

National Geographic Society Alliance Study

Submitted to

National Geographic Society
1145 17th St., N.W.
Washington, D.C. 200-36-46888

Submitted by

Mid-continent Research for Education and Learning
2500 S. Parker Road, Suite 500
Aurora, CO 80014
303-337-0990 (phone)
303-337-3005 (fax)
www.mcrel.org

prepared by

Kerry Englert, Ph.D.
Zoe Barley, Ph.D.

June 18, 2002





© 2002 McREL

To order copies of *National Geographic Society Alliance Study*, contact McREL:

Mid-continent Research for Education and Learning
2550 So. Parker Road, Suite 500
Aurora, CO 80014
Tel: 303-337-0990
Fax: 303-337-3005
Web site: <http://mcrel.org>
e-mail: info@mcrel.org

TABLE OF CONTENTS

Executive Summary	1
Overview of Study	2
Purpose of the Study	2
Methods.....	2
Data Collection Instruments	2
Initial Mailing	3
Matching of School Characteristics	3
Data Collection	5
Data Analysis	5
Data Formatting	5
Matching the Student Sample	5
Results.....	7
Additional Analyses.....	10
Summary	10
APPENDIX A: Initial Mailing	
APPENDIX B: Teacher Background Data	
APPENDIX C: Descriptive Statistics and Significance for Racial/Ethnic Groups	
Sample of Student Assessment	

EXECUTIVE SUMMARY

The National Geographic Society (NGS) Alliance Program has been dedicated to providing training and assistance to geography teachers across the United States for many years. This program has offered teachers intensive professional development in geography education. Although anecdotal evidence shows that teachers have found the program to be very beneficial, previously no data have been collected to assess impact on student achievement. The purpose of this study was to compare the achievement of students of Alliance teachers with a carefully matched sample of students who had taken the 2001 National Assessment of Educational Progress (NAEP) geography assessment.

Eighth-grade Alliance teachers from across the country were asked to participate in the study. The eighth-grade was selected because it represents a point in a students' education when they have typically been exposed to geography and also matches a grade level when NAEP is administered. 62 teachers from 18 states and their students participated in the study. These students were tested on their geography knowledge using a set of items from the 2001 NAEP geography assessment. The results from this test were then compared with the results of a sample of students drawn from the NAEP data. The NAEP students were randomly selected from the larger dataset in order to achieve a sample whose demographic characteristics were similar to those of the students in the Alliance test-taking group.

The major findings of the study include the following:

- Students who had Alliance members as teachers had significantly higher scores overall compared with students in the matched NAEP sample.
- Item-level analyses also show that Alliance students outperformed the NAEP sample and identify relative strengths of these students. Alliance students tended to perform better on questions pertaining to understanding U.S. geography, reading maps, and drawing and interpreting maps.
- Another important consideration for this study is to more directly understand the impact teacher training has on student achievement. This study collected teacher background information such as level of participation in the NGS Alliance program, years of teaching, and hours of general professional development. These variables were merged with student scores and a regression analysis was run to determine if these variables predicted achievement. While there was some correlation between the background variables, the analyses showed that the only variable that significantly impacted student achievement was the frequency that the teacher participated in the Alliance Summer Institutes.

This study found that, overall, students of teachers who have been involved in the NGS Alliance program have statistically significantly higher levels of achievement when compared with students of similar demographic backgrounds from a NAEP sample.

OVERVIEW OF STUDY

Historically, the National Geographic Society (NGS) has been very active in geography education across the country. This interest continues today as the Society provides funding and support for the Alliance program. The Alliance provides opportunities for teachers in state-level chapters to discuss geography education and to participate in NGS professional development programs. Anecdotal evidence showed that teachers had very positive opinions about the services that the Alliance program provided and regarding the impact that the Alliance had on classroom instruction. However, formal research was needed to understand the impact that Alliance participation by teachers had on student achievement in geography.

This research project was initiated to understand more directly the impact of the Alliance geography program on student achievement. Specifically, the study examined the achievement of eighth grade students whose teachers had participated in Alliance activities compared with a demographically similar national sample from the 2001 NAEP geography test administration.

PURPOSE OF THE STUDY

The overall purpose of the study was to measure the impact of the NGS Alliance program on student achievement.

METHODS

DATA COLLECTION INSTRUMENTS

The key sources of data for this study included a pre-screen survey of teachers, a background questionnaire for students, and a student geography assessment.

