SAT—Pacific Class of 2005

The *SAT I: Reasoning Test*, commonly known as the SAT, is the most widely taken college entrance exam in America. According to the College Board, the SAT is a measure of developed verbal and mathematical abilities important for success in college. Individual student scores are reported in 10 point increments that range in value from 200-800 with the national average around 500. The SAT was created to predict success in the first year of college which it does with moderate success. For example, the SAT predicts success as well as the total high school grade point average; however, the validity of other common uses of the test (e.g., making inferences about the quality of instruction in specific schools, districts, or states) is highly questionable. The SAT is developed by the Educational Testing Service for the College Board.

2005 Results

Table 1 provides results for various levels of DoDEA as taken from College Board reports and DoDEA press releases. In this table Guam students are included in the DDESS results.

Table 1: SAT I Results for DoDEA, DDESS, DoDDS, and the Nation

SAT	1998	1999	2000	2001	2002	2003	2004	2005
DoDEA Verbal	508	504	504	505	504	509	506	**514**
DDESS Verbal	483	483	496	497	490	499	496	**493**
DoDDS Verbal	511	506	505	506	506	510	508	**576**
Nation Verbal	505	505	505	505	504	507	508	**508**
DoDEA Math	502	498	500	497	497	498	503	**505**
DDESS Math	481	474	493	486	476	481	483	**474**
DoDDS Math	505	501	501	499	500	500	506	**508**
Nation Math	512	511	514	514	516	519	518	**520**

Table 2 provides the DoDDS-Pacific/DDESS-Guam results for the class of 2004 and 2005. Average scores can vary from year to year for reasons unrelated to students' educational experiences. When small numbers of students are tested (as in the Guam district) the averages can change greatly from year to year because of random influences alone. In order to provide a context for the results and to inhibit over interpretation, a test of statistical significance was performed to determine whether or not each average was different from the national average.

Table 2: Pacific Average 2004 SAT Scores

Unit	2004			2005		
	Verbal	Math	# Tested	Verbal	Math	# Tested
Pacific	499	509	649	511	**509**	685
Guam	531	507	40	511	492	37
Japan	488	494	228	510	**498**	273
Korea	503	**535**	189	516	534	199
Okinawa	503	500	192	504	**501**	176

Values in **red** are below the national average at the .05 level of significance.
Values in **blue** are above the national average at the .05 level of significance.

The scores in red were significantly below the national average for the year in question. Scores in blue were above the national average. As a general rule, students in Japan and Korea tend

to score below and students in Korea tend to score above the national average in math. The reader is encouraged to read further in Appendices A and B to learn how it may be an error to conclude that math instruction is better in Korea than in the other districts.

Overall, the Pacific average in math is below the national average. However, it is evident from Table 1 that DoDEA students as a whole consistently score below average in math on the SAT—from 10 to 21 points below the national average. Relatively low math scores are a systemwide phenomenon. Given that the SAT is relatively insensitive to recent instruction and that DoDEA students only receive a portion of the education in DoDEA schools, the results suggest that factors outside of the school may have a significant influence on the math results; however, this is not to suggest that improving math instruction in DoDEA would not improve the scores only that the factors influencing the scores are not entirely clear. It is also interesting to note that while DoDEA probably has a larger than average population of students for whom English is not their first language, DoDEA scores on the verbal test are relatively stronger than the math scores. In Table 1, we can see that DoDEA verbal scores have ranged from two points below to six points above the national average.

Readers are encouraged to read to Appendices A and B to understand better what the SAT measures and how factors such as self-selection and socioeconomic status impact the average scores. The intent is not to suggest that DoDEA should not be held accountable for student learning, but that the SAT is no an appropriate assessment for accountability purposes.

