DOCUMENT RESUME

ED 273 676 TM 860 531

AUTHOR Ligon, Glynn; Battaile, Richard
INSTITUTION Austin Independent School District, Tex. Office of Research and Evaluation.
REPORT NO AISD-ORE-85-47
PUB DATE Apr 86
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986). Appended materials contain small print.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; Elementary Secondary Education; *Grade Equivalent Scores; Measurement Objectives; Norm Referenced Tests; Raw Scores; School Districts; *Scores; Standardized Tests; Teacher Attitudes; *Testing Problems; *Test Interpretation
IDENTIFIERS *Austin Independent School District TX; Iowa Tests of Basic Skills; *Percentile Ranks; Tests of Achievement and Proficiency; Texas Educational Assessment of Minimum Skills

ABSTRACT

In certain situations, grade equivalent scores are the most appropriate statistic available for reporting achievement test data. It is noted that testing practitioners have found that raw scores, normal curve equivalents, stanines, and standard scores are very useful. However, it is best to convert to either grade equivalents or percentiles before communicating them to lay audiences. In the Austin, Texas Independent School District, both grade equivalents and percentiles are routinely reported to high school students' parents. Elementary school parents receive percentile scores routinely, but may request grade equivalents. Both percentiles and grade equivalents are often misinterpreted; the shortcomings of percentiles are merely less well known. Seven critical questions that can be used to decide whether achievement test results should be reported in grade equivalent or percentile scores are presented and briefly discussed. These questions can be located along two dimensions: comparison standard and time. The five standards of comparison are: (1) grade level; (2) peers; (3) all students, all grades; (4) self; and (5) students at the same achievement level. The two criteria involving time periods are: achievement status to date; and pretest posttest gain. Appendices include a 1977 publication describing myths in interpreting grade equivalent scores and an information kit containing training materials on grade equivalents, percentiles, and norming a standardized test. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED273676

GRADE EQUIVALENTS: WE REPORT THEM, YOU SHOULD TOO

Glynn Ligon, Ph.D.

Richard Battaile

Austin Independent School District
Department of Management Information
Office of Research and Evaluation

Austin, Texas

U S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Ligon
R. Battaile

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper Presented at the Annual Meeting of the
American Educational Research Association,
San Francisco, April, 1986

ORE Publication Number 85.47

TM 860 531

Grade Equivalents: We Report Them, You Should Too

Grade equivalents (GEs) are better than percentiles for answering several key questions about student achievement on standardized tests. Raw scores, normal curve equivalents (NCEs), and standard scores are dysfunctional when trying to explain achievement test scores to most audiences. Now, as a reader, you must be looking for some hard data to back up these statements. This paper contains some data, but is more a position paper on the best statistic for reporting standardized achievement test results based upon the authors' experiences in managing achievement testing in an urban public school system.

GEs have long been the subject of much debate within the educational and psychological arenas. Many teachers, principals, and counselors have shown an acquired distaste for them--perhaps in some cases inspired by a provincial university professor without adequate experience in the public school realm. Sometimes the opposition has come from our own ranks, as evidenced by Attachment A, a nine-year-old exposé from our own testing office. **While common misconceptions and misuses of GEs have been discussed and documented, too poorly documented are the questions for which GEs are the most appropriate statistic available.**

We intend to clarify some issues in the lingering debate over the use of GEs. Moreover, the relative usefulness of different statistics will be examined in relation to their abilities to answer essential questions for assessing student learning. Our objective is to convince practitioners that GEs have a unique and important contribution to make to educational evaluation and the reporting of student achievement progress.

Let us dispense with the utility of raw scores, standard scores, and normal curve equivalents as statistics for the common folk (who make up the majority of our audience for test results). **When we have to spend more time defining and explaining the limited information these statistics carry, they become dysfunctional.** More importantly, they do not communicate with most lay audiences as well as do GEs or percentiles. To be fair, the testing practitioner will find raw scores and NCEs but especially standard scores to be more useful and appropriate for calculating gains and averages and for tests of significance; however, converting them to either GEs or percentiles is practical before communicating the results.

This leaves us with GEs and percentiles as competitors for our endorsement as the best statistic for reporting to general audiences. In this paper, we present the position that neither is omniscient, but each answers specific questions better than the other.

Our Perspective

We shall begin with some background and content. Austin is a central Texas urban community of approximately 350,000 - 450,000 people composed of heterogeneous groups having varied interests. The University of Texas, the State governmental offices and political community, high-tech industry, and many other facets make Austin and the students in the Austin Independent School District (AISD) a sophisticated blend of sun-belt diversity. Citizens of Austin take a healthy interest in the affairs of AISD and the progress of its students.

The School District itself has long been known for innovative programs and strategies and, in many areas, has become a standard for emulation. Students consistently achieve above the national average at grades 1-12 in all achievement areas. Compared to urban districts nationwide, the average AISD student in grades 1-8 achieves higher in all areas than three fourths of the students tested across the country. The average AISD student at grades 9-12 achieves higher than two thirds of the students in urban districts nationwide. On the other hand, one third of AISD's students are from low-income families and almost half are either Hispanic or Black. What this shows is that we have both a sophisticated and varied clientele to which we report achievement test results.

The Office of Research and Evaluation (ORE) is responsible for all types of systemwide testing. Students in grades K-8 take the Iowa Tests of Basic Skills (ITBS), while high school students (grades 9-12) take the Tests of Achievement and Proficiency (TAP). We also coordinate the administration of the Texas Educational Assessment of Minimum Skills (TEAMS), a State-mandated testing battery administered in Texas to students in grades 1, 3, 5, 7, 9, and 11-12.

ORE has been reporting GEs in addition to percentiles to central office and campus staff for 6 years. Since 1984, both GE and percentile scores have been sent home to parents of students in grades 7-12. (Parents of elementary students receive percentile scores only; however, GEs may be provided in conferences with elementary teachers.)

One of our primary responsibilities is doing as much as possible to communicate test (and other evaluation) data to all groups (e.g., the School Board, teachers, parents) who may utilize it. Moreover, our mission is to educate effectively these groups in the understanding, interpretation, and use of test scores and related information. Only by comprehending all test data available can we and others fully (and most cost-effectively) evaluate the state of learning and the progress made by the individual student, the class, the school, or the District.

