The Hippocratic Oath and Grade Equivalents

by Malbert Smith III, Ph.D., President, MetaMetrics

Medical researchers and practitioners often cite Hippocrates' famous adage, "Above all, do no harm." Implicit in this statement is that all actions are not helpful and, in fact, can be harmful. While we typically think of the Hippocratic Oath in terms of medical treatments, the principle also applies to the assessment of students' reading and mathematics abilities. For years, the educational measurement community has lamented the practice of reporting students' abilities as grade equivalents. And yet, some assessments continue to measure performance using these inaccurate and often misleading metrics.

The misconceptions of grade equivalents have been well documented by research organizations, most noting that the metrics create more confusion than clarity (AERA/APA/NCME, 1985; Airasian, 1994; Miller, Linn and Gronlund, 2009; Stiggins, 2009). In 1981, for example, the International Reading Association (IRA) crafted a resolution on the misuse of grade equivalents. In it, the organization "strongly advocates that those who administer standardized reading tests abandon the practice of using grade equivalents to report performance of either individuals or groups of test takers" (1981). In spite of the advice, counsel and warnings from IRA, as well as other leading organizations such as the National Council for Measurement and Education (NCME), the American Educational Research Association (AERA) and the American Psychological Association (APA), about the misuse of grade equivalents, the architects of testing organizations that employ these metrics assume that they provide educators and parents with useful information about students' abilities. Sadly, teachers, parents and, ultimately, students only can be harmed when grade equivalents are used to convey test information.

Before we delve into the numerous problems associated with grade equivalents, let's first review how these metrics are calculated. Typically, grade equivalents are determined based on student performance in a norming group. For our hypothetical scenario, we will assume that 1,000 students in grades 1–12 were tested during the last month of the school year. The median raw score at each grade is assigned a grade level and month. In this example, the median score of fifth graders would be assigned a grade equivalent of 5.9. The month description for each grade is computed by dividing the difference in growth between two grade levels by the number of months in the school year. This provides scores for each month in the school year, or 5.1–5.8 in our example.

On the surface, grade equivalents appear to be an intuitive way to report students' test scores. However, this seemingly simplistic method glosses over some significant limitations that often promote misleading and inaccurate interpretations of the data. Miller et al (2009) highlight six misconceptions concerning these metrics: 1) assume that norms are standards of what should be; 2) assume that grade equivalents indicate the appropriate grade placement for a student; 3) assume that all students should be expected to grow one grade-equivalent unit per year; 4) assume that the units are equal throughout the score range; 5) assume that grade equivalents for different tests are comparable; and 6) assume that the scores that are based on extrapolations to grades well above or below the test level are meaningful (467).

Throughout this paper, we will explore each of these misconceptions in more detail and explain how Lexile® and Quantile® measures differ dramatically from grade equivalents. For example, Lexile and Quantile measures rely on developmental scales—the Lexile scale and the Quantile scale—to measure and monitor student performance in reading and mathematics. In addition, the interpretation of a Lexile or Quantile measure is the same regardless of which test was used. In sharp contrast to grade equivalents, which only denote student ability in comparison to those in a norming group, Lexile and Quantile measures stand alone in their interpretation because they do not rely on who was in the norming group, when the norming test was administered or which assessment was used.

1) Assume that norms are standards of what should be

The structure of grade equivalents (grade.month) makes this a common and unfortunate misinterpretation. Grade equivalents do not represent a grade-level curriculum standard. For example, a grade equivalent of 5.9 does not denote the desired level of achievement for all fifth graders. Rather, it simply represents the norming group's median score, or projected score, for fifth graders in their ninth month of schooling. Achieving the same score as the average student in the norming group may not be an appropriate goal for all students.

Lexile and Quantile measures, in contrast, are not generated from grade-level norms and do not presume a specific grade-level interpretation. Struggling students are not stigmatized with a grade equivalent that labels them as "below grade." Rather, students have an independent Lexile measure that enables them to select appropriately difficult books and other materials within their ability range. Similarly, students use their Quantile measure to identify the mathematical skills and concepts they are ready to learn. MetaMetrics®, developer of The Lexile Framework for Reading and The Quantile Framework for Mathematics, has studied typical Lexile and Quantile ranges for students in specific grades. Educators who are interested in this type of normative comparison can find the information at www.Lexile.com and www.Quantiles.com.