Appendix A

What the SAT Measures

The College Board describes the SAT as a measure of developed verbal and mathematical abilities important for success in college. An interesting research report entitled A Historical Perspective on the SAT provides a description of how the test has changed since it was first given to 8,000 young men in 1926. The current SAT is described as follows:

> The verbal portion of today's SAT can be described as a measure of the fundamental academic skills of constructing meaning out of the English language in such a way as to be able to understand and participate in certain kinds of formal discourse. This section of the test focuses primarily on critical reading. Students are asked to read passages from the sciences, the social sciences, and the humanities, and to reflect on the author's point of view, technique, and logic. . . .
>
> The math portion of today's SAT can be described as a measure of the ability to use mathematical concepts and skills in order to engage in problem solving. The test does not measure advanced math skill such as trigonometry or calculus. But it does challenge students to apply strong problem-solving techniques and use the math they know in flexible ways. It asks that students go beyond applying rules and formulas to think through problems they have not solved before. . . .

A recent report by the National Research Council's Committee on the Foundations of Assessment, Knowing What Students Know, examined assessment in light of recent development in cognitive science and in statistical and psychometric theory. According to the report, contemporary theories distinguish between two major types of knowledge and problem-solving processes—domain-general and domain-specific. The authors see the SAT as a measure of domain-general knowledge and problem-solving processes—those that are applicable to a range of situations.

Because domain-general knowledge and problem-solving processes

> "are not tied to any specific context, they may reveal (and predict) people's underlying ability to solve problems in a wide range of novel situations. In that sense, they can be viewed as the types of processes that are frequently assessed by general aptitude tests such as the SAT I." (p.69)

Domain-specific processes are

> "relatively specific algorithms, particular to the domain, that will make it possible to solve problems efficiently." These processes "are often measured by such tests as the SAT II."

The important point to understand is that the SAT *assumes that all test takers have the domain-specific knowledge necessary to take the test.* As noted above, the SAT (at least prior to the class of 2006 who will be the first to take the new SAT) attempts to measure the students' ability to apply generally available domain-specific knowledge to novel problems so as to obtain a domain-general assessment of their verbal and mathematical reasoning ability. It is not the College Board's intent to measure the domain-specific knowledge and processes that are more sensitive to recent instruction; consequently, the SAT is relatively impervious to both coaching

and to the impact of recent instruction. This focus on domain-general knowledge and processes is one of the reasons the SAT is inappropriate as a measure of school quality.

This is not to argue that the domain-specific knowledge and processes obtained most directly through instruction are not important. All educational tests measure a mixture of domain-general and domain-specific knowledge and processes. For example, students must have the domain-specific skill of factoring polynomials to do well on the SAT. In fact, cognitive scientists refer to domain-general knowledge and process as *weak methods* and domain-specific knowledge and processes as *strong methods.* In most situations, domain-specific knowledge outweighs domain-general knowledge in value.

Think about a person with a broken-down car. High levels of domain-general knowledge will be virtually useless without the domain-specific knowledge of the mechanic. On the other hand, domain-general methods of problem solving are more valuable in obscure or novel situations. "Although one of the hallmarks of expertise is access to a vast store of strong methods in a particular domain, both children and scientists fall back on their repertoire of weak methods when faced with truly novel problems." (p. 70) When both methods are brought together, however, the results can be powerful. For example, consider Bob and Ray Magliozzi, "Click and Clack, the Tappet Brothers" of *Car Talk* fame. One is a former teacher, and the other holds a Ph.D. in engineering from MIT. Listening to their radio show provides a good example of the power of high levels of reasoning ability combined with the specific knowledge of automobiles.

In summary:

1. Attempts to assess school quality should focus heavily on measures of domain-specific knowledge and processes, and assessments like the SAT are inappropriate measures for such purposes.

2. Short-term interventions such as SAT coaching are unlikely to have much impact on students' scores. However, ensuring that SAT takers have the basic mathematical skills required by the SAT could help ensure that the test validly assesses their domain-general mathematical reasoning ability.