As educators, we need to identify specific questions which must be answered when determining progress. What issues are most critical in which situations? What score would best provide the insight needed for an effective evaluation? These and others are the questions which, ultimately, we seek to answer.

Our data set and the resources drawn together for this paper are derived from our years of coordinating and administering standardized tests, and from the experience gained in executing and molding a public school testing program. In addition, in November, 1985, we surveyed elementary school teachers about their attitudes towards the accuracy and reporting of GEs.

Types of Test Scores

Many types of data are available for testing practitioners to use when reporting results to various audiences. We must first examine some characteristics of each, citing strengths and weaknesses, both from a practical and a technical standpoint.

Percentile rank scores are the most widely reported testing statistic. They offer a comparison of each student to a norm group, in this case a representative sampling of thousands of students nationwide. Percentiles can be readily understood at face value by school staff as well as the public; however, they are frequently misunderstood and misused when employed as the sole index by which to examine progress over a period of time. From a technical standpoint, percentile rank scales are frequency distributions having significant inequality of intervals among points on the scale, especially at the extremes of the distribution. Consequently, percentile rank scores show the student's relative position compared to the norm group, while offering limited information on amounts of learning differences between pre- and posttest data and other aspects.

Standard scores (SSs), which include stanines and NCEs, offer a different perspective. Standard scores express the student's distance from the mean in terms of the distribution's standard deviation. The process used to derive standard scores yields a scale with a distribution and properties corresponding to those of the raw score distribution scale. The result is that, without any distortion of results, computations and other manipulations that can be executed using the original raw scores can in most cases also be done using standard scores. Additionally, standard scores are useful to make comparisons between results from different test instruments, providing a common basis and scale for evaluation.

Reporting standard scores in most cases would be meaningless to the public. The numerical values used and their relationships would be confusing and inappropriate for imparting information to the public.

Normal curve equivalents (NCEs) are similar to a stanine scale with the stanines split into ten parts each. NCEs (and stanines) are useful because they provide equal-interval scales that may be averaged; however, they also have drawbacks limiting their use in reporting results in a public school setting. Namely, they are often interpreted as if they are percentiles.

Grade equivalents (GEs) offer the perspective of where the student falls compared to a developmental continuum constructed based on the norming sample. For example, a GE of 6.7 means the student made the same raw score as would the average student in the seventh month of grade six.

Historically, GEs have been somewhat misunderstood by school staff, parents, and other groups. For example, a parent might mistakenly think that his child in fourth grade who scored a GE of 6.7 should actually be enrolled in the sixth grade.

From a technical consideration, GE scales have varying degrees of usefulness. Although the ITBS GE scales are derived to be equal interval scales, providing users the luxury of having flexibility in utilizing the data, this is not true of all GE scales. When deciding whether to report GEs, a careful scrutiny of the norms must be conducted.

Frankly, AISD would not be reporting GEs if we had not investigated and found the ITBS GEs to be reasonably equal interval and to meet certain simple standards (i.e., the GE associated with the 50th percentile must be the grade and month representing the critical norming date; a 50th percentile student must gain 1.0 GE a year to maintain the 50th percentile; students above the 50th percentile must gain more than 1.0 GE in a year to maintain the same percentile; students below the 50th percentile must gain less than 1.0 GE in a year to maintain the same percentile.)

Problems with Grade Equivalents and Percentiles

Unfortunately, our chosen champions among these statistics are both deeply flawed.

A brief overview of the problems most often cited in the use of GEs is provided in Figure 1. Also provided is an overview of the similar problems with percentiles. Because both can be (and are) misinterpreted, a logical conclusion is that the misinterpretation of GEs must be more problematic for educators and influences their discrepant acceptance of GEs and percentiles. This may be plausible considering the misinterpretation of GEs can more easily be used by parents to argue for double promotions to higher grades, or to express concern over how far behind a student is. For teachers, there may be the frustration of wanting from their investment in standardized testing a GE that perfectly diagnoses the best instructional level for students, but not finding that perfect diagnosis.

We believe that the shortcomings of percentiles are merely less well known than are those of GEs. In our experience, percentiles are misinterpreted and misunderstood at least as often as are GEs.

Figure 1. PROBLEMS WITH GRADE EQUIVALENTS AND PERCENTILES.

Grade Equivalents	Percentiles
1. Cannot be averaged (Unless developed as an equal-interval scale)	1. Cannot be averaged
2. Often misinterpreted, for example: a. Fifth grader scoring at 6.5 could succeed in sixth grade. b. Fifth grader scoring at 6.5 should be in a sixth-grade text. c. A GE of 18.5 is more than just a theoretical projection.	2. Often misinterpreted, for example: a. Percentile means the percentage of items answered correctly. b. A 69th percentile is a failing grade. c. Student at 50th percentile two years in a row did not gain (learn). d. Student at 30th percentile two years in a row gained a year, kept place with average student. e. Student going from 15th to 16th percentile is catching up with average student. f. Student at 65th percentile in in both reading and math is equally above grade level in both areas.
3. Most GE score points are either interpolated or extrapolated rather than empirically derived.	

When Should Achievement Test Results Be Reported in Grade Equivalents or Percentiles?

This is the key question in this paper. By examining the questions that parents and educators ask about student achievement, we can determine which can be answered better by either GEs or percentiles. Figure 2 summarizes the seven critical questions most frequently encountered in our achievement testing experience.

After much agonizing over how these seven questions could be conceptually related to each other, we were able to locate each question along two dimensions--"comparison standard" and "time." The five comparison standards include:

- Grade Level--that almost mystical level that indicates average or expected achievement, as in being "on grade level";
- Peers--other students in the same grade level (or age level);
- All Students, All Grades--students from prekindergarten through grade 12 who take standardized tests;
- Self--the student's own past achievement; and
- Students at the Same Achievement Level--students who made the same score on past achievement tests.

The two time periods include:

- Status--achievement to date; and
- Gain--achievement improvement within the last measurement period between a pre- and posttest.

A brief discussion of the rationale for selecting either GEs or percentiles as more appropriate for each of the seven questions follows:

1. Is the student above, at, or below grade level? (Comparison Standard: grade level; Time: status)

   Both GEs and percentiles answer this question. Grade level can be defined as a point representing the exact average score at the critical norming date (i.e., 50th percentile or a GE equivalent to the year and month of testing) or as a range of scores around that point.

