ACT Research Report Series 90-9

Differential Test and Item Performance of Talented Youth and High School Seniors on the ACT

Allen E. Doolittle
Catherine J. Welch

October 1990

ACT

DIFFERENTIAL TEST AND ITEM PERFORMANCE OF TALENTED YOUTH AND HIGH SCHOOL SENIORS ON THE ACT

Allen E. Doolittle
Catherine J. Welch

The American College Testing Program

ABSTRACT

The performance of academically talented seventh grade students on the ACT Assessment (ACT) was investigated with respect to college-bound, twelfth grade examinees. Differential test and item performance were examined based on one form of the ACT, and the results were cross-validated on a different form. Although the twelfth grade examinees outperformed the seventh graders on all tests, the differences were greatest on the Mathematics and Social Studies Reading tests. In addition, systematic performance differences between the two groups of students were found at the item level. The nature and implications of these differences are discussed.

Differential Test and Item Performance of Talented Youth and High School Seniors on the ACT

The early identification of academically talented students has long been of interest to educators. One approach has been to identify able students on the basis of their performance on one of the major standardized achievement batteries that are administered in school systems throughout the country. However, these achievement batteries are typically designed to measure performance for a wide range of students and, as a result, do not differentiate particularly well at the top end of the score scale. Beginning in 1971 with a mathematics talent search sponsored by Johns Hopkins University (Stanley, 1976), the SAT has been used as an "out-of-level" test to provide adequate ceiling for very able seventh graders. Subsequently, four regional, university-based talent searches have been developed in the United States, all using the SAT as an important part of the identification process. Duke University's Talent Identification Program (TIP) is the regional search that focuses on 16 southeastern and midwestern states, bordered by Iowa and North Carolina in the north and Texas and Florida in the south.

Because the ACT Assessment (ACT) is the most commonly used college admissions test in the TIP region, the use of the ACT as an alternative for selecting academically talented seventh graders has recently been explored. Beginning in 1987, TIP and ACT began working cooperatively to investigate the ACT for this

purpose. (Dreyden & Stanley, 1988, Maxey & Dreyden, 1988; Sawyer & Brounstein, 1988; Stanley & York; 1988.)

Dreyden and Stanley (1988) investigated TIP examinees (seventh graders) and found that the population was very homogeneous in terms of test performance and had uniformly positive attitudes toward academic achievement and self perception. The observed homogeneity of the population on these characteristics was consistent for TIP examinees from different gender, race, and income level groups.

Maxey and Dreyden (1988) investigated the statistical relationships between ACT scores and scores on standardized achievement tests used in elementary and junior high schools. Generally they found fairly weak correlations and attributed these results to a lack of differentiation by the achievement tests at the top end of the scale. These results are consistent with the rationale for using either the ACT or the SAT rather than the achievement tests to identify talented students at the top end of the distribution.

Stanley and York (1988) summarized part of the deliberation process for deciding to offer the ACT as an alternative to the SAT for qualifying TIP applicants. They reported that the ACT and SAT seemed to be equally valid for TIP selection and that the ACT was a particularly good alternative because of the accessibility of ACT test centers in the TIP region.

The objective of this research was to examine the performance in detail of very able seventh graders relative to college-bound high school students on the ACT.

METHODOLOGY

Research Plan

There were two stages to this research, each involving an independent data set. The first stage was designed to identify areas of differential performance on one form of the ACT, and the second stage was to confirm or clarify the observations from the first stage on comparable, but different students and a second, equated form of the ACT.

Data Source

The TIP examinee population consisted of seventh-grade students in the 16-state region covered by the program who had accepted an invitation to take the test. These students all had scores in the upper 3 percent of any of several nationally standardized academic achievement tests. Approximately 55% of these students were female.

The high-school examinee population consisted of college-bound students (primarily seniors) who were taking the test as part of the college admissions process. Although the population was national in scope, most examinees were from midwestern and southern states. As for the TIP applicant population, approximately 55% of the high school examinees were female.

The initial stage of this research was based on random samples of 2,000 each of seventh grade (TIP) and high school (HS) students taken from the February 1988 administration of the ACT. Data from comparably drawn samples were taken from the February 1989 administration for the second stage.