Using the SAT as an Accountability Measure

The average or aggregate SAT scores reported here are inappropriate for use as an accountability measure because a variety of factors complicate the use of the SAT in this way.

Self-Selection

The self-selection of students is probably the most significant limiting factor. The fact that students take the SAT to meet their own needs—college admission—means that the scores are not representative of the school as a whole. There are a couple of aspects to self-selection to consider.

The first is the macroscopic issue related to the fact that students in different parts of the country participate in the SAT testing at different rates because of the testing requirements of the colleges they are planning to attend. For example, many Midwestern and Western colleges ask for the ACT Assessment instead of the SAT. Consequently, the participation rate in Iowa, the home of the ACT, is lower than states in the Northeast where the College Board originated and is located. The students in Iowa who take the SAT are more likely than students elsewhere to be applying to elite colleges that require the SAT; consequently, their scores will tend to be higher than those of many students from New York state, for example, who take the SAT for admission to a variety of more and less elite schools in the area. Iowa's average SAT scores are about 100 points higher than New York's. Figures 1 and 2 provide a graphic representation of the relationship between states' SAT participation rates and their average SAT scores. Note that the relationship is very strong. For the verbal scores the relationship is equal to a correlation of .93. The strength of the relationship between participation rate and SAT scores suggests that large changes in SAT scores are unlikely to be achieved. With few exceptions, you know almost all you can about a state's SAT scores from their participation rate.

Table 3 provides the participation rates for DoDEA units.

Table 3: SAT I Participation Rates*

SAT Participation	2005
DoDEA	67%
DDESS	38%
DoDDS	72%
Nation	49%

The DDESS participation rate is low because the ACT is the more common college entrance assessment in that area.

DoDDS is represented in the graphs by the red dot. Note that our scores are very close to the level predicted by the national data. Interestingly, our math scores are exactly as predicted by our participation rate while our verbal scores are slightly higher than predicted. This suggests that DoDEA is relatively strong in the verbal area and average in math which is somewhat different from the text above that focused on differences from the national average rather than the average predicted by the participation rate.

