn-Effect must be considered to avoid undue effects of out-of-date norms. (p. 1347)

While superficially this might “solve” the dilemma, it more than likely does not, as the next section will illustrate. Consequently, this brief manuscript is to serve two purposes: First, to show via two contrived situations—based on the LFE literature—possible “real world” effects of either ignoring the LFE or responding to it via unresearched remedies; and second, to call for a more concentrated effort within the field of school psychology, both to discuss and further the research in practical applications of this effect.

CONTINUED ON PAGE 18
**Example A**

Suppose Student A (SA) was assessed for gifted placement in 2003 and the school psychologist gave him the Wechsler Intelligence Scale for Children-Third Edition (WISC-III) to assess his IQ. Say SA had a true score of 130 on the WISC-III, but due to (random) measurement error, he received an obtained score of 128. Being that the threshold for placement at his school was 130, he was not placed, although he was so close that the district decided to assess him again during the next school year. Fast forward to 2005, after the school district has purchased the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV), with which the school psychologist will now assess SA for gifted placement. As the norms have changed from the WISC-III to the WISC-IV (and thus the LFE is now a factor), SA now has a true score of 125 (Flynn, 1984, 1990), but due to random measurement error receives an obtained score of 127—still high, but it does not cross the placement threshold, even though his obtained score (due to random error) is actually above his true score. Were SA to experience the same (random) error on the WISC-III, his score would have been 132, which would have been high enough for admittance to the gifted program.

**Example B**

Suppose Student B (SB) was assessed for a reading learning disability (RLD) in 2002. At School 1, where SB attended in 2002, the school psychologist used the WISC-III and the Woodcock Johnson-III Tests of Achievement (WJ-3) to do the assessment for the RLD. SB had a true score of 100 on the WISC-III, but obtained an observed score of 101. For the Broad Reading section of the WJ-3, SB had a true score of 85 and scored exactly an 85, giving her a discrepancy of 16 points and thus qualifying her for special assistance from the school. Three years later, SB is up for her triennial re-evaluation, but she has moved to School 2, where they use the WISC-IV and the WJ-3. On the WJ-3, SB had a true score of 85 and scored exactly an 85, giving her a discrepancy of 16 points and thus qualifying her for special assistance from the school.

**Prevalence**

No known research to date has examined the absolute prevalence of the LFE in populations of students who are gifted or have learning disabilities. One study has examined the prevalence with students diagnosed with mental retardation (Kanaya, Scullin, & Ceci, 2003), in which they found both a statistically and politically significant effect:

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 [italics added] and were more likely to be classified MR compared with peers retested on the same test. (p. 778)

While an issue that can be investigated empirically, it is doubtful that epidemiology of students who are gifted or learning disabled would fare much better from their peers with mental retardation.

**School Psychology’s Response**

To date, there has been a minimal response by the field of school psychology to the LFE. While there is occasionally the one-or-two-paragraph description of the LFE and the subsequent solution of “use the newest revisions and most recent norms” preferred (e.g., Reschly & Grimes, 2002), it appears that applied psychology in general, and school psychology in particular, has not given much systematic thought and investigation to this phenomenon. Consequently, as a place to begin, this manuscript advises that research needs to be done in this area. More specifically, three different areas need investigation.

First, the epidemiology of placement effects due to the LFE needs much more investigation. As it stands, there is little knowledge of how many students are given a diagnosis, or have a diagnosis taken away, based, at least in part, on differently-normed IQ instruments being used during different evaluations over the student’s educational career. To that end, Kanaya et al.’s (2003) article can serve as a model study of what school psychologists should look to when studying the LFE and educational diagnoses.

Second, there needs to be more systematic investigation of the positive and negative effects that arise from various LFE “interventions.” For example, what are the pros and cons of keeping a given intellectual assessment with a student for his/her academic career? If a child’s first testing involves norms from students, say, in 2000, then what are the
effects of always using psychometric instruments that were normed circa 2000? Moreover, which of these “interventions” best aligns itself with assessment ethics (AERA/APA/NCME, 1999; American Psychological Association, 2002; National Association of School Psychologists, 2000)?

Third, there needs to be more research in alternative measures of cognitive ability, both general and specific aspects. For example, we know a sufficiently diverse battery of chronometric tasks can be a proxy for general cognitive ability (Jensen, 1998, Chapter 8), and that a specific enough battery of them can discriminate between LD and non-LD populations (Beaujean, Knoop, & Holliday, in press). Moreover, these tasks do not appear to be subject to the LFE (Nettelbeck & Wilson, 2004), so what are the pros and cons to begin using them in a diagnostic battery?

Most likely, there will not be a single right answer for this given dilemma, as various situations will call upon unforeseen variables; but a decision that is definitely wrong is to either continue to ignore the issue or to throw palliative remedies at it.

Please e-mail all submissions for The Commentary Section to LReddy2271@aol.com

References


Washington, DC: Author.


Lynn-Flynn Effect and School Psychology

Nature, 301, 655.
Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. Intelligence, 31, 461–471.
Bethesda, MD: Author.

Footnote

1In other texts, this effect is sometimes referred to as simply the Flynn Effect. This is (mainly) due to the fact that the Herrnstein and Murray (1994) coined the term in their widely-read book on the importance of IQ in determining life outcomes. In actuality, both Richard Lynn and James Flynn deserve credit for the finding, as Lynn (1982) first brough the effect to the world’s attention, even though the effect was seen over a half-century earlier (Smith, 1942; Tuddenham, 1948). This text will follow the recommendation made by Rushton (1997) and keep the effect entitled Lynn-Flynn Effect.