Instrument

The ACT forms used in this study consisted of four, separately timed tests: English Usage, Mathematics Usage, Social Studies Reading, and Natural Sciences Reading. Differential performance between the TIP and the high school students was investigated on all four tests. However, it should be noted that with the October 1989 administration (after the data for this study were collected), the ACT was revised to reflect different test specifications. Four tests are also included in the new ACT: English, Mathematics, Reading, and Science Reasoning. Because of the differences in the test specifications, the results of this study are not completely generalizable to the new test.

Measurement of Differences

The measurement of differential performance was conducted at both the test and individual item levels. Differential test performance was assessed based on mean differences and effect size (mean differences ÷ standard deviation). Differential item performance was evaluated using the Mantel-Haenszel common-odds ratio (Holland & Thayer, 1986; Mantel & Haenszel, 1959). The log of the common-odds ratio transforms it to a symmetric scale centered around zero. The value of the transformed ratio is the

average amount more difficult that members of one group find the item than do comparably scoring members of another group.

RESULTS

Table 1 presents the standard score means, standard deviations, *t*-statistics, and effect sizes found for each test. These results indicate that the high school students for the first sample performed better than TIP candidates on all tests. The largest differences were on the Social Studies Reading and the Mathematics tests. In terms of the magnitude of performance differences on the tests, effect sizes were moderate to fairly sizable. The English and the Natural Sciences Reading effect sizes were .42 and .37, respectively. The effect sizes for the Mathematics and Social Studies Reading scores were larger, .73 and .79 respectively. On the standard score scale, these differences ranged from 2.1 on the English and Natural Sciences Reading tests to 4.6 on the Mathematics test and 4.9 on the Social Studies Reading test.

Insert Table 1 About Here

Gender differences in performance were found, but the differences seemed to be consistent for both populations: females outperformed males on the English test and males scored higher than females on the Mathematics, Social Studies and Natural Sciences Reading tests. Effect sizes for mean differences between the samples, when analyzed separately by gender, tended to be consistent with the differences found for the total samples.

Mantel-Haenszel procedures were used, but not in the typical sense of identifying individual items for differential performance. This use of differential item performance methodology was adapted from Doolittle and Welch (1989) and was exploratory in nature. The objective of this procedure was to identify categories of items that favored either the TIP students or the regular examinee group of high school seniors. A summary of these exploratory analyses, based on the 1988 administration data (Sample 1), is presented in Table 2. In these analyses, a criterion of ± 1.0 on the M-H delta was used to identify items that seemed to perform differently for the two groups. The number of items in each category that seemed to favor the regular or the TIP examinees is shown.

Insert Table 2 About Here

Generally the results in Table 2 portray seemingly random distributions of items favoring TIP or HS examinees in the subcategories. However, there were some item categories that seemed to suggest significant differences in group performance. These categories were:

- English "grammar" items--4 items favoring TIP candidates to 1 item favoring the regular sample of high school (HS) seniors;
- English "punctuation" items--5 to 1 favoring TIP students;
- English Passage 2 (topic: black women authors)--6 to 0 favoring the HS examinees;
- English Passage 5 (topic: comic books and other collectibles)--4 to 0 favoring the TIP students;
- "Arithmetic/algebraic reasoning" items--4 to 0 favoring TIP students;
- "Intermediate algebra" items--3 to 1 favoring the HS examinees;
- "Geometry" items--3 to 0 favoring the HS examinees;
- Social Studies discrete (knowledge) items--3 to 1 favoring the HS examinees;
- Natural Sciences Passage 1 (topic: Pluto's origin in the solar system)--4 to 0 favoring the TIP students;
- Natural Sciences Passage 4 (topic: cilia function)--3 to 0 favoring the HS examinees.

Table 3 presents a summary of the Mantel-Haenszel analyses by item categories, but for different samples of students and a different test form than that summarized in Table 2. To the extent that the results are comparable, the findings in Table 3 may be viewed as confirming those in Table 2.

Insert Table 3 About Here

As in Table 2, there were some item categories that seemed to suggest a leaning toward one group or another, and in many cases they were in the same direction for the data in both tables. In this group of similarly aligned item categories were:

- English "punctuation" items (favoring TIP students);
- "Arithmetic/algebraic reasoning" items (favoring TIP students);
- "Intermediate algebra" items (favoring the HS examinees);
- "Geometry" items (favoring HS examinees).

