

# The Stanford Achievement Test, 9th Edition: National Norming and Performance Standards for Deaf and Hard-of-Hearing Students

Carol Bloomquist Traxler

Gallaudet Research Institute, Gallaudet University

This article presents a rich context of information for interpreting Stanford Achievement Test scores and for describing the achievement of deaf and hard-of-hearing students. The publisher's national norming of the Stanford Achievement Test provides a context of actual performance of hearing students. The publisher's Performance Standards provide a context of expectations for hearing students as determined by a panel of experts. The Gallaudet Research Institute's norming of the test on a national sample of deaf and hard-of-hearing students provides a context of test performance by this special population. A smaller subsample of the deaf and hard-of-hearing students who take the same test levels as hearing students provides an additional reference group with respect to the Performance Standards. Information from these sources is brought together into two graphical contexts to address these questions: Can the normative data from the publisher's national standardization of the test with hearing students, and the normative data from the GRI's national norming of the test with deaf and hard-of-hearing students provide a useful context for the interpretation of individual test scores? Can they provide a useful way to examine achievement of groups of students? Can the new Performance Standards defined by the test publisher offer a useful context for test score interpretation for high-achieving deaf and hard-of-hearing students?

This article is based on a paper presented at the Annual Meeting of the American Educational Research Association, April 19–23, 1999, Montreal. I thank Sue Hotto for constructing the figures used in this article and the editor and two anonymous reviewers for their helpful comments on an earlier version. Funding for the norming of the Stanford 9 was provided by the U.S. Department of Education, Office of Special Education and Rehabilitation Services (Grant #HO23C40030-95). Testing materials and scoring were provided by Harcourt Educational Measurement. Correspondence should be sent to Carol Bloomquist Traxler, Gallaudet Research Institute, 800 Florida Avenue NE, Washington, DC 20002 (e-mail: carol.traxler@gallaudet.edu).

©2000 Oxford University Press

For every new edition of the Stanford Achievement Test since the 6th edition, the Gallaudet Research Institute (GRI) has conducted a norming study that involved administering the test to a national sample of several thousand deaf and hard-of-hearing students. Along with materials to aid teachers in assigning appropriate test levels to individual students, the GRI has prepared age-based percentile norms for deaf and hard-of-hearing students.

The Stanford Achievement Test, 9th edition (Stanford 9; Harcourt Educational Measurement, 1996a), a highly regarded test, reflects content material commonly taught nationally in grades 1 through 9. The test has eight levels, and each subtest is vertically equated, so that the subtest has its own scaled score scale that allows score comparisons irrespective of test level. This feature allows a test user to examine a student's performance over time, as the student takes progressively more difficult levels of the test. It also allows the comparison of achievement of students who have taken different levels of the test. Using the scaled scores, one can report on individual students or groups of students by subtest as though they had taken the same test level and the identical test items.

This feature of vertical equating across test levels has led to screening procedures developed by the GRI for assigning test levels to individual students. Because optimal information cannot be gained from giving a test that is too easy for a student (the student answers all the presented items correctly) or one that is too difficult (a frustrated student may resort to guessing),

screening procedures were developed to allow each student to take a test level of appropriate difficulty. The screening procedures recommended by the GRI allow the teacher to examine objectives for each test level and select a brief screening test of approximately 20 items to verify that the selected level is appropriate for the student (see Gallaudet Research Institute, 1996b, for details). This contrasts with the testing of classes of hearing students, for whom test levels are routinely assigned by grade in school.

Screening procedures are the most salient feature of the administration of this test to deaf and hard-of-hearing students. The test booklets and answer sheets are identical to those used by hearing students, and the same time limits pertain. With deaf and hard-of-hearing students, the GRI recommends that test directions (but not individual items) be communicated using the students' usual mode of communication in the classroom and that additional sample items be used if the directions are not clear with only the samples provided. At the lowest two test levels, Primary 1 and 2, the test is dictated to hearing students, and the dictated portions are communicated to deaf and hard-of-hearing students using the students' regular mode of classroom communication (Gallaudet Research Institute, 1996d).

By using the scaled scores and grade equivalent scores from the national standardization of the Stanford 9 conducted by the test publisher (Harcourt Educational Measurement, 1996b), test users can compare the achievement of deaf and hard-of-hearing students to the achievement of hearing students of approximately the same achievement level. This comparison often involves comparing achievement of older deaf and hard-of-hearing students with that of younger hearing students, and with respect to test material suitable to the age and developmental level of the younger, hearing students.

By using the age-based percentiles for deaf and hard-of-hearing students, the test user can examine the achievement of an individual student with respect to the student's peers, irrespective of the test level taken by the other students.

Both contexts give useful information for the test user. Yet in both score interpretation contexts, there is of necessity a problem with test validity. On the one

hand, the test user examines a deaf student's achievement in comparison to that of younger hearing students on material that may appear juvenile to the deaf test taker. On the other hand, the test user examines a student's performance in the context of other deaf students the same age, even though the other students may have taken test items widely different in difficulty. These limitations are well known to test users and must be taken into consideration in test score interpretation.

A portion of this study focuses on a subgroup of deaf students for whom these validity considerations do not pertain. It focuses on deaf students who are close in age to the hearing peers to whom their achievement is compared and near the age (and developmental level) for which the content was deemed appropriate by the test developers. This group also is of special interest to many educators because they are likely to seek college admission. This group is presented with an additional score format that is new with the Stanford 9: Performance Standards.

