DOCUMENT RESUME

ED 395 009 TM 025 021

AUTHOR Ramos, Cristina
TITLE The Computation, Interpretation, and Limits of Grade Equivalent Scores.
PUB DATE 25 Jan 96
NOTE 19p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, January 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Age Differences; *Comparative Analysis; Computation; *Grade Equivalent Scores; *Grades (Scholastic); Heuristics; Instructional Program Divisions; *Norms; *Scoring; Standardized Tests; Student Evaluation; *Test Interpretation; Test Results

ABSTRACT

Grade-equivalent scores are widely used in the school system in reporting students' performance on standardized achievement tests. This paper explores how grade equivalent scores are calculated and interpreted. In addition, the paper examines the limitations of grade-equivalent scores through the use of small heuristic data sets. A grade-equivalent score is a score indicating the grade level at which this score is the mean performance level. They are created by administering a test to students in various grades. Typically, a test being normed is administered to large groups of students in each of several successive grade levels at different times of the year. Grade norms have several limitations and are often misinterpreted. It must be noted that when grade-equivalent scores are derived, it does not mean that the test was given to all grades or to children at all different points on a grade's continuum. Grade-equivalent scores reported beyond tested grade limits can be misleading. In addition, grade-equivalent scores obtained from tests produced by different publishers often give conflicting results. A further limitation is that grade-equivalent scores should not be used to make comparisons of the grade-equivalent performances of an individual across different subjects. Other incorrect interpretations are reviewed. (Contains one table, three figures, and nine references.) (SLD)

# The Computation, Interpretation, and Limits of Grade Equivalent Scores

Cristina Ramos

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, January 25, 1996.

# The Computation, Interpretation, and Limits of Grade Equivalent Scores

ABSTRACT

Grade equivalent scores are widely used in the school system in reporting performance of students on standardized achievement tests. The present paper explores how grade equivalent scores are calculated and interpreted. Furthermore, the paper examines the limitations of grade equivalent scores through the use of small heuristic data sets.

## The Computation, Interpretation, and Limits of Grade Equivalent Scores

For some test uses, such as those seen in the school system, it is often appropriate to compare an examinee's performance to performance of a current peer norm group (e.g., all examinee's taking this test at this time). Some normative scores used in such cases include percentile ranks, z-scores, scaled scores, and stanines (Crocker & Algina, 1986). However, when children are tested on aptitude or achievement measures, probably the most commonly used score for reporting performance results is the grade-equivalent score (Lyman, 1971).

A grade-equivalent score is a score indicating the grade level at which this score is the mean performance level (i.e., a grade-equivalent score of 7.3 indicates that the test taker had a score equal to the mean score for pupils in the third month of the seventh grade) (Hills, 1981). The grade-equivalent score was introduced to obtain two advantages: first, to make understanding scores easier for educational personnel and parents, and second, to make it easier to evaluate growth and development (Hills, 1981). According to Gay (1980), grade-equivalent scores remain popular because they are considered to be fairly easy for most people to understand. Ligon and Battaile (1986) report that in certain situations grade equivalent scores are the most appropriate score available for reporting achievement test data and that in the Austin, Texas Independent

School District grade equivalents are routinely reported to high school students' parents. However, some researchers point out that although grade-equivalent scores have a high degree of intuitive appeal (Burns, 1980) they are less meaningful and are often misinterpreted by administrators and laypersons (Campbell, 1994; Hills, 1981).

The present paper was written for the purpose of describing how grade-equivalent scores are computed, how they should be interpreted, and what are the limitations of such scores. Researchers and educational personnel need to be aware that there can result severe misinterpretations of grade-equivalent scores if the creation and limits of grade-equivalent scores are not well understood.

## Development of Grade-Equivalent Scores

Grade-equivalent scores are created by administering a test to students in various grades. Grade norms are found by computing the mean raw scores obtained by children in each grade during each month of instruction. For example, if the average number of problems solved correctly on an arithmetic test is 32 for a norm group of students beginning the third grade, then any student who scores a 32 will have a grade equivalent of 3.0, regardless of what grade the student is actually in. To illustrate this computation process, refer to Table 1. Typically, a test being normed is administered to large groups of students in each of several successive grade levels at different times of the year.

Insert Table 1

Once the median score is determined for each grade level, it is plotted on a bivariate axis as seen in Figure 1. A grade equivalent expresses both the grade and the month in the grade. The units on the baseline are usually divided into 10 parts, from grade 3.0 to 3.9, and then 4.0 to 4.9, and so on. The first 9 parts correspond to the months of the school year and the tenth part corresponds to the summer vacation. A score at 3.4 indicates the fourth month of the third grade. Essentially, this procedure assumes that increases in test scores are due to the months that a student is in school and that no learning occurs during the summer.