2. How does the student rank among others in the same grade? (Comparison Standard: peers; Time: status)

   Percentiles provide a rank, GEs do not.

3. How far ahead/behind is the student? (Comparison Standard: peers; Time: status)

   GEs provide an indication of how far (in months and years) a student differs from the average. Percentiles tell how far from average a student is in terms of a percentage of students, but this does not communicate in terms that answer the question.

4. Did the student learn more, the same, or less than the average student? (Comparison Standard: peers; Time: gain)

GEs answer this question because by definition the average student gains one year (1.0 GE) in a year's time. Percentiles answer this question only for certain students, i.e., those who score at the 50th percentile on the pretest (a score below 50 on the posttest indicates a gain less than the average student, a score above 50 indicates a greater gain, and a score at 50 indicates an equal gain), those who score below the 50th percentile on the pretest and the same or lower on the posttest (this represents less than the average student's gain), and those who score above the 50th percentile on the pretest and the same or higher on the posttest (this represents greater than the average student's gain).

5. In which grade would the average student score the same as this student? (Comparison Standard: all students, all grades; Time: status)

   This is the definition of GE.

6. Did the student learn at a rate that is faster, the same, or slower than in the past? (Comparison Standard: self; Time: gain)

   Percentiles easily answer this question because a student who maintains the same percentile rank on the pre- and posttest has learned at the same rate as in the past. A higher posttest percentile indicates a faster rate than in the past; a lower one, a slower rate.

7. Did the student learn at a rate that is faster, the same, or slower than other students who began at the same level? (Comparison Standard: students at the same achievement level; Time: gain)

   Percentiles easily answer this question because a student who maintains the same percentile rank on the pre- and posttest has learned at the same rate as other students who were at the same level. A higher posttest percentile indicates a faster rate than in the past; a lower one, a slower rate.

Figure 2. WHEN SHOULD ACHIEVEMENT TEST RESULTS BE REPORTED IN GRADE EQUIVALENTS OR PERCENTILES?

COMPARISON STANDARD	TIME	
	Status: Achievement To Date	Gain: Within Last Measurement Period
Grade Level	%ile or GE: Is student above, at, or below grade level?	N/A
Peers	%ile: How does student rank among others in the same grade? GE: How far ahead/behind is this student?	GE: Did student learn more, the same, or less than the average student?
All Students, All Grades	GE: In which grade would the average student score the same as this student?	N/A
Self	N/A	%ile: Did student learn at a rate that is faster, the same, or slower than in the past?
Students at the Same Achievement Level	N/A	%ile: Did student learn at a rate that is faster, the same or slower than other students who began at the same level?

Implementation

Having identified the benefits of GEs, the reporting of them in our District was the logical and appropriate course to follow. However, because the statistic is one so frequently misunderstood and misused, a massive training effort had to be undertaken throughout the District. Every segment of potential users, including the School Board, central administration and campus staff, and the students and parents, had to be informed in varying degrees about GEs.

An assortment of materials was created, including a packet of handouts for each teacher. In addition, ORE produced a videotape dealing with understanding and using GEs that was broadcast on AISD's cable television channel for viewing during faculty meetings and at other times. Attachment B contains samples of some materials produced for training campus and central office staff about GEs.

Conversations, notes, and comments over the past few years indicate that misconceptions and misinterpretations about test scores are persistent and resistant to training.

Observations and Results

GEs have now become a commonly used statistic in AISD. Teachers, counselors, and principals appear to have begun to grasp the concepts of GEs and, in discussions with ORE staff, seem to prefer them in many cases (e.g., when discussing issues dealing with pretest and posttest data). However, we wanted some actual data on the acceptance level of GEs.

Therefore, in November, 1985, a sample of AISD elementary teachers was surveyed about their attitudes towards GEs. For two of the questions, Figure 3 provides the percentages of teachers answering with each response.

Of the teachers who responded to Question 1, 69.4% expressed positive support for the reporting of GEs in addition to reporting percentiles. However, this support is confounded by the responses to Question 2, which concerns the misleading properties of GEs. While 31.3% of the teachers responding thought that GEs are not misleading, 39.1% said that GEs are in fact misleading indicators of student achievement levels.

Figure 3. RESULTS OF GRADE EQUIVALENT (GE) SURVEY OF ELEMENTARY SCHOOL TEACHERS, NOVEMBER, 1985.

Question:	1. Having GE scores reported in addition to percentiles is helpful to me.	2. GE scores are misleading indicators of student achievement levels.
Responses:		
Strongly Agree	30.6%	12.0%
Agree	38.8%	27.1%
Neutral	20.9%	29.6%
Disagree	7.5%	24.8%
Strongly Disagree	2.1%	6.5%
Total Number of Responses	425	443
Total Number of Surveys Sent	563	575
Percentage of Teachers Responding	75.5%	77.0%

An additional survey item asked fifth-grade teachers to identify the reading instructional level of a particular student in their class. (One student in each class was selected by us. The total group of students selected spanned the range of low to high achievement, with most students chosen being around the average achievement level.) The response options provided on the survey item covered the achievement range from 2.3 to 8.3, with responses available in-between at increments of .5 GE. This gave teachers the option of choosing "5.3" (on grade level for the date of the survey) or values lower or higher than 5.3 in increments of .5 year. Teachers' responses were compared to actual ITBS Reading Total scores from the previous April's testing (after adding .5 GE to account for the difference between date of testing and date of survey). Figure 4 contains the survey item and the results.

Clearly, the ITBS GE and the teachers' estimates of the students' actual reading levels are close. Inspection of Figure 5 shows that the ITBS GE may overestimate actual classroom performance/instructional levels for the higher scoring students; however, a larger sample with more precise measurement is needed to verify this.

The bottom line appears to be that GEs can provide teachers a gross estimate for a student's instructional level, and the majority of teachers find having both GEs and percentiles to be helpful. However, there is obviously still a high level of skepticism on the part of our teachers as to the accuracy of GEs as indicators of student achievement levels.

Figure 4: COMPARISON OF FIFTH-GRADE TEACHERS' ESTIMATES OF STUDENTS' READING LEVELS AND LATEST ITBS SCORE.