Pre-Screen Protocol. A pre-screen protocol for teachers was used to gather data on teacher preparedness (years of teaching, preparation area of licensure), participation in professional development in the past year (general and geography related), and school characteristics (total student enrollment, percentage of minority students, percentage of students eligible for free or reduced price lunch, student mobility). Teacher responses to these questions allowed for the selection of a final sample of teachers who would participate in the study.

Student Assessment. The student assessment consisted of a block of released items were part of the 2001 NAEP. This block consisted of 17 multiple-choice and constructed-response items and combined constituted a total of 23 score points. These items were selected by and provided to Mid-Continent Research for Education and Learning (McREL) by the National Center for Educational Statistics (NCES). National Computing Systems (NCS Pearson) printed the assessment from the camera-ready copy supplied by NAEP during the printing of the 2001 assessment. A new cover was created that identified the assessment as The National Geographic Society Eighth Grade Geography Study.

Student Background Questionnaire. A section of the assessment also included questions on students' background and home life (e.g., ethnicity, parental education). The background

questions that were used were provided by Educational Testing Service (ETS) personnel and were identical to the questions asked on the 2001 NAEP.

INITIAL MAILING

The primary sample consisted of students whose teachers participated in the Alliance program. In order to collect an appropriate sample of teachers, McREL staff contacted State Geography Alliance Coordinators from across the country. Particular states were selected based on the fact that they had strong Alliance programs and would be able to provide relatively large lists of teachers who had participated in the program. Another consideration for contacting particular states was the geographic location of the states. It was important to select a sample that represented the geographic diversity of the United States, so that there was a closer initial match with the NAEP data. (See Appendix A for the list of contacted states.)

Each of the Alliance teachers in the selected states was sent a letter that asked the teacher to participate in the study and to complete a pre-screen survey. The main purpose of the pre-screen survey was to obtain teacher and school background information that could be used to determine how well the initial NGS sample might reflect the NAEP sample. Items used reflected important characteristics in determining comparability to the NAEP sample, such as location of school (urban, rural, large town/suburb), type of school (public or private), and region of the country in which the school was located. The pre-screen survey also collected teacher- and school-level data on factors such as preparedness (years of teaching, preparation area of licensure), participation in professional development in the past year (general and geography related), and school characteristics (e.g., percentage of minority students, percentage of students eligible for free or reduced price lunch). The pre-screen also collected information regarding the impact of the NGS Alliance program. These questions included items regarding the Alliance’s impact on instructional practices, improving student engagement, and enhancing student performance (see Appendix B).

MATCHING OF SCHOOL CHARACTERISTICS

Teachers were accepted for inclusion in the study based on their pre-screen information. These teachers represented schools with diverse characteristics such as school type, school location (see Table 1.1 and 1.2), and schools with economically disadvantaged students (see Tables 1.3 and 1.4). These schools had similar characteristics to schools in the 1994 NAEP sample (2001 data on school types were not yet available at the time of teacher selection). There was a final sample of 62 teachers used for the study.

Table 1.1 Frequency and Percentage of Teachers by School Type

LOCATION	Final Sample		1994 NAEP Sample
	FREQUENCY	PERCENTAGE	PERCENTAGE
Public	52	84	90
Private	10	16	10
Total	62	100	100

Table 1.2 Frequency and Percentage of Teachers by Type of School Location

LOCATION	Final Sample		1994 NAEP Sample
	FREQUENCY	PERCENTAGE	PERCENTAGE
Urban	18	30	36
Rural	20	33	26
Large Town or Urban Fringe	22	37	38
Total	60 ¹	100	100

Table 1.3 Frequency and Percentages of Teachers by State

State	Final Sample of Teachers	
	FREQUENCY	PERCENTAGE
ARKANSAS	1	1.6
ARIZONA	4	6.5
CALIFORNIA	10	16.1
COLORADO	2	3.2
DELAWARE	1	1.6
LOUISIANA	2	3.2
MARYLAND	1	1.6
MINNESOTA	7	11.3
NORTH DAKOTA	3	4.8
NEW HAMPSHIRE	1	1.6
OREGON	2	3.2
PENNSYLVANIA	8	12.9
SOUTH DAKOTA	2	3.2
TEXAS	6	9.7
UTAH	3	4.8
WASHINGTON	5	8.1
WISCONSIN	3	4.8
WYOMING	1	1.6
TOTAL	62	100.0

Table 1.4 Average Free and Reduced Lunch and Percentage of White Students

	Final Sample		
	N	Mean	Std. Deviation
Percentage free and reduced lunch	51	34.7	28.3
Percentage of White students	55	75.4	29.9

¹ Number of teachers does not always sum to 62 due to incomplete information provided by the teachers.