Figure 1: 2005 State Averages and Percent Participation
SAT Verbal

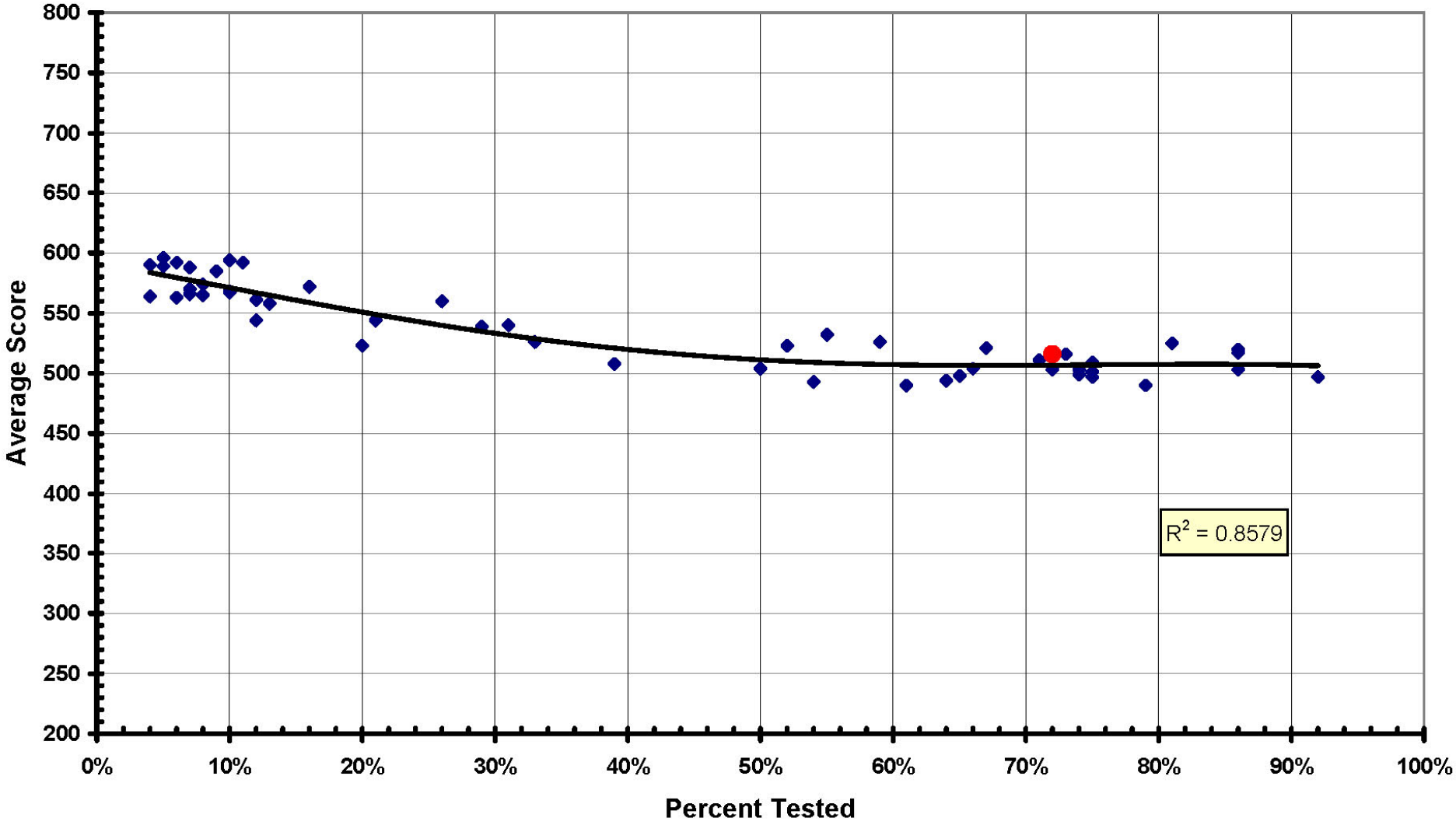
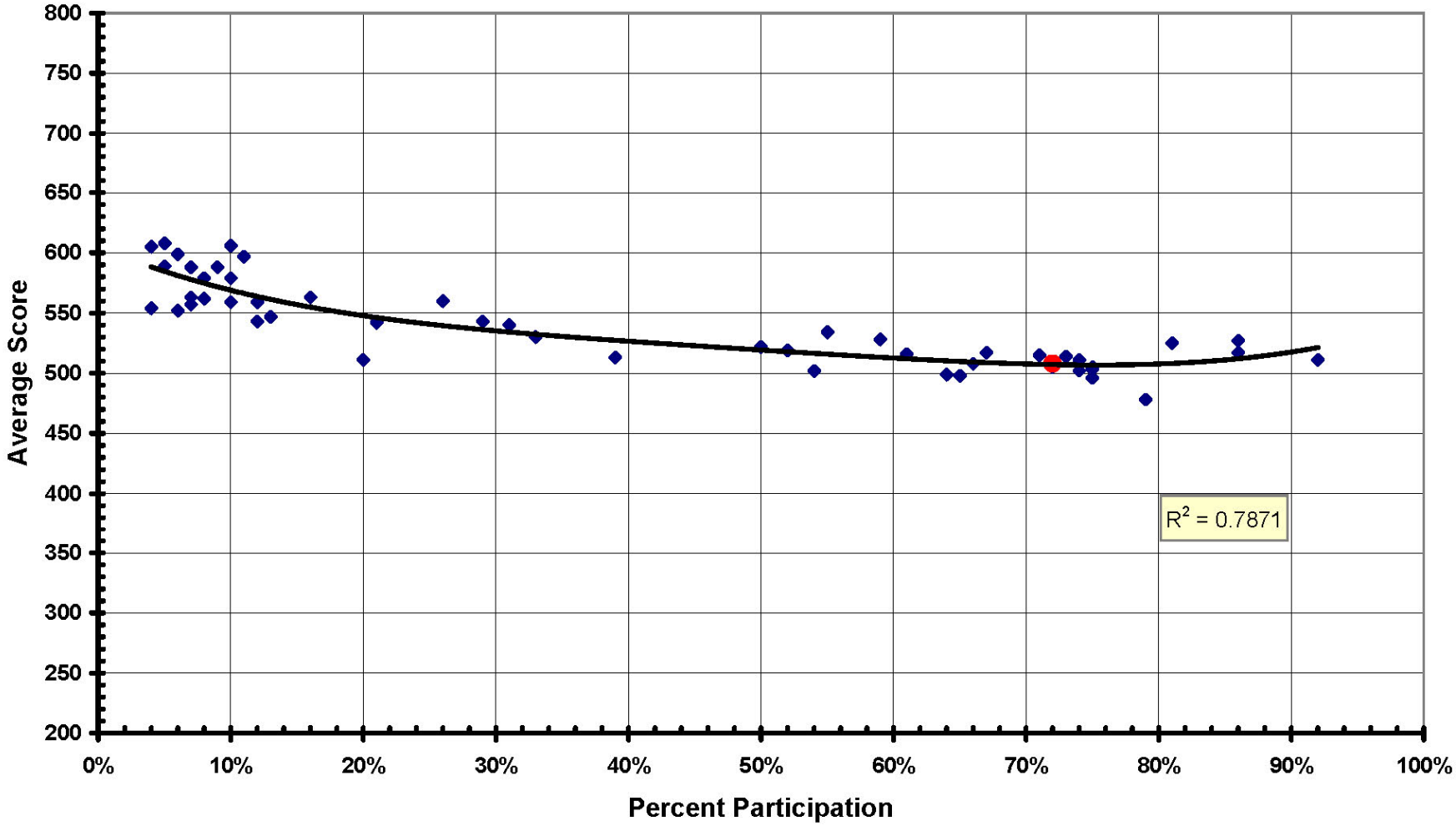