Insert Figure 1

Once the observed mean scores are computed and graphed, these points are connected (interpolated) with a straight line and then extended (extrapolated) with a line going beyond those grades tested to indicate how students in grades below and above the tested grades might have performed if the test had been administered to them. This follows the assumption that learning occurs in a positively linear progression, as illustrated in Figure 2. However, as teachers, counselors, and administrators know, this assumption is not consistent with reality since students sometimes experience losses in learning during vacation periods and may make more rapid gains at some times of the year than at others (Campbell, 1994).

Based on this description, it is apparent that the computation of grade-equivalent scores is a simple procedure. However, grade-equivalent scores become more complicated when they are interpreted.

Insert Figure 2

## Interpretation and Limitations of Grade-Equivalent Scores

Grade norms have several limitations and are often misinterpreted despite their popularity. In the first place, it is important to note that when grade-equivalent scores are derived this does not necessarily mean that the test was given to all of the grades, nor does this mean the test was given to children at all the different points on the grade continuum within a grade. Therefore, at some grades no one ever took the test, but grade-equivalent scores give the impression that students in many different grade levels were tested. For example, there may be grade-equivalent scale for reading which extends to the tenth grade even though reading is not taught in that grade. In this case reporting the grade-equivalent score would be inappropriate (Anastasi, 1988).

Such scores might also be inappropriate if the test is too easy or too hard. For example, a student at a lower grade, say fourth grade, might score very high on an arithmetic test, e.g., a 7.2 grade score. Does this mean this student should be placed in the seventh grade? No. All this means is that the student is above average in math, but it is very unlikely that that the fourth grade student has mastered all of the skills and concepts taught in arithmetic at the fifth and sixth grade levels and that the student could score as well on a test designed for seventh graders. Grade equivalents reported beyond tested grade limits can be misleading, as seen in this example.

Furthermore, grade-equivalent scores obtained from tests produced by different publishers often give conflicting results. Not only do the publishers use different normative samples, but often the publishers emphasize different areas within the same subject matter at the same level. For example, on an arithmetic test, one publisher may focus on addition of fractions more so than another publisher, thus altering the grade-equivalent scores of some students (Lyman, 1971).

Another misinterpretation the users of grade-equivalent scores often make is that of making comparisons of the grade-equivalent performances of an individual student across different subjects. This is wrong! Students as a whole tend to perform very differently across subjects. For example, in arithmetic, students within a grade tend to perform similarly because there is usually little opportunity to learn arithmetic outside of school. Therefore, the variability within a class tends to be small and students tend to progress through grades at relatively the same rate. Reading, on the otherhand, is entirely a different matter. Students vary greatly on the ability to read at any grade; some read fluently while others can barely decode words. In the upper grades, the variability within a class is often very large. Given this, the standard deviations for these two subject matters will be very different. The norm distribution of Reading scores, having a lot more variability between students, will have a larger standard deviation than will the norm distribution of arithmetic scores. Because of the difference in size of standard deviations, the same grade-equivalent, say 7.0, in reading and math may result in a percentile rank of 64 in reading but a 90th percentile rank in arithmetic. Especially for higher elementary grades and beyond, grade-equivalent scores for different tests, even if they're within the same battery ( say the KTEA), cannot be meaningfully compared ( Hills, 1981; Lyman, 1971).

Not only do grade-equivalent scores differ in variability form one subject to another, but also within the same subject at different grades. In general, children at lower grades tend to perform more similarly due to their lack of educational experience. In later grades, however, students tend to perform more differently from one another, thus increasing score variability. For example, according to Hills (1981), the spread of scores expressed as standard deviations on the Reading Score of the Metropolitan Achievement Test increases from 0.5 at the beginning of grade 1 to 3.7 at the beginning of grade 6.

Insert Figure 3

To help explain the change in size of standard deviations as grade level increases, Figure 3 is provided. In Figure 3, it is evident that as one goes from grade 4 to grade 8 the size in standard deviation gets larger. In Figure 3, the diagonal line represents the mean performance at each grade level. The standard deviation is relatively small at grade 4. A point on the graph indicates one standard deviation below the mean, and there is a projection from that score over to the diagonal line and then down to the line showing grade level. It can be seen that a student with a score of 1 SD below the mean at grade 4 would have a grade-equivalent score of 3, or one year below grade level.

Now look at the grade 8 student who also performs 1 SD below the mean, the same relative performance level as the grade 4 student, with neither gain nor loss in performance compared to the norm group. The larger standard deviation at grade 8 results in a grade-equivalent score of 6 when we project over to the diagonal line and down to the corresponding grade. The percentile score for the two students in grade 4 and grade 8 will remain the same. The only change is the size of the standard deviation as one goes from lower to higher grades, which is characteristic of grade-equivalent scores.