DATA COLLECTION

Each of the teachers selected for the study received a letter explaining the study in more detail, along with pencils and rulers that students would need to complete the assessment. Teachers also received a copy of the test administration procedures, which provided a script for teachers to read to the students while administering the test. These procedures were based on the NAEP administration procedures but modified slightly to accommodate some slight differences in the format of the test. The intent of the procedures remained the same: to maintain testing procedures that were consistent with the 2001 NAEP.

Teachers then received copies of the assessment directly from NCS Pearson. The teachers were instructed to administer the assessment during a specified testing window, March 4th to March 15th, using the provided administration procedures. Teachers returned the assessments to NCS Pearson for scoring. Papers were scored by NCS using established NAEP scoring guides.

NCS scoring procedures were followed closely and aligned to the methods used when scoring NAEP. A trained scorer read all constructed-response items once, and a table leader conducted a second reading on approximately 25 percent of the papers. This “read behind” allowed the table leader to assess the accuracy and consistency of the scorer. If a table leader determined that the scores were drifting, retraining was conducted. Using these procedures resulted in a high degree of accuracy in the data and consistency of scoring across students and with the NAEP items.

DATA ANALYSIS

DATA FORMATTING

Because two data sets were used for the analyses, each data set went through an extensive process of verification and cleaning. The NAEP data were e-mailed to McREL from ETS, along with information about the percentage correct for the multiple-choice items or the percentage of students obtaining each point for the constructed response items (p-values). Analyses were run on the data file to ensure that McREL received the correct data file and that the items were coded accurately.

McREL personnel then formatted the NGS data that were sent from the scoring contractor. Similar calculations were conducted on the items to ensure accuracy in the coding and scoring of the results to the NAEP data. The results showed that the NGS data were consistent with NAEP in terms of the numbers of items, the coding of the items, and the percentage of students answering the item correctly.

MATCHING THE STUDENT SAMPLE

The next phase of the analyses consisted of drawing a sample from the NAEP dataset whose demographics were comparable with the NGS sample. Although it is practically impossible to match the two datasets on all 13 of the common demographic variables, a procedure was developed to establish a priority for matching the variables. This procedure involved first calculating an overall score for each student in the NAEP file. A regression

analysis then determined the variables that would best predict achievement on the total score. By using this method, the following six variables were determined to be the most important on which to match. In order of predictive value, these were (1) number of books in the home, (2) race/ethnicity, (3) the level of mother’s education, (4) presence of a computer in the home, (5) time spent talking to someone in the family about school, and (6) the number of days a student was absent from school. This procedure allowed for a very close match on the background variables (see Table 1.5) and numbers of students in each sample. After the matching process, the final Alliance sample contained 878 students; the NAEP sample contained 986 students.

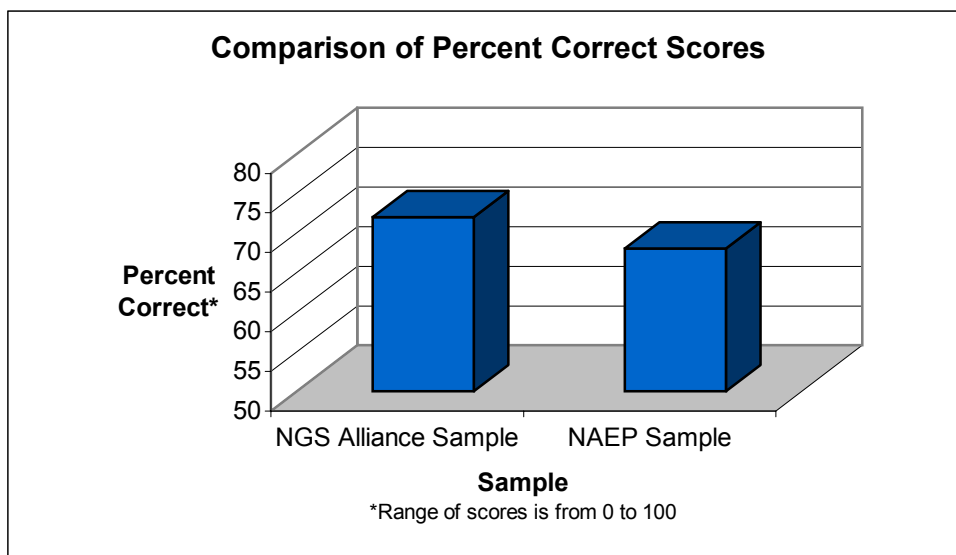
Table 1.5 Comparisons of Percentages on Matching Variables between NAEP and NGS Datasets

