

White Paper

Why Must Tests Be Periodically Restandardized?

Richard D. Baer, Ph.D.

In order to answer this question it is helpful to define two terms. First, a “population” is any defined group of people. For example, all kindergarten children in Utah or all school age children in the United States would define populations. Second a “representative sample” is a sample drawn from a population that has a high probability of exhibiting the same characteristics as the population. That is, the sample would have the same average height, weight, intelligence, and other characteristics as the population.

Tests are standardized by giving them to representative samples of people from defined populations. Many intelligence and achievement tests are standardized by giving them to representative samples of the population of school age children across the nation. Standardization allows us to compare the performance of individual children to the performance of the average child his/her age or in his/her grade in the population. Thus, for example, we could determine that Dick’s intelligence is below that of the average child his age, or that Mary’s reading achievement is like that of the average child her age or in her grade, or that Pam’s math achievement is better than the average child her age or in her grade.

It has been observed that the average score of populations on intelligence tests increases over time (Graham & Plucker, 2001-02). The researcher who discovered this phenomenon is James R. Flynn and it is referred to as the “Flynn Effect.” The Flynn Effect has been observed consistently in every country it has been studied in across the world (Flynn, 1987). It is because of the Flynn Effect, this increase in population intelligence test scores over time, that intelligence tests must be periodically restandardized.

By how much does the average score of populations on intelligence tests increase over time? The answer here depends on what aspect of intelligence is being measured. In general, intelligence can be divided into two components. Fluid intelligence refers to people’s problem solving ability and is relatively independent of learning. Crystallized intelligence refers to people’s ability to benefit from experience and is relatively dependent on learning. On average, fluid intelligence increases about 15 points over a generation while crystallized intelligence increases about 9 points. Why this happens is not known but is the subject of much research and debate (Graham & Plucker, 2001-02).

As noted, because of the Flynn Effect, intelligence tests need to be periodically restandardized. This is necessary so that the scores of individual subjects can be compared to scores of their contemporary peer groups. We would not, for example, want to give a child a test that had been standardized twenty years ago and compare his/her score to the twenty year old standardization group. If we did, we might find, for example, that a student scored 115 and conclude that his/her score was above average. Such a conclusion would be incorrect because of the Flynn Effect that would predict that after twenty years the average score on the test would have increased by 10 to 15 points.

Achievement test scores also change over time although they do not necessarily increase. For example, Stedman and Kastle (1991) cite a number of reviews indicating that reading achievement in the United States increased from about 1900 to 1960 then decreased through the 1970s. As with intelligence tests, because it is desirable to be able to compare the scores of individual subjects with their contemporary peers, achievement tests need to be periodically restandardized.

January 2011

References

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

Graham, C. & Plucker, J (2001-02). The Flynn effect. **Retrieved September 26,2010 from <http://www.indiana.edu/~intell/flynneffect.shtml>.**

Stedman, L. C. & Kastle, C. F. (1991). Literacy and Reading Performance in the United States from 1880 to the Present. In H. Damon-Moore, L. C. Stedman, K. Tinsley, & W. V. Tollinger, Jr. Readers and reading since 1880: Literacy in the United States (pp. 75-128). New Haven CT: Yale University Press.