This study, then, examines these questions of interest to researchers, educators, and administrators: Can the normative data from the publisher's national standardization of the test with hearing students, and the normative data from the GRI's national norming of the test with deaf and hard of hearing students, taken together, provide a useful context for the interpretation of individual test scores? Can they provide a useful way to examine achievement of groups of students? Can the new Performance Standards defined by the test publisher offer a useful context for test score interpretation for high-achieving deaf and hard-of-hearing students?

### **The Stanford 9 Norming Samples and Performance Standards Study Sample**

Table 1 summarizes information about the levels of the Stanford 9 from Primary 1 (P1) through Advanced 2 (A2). In this table the first five columns refer to information that is provided by the test publisher and is related to the customary use of the test with hearing students. The last two columns refer to the deaf and hard-of-hearing students who took the test in the 1996 norming conducted by the GRI.

An examination of the information presented in Table 1 for the Primary 3 level of the Stanford 9 shows

**Table 1** Stanford Achievement Test, 9th edition, recommended grade ranges, content level, performance standards target groups, and age of student samples

Test level	Test publisher's national samples and targets					Deaf and hard of hearing samples	
	Recommended grade ranges	Grade level of content covered by test	Grades for testing national standardization spring sample	Performance standards target group	Age of hearing students used for performance standard comparisons	Age of norming sample	Age of performance standards study sample
P1	1.5–2.9	1.5–2.5	1.8, 2.8	Spring, Grade 1	7	8–18	7–8
P2	2.5–3.9	2.5–3.5	2.8, 3.8	Spring, Grade 2	8	8–18	7–9
P3	3.5–4.9	3.5–4.5	3.8, 4.8	Spring, Grade 3	9	8–18	8–10
I1	4.5–5.9	4.5–5.5	4.8, 5.8	Spring, Grade 4	10	8–18	9–11
I2	5.5–6.9	5.5–6.5	5.8, 6.8	Spring, Grade 5	11	8–18	10–12
I3	6.5–7.9	6.5–7.5	6.8, 7.8	Spring, Grade 6	12	8–18	11–13
A1	7.5–8.9	7.5–8.5	7.8, 8.8	Spring, Grade 7	13	12–18	12–14
A2	8.5–9.9	8.5–9.9	8.8, 9.8	Spring, Grade 8	14	11–18	13–15

that the publisher recommends its use with students from the middle of third grade to the end of fourth grade. For testing conducted in spring, Primary 3 would generally be used with students in third grade. The test content for Primary 3 would generally be presented in the second half of third grade and the first half of fourth grade. The test publisher administered the Primary 3 test level in its spring standardization to students in grades 3 and 4. (The grade 4 students took subtests at Primary 3 and Intermediate 1 level, as participants in the vertical equating to help determine the scaled score scale.) The group of expert judges who set the Performance Standards for the Primary 3 test level considered students in spring semester of Grade 3 as their target group. These students would be approximately 9 years old.

The last two columns of information in Table 1 refer to the deaf and hard-of-hearing samples of this study. The norming sample is the large group of deaf and hard-of-hearing students who participated in the special norming conducted by the GRI in 1996. The Performance Standards (PS) study sample is a special subgroup examined in the study reported here. The

students in the PS study sample were judged by their teachers to be on or near grade level with their hearing peers and were assigned a corresponding test level. The deaf and hard-of-hearing students in the GRI norming sample who took Primary 3 test level ranged in age from 8 to 18. The PS study sample for Primary 3 includes only those deaf and hard-of-hearing students who are 8 to 10.

It may be seen as a limitation of this study that older deaf and hard-of-hearing students were not included in the PS study sample. These students were not available in the Stanford 9 norming database, for their more age-appropriate test levels were not selected by the GRI for norming on this population. The decision not to include the high school levels of the Stanford 9, the Test of Academic Skills (TASK), in the norming was based on the relatively small numbers of students at those levels and the less compelling need for separate norms for this group who are performing largely at grade level.

Table 2 shows the number of students, by age and test level, in the GRI's national norming sample and in the PS study sample. The norming sample consists of

**Table 2** Stanford Achievement Test, 9th edition, deaf and hard of hearing norming sample and performance standards study sample

Age	Test level								Sample		
	P1	P2	P3	I1	I2	I3	A1	A2	Norm <i>n</i>	Performance standards	
										<i>n</i>	%
7	(90) <sup>a</sup>	(14) <sup>a</sup>								104 <sup>a</sup>	(100)
8	198 <sup>a</sup>	59 <sup>a</sup>	10 <sup>a</sup>	1	1	1			270	267 <sup>a</sup>	99
9	211	110 <sup>a</sup>	32 <sup>a</sup>	4 <sup>a</sup>	1				358	146 <sup>a</sup>	41
10	205	121	80 <sup>a</sup>	28 <sup>a</sup>	5 <sup>a</sup>	1			440	113 <sup>a</sup>	26
11	105	126	93	48 <sup>a</sup>	28 <sup>a</sup>	7 <sup>a</sup>		1	408	83 <sup>a</sup>	20
12	105	118	86	58	30 <sup>a</sup>	29 <sup>a</sup>	14 <sup>a</sup>	2	442	73 <sup>a</sup>	17
13	100	89	111	57	46	30 <sup>a</sup>	25 <sup>a</sup>	18 <sup>a</sup>	476	73 <sup>a</sup>	15
14	79	97	82	64	52	40	37 <sup>a</sup>	27 <sup>a</sup>	478	64 <sup>a</sup>	13
15	62	102	86	57	46	49	47	48 <sup>a</sup>	497	48 <sup>a</sup>	10
16	67	81	98	71	53	41	53	49	513	—	—
17	74	76	67	56	40	46	58	91	508	—	—
18	46	42	65	69	38	38	51	69	418	—	—
Norming sample	1252	1021	810	513	340	282	285	305	4808	971 <sup>a</sup>	
PS study sample	288 <sup>a</sup>	183 <sup>a</sup>	122 <sup>a</sup>	80 <sup>a</sup>	63 <sup>a</sup>	66 <sup>a</sup>	76 <sup>a</sup>	93 <sup>a</sup>	971 <sup>a</sup>		

The norming sample cases are ages 8–18 only.

