PEARSON

The Value of the Stanford Scale as a Common Metric

Margaret A. Jorgensen, Ph.D.

March 2004

(Revision 1, May 2004)

The Value of the Stanford Scale as a Common Metric

Introduction

As states and school districts report on the adequate yearly progress (AYP) of student performance as required by the *No Child Left Behind Act* of 2001 (NCLB), there will be increasing emphasis on documenting student progress along a developmental continuum within selected subject areas. The vertical standard score scale developed by Pearson Inc. (Pearson) for *The Stanford Achievement Test Series* and other Pearson products meets this need by providing a common interpretive framework for test results across grades. The use of an equal-interval "vertical" scale to evaluate student progress over a period of years yields important information that informs individual and classroom instruction.

What is Vertical Scaling?

A vertical scale is an *expanded normalized standard score scale*—an equal-interval scale that spans all grades from kindergarten through Grade 12 (Nitko, 2004). Vertical scales are also referred to as "developmental scales." A separate vertical scale may be established for each subtest or content area. The benefit is the ability to directly compare scores earned by students on different assessment instruments administered at different times.

A vertical scale can provide a common vocabulary and metric for describing a student's progress throughout his or her educational journey. Analogies to the vertical scale are found in the fields of healthcare and business. For example, in the healthcare field, there are common American measurement scales for height and weight (inches and pounds). These scales are understood and used in the same way throughout the network of health-care providers, thereby enabling the evaluation of patients by different clinicians in ways that are directly comparable.

Figure 1 illustrates how an equal-interval scale divided into feet describes the growth of a child over a period of years. In other words, a three-year-old who grows two inches in a year, and a nine-year-old who also grows two inches in the same year, have both increased in height by equal amounts.
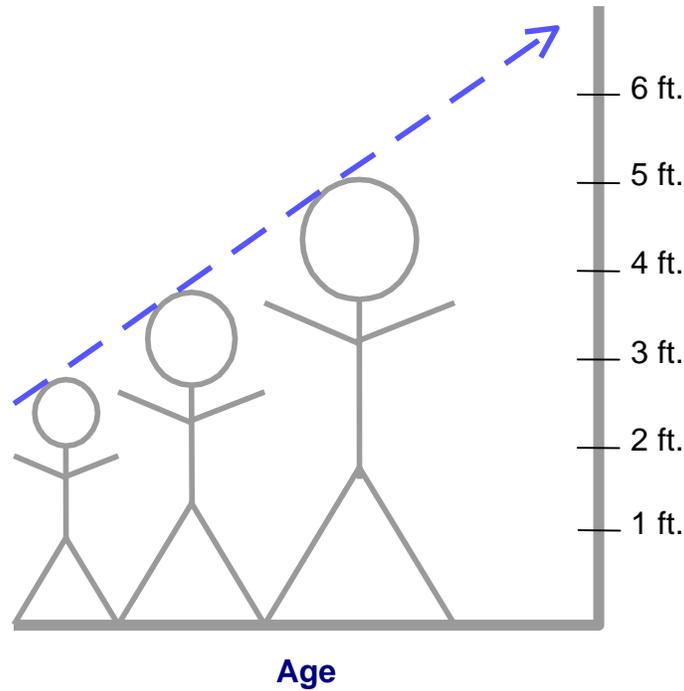

Age

Figure 1. Illustration of height increasing with age

In the worlds of business and economics, currency types serve as links to equal-interval scales. Countries that use the same currency can directly compare the value of their goods and the performance of their markets. Countries using different currencies can use accepted exchange rates to enable direct comparisons, as shown in Figure 2.

U.S. Dollars	Euro	Japan Yen	Fiji Dollars	Turkey Liras
100 USD	79.60 EUR	10,833.56 JPK	169.52 FJD	132,760,000.00 TRL

Figure 2. Examples of international monetary exchange rates*

* Source: Universal Currency Converter®, XE Corporation, http://www.xe.com/ucc, converted 2-23-04.
  (Note: This information serves as an example only; the values shown may not be accurate.)

Regardless of which currency one is trading, links to the other systems are straightforward. One can readily interpret Japanese yen relative to U.S. dollars and understand prices. This ease of metric translation does not currently exist in achievement testing.