ITBS Achievement Level Minus Teacher's Estimate	Number of Teachers	Percentage of Teachers
+3.0	0	0.0%
+2.5	1	1.7%
+2.0	5	8.6%
+1.5	3	5.2%
+1.0	8	13.8%
+0.5	11	19.0%
0.0	11	19.0%
-0.5	16	27.6%
-1.0	2	3.4%
-1.5	0	0.0%
-2.0	1	1.7%
-2.5	0	0.0%
-3.0	0	0.0%

N = 58

Teacher's estimate and actual ITBS GE were both expressed to the nearest .5 GE within a range from 2.3 to 8.3.

Correlation Coefficient
Teacher's Estimate and Actual ITBS GE = .760
Teacher's Estimate and Rounded ITBS GE = .738
($P < .0001$ for both)

QUESTION: Which grade equivalent listed below is closest to the reading instructional level of _____?

A. 2.3 or lower
B. 2.8
C. 3.3 (two years below grade level)
D. 3.8
E. 4.3 (one year below grade level)
F. 4.8
G. 5th grade 3rd month (on grade level for this date)
H. 5.8
I. 6.3 (one year above grade level)
J. 6.8
K. 7.3 (two years above grade level)
L. 7.8
M. 8.3 or higher

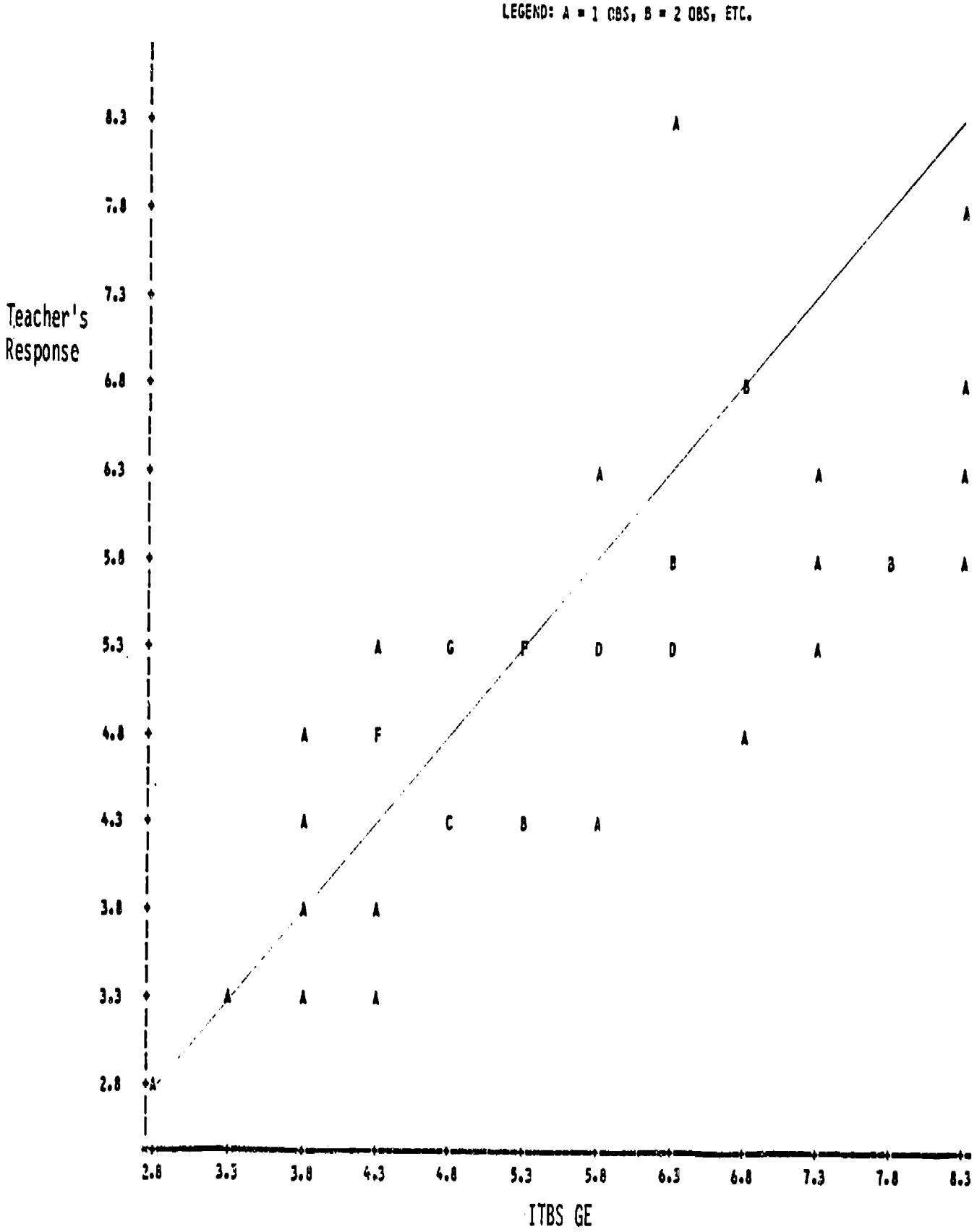


Figure 5. PLOT OF TEACHERS' RESPONSES TO SURVEY QUESTION.

Conclusion

The authors are not taking the position that GEs are the best statistic (they are not in many circumstances), nor the position that GEs will be used without misinterpretation (they seldom have been), but the position that GEs provide a perspective that is missing from other statistics and is necessary for complete understanding of a student's achievement level.

Some data have been presented to show that the majority of teachers probably find having GEs reported to be useful and to show that GEs can predict a student's approximate instructional level in reading.

Clearly GEs have their technical shortcomings, but equally as clear is the conclusion that **GEs help us answer questions that other statistics respond to inadequately.**

Challenge

We challenge test publishers to develop grade equivalent scales that are empirically derived with a minimum of interpolation and extrapolation. We challenge testing practitioners to recognize the usefulness of GEs and to use them wisely and cautiously to answer the questions for which percentiles are inadequate.

References

Anastasi, A. Psychological testing. New York: MacMillan, 1976.

Hoover, H. D., The most appropriate scores for measuring educational development in the elementary schools: ge's. Educational Measurement: Issues and Practice. Winter, 1984, (pp. 8-14).

Ligon, G. and Matter, K. Anomalies in achievement analyses. Paper presented at the annual meeting of the American Educational Research Association, New York, March 1982. (ORE Publication No. 81.60).