<sup>a</sup>These are the performance standards study sample cases. The right columns give the number of students in the performance standards study sample, by age, and the percentage of the norming sample they comprise.

4,808 students ages 8 through 18 who took Stanford 9 test levels Primary 1 through Advanced 2. The PS study sample includes 971 students ages 7 through 15 who took the same test levels. Only the 104 students age 7 in the study sample were not included in the national norming of the test. For each test level, the PS study sample includes those deaf and hard-of-hearing students in the norming sample who were the same age as the hearing students for whom the test level was designed. For example, the students in the PS study sample for the Intermediate 3 test level were the 66 students ages 11 through 13.

The right columns of Table 2 show the number of students in the PS study sample, by age, and the percentage of the norming sample they comprise. Nearly all (99%) of the 8-year-olds in the norming sample were included in the PS study sample, but only 10% of the 15-year-olds were included. Only 8 of the 4,808 deaf students were assigned test levels that were higher than those given to their hearing peers (ranging from 0 to 1%). But well over half (as many as 90% of the 15-year-olds) were assigned test levels clearly lower than

those of their hearing peers. Table 2 allows the reader to see that nearly all the 7- and 8-year-olds are in the PS study sample, with smaller proportions included as age increases.

The norming sample is described in greater detail elsewhere in terms of individual student characteristics, such as gender (53% female), ethnic background (54% white, 18% Black, 19% Hispanic), level of hearing loss (28% less than severe, 21% severe, 51% profound), additional physical (8%) or cognitive (24%) disabilities, age at onset of hearing loss (96% at birth or before age 3) and cause of hearing loss (see Holt, Traxler, & Allen, 1996, p. 5). The sample was drawn to be representative of regions of the country and program type (Holt et al., 1996, pp. 2–4). For the norming study, 23% of the weighted sample were in special programs (residential or day schools for the deaf) and 77% in local schools (public or private local school programs with full-time or part-time special education classes). The norming sample was weighted in these individual student characteristics as well as by region and program type to be similar to the database developed by

the GRI through its Annual Survey of Deaf and Hard-of-Hearing Children and Youth, which was used as a sampling framework.

The smaller PS study sample is *not* representative of the national database, however, and it should not be regarded as a random sample. The PS study sample happens to be quite similar to the norming sample in gender and in age at onset of hearing loss, but it is not similar in other aspects. For example, the PS study sample has a larger proportion of students with no additional physical or cognitive conditions. Because it is not the purpose of this study to compare the PS study sample and the norming sample, note that there are differences between these samples in characteristics frequently taken into consideration in examining student achievement.

### Performance Standards

The Performance Standards used in this study are those developed by the test publisher as a reference for test score interpretation (Harcourt Educational Measurement, 1997a, 1997b). To establish the Performance Standards, approximately 200 teachers representing the nation's school districts with respect to "all content areas and grade levels, important school district demographic variables, and major ethnicities and cultures" were convened for a 3-week series of standard-setting meetings. Using a modified Angoff (1971) procedure, they reviewed every item on the Stanford battery and made judgments about how they would expect students at various ability levels to perform on the items. After several rounds of judgments and feedback, final judgments were made. The result of these judgments was the establishment of four Performance Standards labeled "advanced," "proficient," "basic," and "below basic." It is important to remember that these standards were determined with hearing students, not deaf students, in mind. The four Performance Standards levels appear in Table 3. The point in the school year for which the standards were established is given in Table 1 as spring semester of grades 1 through 8. The corresponding ages of the hearing students whose performance with respect to the Performance Standards is shown is given in Table 3 as ages 7 through 14. It should be noted that the reference to high school level

**Table 3** Description of performance standards levels for the Stanford Achievement Test, 9th edition

Level	Description
4	Advanced, represents <i>superior performance</i> beyond grade-level mastery. At the high school levels, students achieving at this level show readiness for advanced academic courses, advanced technical training, or career-oriented employment.
3	Proficient, represents <i>solid academic performance</i> , indicating that students are prepared for the next grade. At high school, this level reflects competency in a body of subject-matter knowledge and skills that prepares students for responsible adulthood and productive work.
2	Basic, denotes <i>partial mastery</i> of the knowledge and skills that are fundamental for satisfactory work. At the high school level, this is higher than minimum competency skills.
1	Below basic, indicates <i>less than partial mastery</i> .

Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, p. 41. San Antonio: Harcourt Educational Measurement, 1997.

in the descriptions in Table 3 does not pertain to most of the deaf and hard-of-hearing students in this study. The highest level of the Stanford 9 used in the development of norms for deaf and hard of hearing students (Advanced 2) extends no higher than ninth grade curriculum content.

### Method

Scaled scores on the six Stanford 9 subtests normed for deaf and hard-of-hearing students were examined for the norming sample to provide a context in which to interpret individual student scores (Gallaudet Research Institute, 1996a). The six normed subtests are Reading Comprehension, Reading Vocabulary, Mathematics: Problem Solving, Mathematics: Procedures, Language, and Spelling.