Use of Vertical Scaling in Assessment

In achievement testing, a vertical scale that spans kindergarten through Grade 12 can be the common metric enabling consistent and comparable evaluations of student progress regardless of age, school, or evaluating teacher.

A student's raw score (number correct) must first be converted to its corresponding scale score through the use of a conversion table developed by the test publisher. Once converted, the scale score describes the student's placement on an achievement continuum. A scale score is independent of the child's grade level, time of testing, or the test level administered. This allows direct comparisons of a student or group's current scores with scores obtained in previous testing programs. It is the incremental change along the vertical scale that is of interest when demonstrating the progress of an individual, class, or school.

When a test publisher develops a new edition of an assessment product, a research program is conducted to equate the two adjacent editions. As a result, a vertical scale is developed for the new edition that allows scores from either edition to be directly compared as though on the same scale.

Norm-referenced Scores

Scores from norm-referenced standardized assessments are frequently reported on a vertical developmental scale developed by the test publisher. The norm-referenced interpretive framework ranks an individual's performance by comparing his or her score to the distribution of scores from a larger "norm" group. The evaluation of norm-referenced scores on a vertical scale describes achievement and growth over time in relation to the growth of the norm group.

Criterion-referenced Scores

Criterion-referenced tests yield scores that describe a student's performance in comparison to learning targets (i.e., instructional standards) specified in the curriculum (Nitko, 2004). Criterion-referenced scores are typically reported in terms of achievement thresholds or performance levels defined by "cut-off" scores. The norm-referenced vertical scale described above can be used to set cut-off scores for standardized assessments that also measure defined instructional standards. Using the criterion-referenced interpretive framework, growth is demonstrated by measuring gains in the amount of actual knowledge possessed by the student, rather than the student's change in rank relative to a separate reference group.

The Stanford Scale

Pearson has developed a vertical developmental scale for the *Stanford Achievement Test* Series and other Pearson assessment products. Stanford vertical scales exist for each subject area included in a test. Pearson is the only assessment publisher offering this common metric and ease of score interpretation across multiple products, which include:

- *Stanford Achievement Test,* Tenth Edition (Stanford 10)

- *Stanford English Language Proficiency Test* (Stanford ELP)

- *Stanford Diagnostic Reading Test,* Fourth Edition (SDRT 4)

- *Stanford Diagnostic Mathematics Test,* Fourth Edition (SDMT 4)

As each new edition of a Pearson assessment product is published, extensive research programs ensure the continuity, stability, and validity of the overarching Stanford scale. As a result, scores can be compared across different levels and editions of the same test as well as between Pearson products assessing the same content area.

The Stanford scale meets rigorous industry standards, including suggestions in *Standards for Educational and Psychological Testing* that test publishers conduct periodic checks of the stability of the scale on which scores are reported (American Educational Research Association, 1999).

The Stanford scale has been constructed based on a *common person equating* design in which a large number of on-level and off-level test items are administered to several hundred thousand students during periodic national research programs. The resulting data allow for the vertical linking of all test editions and levels along a common continuous scale.

The high quality of the Stanford scale is especially critical for high-stakes assessment programs that determine grade placement for individuals. For decades, the Stanford scale has been the foundation of a comprehensive, valid, and reliable suite of assessment products that span all grades.

Conclusion

Accurately evaluating and reporting student progress in selected academic areas is very important in today's climate of educational accountability. A useful tool for demonstrating adequate yearly progress is the vertical scale. When a student's raw test scores are transformed into scale scores, academic progress can be measured clearly on a continuous scale that spans all grades. Pearson's products are linked by a research-based set of common vertical scales that conform to the most rigorous industry standards of stability and validity.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Nitko, A. J., (2004). *The educational assessment of students* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Pearson Inc. (2004). *Stanford Achievement Test series, Tenth Edition technical data report*. San Antonio, TX: Author.

Pearson Inc. (2003). *Stanford English Language Proficiency test technical manual*. San Antonio, TX: Author.

Additional copies of this and related documents are available from:
Pearson Inc.
19500 Bulverde Rd.
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
http://www.pearsonassess.com