
Program Efficacy Research
Prentice Hall Language Arts

Writing and Grammar:
Communication in Action



PEARSON
Prentice
Hall

TEACH • INSPIRE • CONNECT

SUMMARY

This study investigated the effects of writing and grammar textbook programs on student performance. Tenth grade language arts students were assigned to either a treatment group (using the Prentice Hall *Writing and Grammar: Communication in Action* ©2001 program) or a control group (using their current textbook program). Students were tested at the start of the school year with two assessments: for grammar, a nationally normed standardized test was used—the language mechanics portion of the TerraNova CTBS Complete Battery Plus exam; for writing, a 10th grade-level persuasive prompt of the CTB Writing Assessment was used. At the end of the full school year treatment period, students were re-tested with the same tests. The 10th grade treatment students utilizing the Prentice Hall *Writing and Grammar* program showed a significant overall increase in both grammar and writing test scores over the course of a full school year, with significant learning improvement in the diagnostic skills areas of Sentences, Phrases, Clauses and Writing Conventions. The 10th grade control students using other writing and grammar programs did not show significant overall increases in grammar nor writing test scores over the same time period. Control students had a significant gain in only one diagnostic skill area: Sentences, Phrases, Clauses. Effect size data support the findings of improvement among students using the Prentice Hall program.

OBJECTIVE

The main objective of this research was to determine whether students who were enrolled in classes using Prentice Hall *Writing and Grammar* ©2001 significantly increased their writing and grammar knowledge and skills after using the program for one full school year. The measures that were used were the NCE (Normal Curve Equivalent) and the OPI (Objectives Performance Index) of the TerraNova CTBS Complete Battery Plus exam, and a six-point holistic scoring scale (developed by ETS Technologies) for the writing assessment.

METHODOLOGY

This study followed a quasi-experimental research design. There was a treatment group to whom the Prentice Hall textbook program was administered and a control group to whom the program was not administered. The treatment group used the Prentice Hall *Writing and Grammar* ©2001 program. The control group used the writing and grammar textbook program that had previously been adopted for use in the school.

Both groups were tested in September 2001, prior to the program's introduction, and then again at the end of the school year in May 2002. Therefore, for both the treatment group and the control group there was a pre-test score and a post-test score. A total of three teachers and six 10th grade writing and grammar classes (three treatment classes, and three control classes) participated in the study. Only students who completed both the pre- and post-tests were included in this analysis. (*See below 1 for participating students by test given*).

10 th Grade	Treatment Group Prentice Hall <i>Writing and Grammar</i> Textbook Program		Control Group Other Writing and Grammar Textbook Programs	
Tests Given:	Standardized	Writing	Standardized	Writing
Pre-Test (administered during 1 st week of school)	43 students	48 students	36 students	35 students
Post-Test (administered at end of school year)	43 students	48 students	36 students	35 students

In each case, one teacher taught both treatment and control classes. The intervention being studied comprised a mainstream textbook program with which teachers were generally familiar. Therefore, training was limited to on-site in-service designed to ensure that teachers understood the treatment program and could instruct with it as its designers intended. Additionally, teachers (and administrators) were repeatedly briefed and provided with materials explaining the study and reinforcing the need to implement it as designed, without, for example, contamination of the program materials or approaches between their treatment and control classes. Discussions with and surveys of the teachers helped establish their understanding of the study's parameters.*

Both treatment and control classes were in the same school building. Treatment and control classes were selected by the teachers to be similar in student ability levels. Of the three schools participating in this research, two were in urban settings and one was in a rural setting. These schools contained a mix of socioeconomic characteristics. Study participants were from two states: Ohio and Virginia.

School Profiles*

	School 1 – OH	School 2 – VA	School 3 - VA
Enrollment	850	2,030	1,330
Ethnicity	Caucasian 98%, Asian 1%, African-American 1%	African-American 57%, Caucasian 38%, Asian 3%, Hispanic 2%	African-American 60%, Caucasian 37%, Hispanic 2%, Asian 1%
Metro Status	Rural	Urban	Urban
Poverty Level	High Income (0.0 - 5.9% Poverty)	Medium-High Income (6.0 - 15.9% Poverty)	Medium-High Income (6.0 - 15.9% Poverty)
College Bound Students (District)	67%	70%	70%

*Source: Market Data Retrieval

This research was facilitated by Pearson Education School Group, with assistance from the following parties:

- Prentice Hall Consultants conducted in-services of the intervention and maintained regular contact with the teachers via telephone to monitor implementation and to answer questions about the study process and materials.
- Statistical analyses and conclusions were completed by an independent statistician with experience in educational research methodologies and analyses (*see Appendix on page 14 for profile*).
- Scoring of the standardized grammar test was completed by CTB/McGraw Hill (publisher of the TerraNova), and scoring of the writing test was completed by Pearson Educational Measurement (*see Appendix on page 14 for company profiles*).