In contrast, there were two categories of items--the English "grammar" and Social Studies discrete items--for which the results on the second form did not confirm findings from the first.

Because two unique forms were used for the two analyses, the relative performance of the populations on items from individual passage sets could not be confirmed. However, the results are still useful in that they may suggest reasons for apparent differential performance. Table 4 summarizes the topics of the passage sets that contained a disproportionate number of items favoring one group or another.

Insert Table 4 About Here

DISCUSSION

The outcomes of this research suggest that there are some types of test items in the ACT Assessment that tend to perform differently for talented seventh graders than they do for college-bound high school students. There is also some evidence that passage topic may affect subgroup performance on associated item sets. Because the ACT Assessment is designed to be sensitive to educational achievement, however, it should not be surprising that seventh grade students, regardless of how talented they are, may have difficulty with test items that cover material that hasn't been taught to them. Good examples here are Intermediate Algebra and Geometry items. These item categories, covering material typically taught in the 9th, 10th, or 11th grades, were disproportionately difficult for the 7th-grade TIP students. Figure 1 presents examples of these items.

Insert Figure 1 About Here

There are other categories of items on the ACT Assessment that, though challenging for its target population of college-bound high school students, deal with concepts with which many seventh graders are familiar. Many of these items are identified as tending to favor the TIP students--that is, with respect to the high school population, the TIP students perform disproportionately better on these items than on other items in the test. Within mathematics, "Arithmetic and Algebraic Reasoning" items (word problems) tend to favor TIP students.

These items do not typically require more than basic arithmetic or algebraic concepts, but they do require substantial problem solving skills. Bright seventh graders might be predicted to do relatively well on these items, and they do. Figure 2 presents two examples of Arithmetic and Algebraic Reasoning items that seem to favor the TIP examinee population.

Insert Figure 2 About Here

Another item category that seemed to favor the TIP students was the "Punctuation" category in the English Usage Test. Because punctuation is a general topic that is initially presented to students during elementary school, the strong performance of the TIP students on these items relative to more complex language usage skills, seems predictable. Two examples of these items appear in Figure 3.

Insert Figure 3 About Here

The October 1989 revision of the ACT Assessment differs somewhat from the version represented by the two forms used in this study. The new ACT increases the emphasis on rhetorical skills in the measurement of writing proficiency, increases the number of advanced math items, and includes a new reading test which features inferential and reasoning skills and a test designed to measure science reasoning. However, items comparable to those highlighted in this study as tending to favor either the TIP or the HS examinees, continue to be represented in the new ACT. Though the data haven't been collected and analyzed yet,

differential test and item performance results, comparable to those presented here, would be expected. That is, based on the results of this study, punctuation and mathematics word problems would be expected to favor TIP students, and intermediate algebra and geometry skill items would be predicted to favor HS examinees, on the new ACT Assessment. These predictions can be confirmed when data are available.

Despite the fact that the specifications for the ACT have changed, there seems to be enough evidence about types of items common to both versions of the ACT to suggest that the findings from this study will be generalizable to the new tests. It seems clear that the ACT is capable of functioning well as an out-of-level test for TIP candidates by providing substantial ceiling for measurement. A helpful focus for future research would seem to be in evaluating the degree to which the ACT tests, or relevant subscores of the ACT test, enhance the identification of talented seventh graders.

The seventh graders who eventually get selected for TIP are a very select population. They are already preselected just to be invited to take the ACT, and only about 20% of them are eventually selected for active involvement in TIP (Stanley, 1990). If one were to look at just the upper 20% of the TIP distribution, the selected group would be scoring at about the same mean score level as the high school examinees. Clearly these talented seventh graders would seem capable of learning more and faster if the opportunity were there. In fact, one of

the goals of TIP is to provide such enriched educational opportunities for these bright students, and the ACT seems to be a promising vehicle for use in selection.

REFERENCES

Doolittle, A., and Welch, C. (1989). Gender differences in performance on a college-level achievement test. (ACT Research Report 89-90). Iowa City: American College Testing Program.

Dreyden, J. I., and Stanley G. E. (1988, April). College entrance test score and demographic profile information for talented seventh grade youth. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Maxey, E. J., and Dreyden, J. I. (1988, April). Measures of validity between the ACT Assessment and other achievement variables for talented seventh grade youth. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Sawyer, R., and Brounstein, P. (1988, April). The relationship between ACT and SAT scores among academically talented seventh grade students. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Stanley, G. E. (1990). Personal communication.