Variable	NGS Data	NAEP Data
Books in Home		
Few (0–10)	5	4
Enough to fill one shelf (11–25)	10	11
Enough to fill one bookcase (26–100)	34	35
Enough to fill several bookcases (more than 100)	51	50
Race/Ethnicity		
White and Asian students	75	75
Non-Asian minorities	25	25
Mother’s education		
Did not finish high school	7	9
Graduated from high school	23	23
Had some education after high school	24	22
Graduated from college	47	46
Computer in Home		
Yes	92	88
No	8	12
Talking about school		
Every day	27	27
Two or three times a week	32	32
About once a week	16	17
Once every few weeks	11	11
Never or hardly ever	15	13
Number of days absent		
None	48	48
1 or 2 days	35	34
3 or 4 days	12	12
5 to 10 days	3	5
More than 10 days	2	1

RESULTS

The first step of the analysis was to compare overall scores. These scores were calculated for each student by summing the score points. Students received one point for each correct answer on a multiple-choice item and from 1 to 3 points for each correct constructed response item. The highest possible total score for the set of items was 23. Each student's score was divided by 23 to obtain a percentage correct. The mean score for the Alliance students was 72 percent, with a standard deviation of 0.19. For the NAEP sample, students scored an average of 68 percent, with a standard deviation of 0.19 (See chart 1.0).

Chart 1.0 Overall Means for Both Samples



Mean comparisons and corresponding significance testing were then conducted on the scores. The significance tests (see Table 1.6) showed that the mean differences in the percent correct scores between the two groups of students were in fact significant, with the NGS students scoring higher ($T=3.63$ and $p<0.001$).

A common practice when conducting mean comparisons is to calculate the effect size between the two groups of interest. The effect size statistic can validate the statistical significance, which is often inflated when working with large sample sizes. Using the conventional cutoffs for the effect size shows small but meaningful differences between the two samples of 0.21.

Although the effect size difference between the NGS and NAEP samples would be categorized under the traditional classification of “small” (effect size greater than or equal to 0.20), this effect size represents meaningful differences in achievement between the NGS and NAEP students. When this effect size is converted into a percentile gain for ease of interpretation, the impact can be more apparent. This conversion shows that this small effect size

actually represents an eight-point percentile difference between the two groups of students. In other words, the average NGS student would be equivalent to the 58th percentile of the NAEP sample.

Table 1.6 Descriptive Statistics and Significance Testing for the Overall Scores for All Students

	NGS	NAEP	Test of Significance
Mean	0.72	0.68	
Standard Deviation	0.19	0.19	
Sample Size	878	986	
T Statistic			3.63
Significance			p<0.001
Effect Size			0.21

The other analysis that was conducted on these data was a calculation of the percentage of students scoring each item correctly (p-values) for the two groups of students. These data indicated which items the Alliance students answered correctly more often relative to the NAEP sample (see Tables 1.7 and 1.8 for multiple choice and constructed response items, respectively). Overall, the Alliance students outperformed students in the NAEP sample on most items. However, there were seven items with large differences (greater than four percentage points) between the two groups of students. McREL staff familiar with geography content knowledge examined the items to determine if they represented a certain set of skills. In general, the items measure particular geography skills related to map reading such as a familiarity with U.S. geological features, ability to use a mental map to answer questions, ability to accurately draw maps from a written description, and ability to use information on maps to solve problems. For both the NAEP and the Alliance eighth-grade students, the typical course studied is U.S. History. The NGS students outperformed the NAEP sample in geography questions related to the U.S.; this finding suggests that Alliance teachers may place a greater emphasis on geography when teaching History classes than did teachers of students in the NAEP sample. While this finding is shows some item level differences, it is important to note that these items are not necessarily representative of the entire domain of geography but were rather a convenient sample of items that were to be released by NAEP.