Statistical controls and tests were used to examine the following issues with regard to program effectiveness:

- 1) Whether the pre-test scores for the treatment group and the control group showed significant differences at the starting point of the study (despite being pre-selected as evenly matched);

- 2) Whether overall writing and grammar knowledge and skills increased, decreased, or stayed the same from the pre-test to the post-test among students using Prentice Hall *Writing and Grammar* (the treatment group) and among students using their incumbent textbook program (the control group); and
- 3) Whether the students in the treatment and control groups showed significant learning improvement in key diagnostic areas for writing and grammar.

ANALYSIS: Part I: Grammar Test

To evaluate student grammar skills, a nationally normed standardized test was used—the language mechanics portion of the TerraNova CTBS Complete Battery Plus exam. Analysis was conducted only for students who remained in the study from pre- through post-testing. A total of 46 students were excluded due to attrition from the pre- to the post-test. To ensure that the groups remained balanced after the loss of these students, an analysis was completed comparing the mean pre-test scores of all Treatment students with the mean pre-test scores of all Control students (including both attritted students and those who remained in the study through its conclusion). Results of this analysis revealed no evidence of differential mortality between treatment and control groups in terms of student ability levels that might skew study results.

Step 1a. To remove any spurious student data, an outlier analysis was completed in order to eliminate student scores that fell outside the normal curve distribution. The measure used in this analysis was the overall score, the NCE (Normal Curve Equivalent) of the TerraNova CTBS® Complete Battery Plus exam (*see Technical Post Script on page 11 for a full description of this measure*). An outlier was defined as any student with an NCE score falling two or more standard deviations from the mean. A total of eight students had a pre-test and/or post-test score that qualified as an outlier and were thus removed from the data set (four from the treatment group, and four from the control group).

Step 1b. Once the outlying scores were removed from the data set, an analysis of whether or not there was a statistically significant difference between the treatment group and the control group on the overall pre-test score was performed. For this, a *t-test* on the difference between the treatment and control mean pre-test scores was used. The measure used in this analysis was the NCE (Normal Curve Equivalent) of the TerraNova CTBS® Complete Battery Plus exam. Only students who completed both the pre-and post-tests were included in this analysis.

The hypothesis of the *t-tests* shown below is that the pre-test means of the treatment and control groups are equal; the alternative hypothesis is that they are not equal, for which a two-tail test is appropriate. If the significance of the *t-value* is less than or equal to .10, then the hypothesis is rejected at the 90% confidence level and the alternate hypothesis is accepted.

Since the significance of the *t-value* is greater than .10 (that is, 0.835, as shown below), we accept the hypothesis; that is, that the pre-test means of the treatment and control groups are equal at the 90% level of confidence, showing **no significant differences at the starting point of the study**.

Assessment:	Treatment pre-test NCE (base)	Control pre-test NCE (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>)
Standardized Grammar Test	56.26 (43)	55.67 (36)	0.59	0.209	0.835

Step 2. This step consisted of an analysis of whether or not there was a statistically significant difference between the pre-test and post-test scores within the treatment group and within the control group. For each group, a *t-test* on the difference between pre-test and post-test scores was used. The NCE was the measure used in this analysis.

The hypothesis of the *t-tests* shown below is that the pre-test and post-test means are equal; the alternative hypothesis is that they are not equal. If the significance of the *t-value* is less than or equal to .10, then the hypothesis is rejected at the 90% confidence level and the alternate hypothesis is accepted.

For the treatment group, the significance of the *t-value* is less than .10 (that is, 0.015, as shown on page 4). Therefore, we accept the alternative hypothesis.