The two mathematics subtests on the Stanford 9 are dramatically different from the three mathematics subtests that appeared on the previous edition of the Stanford, and student performance can be expected to reveal the absence of the new test content from the curriculum for these students. Because the publisher's norming study and the GRI norming study both occurred in spring 1996, however, the groups of students can be expected to be at equal disadvantage with re-

spect to exposure to content and skills measured in the new mathematics subtests. The hearing students as well as the deaf and hard-of-hearing students whose achievement is reported here were all facing a new test that may not have reflected their school's mathematics curriculum.

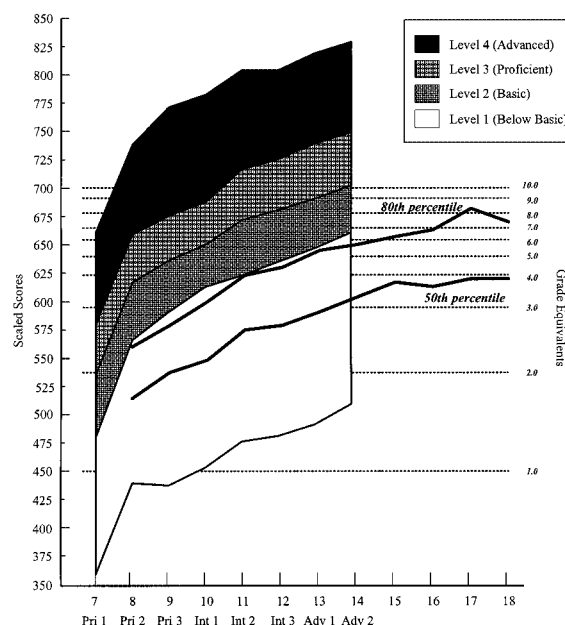
The performance of the deaf and hard-of-hearing students in the GRI's norming sample and of the PS study sample was studied in the context of the Performance Standards. In addition, the percentage of the PS study sample achieving each of the four Performance Standard levels was examined in tandem with the parallel results reported for the hearing students ages 7 through 14 in the test publisher's standardization of the Stanford 9 (see Table 1). To enable these comparisons, the percentages of hearing students ages 7 through 14 whose scores fell into the four Performance Standards levels were obtained, and the corresponding percentages of deaf and hard-of-hearing students in the PS study sample were computed (Harcourt Educational Measurement, 1997a, 1997b). The information was laid out in graphical form to evaluate its usefulness for describing and interpreting test scores and student achievement.

## Results

### Deaf and Hard-of-Hearing Students in the Norming Sample

Figure 1 shows the median (50th percentile) and 80th percentile performance of all of the deaf and hard-of-hearing students in the Stanford 9 norming for ages 8 through 18 on the Reading Comprehension subtest. This performance is shown in the context of other information related to the performance of hearing students.

The vertical axis on the left shows scaled scores for the Reading Comprehension subtest. The vertical axis on the right and the corresponding dotted horizontal lines show the grade equivalents associated with those scaled scores. Grade equivalent scores are the median scores for hearing students at those points (where 3.0 is the first month of third grade, for example). Along the horizontal axis, the age of the deaf students in the

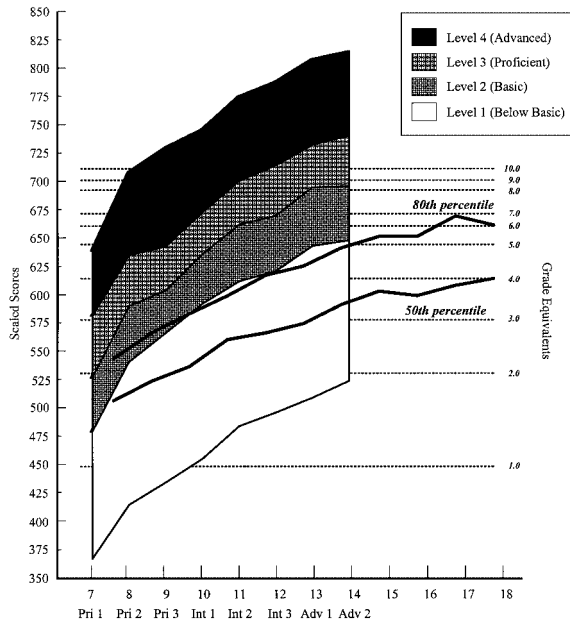


**Figure 1** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Reading Comprehension With Performance Standard Levels.

norming sample is given, for whom the 50th and 80th percentile scores are plotted. The age is also that of the target group of hearing students for whom the Performance Standards were set. Below each age is shown the associated test level.

The Performance Standards levels are shown in four bands that form a background for the figure. That is, above the age 9 marker on the horizontal axis, the Performance Standards levels for Primary 3 (as judged for 9-year-old hearing students in spring of grade 3) are shown. The lowest (and lightest) band shows the area judged to be Level 1: Below Basic. The next band shows the area judged to be Level 2: Basic. The next higher band shows the area judged to be Level 3: Proficient, and the highest (and darkest) band shows Level 4: Advanced.