Table 1.7 Frequencies of Correct Responses for Multiple Choice Items²

Item Number and Item Description	NGS Frequency	NAEP Frequency	Difference
2. The number 1 on the map shows	85.3	80.3	5.0
3. The number 2 on the map is on the	90.1	85.0	5.1
4. About how much of South America has a growing season of over 240 days?	90.7	90.8	-0.1
5. The information on the map shows that	54.2	52.8	1.4
6. The world's largest ocean is the	76.3	71.7	4.6
7. Where would a large city be most likely to develop?	46.6	42.0	4.6
8. Which two nations are most likely to have a conflict over mineral resources	60.9	58.7	2.2
10. Florida is an example of	78.4	78.6	-0.2
11. What is an important reason that countries join international organizations like the United Nations?	64.5	61.3	3.2
12. Look at the photograph above. What would help farmers in this area grow more food?	88.4	86.0	2.4
13. Which question could you answer based only on the information in the map?	78.0	74.7	3.3
16. Switzerland is located in the middle of what mountain range	67.1	60.1	7.0

A final analysis was conducted to more directly examine the impact that teacher participation in the Alliance, as well as other teacher background variables had on student achievement. Teacher background information had been collected from the teacher pre-screen, such as level of participation in the NGS Alliance program, years of teaching, and hours of general professional development. These variables were merged with student scores and a regression analysis was run to determine if these variables predicted achievement. The results showed a correlation between many of the variables but the only variable that significantly predicted achievement ($F=8.69$, $p<0.01$) was the number of times a teacher had attended the NGS Summer Institutes. This finding indicates that there is a positive relationship between how often a teacher participates in the summer programs and the achievement of their students.

² See attached Student Test Booklet for complete items

Table 1.8 Cumulative Frequencies of Correct Responses for Constructed Response Items

Item Number and Item Description	Score Points*	NGS Frequency	NAEP Frequency	Difference
1. Write the name of the state or district where you live. Directly on the map draw an “X” on the state or district where you live.	1	94.5	93.4	1.1
	2	91.7	90.1	1.6
9. Which nation is likely to have a steel industry?	1	86.8	88.9	-2.1
	2	27.4	25.7	1.7
14. Directly on the map, draw the most direct public train route you would take to get from Boston College to Braintree	1	76.5	70.1	6.4
	2	71.7	68.9	2.8
15. Draw a map of Little Town on the grid.	1	84.4	78.2	6.2
	2	76.7	70.7	6.0
	3	57.7	50.9	6.8
17. Read the passage above. What does Chief Seattle believe about owning land?	1	63.9	61.8	2.1
	2	37.1	34.8	2.3

Note: This table shows the percentage of students obtaining each score point for each question answered correctly. Note that the frequencies are cumulative. For example, a student who scored three points on item #4 (draw a map of Little Town on the grid) was included in the percentages for both one point and two points.

*The highest number of points for a correct response varied from two to three points.

ADDITIONAL ANALYSES

Mean comparisons were also calculated for White, Black, and Hispanic students across the datasets (see Appendix C for descriptive statistics). These results showed that the White students in the NGS sample scored significantly higher than in the NAEP sample ($T=3.92$ and $p<0.001$). An effect size of 0.24 also indicated the presence of meaningful differences between the groups. This effect size translated into a nine percentile-point difference for the Alliance students relative to the NAEP sample. Analyses for Black and Hispanic students, however, showed no significant differences for the minority students across the two samples.

SUMMARY

The overall results of this study show meaningful and statistically significant differences in achievement between the students of teachers who participated in the Alliance professional

development and the sample of students extracted from the NAEP data. These data are particularly meaningful given the sample of students in the study. These students represented 18 geographically dispersed states, lived in a variety of locations (urban, rural, and suburban), and represented a range of racial and economic diversity. This sample of students was then matched to the NAEP sample, resulting in a high degree of comparability across demographic variables.

The success of the matching process allows for a high degree of confidence that the statistical comparisons have yielded accurate and meaningful results. T-tests were conducted to determine the statistical significance of the mean differences, which showed that, overall, students of Alliance teachers scored significantly higher than did their peers in the NAEP sample. Effect size results also translated into an eight percentile-point difference between the Alliance students and the NAEP sample. Examining mean differences for racial or ethnic groups showed that White NGS students scored higher than their peers from the NAEP sample, but no other significant ethnic or racial differences were present.

The item-level analyses support the overall finding that Alliance students outperformed the students in the NAEP sample. These results also indicate that Alliance students tended to do better on items that might have been taught during the typical eighth-grade Social Studies curriculum. These results might be due to the fact that these students were doing well on geography topics that were being taught to them by teachers who were Alliance members.