Furthermore, a student who is at the same percentile score each successive year, again indicating neither gain nor loss in performance relative to the norming group, would appear to be at a lower grade-equivalent score each year if the student is below the mean. The naive interpreter of grade-equivalent scores would perceive this to mean that the student was falling further and further behind peers when this in fact is not the case. As Reynolds (1981) indicates, this type of interpretation of grade-equivalent scores is often used as a diagnostic criterion for diagnosing reading disorders which results in substantial overestimations of disabilities in upper grade levels and underestimation of reading difficulties in the lower grade levels.

The corollary of this is that students who consistently perform 1 SD above the mean will appear to perform substantially better at each grade they are tested even though their performance is consistent. In this case, students may be placed in more advanced classes or even promoted a grade when in reality they are not academically prepared.

Another problem with the interpretation of grade-equivalent scores results when administrators and parents consider a particular grade-equivalent score as a standard to be reached by all of the students in a class or school. Recall that in the development of grade-equivalent scores a mean score was computed and therefore there will always be people below the mean as well as above. The goal of "getting everyone up to grade level" can be very frustrating for teachers and students. The failure to recognize that grade-equivalent scores are by definition an average causes administrators to impose unreachable goals for teachers and students. If everyone in a grade were to perform "at grade level," then the grade-equivalent score would just move up because, by definition, it would still be the mean score for students in that grade.

## Review of Fallacious Interpretations

Hills (1981) provides a good review of incorrect interpretations often made of grade-equivalent scores, some of which are repeated here:

1. A grade-equivalent score for sixth grader Tim of 9.2 in reading means that he can read as well as ninth graders in the second month of the school year.
2. A grade-equivalent score of 9.2 in reading for Tim and of 7.3 in arithmetic means that in his reading Tim is nearly two years ahead of his performance in arithmetic.
3. Grade-equivalency scores of 9.2 in reading for Tim and of 7.3 in arithmetic indicate that Tim is farther ahead of his group in reading than in arithmetic.
4. Since 30% of the students in Mr. Brown's fifth-grade class got grade-equivalency scores below 5.0, something needs to be done to improve the

instruction in his class and perhaps in the instruction given to students before they reach Mr. Brown.

5. When Tim was tested in the fall of the sixth grade, he received a grade-equivalent score of 9.2 in reading. Tested in the spring, he received a grade-equivalent score of 8.0. This indicates that he lost a lot of his reading skill during the school year, and some effort should be made to find out why and whether such losses can be expected to continue.

Note that although all of the above interpretive statements are false, they are all too common.

## Alternatives to Grade-Equivalent Scores

Because of the likelihood that grade-equivalent scores will be misinterpreted, some researchers have proposed that standard scores or percentile ranks be used instead of grade-equivalent scores when giving performance results to parents (Campbell, 1994; Lyman, 1971). A percentile rank indicates a student's relative standing within a specified group. Although percentile ranks do have limitations, they are reasonably easy to understand (Campbell, 1994). Percentile ranks have been described as "probably the best single derived score for general use in expressing test results" (Lyman, 1971, p.101).

Reynolds (1981) advocates that standard scores be used as a substitute for grade-equivalent scores, stating that they are more precise. Standard scores express the distance between a raw score and the mean in terms of standard deviation units. The standard score is often reported as a z score, T score, or deviation IQ score, and it is a measure of relative position (Crocker & Algina, 1986). Unlike percentile ranks, standard scores have equal units and can be averaged; however, they are more difficult to explain to parents and students, because of the concept of standard deviation, which is often confusing to educators (Campbell, 1994).

## Summary

Knowledge about the development and interpretation of grade-equivalent scores is important for educators, parents, and students since scores are often reported in the schools as a way of describing a student's relative performance on achievement tests and in relation to peers. Furthermore, grade-equivalent scores possess inherent limitations and can often be misinterpreted by naive users. Finally, the reporting of percentile ranks and standard scores were proposed as possible alternatives to grade-equivalent scores.

## References

Anastasi, A. (1988). Psychological testing. New York: Macmillan Publishing Company.

Burns, E. (1982). The use and interpretation of standardized grade equivalents. Journal of Learning Disabilities, 15, 1, 17-18.

Campbell, N.J. (1994). Interpreting scores from standardized tests. The Clearing House, 67, 6, 314-316.

Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. Orlando: Harcourt Brace Jovanovich, Inc.

Gay, L. R. (1980). Educational evaluation and measurement: Competencies for analysis and application. Columbus, Ohio: Merrill Publishing Co.

Hills, J.R. (1981). Measurement and evaluation in the classroom (2nd ed.) Columbus, OH: Bell & Howell Company.