Ligon, G. and Wilkinson, D. The average achievement test score: a demagogue statistic. Paper presented at the annual meeting of the American Educational Research Association, New York, March 1985. (ORE Publication No. 84.42).

ATTACHMENT A
EARLY ORE PUBLICATION
CONCERNING GEs

THE TRUTH ABOUT GRADE EQUIVALENT SCORES !

OFFICE OF RESEARCH AND EVALUATION
Austin Independent School District

Pub. No. 76-18 January, 1977

Grade equivalent scores have a number of real technical drawbacks which lead to practical problems when they are used to interpret a student's performance on a standardized test. The following are some fables about grade equivalent scores that O.R.E. has compiled in the tradition of Aesop.

FABLE 1

Johnnie, a 5th grade student, received a 4.0 grade equivalent score in both math and reading. Conclusion: Johnnie's teacher does not need to concentrate on one subject more than the other.

What is wrong with this interpretation of grade equivalent scores?

As this table shows, even though Johnnie scored the same grade equivalent in reading and math, his percentile in a national sample of 5th grade students for these subjects indicates a lower rank for math than for reading.

%iles for GE of 4.0	
Reading	Math
20	13

MORAL : Comparing a student's grade equivalent scores in two different content areas is inappropriate and can lead to incorrect instructional decisions.

FABLE 2

Emma, a 4th grade student, scored a 7.5 grade equivalent on the reading section of a standardized test. Obviously, Emma is a much more advanced student than most 4th graders and should possibly be provided with enrichment materials in reading. On the basis of this test score, her teacher, Ms. Wagner, decides that the appropriate level of reading enrichment materials that would best suit Emma's capability is 7th grade reading materials.

Is this use of grade equivalents correct?

A grade equivalent score of 7.5 on this test indicates that Emma reads the material on this particular test which was designed for 4th grade students as well as would the average 7th grader. We do not know how well Emma might read materials on a test designed for 7th graders. Accordingly, there is no basis for assuming that Emma can adequately read and comprehend 7th grade materials.

MORAL : Grade equivalent scores should not be the sole basis for deciding which instructional materials a student should use.

TABLE 3

On a standardized test for the 4th grade level, grade equivalent scores in reading of non-Title I students were higher by 2.3 years than those of Title I students. For 5th graders, the difference was 2.6; and in the 6th grade, the difference in average scores was 3.0 years (Figure 1). The conclusion drawn from this is that "Title I students are losing ground every year."

Is this conclusion correct?

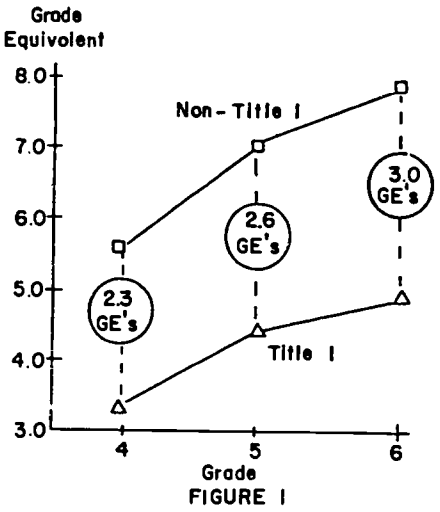


FIGURE 1

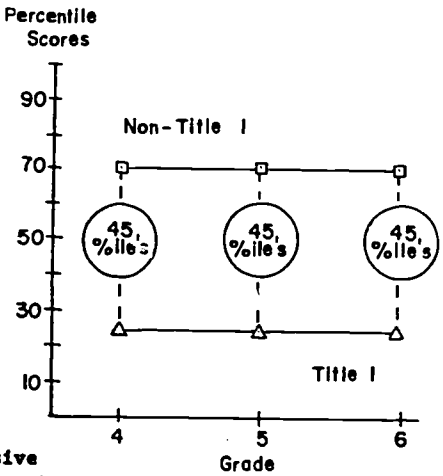


FIGURE 2

Not entirely. The problem is that the grade equivalent scores do not provide all of the information that is needed. There is evidently a progressive gap in absolute achievement between the group of Title I students and the group of non-Title I students. But, as Figure 2 illustrates, these Title I students are not losing ground in terms of their relative ranking with other students. Actually, the Title I student average for each of the Grades 4, 5, and 6 corresponds to the same percentile score (25%ile) for each grade.

MORAL : Comparing grade equivalent scores of a high-achieving group of students with a low-achieving group of students, without considering the corresponding percentile scores, can lead to incorrect conclusions.

TABLE 4

Marvin Howard, a second grader, is given Level 2 of a standardized test. He scores an 8.4 grade equivalent on the math test. The counselor at Marvin's school tells Mrs. Howard that "Marvin did as well in this test as did the typical 8th grade student who took the test in the fourth month of the school year."

Is there anything wrong in this statement?

When the publishers of this test tried out their Level 2 test, they gave it to 2nd, 3rd, and 4th graders. They did not give the test to 8th graders. Using the scores from the 2nd, 3rd, and 4th graders, they estimated what score an 8th grader would achieve. Therefore, all that can truthfully be said about Marvin's score of 8.4 is that "If a typical 8th grader had been given this Level 2 test during the fourth month of the school year, it is a good guess to say that he would have made about the same score as did Marvin."

MORAL : In order to interpret grade equivalent scores correctly, one must be aware of how they are derived and be justly wary of extreme projections such as the one in this illustration.

AUSTIN INDEPENDENT SCHOOL DISTRICT
Office of Research and Evaluation

UNDERSTANDING GRADE EQUIVALENTS AND PERCENTILES

TYPE OF SCORE: Grade Equivalents

DEFINITION: A grade equivalent represents the grade placement (year and month) for which a given raw score is average (the median).

QUESTIONS ADDRESSED:
1. How close to grade level is a student achieving?
2. Did a student learn as much from one year to the next as did the average student?

TYPE OF SCORE: Percentile

DEFINITION: A percentile represents a student's rank--the percentage of students which scored lower than a student.

QUESTIONS ADDRESSED:
1. What proportion of students did a student score higher or lower than?
2. Did a student learn as much from one year to the next as other students who were at the same percentile rank the first year?

(Match the ◀ at the top and bottom of this page with the ◀ on the next page.)