Figure 2: 2005 State Averages and Percent Participation
SAT Math



From the relationship described above, one would expect DDESS scores to exceed those of DoDDS because their participation rate is much lower. However, their scores are actually lower. The same selection factor that affects Iowa does not seem to be evident when comparing DoDDS and DDESS. While fewer DDESS students take the SAT, it seems unlikely that tend to be applying only to selective colleges like the students in Iowa. It seems likely that DDESS students who take the SAT are more like their New York counterparts and are applying to a wider variety of schools that require the SAT. The rule of thumb that says the higher the participation rate the lower the SAT scores cannot be applied equally in all situations.

Finally, with regard to self-selection, it should be noted that the degree to which a school encourage students to take the SAT should have an influence on school averages. The higher the percentage of students in a school who take the SAT, the lower the average score should be.

Other Factors Affecting Test Scores

The lower scores of DDESS students may stem from significant factors other than classroom instruction. The College Board publication, *Guidelines on the Uses of College Board Test Scores and Related Data (Guidelines on the Web)* contains the following:

> In looking at average SAT scores, the user must understand the context in which the particular test scores were earned. Other factors variously related to performance on the SAT include academic courses studied in high school, family background, and education of parents. These factors and others of a less tangible nature could very well have a significant influence on average scores.

Socioeconomic Status

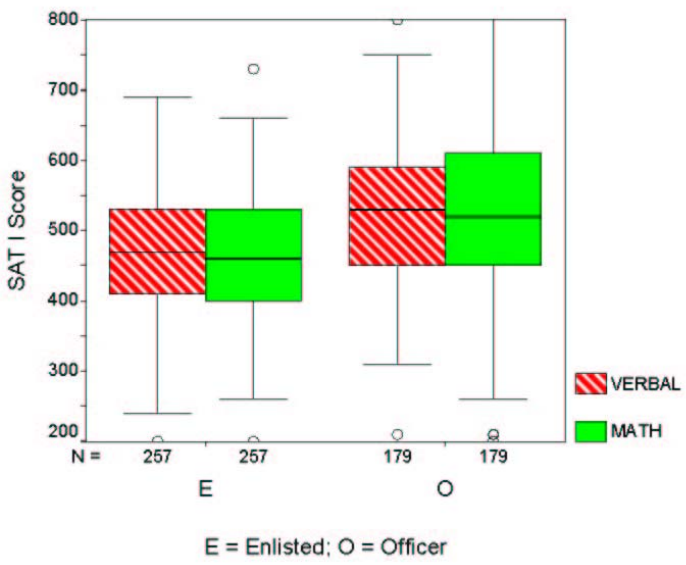
Analyses of DoDDS-Pacific scores from 2001 showed evidence for the impact of one of these additional factors. The table below illustrates the relationship between socioeconomic status (SES) and average SAT scores. For these analyses, SES was defined by military ranks and GS pay grades. The children of warrant officers were included with enlisted dependents, and for DoD civilians, an arbitrary division was made at GS-10. Those below GS-10 were included with the enlisted dependents.

Table 4 shows that students who were the dependents of lower ranking sponsors scored significantly below the national average in both areas. The higher SES students scored above the national average on the verbal test and at the national average on the math test. The two groups also differed significantly from each other on both verbal and math scores; i.e., the dependents of officers scored significantly higher than the dependents of enlisted personnel. Unfortunately, determining *why* the differences are there and how to overcome them is a much more difficult proposition.

Table 4: Average SAT Scores by Sponsor's Pay Grade Compared with National Average

Pay Grade	Verbal	Evaluation	Math	Evaluation	Tests
Enlisted or Below GS-10	471	Below Average	463	Below Average	257
Officer or GS-10 & Above	523	Above Average	523	Average	179

Figure 3: Distribution of Scores by Sponsor's Pay Grade



As informative as average scores are, they say nothing about the distribution of scores. It is almost always true that the range of scores within any group is greater than the differences between any two groups. The figure to the right illustrates the point.

Each "box and whisker" in the graph represents the distribution of scores for a group of students. The box shows the range of scores that includes the middle 50% of the group. For example, the left-most graph shows that the verbal scores of the middle 50% of the dependents of enlisted sponsors ranged from a little over 400 to about 525—between 405 and 535 to be exact. The whiskers extend upward and downward to include most of the remaining students. In this case, about 25% of the students scored above 535. The circles and half circles at the ends of the whiskers are extreme scores that fall outside the normal range of scores. Note that both groups had one or more students receiving the minimum score of 200.

Even though the groups differed in their average scores, both groups had students scoring at all levels of achievement. Therefore, the high average of a school with large numbers of students who are the dependents of high-ranking military members and civilian employees probably says more about the student population than about the quality of the school. The consistency of the correlation between income and SAT scores over the years indicates the importance of outside-of-school factors on SAT scores, and education has not yet learned how to eliminate this relationship in general.

When scores are aggregated over large number of schools, the impact of factors like socioeconomic status is reduced; however, they are still evident when there are large differences between groups. Only 4% of the students in Mississippi and North Dakota took the SAT in 2005, but Mississippi students scored 50-60 points below the North Dakotans.

The differences in scores between smaller groupings of students such as districts and schools can reflect differences in SES just as dramatically. The table below provides the percentage of the Pacific's 2005 seniors in various groups based on their sponsor's pay grade. The method described above was used to group the first two groups in the table. The *Other* group includes students with sponsors in the Foreign Service and those whose sponsors are not associated with the military or the government such as foreign nationals and business executives.