Stanley, G. E., and York, A. V. (1988, April). The ACT Assessment as a measure for identifying talented seventh grade youth. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Stanley, J. C. (1976). Identifying and nurturing the academically gifted. Phi Delta Kappan, 59, 3, 234-237.

Table 1
Standard Score Mean Comparisons of High School and TIP Examinees (Sample 1)

	HS			TIP					
Test	N	Mean	S.D.	N	Mean	S.D.	*t*	Prob.	Effect Size
English	2000	17.4	5.2	2000	15.3	4.9	13.14	.000	.42
Males	865	16.4	5.3	934	14.2	4.9			.43
Females	1135	18.1	5.0	1066	16.2	4.7			.39
Mathematics	2000	14.9	7.3	2000	10.3	5.0	23.25	.000	.73
Males	865	15.9	7.6	934	10.9	5.2			.77
Females	1135	14.1	7.0	1066	9.7	4.8			.73
Social Studies Reading	2000	15.4	7.1	2000	10.5	5.1	25.07	.000	.79
Males	865	15.9	7.3	934	10.7	5.3			.82
Females	1135	15.0	6.9	1066	10.4	4.9			.77
Natural Sciences Reading	2000	19.5	6.3	2000	17.4	5.0	11.68	.000	.37
Males	865	20.7	6.4	934	18.3	5.2			.41
Females	1135	18.6	6.0	1066	16.7	4.7			.35

Table 2
Differential Item Performance (by Favored Group) for Item/Passage Categories (Sample 1)

Test	Subcategory	Total Items	Number Favoring HS Examinees	Number Favoring TIP Examinees
English (75 items)	Grammar	14	1	4
	Sentence Structure	17	2	1
	Logic & Organization	14	3	1
	Diction & Style	16	1	0
	Punctuation	14	1	5
	Passage 1	9	0	1
	Passage 2	17	6	0
	Passage 3	15	2	2
	Passage 4	9	0	1
	Passage 5	7	0	4
	Passage 6	9	1	2
	Passage 7	9	0	1
Mathematics (40 items)	Arith./Alg. Operations	4	1	0
	Arith./Alg. Reasoning	14	0	4
	Int. Algebra	8	3	1
	Geometry	8	3	0
	Number/Numeration Systems	4	1	0
	Advanced Topics	2	0	1
Social Studies Reading (52 items)	Passage 1 (Government)	9	0	0
	Passage 2 (Psychology)	9	0	1
	Passage 3 (History)	10	0	1
	Passage 4 (Economics)	9	0	0
	Discrete (SS information)	15	3	1

Table 2 cont.
Differential Item Performance (by Favored Group) for Item/Passage Categories (Sample 1)

Test	Subcategory	Total Items	Number Favoring HS Examinees	Number Favoring TIP Examinees
Natural Science Reading (52 items)	Passage 1 (Phy Sci)	9	0	4
	Passage 2 (Chemistry)	9	2	1
	Passage 3 (Physics)	9	0	0
	Passage 4 (Biology)	10	3	0
	Discrete (NS information)	15	3	2

Table 3
Differential Item Performance (by Favored Group) for Item/Passage Categories (Sample 2)

Test	Subcategory	Total Items	Number Favoring HS Examinees	Number Favoring TIP Examinees
English (75 items)	Grammar	8	2	0
	Sentence Structure	19	2	1
	Logic & Organization	17	0	0
	Diction & Style	14	2	2
	Punctuation	17	0	3
	Passage 1	14	1	1
	Passage 2	9	1	2
	Passage 3	9	0	1
	Passage 4	10	2	0
	Passage 5	9	1	1
	Passage 6	9	1	1
	Passage 7	15	0	1
Mathematics (40 items)	Arith./Alg. Operations	4	1	1
	Arith./Alg/ Reasoning	14	2	4
	Int. Algebra	8	4	0
	Geometry	8	3	1
	Numbers/Numeration Systems	4	0	2
	Advanced Topics	2	0	0
Social Studies Reading (52 items)	Passage 1 (History)	10	0	3
	Passage 2 (Psychology)	9	0	0
	Passage 3 (Economics)	9	0	1
	Passage 4 (Government)	9	0	0
	Discrete (SS information)	15	2	3