To the left is a scale which represents the full range of possible grade equivalent scores for the ITBS Vocabulary Test, Levels 5-14.

This part of the scale represents the sixth grade: 6.0-6.9. The ten decimal places are equivalent to one month each. The three summer months together equal one month, and the nine months of the school year equal one month each.

Over a 12-month period, how much does this grade equivalent scale go up? ___

What grade equivalent represents the national average for the time that AISD tests students in grade 8? ____

What grade equivalent represents the national average for the time that AISD tests students in grade 3? ____

(Match the ◁ at the top and bottom of this page with the ◁ on the next page.)

The percentiles shown to the left are for the months of testing in AISD (April, grades K-6; February, grades 7 and 8).

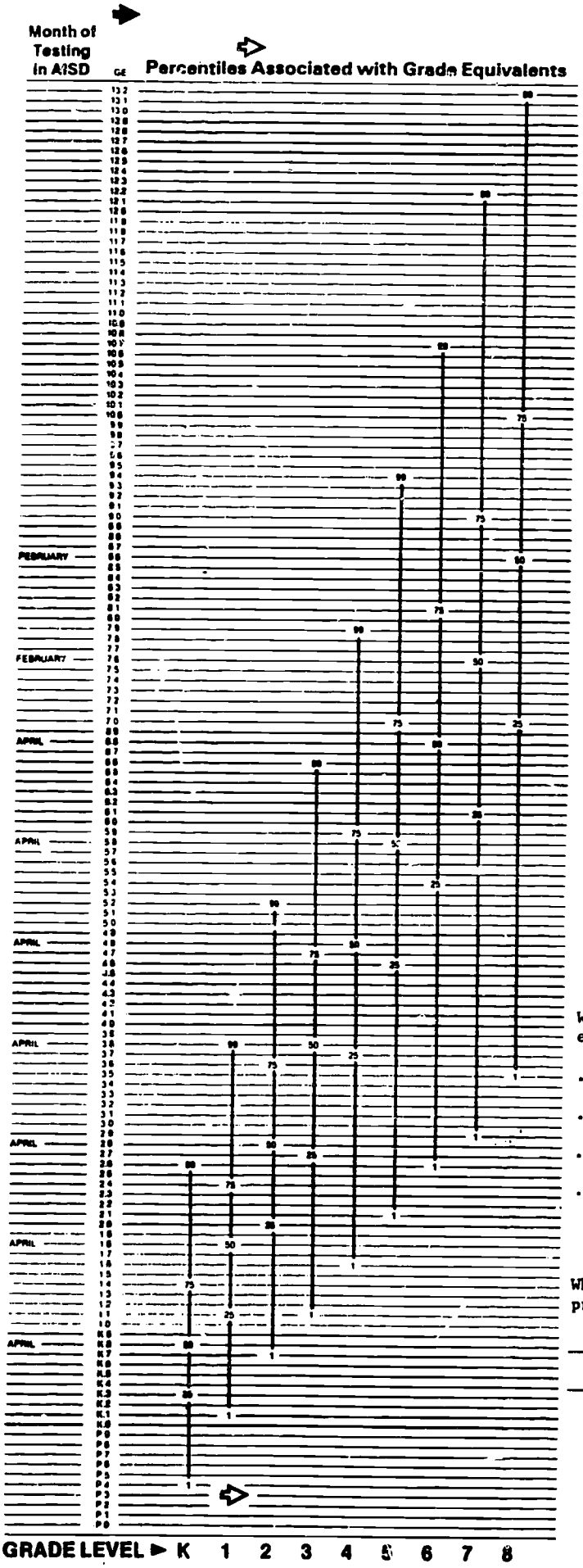
What percentile always corresponds to the grade equivalent for the month of testing?____

What grade equivalent gain is made by a student who scores in both kindergarten and first grade at the 50th percentile? ____

What grade equivalent gain is made by a student who scores in both kindergarten and first grade at the 25th percentile? ____

What grade equivalent gain is made by a student who scores in both kindergarten and first grade at the 99th percentile? ____

CAUTION: Grade equivalents do not necessarily represent the best instructional level for a student. A fourth grader scoring at 7.8 should not necessarily be working in seventh-grade texts. Although this fourth grader knows fourth-grade material as well as the average seventh grader, this student may not know seventh-grade material as well as seventh graders do.



What is the range of grade equivalents between the 1st and the 99th percentile in kindergarten? _____

... in grade 8? _____

The table below shows the grade equivalent gain needed to maintain the same percentile rank from K through 8.

Percentile Rank	K GE	8th GE	GE Gain
1	P.4	3.5	4.1
25	K.3	7.0	6.7
50	K.8	8.6	7.8
75	1.4	10.0	8.6
99	2.6	13.2	10.6

What percentile goes with a grade equivalent of 3.8 in grade 1? _____

... in grade 2? _____

... in grade 3? _____

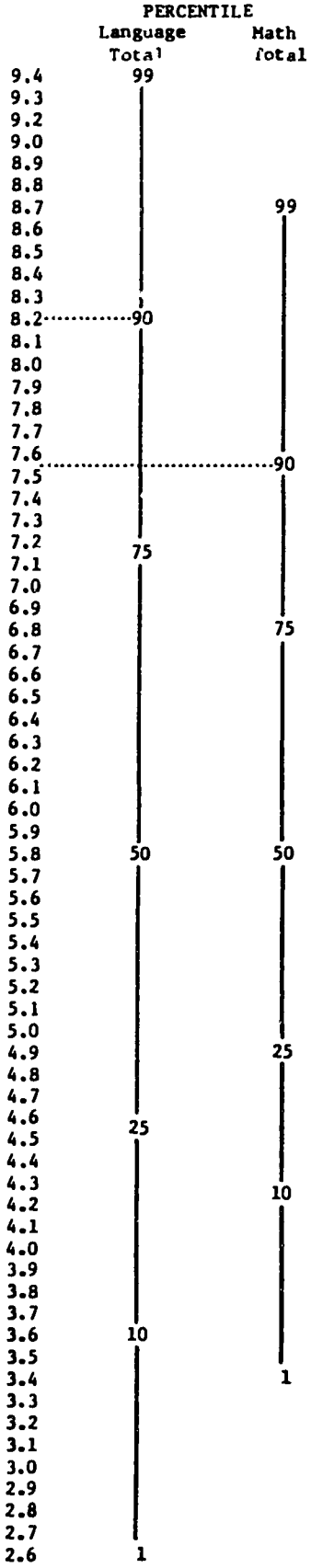
... in grade 4? _____

... in grade 8? _____

Which type of score can be interpreted with fewer cautions?