*It is commonly understood that teacher variation in experience, style, and effectiveness can have a profound impact on student outcomes. Few, if any, instruments have been shown to be reliable in measuring teacher variability, such that it would be possible to select a priori “matched” teachers for treatment and control groups, or to effectively “balance” the groups on this factor after the study. Therefore, to neutralize this factor between the treatment and control groups, a single teacher taught both groups, with a clear understanding of the necessity to stay true to the instruction in each program.

- Thus, students using Prentice Hall *Writing and Grammar* program showed significant improvement in grammar test scores from the pre-test to the post-test.*

Assessment:	Treatment pre-test NCE (base)	Treatment post-test NCE (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>) ¹
Standardized Grammar Test	56.26 (43)	60.16 (43)	3.90	2.54	0.015

¹Shaded values are significant at the 90% level of confidence or higher (98.5%).

*No judgment is made here as to whether these should or should not be considered “meaningful” improvements, a distinction sometimes applied to observed real-world changes in school/district performance on standardized tests.

For the control group, the significance of the *t-value* is greater than .10 (that is, 0.907, as shown below). Therefore, we accept the original hypothesis.

- Thus, students using other textbook programs did not show significant improvement in grammar test scores from the pre-test to the post-test.

Assessment:	Control pre-test NCE (base)	Control post-test NCE (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>)
Standardized Grammar Test	55.67 (36)	55.81 (36)	0.14	0.12	0.907

Step 3. In addition to the test of significance, an effect size (ES) analysis was also used to determine whether there were improvements in pre- to post-test scores for the treatment and control groups. (see *Technical Post Script on pages 11-12 for a full description of effect size and its interpretation*). The NCE was the measure used in this analysis.

- For the treatment group, using Cohen’s *d*, we see an effect size of 0.31, as shown on page 6. This indicates that, on average, the **treatment students increased from the 50th percentile on the pre-test to the 62nd percentile on the post-test. Thus, treatment students showed improvement in grammar test scores from pre to post.**

Assessment:	Treatment pre-test NCE (base)	Treatment post-test NCE (base)	Cohen's d^2
Standardized	56.26 (43)	60.16 (43)	0.31

²Shaded values are educationally significant (0.25 or higher)[†]

[†]Tallmadge, G. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education and the U.S. Office of Education.

- For the control group, using Cohen's d , we see an effect size of 0.01, as shown below. This indicates that, on average, the **control students remained at about the 50th percentile from the pre-test to the post-test. Thus, control students did not show improvement in grammar test scores from pre to post.**

Assessment:	Control pre-test NCE (base)	Control post-test NCE (base)	Cohen's d
Standardized	55.67 (36)	55.81 (36)	0.01

Step 4*. An effect size (ES) analysis was also used to determine whether there was a difference between the treatment and control groups on the post-test score. The NCE was the measure used in this analysis.

Using Cohen's d , we see an effect size of 0.35, as shown below. This indicates that the post-test mean of the treatment group is at about the 64th percentile of the control group. Stated another way, **the score of the average student in the treatment group exceeded the scores of 64% of the control group students on the post-test. Thus, the treatment group showed higher grammar post-test scores than the control group.**

Assessment:	Treatment post-test NCE (base)	Control post-test NCE (base)	Cohen's d^2
Standardized	60.16 (43)	55.81 (36)	0.35

²Shaded values are educationally significant (0.25 or higher)[†]

Statistical Note: Due to a smaller than desired response base in post-test scores for both the Treatment and Control Groups, as well as at the teacher level, it was not prudent to perform analysis of covariance (ANCOVA). The small response base did not affect any of the tests of significance, since the t-test is robust enough to control for this. As an added check, the t-test results were confirmed using appropriate nonparametric tests (Mann-Whitney and Wilcoxon), which are designed to support cases of small base size. The effect size analysis was also not affected, since ES is not sensitive to sample size.

Tallmadge, G. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education and the U.S. Office of Education.

Step 5. In addition to NCE scores, performance on language mechanics instructional objectives for the treatment and control groups was also analyzed. The measure used in this analysis was the OPI (Objectives Performance Index) of the TerraNova CTBS® Complete Battery Plus exam (see *Technical Post Script on page 11 for a full description of this measure*).

Consistent with their overall improvement in test scores, the treatment students showed significant gains in both language mechanics diagnostic areas (as defined by CTB/McGraw Hill): Sentences, Phrases, Clauses and Writing Conventions. Control students showed a significant gain in Sentences, Phrases, Clauses, but did not show significant improvement in Writing Conventions skills.