The median Reading Comprehension scores, by age, for the entire group of deaf and hard-of-hearing students in the norming sample fall largely in the Level 4: Below Basic area. Many of these students are indeed placed below grade level in school, when compared to hearing students of the same age. The 80th percentile line, which lies just below the border between Level 1 and Level 2, shows that many of the top fifth of the deaf

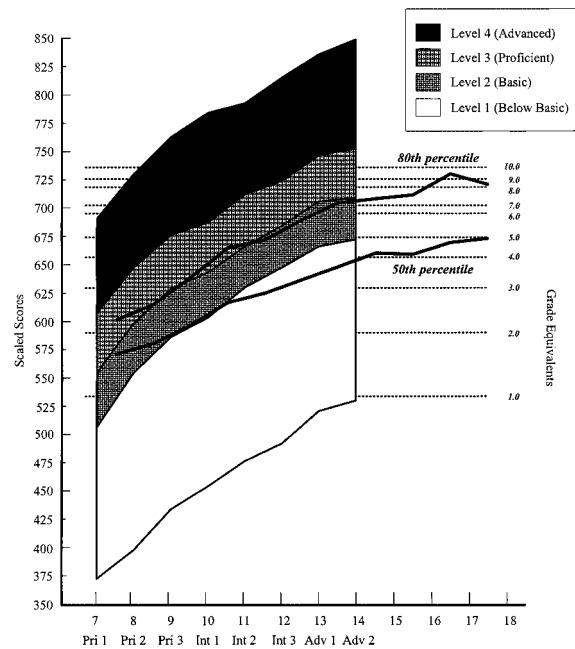


**Figure 2** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Reading Vocabulary With Performance Standard Levels.

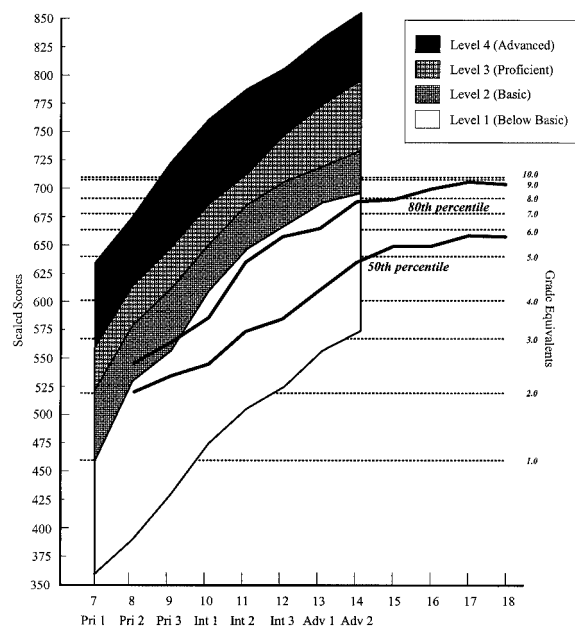
and hard-of-hearing students in the national norming sample (whose scores lie above the 80th percentile) were likely functioning at about the Level 2: Basic level or higher.

Figures 2 through 6 show the performance of the GRI norming sample on the other subtests normed for deaf and hard-of-hearing students: Reading Vocabulary, Mathematics: Problem Solving, Mathematics: Procedures, Language, and Spelling, respectively. The Reading Vocabulary and Mathematics: Procedures performance is very similar to Reading Comprehension, with both the 50th and 80th percentile lines falling largely in Level 1 (Below Basic). Mathematics: Problem Solving and Spelling show the 80th percentile to fall mostly in Level 2 (Basic). In Language the 80th percentile line falls on the border between Level 1 and Level 2. The performance of the entire top fifth of this national sample of deaf and hard-of-hearing students appears to be clearly and consistently above Level 1 only in Mathematics: Problem Solving and in Spelling.

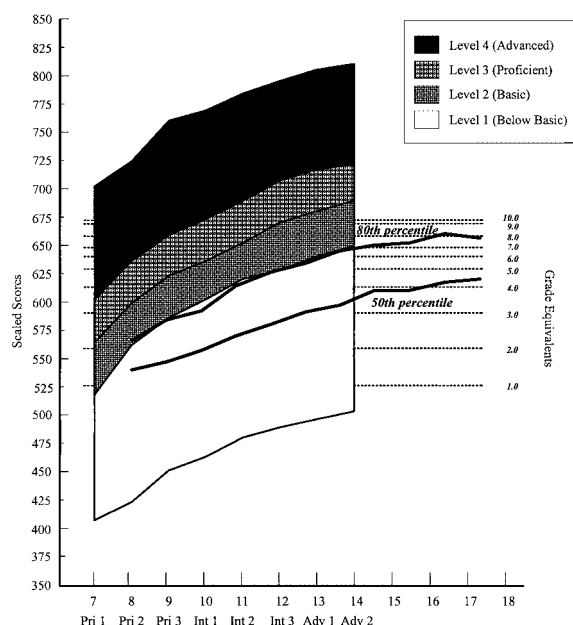
These figures build on those in the score summary folders developed by the GRI (Gallaudet Research Institute, 1996c), which provide a context for tracking student achievement over time with respect to norma-



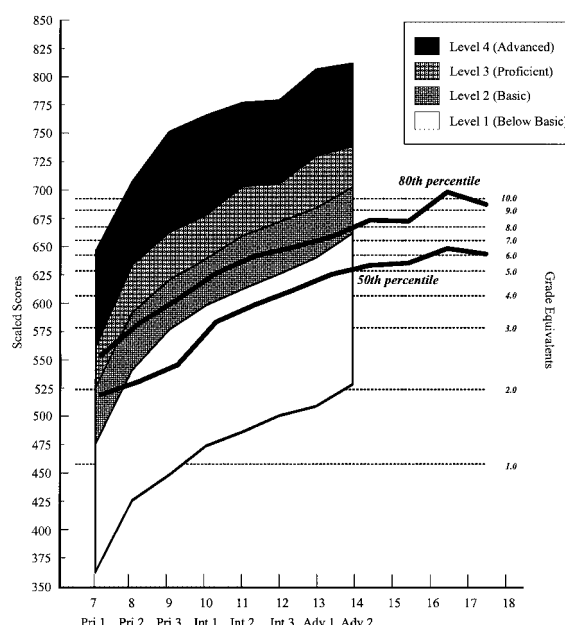
**Figure 3** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Mathematics: Problem Solving With Performance Standard Levels.