_____ Percentiles

_____ Grade Equivalents



DOES A STUDENT WHO MAKES THE SAME GRADE EQUIVALENT IN TWO AREAS ALSO MAKE THE SAME PERCENTILE IN BOTH?

In this example, Math Total and Language Total are compared for fifth graders taking the ITBS in April.

For a student who scores at the 90th percentile in both language and math, what are this student's grade equivalents?

90th %ile = ____ GE for Language Total

90th %ile = ____ GE for Math Total

AUSTIN INDEPENDENT SCHOOL DISTRICT
Office of Research and Evaluation

Norming a Standardized Test

STANDARDIZED, NORM-REFERENCED TEST: a test with--

1. standardized administration and scoring procedures, and
2. national norms.

NATIONAL NORMS: scales on which we can compare the scores made by our students to the scores made by students across the nation.

NATIONAL NORM GROUP: the group of students who took the test when it was normed.

RAW SCORE: the number of items answered correctly.

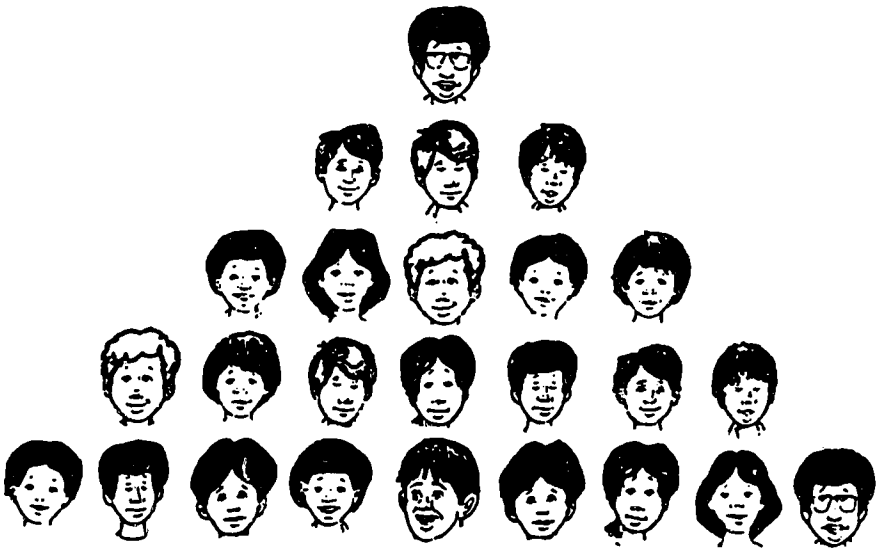
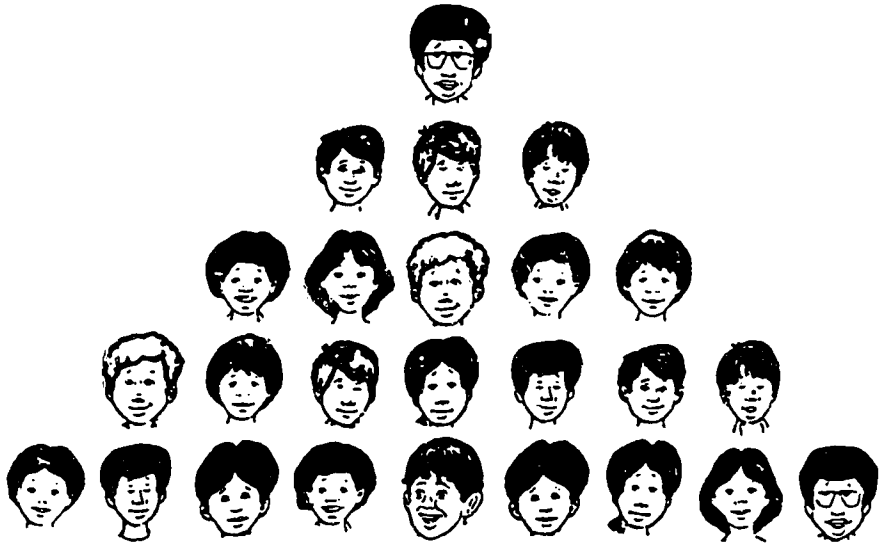


Figure 1. Raw scores made by our national norm group.

TEST: Vocabulary
NUMBER OF ITEMS: 20
NUMBER OF STUDENTS: 25
TIME OF TESTING (NORMING): April of Grade 5

MEDIAN: the score which divides the students into two equal parts. (Always the 50th percentile in the national norm group)

PERCENTILE RANK: the percentage of students scoring below a raw score point.



RAW SCORE	12	13	14	15	16	17	18	19	20
PERCENTILE	—	—	—	—	—	—	—	—	—

Figure 2. Percentiles Associated with Each Raw Score.

Which raw score represents the median?

Which raw score represents the 50th percentile?

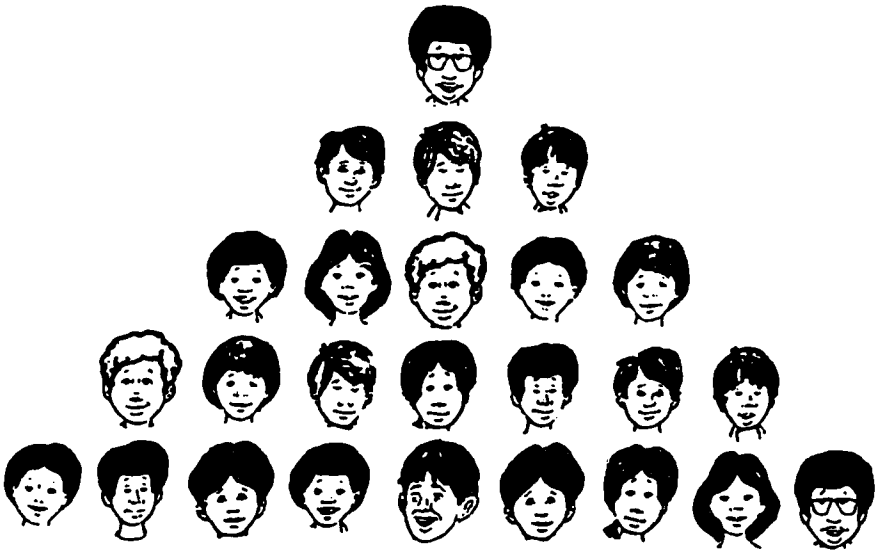
What are the percentiles for the other raw scores?