CTB/McGraw Hill Treatment Control TerraNova® Language Mechanics Instructional Objectives:	Treatment Mean Point Gain ¹ (pre to post)	Control Mean Point Gain ¹ (pre to post)
Sentences, Phrases, Clauses	+4.72	+3.17
Writing Conventions	+5.44	+1.50
TOTAL GAIN	+10.16	+4.67

¹Shaded values are significant at the 90% level of confidence or higher

ANALYSIS: Part II: Writing Assessment

In addition to the standardized test, a writing test was also given to each student: the 10th grade level persuasive prompt of the CTB Writing Assessment. Initial scoring was completed by the publisher of the CTB Writing Assessment, CTB/McGraw Hill. Analysis of the CTB scoring, however, showed a decrease in mean NCE scores from the pre-test to the post-test for both the treatment and control groups. These results seemed counter-intuitive, and possible errors in the CTB scoring process were suspected. Therefore, researchers and language arts editorial staff examined individual student responses from both the treatment and control groups to see if scoring errors could be detected. After looking at a sample of 14 cases where student writing scores had dropped significantly from pre to post (including both treatment and control students), they concluded that most of these scores should not have decreased. As a result, all of the student writing tests (both pre and post) were re-scored by Pearson Educational Measurement. A holistic six-point scoring rubric developed by ETS Technologies (a subsidiary of Educational Testing Service) for use with Pearson’s e-rater™ product (an electronic essay-scoring system) was used to score the tests (see *Technical Post Script on page 13 for a full description of the rubric used*).

Analysis was conducted only for students who remained in the study from pre- through post-testing. A total of 46 students were excluded due to attrition from the pre- to the post-test. As mentioned in Part I, results of the analysis of attrited vs. retained students revealed no evidence of differential mortality between treatment and control groups in terms of student ability levels that might skew study results.

Step 1a. To remove any spurious student data, an outlier analysis was completed in order to eliminate student scores that fell outside the normal curve distribution. The measure used in this analysis was a holistic six-point scoring scale. An outlier was defined as any student with a score falling two or more standard deviations from the mean. A total of three students had a pre-test and/or post-test score that qualified as an outlier and were thus removed from the data set (all three outliers were from the control group).

Step 1b. Once the outlying scores were removed from the data set, an analysis of whether or not there was a statistically significant difference between the treatment group and the control group on the overall pre-test score was performed. For this, a *t-test* on the difference between the treatment and control mean pre-test scores was used. The measure used in this analysis was a holistic six-point scoring scale. Only students who completed both the pre- and post-tests were included in this analysis.

The hypothesis of the *t-tests* shown below is that the pre-test means of the treatment and control groups are equal; the alternative hypothesis is that they are not equal, for which a two-tail test is appropriate. If the significance of the *t-value* is less than or equal to .10, then the hypothesis is rejected at the 90% confidence level and the alternate hypothesis is accepted.

Since the significance of the *t-value* is greater than .10 (that is, 0.552, as shown below), we accept the hypothesis; that is, that the pre-test means of the treatment and control groups are equal at the 90% level of confidence, showing **no significant differences at the starting point of the study.**

Assessment:	Treatment pre-test mean score (base)	Control pre-test mean score (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>)
Writing	3.0104 (48)	3.1000 (35)	.0896	0.598	0.552

Step 2. This step consisted of an analysis of whether or not there was a statistically significant difference between the pre-test and post-test scores within the treatment group and within the control group. For each group, a *t-test* on the difference between pre-test and post-test scores was used. The six-point holistic scale was also used in this analysis.

The hypothesis of the *t-tests* shown below is that the pre-test and post-test means are equal; the alternative hypothesis is that they are not equal. If the significance of the *t-value* is less than or equal to .10, then the hypothesis is rejected at the 90% confidence level and the alternate hypothesis is accepted.

For the treatment group, the significance of the *t-value* is less than .10 (that is, 0.000, as shown on page 8). Therefore, we accept the alternative hypothesis.

- Thus, students using the Prentice Hall *Writing and Grammar* program showed **significant improvement in writing test scores from the pre-test to the post-test.***

Assessment:	Treatment pre-test mean score (base)	Treatment post-test mean score (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>) ¹
Writing	3.0104 (48)	3.3854 (48)	0.3750	4.169	0.000

¹Shaded values are significant at the 90% level of confidence or higher (100%).