**Figure 4** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Mathematics: Procedures With Performance Standard Levels.



**Figure 5** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Language With Performance Standard Levels.



**Figure 6** National Median and 80th Percentile Scores for Deaf and Hard-of-Hearing Students on Stanford 9 Spelling With Performance Standard Levels.

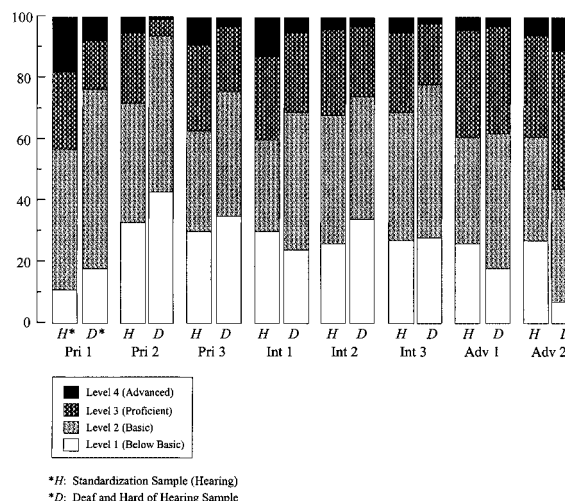
tive information both for hearing students and for deaf and hard-of-hearing students. The score summary folders do not, however, include the 80th percentile reference or the Performance Standards context examined here.

Figures 1 through 6 summarize a great deal of information. They enable the test score interpreter simultaneously to take into account relevant information in discussing individual or group performance. However, these figures do not allow for easy comparisons of the subgroup of deaf and hard-of-hearing students who are judged by their teachers to be functioning at or close to grade level with their hearing peers. For that information, test performance results for the PS study sample are presented.

### Performance Standards Study Sample

Figure 7 shows the percentage of students whose Reading Comprehension subtest scores put them at each of the four Performance Standards levels. Each bar in the stacked bar chart represents 100% of the students labeled. For each pair of bars, the bars on the left, marked H, refer to the hearing students in the publisher's standardization sample for the Stanford 9. The

bars on the right in each pair, marked D, refer to the deaf and hard-of-hearing students in the PS study sample, who are about the same age as the hearing students. That is, for each test level the pair of stacked



**Figure 7** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Reading Comprehension. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.



bars summarizes the performance of the two groups of students, with hearing students shown on the left as indicated.

The lowest (and lightest) indicator bar represents Level 1 (Below Basic). Level 2 (Basic) is shown next. The Level 3 (Proficient) performance is next and the Level 4 (Advanced) performance is the topmost (and darkest) bar.

The pair of bars at the right in Figure 7 show the Advanced 2 test level in Reading Comprehension (taken by deaf and hard-of-hearing students ages 13 through 15). The deaf and hard-of-hearing students whose performance is depicted outperform the hearing students (age 14) with respect to the Performance Standard levels achieved. That is, the two topmost bars are wider for deaf and hard-of-hearing students, indicating that a higher percentage of them had achieved scores falling in Level 3 (Proficient) and Level 4 (Advanced) than the hearing students at Advanced 2. More than half of these deaf and hard-of-hearing students are at Level 3 (Proficient) or higher, as compared to only approximately 40% of the hearing students with that level of achievement. At Intermediate 1 through Intermediate 3 test levels the hearing and deaf students attain the four Performance Standard levels in roughly similar percentages. At the lowest three Stanford 9 test levels, Primary 1 through Primary 3, the youngest deaf and hard-of-hearing students are clearly outperformed by their same-age hearing peers in Reading Comprehension.

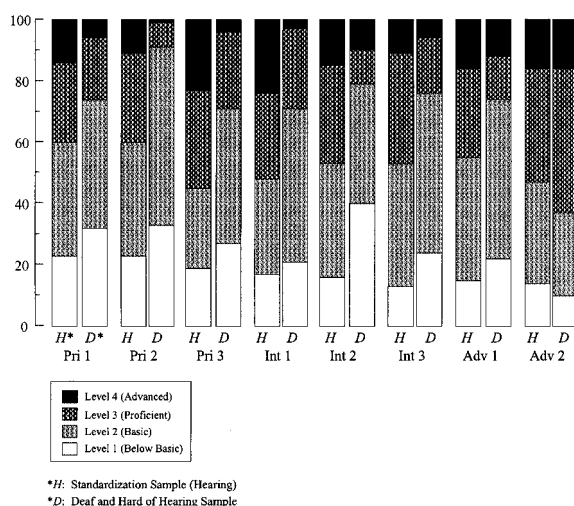
These figures are presented as a possible context for score interpretation, not as a representation for statistical comparison of group performance. The differences in Performance Standard categories attained by the hearing and the deaf and hard-of-hearing students are not examined statistically, for the PS study sample was not randomly drawn and is not considered representative of a larger population of deaf and hard-of-hearing students. The PS study sample was selected specifically because these students were given the same test levels as their hearing same-age peers.