Ligon, G. & Battaile, R. (1986). Grade equivalents: We report them, you should too. Austin, Tx: Austin Independent School District, Office of Research and Evaluation.

Lyman, H.B. (1971). Test scores and what they mean (2nd ed.) Englewood Cliffs, NJ: Prentice-Hall, Inc.

Reynolds, C.R. (1981). The fallacy of "two years below grade level for age" as a diagnostic criterion for reading disorders. The Journal of School Psychology, 19, 4, 350-358.

Table 1: Computation of Mean Score for various grade levels on an arithmetic test administered in September

| Grade 2 Scores: | Grade 3 Scores: | Grade 4 Scores: | Grade 5 Scores: |
|---|---|---|---|
| 17 | 27 | 35 | 46 |
| 18 | 28 | 36 | 46 |
| 19 | 29 | 38 | 47 |
| 20 | 29 | 39 | 48 |
| 20 | 30 | 39 | 49 |
| 21 | 31 | 40 | 49 |
| 21 | 31 | 40 | 50 |
| 21 | 32 | 41 | 52 |
| 22 | 32 | 41 | 52 |
| 23 | 33 | 42 | 56 |
| 24 | 34 | 43 | 57 |
| 24 | 34 | 43 | 60 |
| 24 | 34 | 44 | 62 |
| 25 | 35 | 45 | 63 |
| 25 | 35 | 45 | 65 |
| 26 | 36 | 48 | 66 |
| 27 | 37 | 49 | 68 |
| 396 | 576 | 756 | 936 |
| x = 22 | x=32 | x=42 | x=52 |

## Figure 1

Observed Mean Score for Grades 2-5

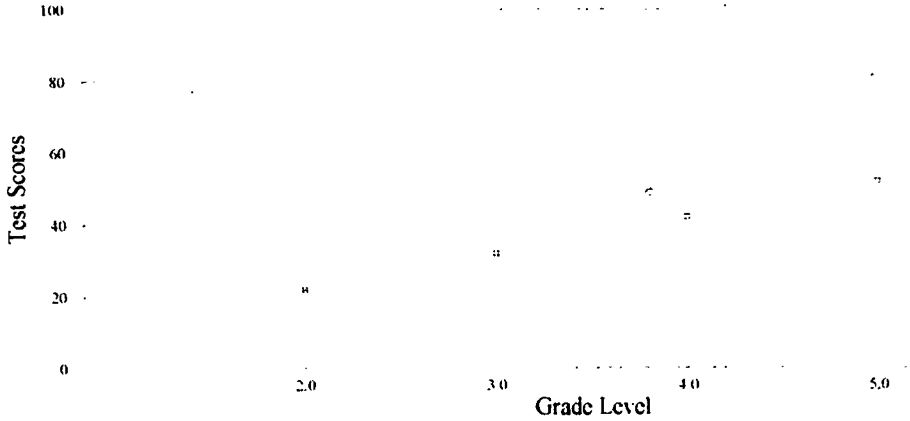

Figure 2
Extrapolated Grade Equivalent Values

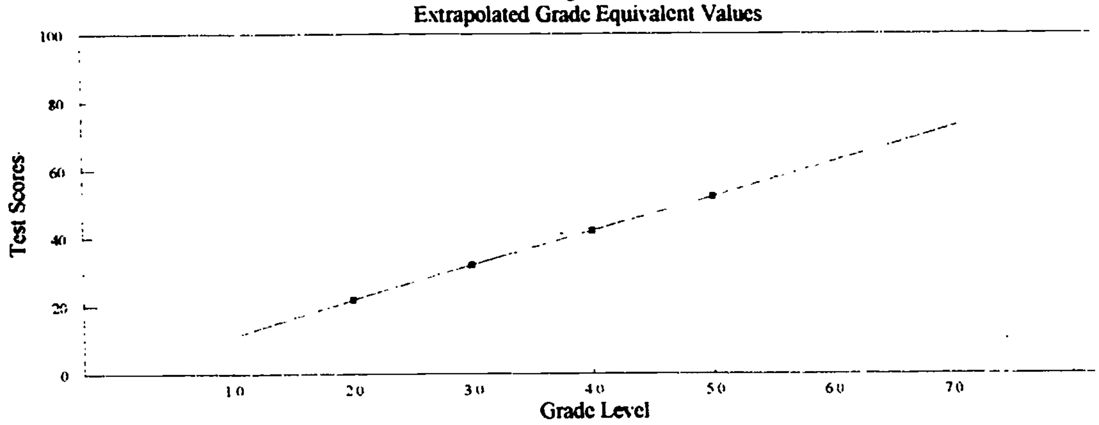

Figure 3: Standard Deviation of Grade-Equivalent Scores