*No judgment is made here as to whether these should or should not be considered “meaningful” improvements, a distinction sometimes applied to observed real-world changes in school/district performance on standardized tests.

For the control group, the significance of the *t-value* is greater than .10 (that is, 0.902, as shown below). Therefore, we accept the original hypothesis.

- Thus, students using other textbook programs did not show significant improvement in writing test scores from the pre-test to the post-test.

Assessment:	Control pre-test mean score (base)	Control post-test mean score (base)	Absolute difference	<i>t-value</i>	Sig (<i>t-value</i>)
Writing	3.1000 (35)	3.1143 (35)	0.0143	0.124	0.902

Step 3. In addition to the test of significance, an effect size (ES) analysis was also used to determine whether there were improvements in pre- to post-test scores for the treatment and control groups. The six-point holistic scale was the measure used in this analysis.

- For the treatment group, using Cohen's *d*, we see an effect size of 0.51, as shown below. This indicates that, on average, the **treatment students increased from the 50th percentile on the pre-test to the 69th percentile on the post-test. Thus, treatment students showed improvement in writing test scores from pre to post.**

Assessment:	Treatment pre-test mean score (base)	Treatment post-test mean score (base)	<i>Cohen's d</i> ²
Writing	3.0104 (48)	3.3854 (48)	0.51

²Shaded values are educationally significant (0.25 or higher)[†]

Tallmadge, G. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education and the U.S. Office of Education.

- For the control group, using Cohen's *d*, we see an effect size of 0.02, as shown below. This indicates that, on average, the **control students remained at about the 50th percentile from the pre-test to the post-test. Thus, control students did not show improvement in writing test scores from pre to post.**

Assessment:	Control pre-test mean score (base)	Control post-test mean score (base)	<i>Cohen's d</i>
Writing	3.1000 (35)	3.1143 (35)	0.02

Step 4*. An effect size (ES) analysis was also used to determine whether there was a difference between the treatment and control groups on the post-test score. The 6-point holistic scale was used in this analysis.

Using Cohen's *d*, we see an effect size of 0.31, as shown below. This indicates that the post-test mean of the treatment group is at about the 62nd percentile of the control group. Stated another way, **the score of the average student in the treatment group exceeded the scores of 62% of the control group students on the post-test. Thus, the treatment group showed higher writing post-test scores than the control group.**

Assessment:	Treatment post-test mean score (base)	Control post-test mean score (base)	Cohen's <i>d</i> ²
Writing	3.3854 (48)	3.1143 (35)	0.31

²Shaded values are educationally significant (0.25 or higher)[†]

Statistical Note: Due to a smaller than desired response base in post-test scores for both the Treatment and Control Groups, as well as at the teacher level, it was not prudent to perform analysis of covariance (ANCOVA). The small response base did not affect any of the tests of significance, since the t-test is robust enough to control for this. As an added check, the t-test results were confirmed using appropriate nonparametric tests (Mann-Whitney and Wilcoxon), which are designed to support cases of small base size. The effect size analysis was also not affected, since ES is not sensitive to sample size.

Tallmadge, G. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education and the U.S. Office of Education.

CONCLUSIONS

The 10th grade treatment students utilizing the Prentice Hall *Writing and Grammar* ©2001 program showed significant overall learning improvement in both grammar skills (as measured by the language mechanics portion of the CTBS TerraNova Complete Battery Plus exam) and writing skills (as measured by the Pearson Education Measurement-scored CTB Writing Assessment) over the course of a full school year. Results of the effect size analysis indicate that the score of the average student in the treatment group:

- Increased from the 50th percentile on the grammar pre-test to the 62nd percentile on the grammar post-test.
- Increased from the 50th percentile on the writing pre-test to the 69th percentile on the writing post-test.

In contrast, the 10th grade control students did not achieve significant learning improvement in overall grammar or writing skills, remaining at about the 50th percentile from pre- to post-test.

In addition, the treatment group outperformed the control group on the overall post-test scores for both the grammar test and the writing assessment. Results of the effect size analysis indicate that the score of the average student in the treatment group:

- Exceeded the scores of 64% of the control group students on the grammar post-test.
- Exceeded the scores of 62% of the control group students on the writing post-test.