The PS study sample is composed of varying proportions of the deaf and hard-of-hearing norming sample, as shown in the far right column of Table 2. The PS study sample, the select group of deaf and hard-of-hearing students who were judged by their

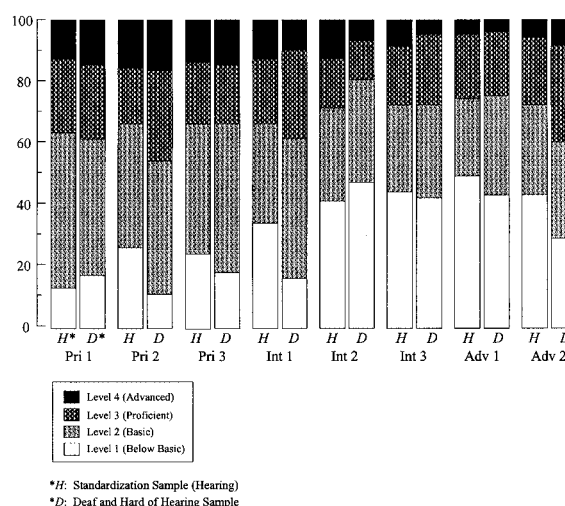
teachers to be achieving largely at grade level (and thus given the same grade-appropriate test level as is given to hearing students), contains ever decreasing percentages of the students available at each age. Some of the relatively low performance of the youngest deaf and hard-of-hearing students, especially when compared to the performance of the older students, may be explained by the fact that there were no easier tests available in the norming study to give these children. Their teachers elected to include them in the testing, assigning to 198 of the 270 8-year-olds the level Primary 1, the lowest level available. Virtually all (267 of 270, or 99%) of the 8-year-olds with Stanford 9 data available (see Table 2) were included in this PS study sample for test levels Primary 1 through Primary 3. Although 7-year-olds were not specifically included in the norming sample (the requested age range was 8 through 18), the 104 7-year-olds in the PS study sample probably were in classes with other students who were being tested, and their teachers chose not to exclude them from the testing. It cannot be said whether these 7-year-olds are typical of other deaf and hard-of-hearing students their age. In contrast to the large percentage of the youngest students included in the PS study sample, only the top 10% to 15% of the deaf and hard-of-hearing 13- to 15-year olds in the norming sample were in the PS study sample (see Table 2).

Figures 8 through 12 show the percentage of students at each of the four Performance Standards levels on the Reading Vocabulary, Mathematics: Problem Solving, Mathematics: Procedures, Language, and Spelling Stanford 9 subtests, respectively. In general, hearing and deaf students had fairly similar proportions in the four Performance Level categories except for the older students at Advanced 2, where consistently higher percentages of the select group of deaf and hard-of-hearing students achieved Level 3 (Proficient) or higher.

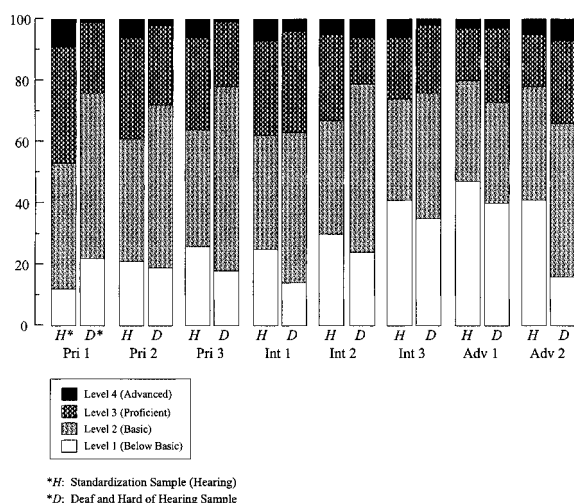
Test scores of the select group of deaf and hard-of-hearing students in the PS study sample did fall into the four Performance Standards categories, just as they did for hearing students. That means it is possible to discuss these students' scores in the context of expectations, irrespective of the performance of other students. It is also possible to consider the performance of deaf and hard-of-hearing students with respect to these Per-



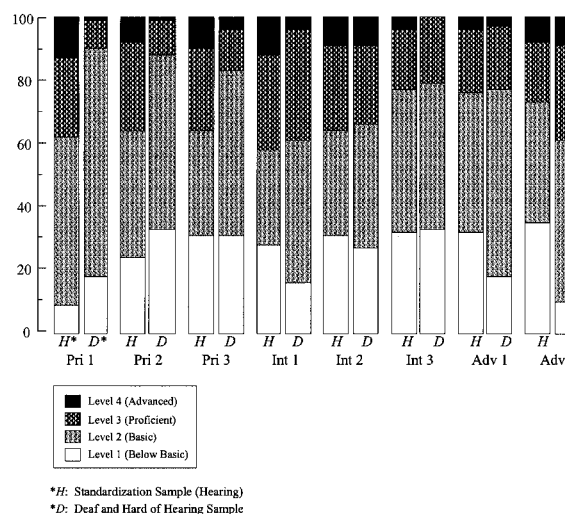
**Figure 8** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Reading Vocabulary. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.



**Figure 10** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Mathematics: Procedures. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.



**Figure 9** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Mathematics: Problem Solving. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.

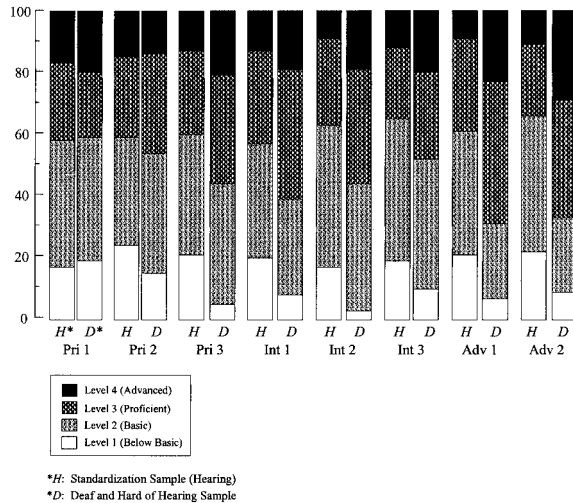


**Figure 11** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Language. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.

formance Standards expectations while comparing them to selected hearing and deaf peers (Figures 7–12).