2) Assume that grade equivalents indicate the appropriate grade placement for a student

Grade equivalents only should be interpreted as rough estimates of grade-level performance. Imagine a student who scores a 6.9 on a fourth grade mathematics test. Educators should not assume that the student has mastered sixth grade mathematics content when, in fact, it may be unknown how sixth graders would perform on the fourth

grade test. Additionally, educators can not assume that the student has the prerequisites for seventh grade mathematics. All that is known for sure is that the student scored well above the average fourth grade student in the mathematics norming group. Because Lexile and Quantile measures do not suggest grade-level placement, the measures eliminate this type of misinterpretation.

3) Assume that all students should be expected to grow one grade-equivalent unit per year

Grade equivalents assume that the typical student will grow one grade equivalent each school year. For example, if Tom and Jane are "average" students and have grade equivalents of 4.7 and 2.5, respectively, we would assume that in one year they will have scores of 5.7 and 3.5. But what if both students are in the third grade? Would we still want and expect them to grow one grade equivalent? For Tom, who is scoring above average for a third grade student, we may not expect him to grow as much as a typical third grader (he likely will be learning harder content and need more time and practice to understand the materials). For Jane, who is scoring below average for a third grader, we probably would expect (and want) her to grow more than the typical third grade student. In order to discuss growth, we need to examine the underlying growth of the ability being measured, not the relationship with a comparison group. Lexile and Quantile measures monitor development along vertical scales that directly measure reading and mathematics abilities. By using the Lexile and Quantile scales to measure growth, we easily can see how much a student is growing and set realistic goals for each of them.

4) Assume that the units are equal throughout the score range

The grade-equivalent scale is not an equal-interval scale; it is like a ruler with inches of varying lengths. Grade-equivalent units do not represent equal amounts of ability at different points along the scale. A student who moves the same number of grade equivalents at one level on the scale (e.g., 2.5–2.9), has not necessarily "grown" in ability the same amount as a student who moves the same number of grade equivalents at a different level on the scale (e.g., 8.5–8.9). The amount of growth in ability required to move from 2.5–2.9 is much greater than that needed to move from 8.5–8.9. Because grade equivalents are not equal-interval units, they should not be used in mathematical calculations, such as averaging. The Lexile and Quantile scales, in contrast, are equal-interval scales. Regardless of the point on the scale, the amount of growth in ability required to move between two points is the same. In other words, a move from 240L–340L on the Lexile scale represents the same increase in ability as a move from 840L–940L. As such, Lexile and Quantile measures can be used in mathematical calculations.

5) Assume that grade equivalents for different tests are comparable

Grade equivalents obtained by different test publishers often provide conflicting results. This is a natural consequence of the way by which the metrics are developed. Grade equivalents are determined based on the norming group of the publisher; each grade-equivalent study reflects the characteristics of that unique norming group.

6) Assume that the scores that are based on extrapolations to grades well above or below the test level are meaningful

This may be the most harmful misuse of grade equivalents. Students never should be placed in a grade or curriculum based on grade equivalents; we can not assume that a second grader with a score of 5.2 is reading like a fifth grader. The only safe assumption we can make is that the student is reading much better than the average second grader. Because grade equivalents are calculated based on interpolations and extrapolations, this hypothetical second grader never was asked any fifth grade questions.

Just imagine the confusion—and discomfort—that would result if the shoe industry adopted the same grade-equivalent methodology to match students with appropriate footwear. Industry representatives would use the same basic approach to construct their grade-level equivalents: conduct a field study in which they measure students in grades 1–12 in the first month of the school year. Then they would take the average at each grade, subtract the difference and divide the difference by the number of months in school year.

The problems and limitations that are obvious in this grade-equivalent shoe scenario are just as glaring in the measurement of educational constructs. A 4.2 shoe carries about as much useful information as stating that a student has a grade equivalent of 4.2. Just as we would never buy a pair of shoes based on grade equivalents, we need to be more discriminating about the metrics we use to describe student performance in reading and mathematics. Nearly all educational measurement organizations have opined on the negative impact of grade equivalents. It is time that we all abide by the Hippocratic Oath and simply "say no" to grade equivalents as measures of student ability.

REFERENCES

AERA/APA/NCME Joint Committee. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Association News: International Reading Association. (1981). Journal of Learning Disabilities, 14 (9), 558.

Airasian, P. W. (1994). Classroom assessment (2nd ed.). New York: McGraw-Hill, Inc.

Miller, M., Linn, R., & Gronlund, N. (2009). Measurement and assessment in teaching (10th ed.). New Jersey: Pearson Education, Inc.

Stiggins, R. J. (2009). Student-centered classroom assessment (Second Edition). Upper Saddle River, NJ: Merrill, an imprint of Prentice Hall.

MetaMetrics®
LINKING ASSESSMENT WITH INSTRUCTION