Treatment students made significant gains in both language mechanics diagnostic areas (as defined by CTB/McGraw Hill for their TerraNova CTBS Complete Battery Plus exam): Sentences, Phrases, Clauses and Writing Conventions. Control students showed significant improvement in Sentences, Phrases, Clauses, but did not make significant gains in Writing Conventions.

TECHNICAL POST SCRIPT

For the purpose of completeness, we include here a brief discussion of the statistical tests and measures that were used. All of the statistical tests are in the classical statistical domain, and are broadly used across all disciplines including psychometrics. The program used in the computational steps of this project was SPSS version 11.5.1.

The *t-test* that was used states: *The mean of one population is equal to the mean of another population.* The hypothesis is that the means are equal; and the alternative hypothesis is that they are unequal; for which we use a two-sided test. The means and the variance are calculated and the t-statistic is computed. The significance of the t-statistic then computed. If the significance of the t-statistic is less than .10 the hypothesis is rejected; otherwise, the alternative hypothesis is accepted.

Normal Curve Equivalent: “Comparison of Scores across Tests. The normal curve equivalent (NCE) scale, ranging from 1 to 99, coincides with the national percentile scale (NP) at 1, 50, and 99. NCEs have many of the same characteristics as percentile ranks, but have the additional advantage of being based on an equal-interval scale. The difference between two successive scores on the scale has the same meaning throughout the scale. This property allows you to make meaningful comparisons among different achievement test batteries and among different tests within the same battery. You can compare NCEs obtained by different groups of students on the same test or test battery by averaging the test scores for the groups.”¹

¹Quoted directly from: Teacher’s Guide to TerraNova, page 138, CTB/McGraw-Hill 1999.

Objectives Performance Index: “The OPI makes test results both understandable and useful for the teacher in planning effective learning strategies and activities. The OPI is an estimate of the number of items that a student would be expected to answer correctly if there had been 100 similar items for that objective... The OPI scale runs from ‘0’ for total lack of mastery to ‘100’ for complete mastery. For CTB achievement tests, OPI scores between 0 and 49 are regarded as the Non-Mastery level. Scores between 50 and 74 are regarded as indications of Partial Mastery. Scores of 75 and above are regarded as the Mastery level.”²

²Quoted directly from: Beyond the Numbers, A Guide to Interpreting and Using the Results of Standardized Achievement Tests, page 11, CTB/McGraw-Hill 1997.

Effect Size: “Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect.”³ Effect size is “simply a way of quantifying the effectiveness of a particular intervention, relative to some comparison intervention. It is easy to calculate, readily understood and can be applied to any measured outcome in Education or Social Science.”⁴ Cohen’s *d* is commonly used to calculate effect size. Cohen’s *d* is a “standardized” measure found by dividing the difference in the group mean scores by the standard deviation (adjusted for sample size differences known as the “pooled” estimate of the standard deviation). Whereas “statistical significance” is based on a specific probability [“p-value”], it can be affected much

more by sample size than effect size's interpretation. In educational research, an effect size of 0.25 or more is commonly considered to be "educationally significant."⁵

³Quoted directly from: University of Colorado at Colorado Springs website, link: <http://web.uccs.edu/lbecker/Psy590/es.htm>.

⁴Quoted directly from: Curriculum, Evaluation and Management Centre (CEM) website, link: <http://cem.dur.ac.uk/ebeuk/research/effecsize/ESguide.htm>.

⁵Tallmadge, G. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education and the U.S. Office of Education.

TECHNICAL POST SCRIPT (continued)

The interpretation of Cohen's d ⁶

"Cohen (1988) hesitantly defined effect sizes as 'small, $d = .2$,' 'medium, $d = .5$,' and 'large, $d = .8$,' stating that 'there is a certain risk in inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science.'⁷

Effect sizes can also be thought of as the average percentile standing of the average treated (or experimental) participant relative to the average untreated (or control) participant. An ES of 0.0 indicates that the mean of the treated group is at the 50th percentile of the untreated group. An ES of 0.8 indicates that the mean of the treated group is at the 79th percentile of the untreated group. An effect size of 1.7 indicates that the mean of the treated group is at the 95.5 percentile of the untreated group."