Further analyses examined the impact that teacher participation in the Alliance program had on student achievement. These analyses showed that the number of times a teacher participated in the NGS Summer Institutes, significantly predicted students scores on the assessment. This finding suggests that intensity of participation has a positive impact on student scores.

These results indicate that the Alliance program is having an important impact on student achievement in geography. This is evidenced by both the statistical and practical significance of the results. There were important overall point differences in the total scores, which translated into a percentile difference of eight points. Although the results indicate higher scores overall and higher scores for White students, there were no statistically significant differences for minority students between the NAEP and NGS samples. Further research might be conducted to examine reasons that might influence the performance of minority students. These racial and ethnic achievement differences present an opportunity and a challenge for the Alliance to focus on closing the gap between White students and students of color.

APPENDIX A

INITIAL MAILING

Twenty-four states sent lists of their Alliance teachers who taught eighth grade. This resulted in a total of 1,362 teachers who were asked to participate in the study (see Table A1).

A1. States that were Selected for Participation and the Number of Eighth-Grade Alliance Teachers in Each State

State	Number of Eighth-Grade Alliance Teachers
Arizona	71
Arkansas	15
California	606
Colorado	16
Connecticut	16
Delaware	38
Louisiana	9
Maryland	17
Minnesota	26
Missouri	14
Montana	12
North Carolina	11
North Dakota	23
New Hampshire	10
New Jersey	10
Oklahoma	13
Oregon	101
Pennsylvania	86
South Dakota	16
Texas	142
Utah	75
Washington	12
Wisconsin	12
Wyoming	11

APPENDIX B

TEACHER BACKGROUND DATA

The following tables summarize the backgrounds of the teachers in the study and their responses to questions concerning Alliance professional development.

B1. Descriptive Statistics of Teacher Preparation

Characteristic	Mean	Std. Deviation
Years teaching	17.1	8.8
Total hours of professional development	53.6	52.0
Hours of geography professional development	14.4	20.2

B2. How many years have you been a member?

Time	Frequency	Percentage
Less than 1 year	3	5.6
1–3 years	12	22.2
4–8 years	22	40.7
Greater than 9 years	17	31.5
Total	54	100.0

B3. How often do you participate in National Geographic Alliance professional development opportunities?

Number of Activities	Frequency	Percentage
Never	3	5.3
1 activity per year	27	47.4
2 activities per year	11	19.3
3 or more activities per year	16	28.1
Total	57	100.0

B4. Approximately how many hours do you spend in Alliance activities in a typical year?

Number of Hours	Frequency	Percentage
3 hours or less	9	16.1
4–7 hours	15	26.8
8–11 hours	6	10.7
12 or more hours	26	46.4
Total	56	100.0

B5. Have you participated in NGS Summer Institutes?

Number of Times Participating	Frequency	Percentage
Never	5	8.5
Once	25	42.4
2–3 times	19	32.2
4 or more times	10	16.9
Total	59	100.0

B6. Please indicate the extent to which the National Geographic Alliance professional development/training and/or publications has enabled you to:

Product Impact	Percentage Who Responded “Moderately” or “Extensively”
Improve instructional or job related practices in geography	93
Improve your ability to meet the needs of at-risk students in geography	72
Improve student engagement in geography	93
Enhance student performance in geography	89

APPENDIX C

DESCRIPTIVE STATISTICS AND SIGNIFICANCE FOR RACIAL/ETHNIC GROUPS

C1. Descriptive Statistics and Significance Testing for the Overall Scores for White Students

	NGS	NAEP	Test of Significance
Mean	0.76	0.72	
Standard Deviation	0.16	0.17	
Sample Size	634	701	
T Statistic			3.92
Significance			p< 0.001
Effect Size			0.24

C2. Descriptive Statistics and Significance Testing for the Overall Scores for Black Students

	NGS	NAEP	Test of Significance
Mean	0.56	0.52	
Standard Deviation	0.21	0.21	
Sample Size	42	51	
T Statistic			0.81
Significance			NS
Effect Size			0.19

C3. Descriptive Statistics and Significance Testing for the Overall Scores for Hispanic Students

	NGS	NAEP	Test of Significance
Mean	0.57	0.58	
Standard Deviation	0.20	0.23	
Observations	125	136	
T Statistic			-0.49
Significance			NS
Effect Size			0.05

SAMPLE OF STUDENT ASSESSMENT