Table 3 cont.
Differential Item Performance (by Favored Group) for Item/Passage Categories (Sample 2)

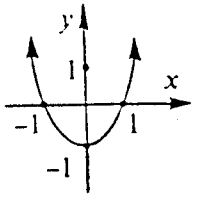
Test	Subcategory	Total Items	Number Favoring HS Examinees	Number Favoring TIP Examinees
Natural Sciences Reading (52 items)	Passage 1 (Phy Sci)	9	1	0
	Passage 2 (Biology)	10	0	1
	Passage 3 (Chemistry)	9	1	0
	Passage 4 (Physics)	9	0	0
	Discrete (NS information)	15	1	2

Table 4
Topics of Passage Sets by Favored Group

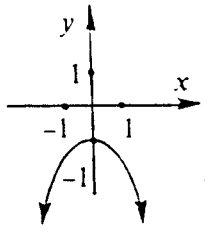
Form	Test	Passage	Topic	Favored Group
1	English	2	Black women authors	HS
1	English	5	Comics/collectibles	TIP
1	Natural Science	1	Pluto/Solar System	TIP
1	Natural Science	4	Cilia function	HS
2	Social Studies	1	History migration of Black Americans	TIP
2	Social Studies	3	Business organization	HS
2	Social Studies	4	Presidential impeachment	HS

28. Which of the following most nearly represents the graph of $y = x^2 - 1$ in the standard coordinate plane?

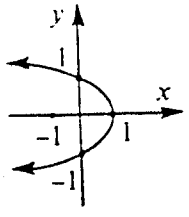
*F.



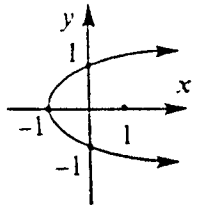
J.



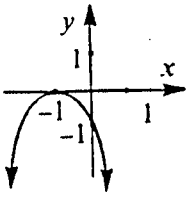
G.



K.

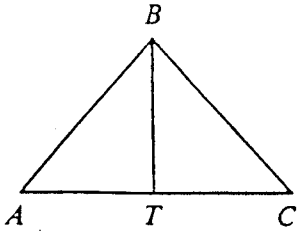


H.



"Intermediate Algebra"

13. Triangle ABC below is an isosceles triangle with the length of $\overline{AB}$ equal to the length of $\overline{BC}$. If $\overline{BT}$ bisects $\angle ABC$ and the measure of $\angle ABT$ is 40°, what is the measure of $\angle BAC$?



A. 40°
*B. 50°
C. 60°
D. 70°
E. 80°

"Geometry"

FIGURE 1. Examples of mathematics item types, designed for college-bound high school students, that were found to be disproportionately difficult for talented seventh graders.

16. A book on physical fitness contains information on "ideal" body weights for men and women. It states that the ideal weight for women is 100 pounds at 5 feet plus 5 pounds for every additional 1 inch, and the ideal weight for men is 106 pounds at 5 feet plus 6 pounds for every additional 1 inch. How many more pounds is the ideal weight for a 6-foot man than for a 6-foot woman?

F. 6
G. 7
H. 11
J. 12
***K.** 18

34. The makers of a low-calorie gumdrop claim that 1 of their low-calorie gumdrops contains only 12% of the calories in 1 of their regular gumdrops. If 1 low-calorie gumdrop contains 5.4 calories, how many calories are there in 1 of their regular gumdrops?

F. 2.2
G. 17.4
H. 22.2
***J.** 45.0
K. 64.8

FIGURE 2. Examples of "Arithmetic and Algebraic Reasoning" items that were found to be disproportionately easy for talented seventh graders.

Herein lies the challenge for the novice collector:

which items in <u>todays trash are</u> tomorrow's treasure?
55

55. A. NO CHANGE
B. today's trash is
*C. today's trash will be
D. todays trash will be

Nobody is <u>certain however</u> experts discuss such
56
matters in numerous periodicals devoted to collecting.

56. F. NO CHANGE
G. certain, however,
H. certain however,
*J. certain. However,

FIGURE 3. Examples of "Punctuation" items that were found to be disproportionately easy for talented seventh graders.