Cohen's Standard	Effect Size	Percentile Standing	Percent of Nonoverlap
	2.0	97.7	81.1%
	1.9	97.1	79.4%
	1.8	96.4	77.4%
	1.7	95.5	75.4%
	1.6	94.5	73.1%
	1.5	93.3	70.7%
	1.4	91.9	68.1%
	1.3	90	65.3%
	1.2	88	62.2%
	1.1	86	58.9%
	1.0	84	55.4%
	0.9	82	51.6%
LARGE	0.8	79	47.4%
	0.7	76	43.0%
	0.6	73	38.2%
MEDIUM	0.5	69	33.0%
	0.4	66	27.4%
	0.3	62	21.3%
SMALL	0.2	58	14.7%
	0.1	54	7.7%
	0.0	50	0%

⁶Quoted directly from: University of Colorado at Colorado Springs website, link: <http://web.uccs.edu/lbecker/Psy590/es.htm>.

⁷Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates, p. 25.

TECHNICAL POST SCRIPT (continued)

Writing Assessment – Scoring Rubric

Developed by ETS Technologies, a subsidiary of Educational Testing Service, for Pearson Education’s e-rater™ product, an electronic essay-scoring system.

	Audience and Purpose	Organization	Elaboration	Use of Language
Score 6	Clearly addresses the writing prompt; a main idea is clearly presented	Well-organized, with strong transitions helping to link words and ideas	The thesis is effectively developed with elaborated support and specific details and ideas	Varies sentence structures and makes good word choices; very few errors in spelling, grammar, or punctuation
Score 5	Clearly addresses the writing prompt; a main idea is presented	Clearly organized, although an occasional lapse may occur	The thesis is developed with elaborated support and details	Some sentence variety and good word choices; some errors in spelling, grammar or punctuation, but they do not interfere with reader understanding
Score 4	Addresses the writing prompt; a main idea is presented	Is consistently organized, although perhaps simplistically	The thesis is adequately supported	Sentence structures and word choices are appropriate; errors in spelling, grammar, or punctuation may occur but do not interfere with reader understanding
Score 3	While the prompt may be addressed, the main idea may not be clear	May have organization in parts, but lacks organization in other parts	The support given for the thesis may be unclear or undeveloped	Inconsistent control of sentence structure, and incorrect word choices; errors in spelling, grammar, or punctuation occasionally interfere with reader understanding
Score 2	An attempt is made to address the prompt; however, either the topic is unclear or the support is limited	Very disorganized, and not easy to follow	Limited support, or support that does not support a clear main idea	Problematic sentence structure and frequent inaccuracies in word choice; errors in spelling, grammar, and punctuation hinder reader understanding
Score 1	Little or no attempt is made to address the prompt; response is unfocused or undeveloped	Lacks organization, and is confused and difficult to follow; may be too brief to assess organization	Lacks elaboration of ideas	Little or no control over sentences, and incorrect word choices, may cause confusion; many errors in spelling, grammar, and punctuation severely hinder reader understanding

APPENDIX

William M. Bailey, Statistician, WMB & Associates, Orlando, FL

William M. Bailey is an independent statistician and market analyst with 15+ years of experience in the design, execution, interpretation, and reporting of qualitative and quantitative research studies both in the education and consumer fields.

His experience with the U.S. Department of Education includes:

- Statistical analysis on several studies tracking pre and post course design among a broad-based selection of community colleges and universities
- Serving as an on-call statistical and methodological consultant to the Research to Practice Division
- Serving as a reviewer of proposals to the Office of Special Education & Programs (OSEP)

Mr. Bailey also serves as a beta test site for SPSS, working with development of their core statistical package since version 7 and is now involved with version 12. He also is assisting in the design and testing of the new complex sampling module.

CTB/McGraw Hill, Monterey, CA

CTB/McGraw Hill, publisher of the TerraNova, is a division of the McGraw Hill Companies. It was founded in 1926 and provides a number of standardized achievement tests for both children and adults. The company scores over 20 million test documents each year.⁸

⁸ Source: www.CTB.com

Pearson Educational Measurement, Iowa City, IA

Pearson Educational Measurement is a division of Pearson Education. It is the largest commercial processor of student assessments. The company provides services and products in support of assessment programs to local and state education agencies as well as other assessment organizations and publishers.⁹

⁹ Source: www.pearsonedmeasurement.com

**For more information on Prentice Hall's research initiatives,
visit PHSchool.com/Research**