GRADE EQUIVALENT: the grade and month for which a particular raw score is the median.

April of Grade 5 = 5.8

April of Grade 7 = _____

February of Grade 2 = _____



RAW SCORE	12	13	14	15	16	17	18	19	20
GRADE EQUIVALENT	—	—	—	—	—	—	—	—	—

Figure 3. Grade Equivalents Associated with Each Raw Score.

What grade equivalent goes with a raw score of 16? _____

What grade equivalent goes with the other raw scores below?

RESULTS OF TESTING NATIONAL NORM GROUPS AT EACH GRADE

Grade	Median Raw Score in April	Grade Equivalent assigned
3	12	
4	14	_____
5	16	_____
6	18	_____
7	20	_____
		_____

What grade equivalents go with the other raw scores?

FOR THE MORE ADVANCED READER

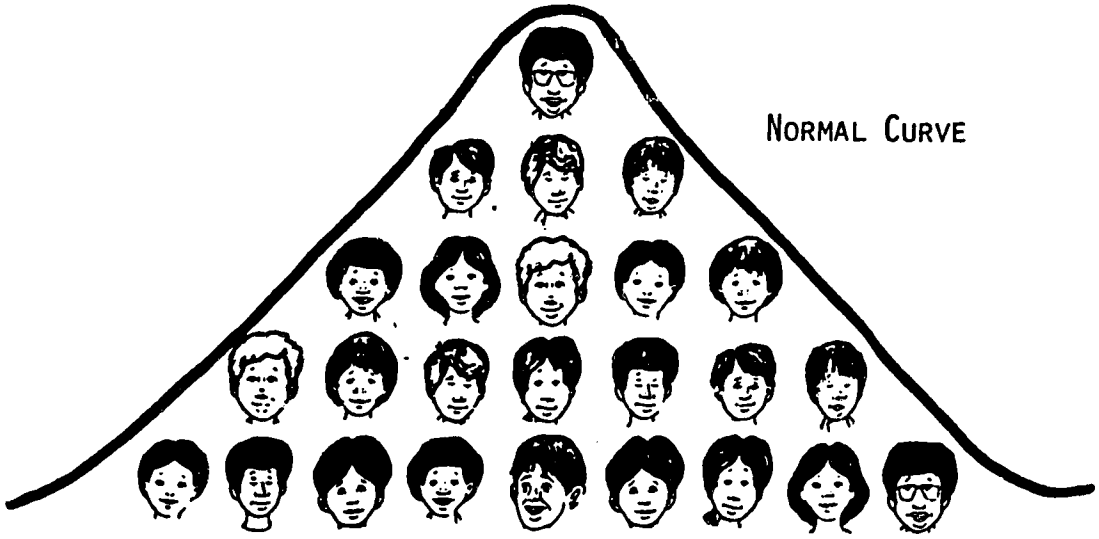
Figure 4 adds on several other terms which are sometimes encountered.

NORMAL CURVE: a distribution of scores that is shaped about the same as our distribution. This bell-shaped curve represents how skills such as vocabulary are distributed--most people being average with fewer and fewer people being at each point as we go higher or lower in vocabulary skill.

STANINES: a scale divided into nine equal parts. Notice that with our normal curve there are many more persons in the middle stanines and very few in the high and low stanines. Stanines and percentiles are always related in the same way. For example, the third stanine always includes persons scoring from the 11th to the 22nd percentile.

NORMAL CURVE EQUIVALENTS (NCE'S): roughly equal to stanines broken down into ten parts each. Stanines and NCE's are useful because they represent equal-interval scales and may be averaged. NCE's are sometimes used with test scores for the Title I Program.

PERCENTAGE OF ITEMS CORRECT: the raw score divided by the total number of items on a test (multiplied by 100 to remove the decimal point). This is often confused with percentile.



RAW SCORE	0-12	13	14	15	16	17	18	19	20
PERCENTILE	1 2 4	8	11 18	23 32	40 50	60 68	77 82	89 92	96 98 99
STAINE	1	2	3	4	5	6	7	8	9
NCE	10	20	30	40	50	60	70	80	90
% ITEMS CORRECT	0-60	65	70	75	80	85	90	95	100
GRADE EQUIVALENT	3.8	4.3	4.8	5.3	5.8	6.3	6.8	7.3	7.8

Figure 4. Norms associated with each raw score.

AUSTIN INDEPENDENT SCHOOL DISTRICT
Office of Research and Evaluation

Interpretation of Grade Equivalents and Percentile Scores

1. In grade equivalents, how much do the 3 months of summer count?

   a. 3 months
   ✓b. 1 month
   c. 0 month
   d. nothing

2. In which area is a third-grade student farther behind more of his peers if his grade equivalent scores are 2.8 in reading and 2.9 in math?

   a. Reading
   ✓b. Math
   c. Not known from this information

3. If a second grader gains in one year from a GE of 4.8 to a GE of 5.8, her percentile scores will most likely -

   a. go up
   b. stay the same
   ✓c. go down
   d. not known from this information

4. Which most likely represents the greatest gain in grade equivalents over a one-year period?

   a. 3.8 GE to 4.8 GE
   b. 20th %ile to 21st %ile
   ✓c. 90th %ile to 91st %ile
   d. Not known from this information

(More on back)

Would you use grade equivalents or percentiles to answer each of these questions? Think in terms of a low-achieving student.

5. How close to grade level is this student achieving?

   ✓a. Grade equivalents
   b. Percentiles
   c. Both

6. Did this student learn as much last year as did an average student?

   ✓a. Grade equivalents
   b. Percentiles
   c. Both

7. What proportion of students did this student score higher than?

   a. Grade equivalents
   ✓b. Percentiles
   c. Both

8. Did this student learn as much last year as did other students who were at the same low level?

   a. Grade equivalents
   ✓b. Percentiles
   c. Both