Note that the percentage of students whose sponsor's were enlisted or employed in positions at the grade of GS10 or lower is much lower in Korea than in the other districts. Also note that the percentage who were DoDDS educators, military officers, or civilians above grade GS09 was much greater. The mission and military basing policies can create significant differences in SES

among schools. Table 5 illustrates how the students in Korea were grouped compared with students in the other Pacific districts combined. If the *Other* students are combined with the officer group, then 69% of the students in Korea are in the higher SES group compared with 36% in the other districts. The generally higher SAT scores of students in Korea are undoubtedly influenced to some extent by large difference in SES.

Table 4: Percentage of Seniors by Sponsor's Pay Grade
April 2005

Sponsor's Pay Grade	Guam	Japan	Korea	Okinawa
Enlisted or Below GS10	67%	62%	31%	64%
Warrant Officer, Officer, DoDDS Educator, or Above GS09	33%	34%	52%	34%
Other	0%	4%	17%	1%
Invalid	0%	<1%	0%	2%
Number of Seniors	110	425	239	371

Table 5: Percentage of Seniors by Sponsor's Pay Grade:
Korea vs. All Other Pacific Districts

Sponsor's Pay Grade	Korea	All But Korea
Enlisted or Below GS10	31%	63%
Warrant Officer, Officer, DoDDS Educator, or Above GS9	52%	34%
Other	17%	2%
Invalid	0%	1%
Number of Seniors	239	933

The Inclusion of Nonstudents

The College Board reports students' **last** SAT scores and places them in the schools in which they took the test. Therefore, if a student took the SAT as a junior and then moved to another school and never took the test again, the College Board would consider the student to have graduated from the school he or she attended as a junior.

This is a particular problem for DoDDS because our students change schools frequently. The summary results may also include students who used the code of one of our schools but never went to school there. Consequently, a school's mean score may not accurately reflect the scores of their senior class.

Previous analyses have shown that removing students who did not graduate from the school in which they were tested does not have much effect on average scores at the district or area but can have an impact on the scores of schools.

Table 3: Mean SAT Reasoning Test™ Verbal and Math Scores by State, with Changes for Selected Years

(States are listed by percent of high school graduates who took the SAT Reasoning Test*.)
The College Board strongly discourages the comparison or ranking of states on the basis of SAT scores alone.