These Performance Standards figures can be used as a reference point in interpreting the scores of specific deaf and hard-of-hearing students who are as-

signed test levels corresponding to the PS study sample (see Table 2), that is, similar to hearing students. It should be remembered that the figures show results for a selected group of students, ranging from 99% of the 8-year-olds to the top 10% of the 15-year-olds in the



**Figure 12** Percentage of Students in Each Performance Standard Category for the Stanford 9 Standardization Sample and the Deaf and Hard-of-Hearing Study Sample in Spelling. Source: Stanford Achievement Test Series, Ninth Edition, Technical Data Report, pp. 459–465.

norming sample, with as few as 63 students at one test level (Intermediate 2; see Table 2). Performance of the deaf and hard-of-hearing students relative to their same-age hearing peers is noticeably favorable in Spelling and unfavorable in Reading Vocabulary (Primary 1 through Advanced 1) and Language (Primary 1 through Primary 3). In each case, the selectivity of the PS study sample should be remembered.

## Discussion

The achievement of a representative sample of deaf and hard-of-hearing students obtained from the national norming of the Stanford 9 is summarized in graphical form. This summary allows test scores to be interpreted in the context of a variety of normative information. Scaled scores allow the tracking of performance of an individual or group longitudinally for each of six subtests. Associated with each scaled score is a grade equivalent score that estimates the grade and month in school at which it is the median score for hearing students. These grade equivalent scores indicating typical performance of hearing students at the beginning of each grade (3.0, for example) are also indicated with horizontal dotted lines.

Beyond these norms for hearing students, there are

norms for deaf and hard-of-hearing students to enrich the context for score interpretation. These norms include the 50th percentile and 80th percentile performance of deaf and hard-of-hearing students from ages 8 to 18.

In addition to this normative information, there is an additional context available for interpreting individual scores and for describing student achievement: Performance Standards, four levels of performance defined by an expert panel who set the levels with respect to hearing students in the spring of grades 1 through 8 (corresponding roughly to ages 7 through 14). Test score interpretation can be conducted considering the information from the publisher's national norming of the Stanford 9 on hearing students, the GRI's national norming of the test on deaf and hard-of-hearing students, and the publisher's expert panel who set Performance Standards.

The performance of deaf and hard-of-hearing students who perform at a level similar to the hearing peers may be examined in the context of these expectations set by the independent expert panel. Selected deaf and hard-of-hearing students who comprised a Performance Standards study sample achieved quite similarly to their hearing peers for the Performance Standards levels in Reading Comprehension. In general, for both the hearing and deaf students, approximately 60% achieved no higher than Level 2 (Basic). There is room for improvement for all our students in reading comprehension if they are to be judged as Proficient (Level 3) or higher by those who set the Performance Standards. In Reading Vocabulary, Mathematics: Problem Solving, and Mathematics: Procedures, the deaf and hard-of-hearing students in this sample performed similarly to their hearing peers on the Performance Standards. In Language they clearly performed lower, but in Spelling clearly higher. The context of Performance Standards appears to be a reasonable one for interpreting scores of deaf and hard-of-hearing students who perform at grade level.

These results are based on a small number of students who were purposefully selected from a larger random sample (the norming sample). The selected PS study sample is not a random sample, and the two groups differ in characteristics often associated with achievement.

This new context for the interpretation of the academic performance of selected deaf students (Performance Standards levels) appears to be meaningful, allowing test users to have the advantage of understanding the student's achievement in the performance terms also employed with hearing students the same age. Researchers, educators, and administrators working with deaf and hard-of-hearing students may interpret the academic achievement of individual students and of groups of students in the context of the Performance Standards expected of hearing students nationally.

Received January 19, 2000; revised May 11, 2000; accepted May 12, 2000

## References

- Angoff, W. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508–600). Washington, DC: American Council on Education.
- Gallaudet Research Institute. (1996a). Stanford Achievement Test, 9th Edition, Form S: Norms booklet for deaf and hard-of-hearing students. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (1996b). Stanford Achievement Test, 9th Edition: Screening procedures for deaf and hard-of-hearing students. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (1996c). Stanford Achievement Test, 9th Edition: Score summary folder for deaf and hard-of-hearing students. Washington, DC: Gallaudet University.
- Gallaudet Research Institute. (1996d). Stanford Achievement Test, 9th Edition: Administration procedures for deaf and hard-of-hearing students. Washington, DC: Gallaudet University.
- Harcourt Educational Measurement. (1996a). Stanford Achievement Test Series, Ninth Ed. San Antonio, TX: Harcourt Educational Measurement.
- Harcourt Educational Measurement. (1996b). Stanford Achievement Test Series, Ninth Ed. Spring norms book. San Antonio, TX: Harcourt Educational Measurement.
- Harcourt Educational Measurement. (1997a). Stanford Achievement Test Series, Ninth Ed. Technical data report. San Antonio, TX: Harcourt Educational Measurement.
- Harcourt Educational Measurement. (1997b). Stanford 9 special report: Performance Standard scores. San Antonio, TX: Harcourt Educational Measurement.
- Holt, J. A., Traxler, C. B., & Allen, T. E. (1997). Interpreting the scores: A user's guide to the 9th Edition Stanford Achievement Test for educators of deaf and hard-of-hearing students. Gallaudet Research Institute Technical Report 97-1. Washington, DC: Gallaudet University.