		2005		2004		One-Year Change		2000		Five-Year Change		1995		Ten-Year Change	
STATE	Participation Rate 2005*	Verbal Mean	Math Mean	Verbal Mean	Math Mean	Verbal	Math	Verbal Mean	Math Mean	Verbal	Math	Verbal Mean	Math Mean	Verbal	Math
New York	92%	497	511	497	510	0	1	494	506	3	5	495	498	2	13
Connecticut	86%	517	517	515	515	2	2	508	509	9	8	507	502	10	15
Massachusetts	86%	520	527	518	523	2	4	511	513	9	14	505	502	15	25
New Jersey	86%	503	517	501	514	2	3	498	513	5	4	496	503	7	14
New Hampshire	81%	525	525	522	521	3	4	520	519	5	6	520	515	5	10
District of Columbia**	79%	490	478	489	476	1	2	494	486	-4	-8	485	471	5	7
Georgia	75%	497	496	494	493	3	3	488	486	9	10	483	477	14	19
Maine	75%	509	505	505	501	4	4	504	500	5	5	504	497	5	8
Pennsylvania	75%	501	503	501	502	0	1	498	497	3	6	496	489	5	14
Delaware	74%	503	502	500	499	3	3	502	496	1	6	505	494	-2	8
North Carolina	74%	499	511	499	507	0	4	492	496	7	15	488	482	11	29
Virginia	73%	516	514	515	509	1	5	509	500	7	14	504	494	12	20
Rhode Island	72%	503	505	503	502	0	3	505	500	-2	5	502	490	1	15
Maryland	71%	511	515	511	515	0	0	507	509	4	6	506	503	5	12
Vermont	67%	521	517	516	512	5	5	513	508	8	9	506	499	15	18
Indiana	66%	504	508	501	506	3	2	498	501	6	7	492	494	12	14
Florida	65%	498	498	499	499	-1	-1	498	500	0	-2	497	496	1	2
South Carolina	64%	494	499	491	495	3	4	484	482	10	17	478	473	16	26
Hawaii	61%	490	516	487	514	3	2	488	519	2	-3	483	507	7	9
Oregon	59%	526	528	527	528	-1	0	527	527	-1	1	525	522	1	6
Washington	55%	532	534	528	531	4	3	526	528	6	6	519	517	13	17
Texas	54%	493	502	493	499	0	3	493	500	0	2	495	501	-2	1
Alaska	52%	523	519	518	514	5	5	519	515	4	4	521	513	2	6
California	50%	504	522	501	519	3	3	497	518	7	4	492	509	12	13
Nevada	39%	508	513	507	514	1	-1	510	517	-2	-4	511	508	-3	5
Arizona	33%	526	530	523	524	3	6	521	523	5	7	524	520	2	10
Montana	31%	540	540	537	539	3	1	543	546	-3	-6	549	553	-9	-13
Ohio	29%	539	543	538	542	1	1	533	539	6	4	536	535	3	8
Colorado	26%	560	560	554	553	6	7	534	537	26	23	538	538	22	22
Idaho	21%	544	542	540	539	4	3	540	541	4	1	544	532	0	10
West Virginia	20%	523	511	524	514	-1	-3	526	511	-3	0	525	509	-2	2
Tennessee	16%	572	563	567	557	5	6	563	553	9	10	571	560	1	3
New Mexico	13%	558	547	554	543	4	4	549	543	9	4	559	549	-1	-2
Kentucky	12%	561	559	559	557	2	2	548	550	13	9	552	542	9	17
Wyoming	12%	544	543	551	546	-7	-3	545	545	-1	-2	551	544	-7	-1
Minnesota	11%	592	597	587	593	5	4	581	594	11	3	580	591	12	6
Alabama	10%	567	559	560	553	7	6	559	555	8	4	565	555	2	4
Illinois	10%	594	606	585	597	9	9	568	586	26	20	563	574	31	32
Michigan	10%	568	579	563	573	5	6	557	569	11	10	559	565	9	14
Kansas	9%	585	588	584	585	1	3	574	580	11	8	576	571	9	17
Louisiana	8%	565	562	564	561	1	1	562	558	3	4	560	552	5	10
Nebraska	8%	574	579	569	576	5	3	560	571	14	8	568	570	6	9
Missouri	7%	588	588	587	585	1	3	572	577	16	11	569	566	19	22
Oklahoma	7%	570	563	569	566	1	-3	563	560	7	3	565	553	5	10
Utah	7%	566	557	565	556	1	1	570	569	-4	-12	585	576	-19	-19
Arkansas	6%	563	552	569	555	-6	-3	563	554	0	-2	556	542	7	10
Wisconsin	6%	592	599	587	596	5	3	584	597	8	2	574	585	18	14
Iowa	5%	596	608	593	602	3	6	589	600	7	8	589	595	7	13
South Dakota	5%	589	589	594	597	-5	-8	587	588	2	1	579	576	10	13
Mississippi	4%	564	554	562	547	2	7	562	549	2	5	572	557	-8	-3
North Dakota	4%	590	605	582	601	8	4	588	609	2	-4	587	602	3	3
All Students	**49%**	**508**	**520**	**508**	**518**	**0**	**2**	**505**	**514**	**3**	**6**	**504**	**506**	**4**	**14**

* Based on the projection of high school graduates in 2005 by the Western Interstate Commission for Higher Education (WICHE), and the number of students in the class of 2005 who took the SAT Reasoning Test.

** 2002 self-reported twelfth-grade enrollment from D.C.'s public and nonpublic schools was used since WICHE estimated fewer graduating seniors than actual SAT Reasoning Test takers.