

Rowan University

## Rowan Digital Works

---

Theses and Dissertations

---

4-24-2002

# The effects of authentic assessment in elementary school science

Timothy L. Trow  
*Rowan University*

Follow this and additional works at: <https://rdw.rowan.edu/etd>



Part of the [Elementary and Middle and Secondary Education Administration Commons](#)

---

### Recommended Citation

Trow, Timothy L., "The effects of authentic assessment in elementary school science" (2002). *Theses and Dissertations*. 1521.

<https://rdw.rowan.edu/etd/1521>

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact [graduateresearch@rowan.edu](mailto:graduateresearch@rowan.edu).

THE EFFECTS OF AUTHENTIC ASSESSMENT  
IN ELEMENTARY SCHOOL SCIENCE

by  
Timothy L. Trow

A Thesis

Submitted in partial fulfillment of the requirements of the  
Master of Arts Degree  
of  
The Graduate School  
at  
Rowan University  
4/23/02

Approved by \_\_\_\_\_  
Professor

Date Approved April 24, 2002

## Abstract

Timothy L. Trow

The Effects of Authentic Assessment  
in Elementary School Science  
2002  
Dr. Gerald Lysik  
School Administration

The purpose of this study was to determine the effectiveness of authentic assessments on Bells Elementary School students' understanding of science concepts. In this study the Washington Township Public School's Elementary School Science Committee created open-ended authentic assessments to enhance the district's science curriculum. The subjects of the study were of various ethnic and economic backgrounds. They were heterogeneously grouped members of self-contained fourth grade classrooms. Surveys of the classroom teachers at Bells Elementary School found that the teachers believed their instruction had changed as a result of the new assessments. They also felt that students were better able to demonstrate understanding of science concepts. Due to the time constraints of the study, no numerical data could be collected to support their belief.

## Mini-Abstract

Timothy L. Trow

The Effects of Authentic Assessment  
in Elementary School Science  
2002  
Dr. Gerald Lysik  
School Administration

Open-ended science assessments were created and used to determine how they effected student understanding of science concepts at Bells Elementary School in Washington Township, New Jersey. The surveyed teachers felt the use of the authentic assessments had positively effected their instruction and student learning

## Acknowledgments

At the conclusion of this study I have many to thank for their indulgence and assistance. Thank you Bobbie Travaline, Principal of Bells Elementary School for allowing me the freedom to complete my projects and for letting me in on so many valuable experiences. Thank you Jeff Pollack, Principal of Thomas Jefferson Elementary School and Chairperson of the science committee, for yielding the committee to me and supporting me along the way. Thank you science committee. I can not mention all of you by name, but I appreciate your hard work, and I am proud of you. Thank you to Sue Knorr, Jeanne Maloney, Denise Scotto, and Tracey Tyler my grade level partners. Your willingness to put up with my surveys and timelines was greatly appreciated. Thank you Sarah Kyle for editing this thesis. I thought I was a better writer than you found. I thank my children, Karley and Dylan. I am sorry for the times you couldn't play with Daddy when you wanted. I thank my wife, Kathy, for filling all of the gaps I left at home when I was working on my studies. Finally, I would like to thank my God for giving me the gifts and the strength I needed to complete my studies. In Him, all things are possible.

## Table of Contents

	Page
Acknowledgements.....	ii
List of Figures.....	v
Chapter 1 Introduction.....	1
Purpose of the Study.....	2
Definitions.....	2
Limitations of the Study.....	3
Setting of the Study.....	3
Organization of the Study.....	5
Chapter 2 Review of Literature.....	6
Introduction.....	6
Authentic Assessment.....	7
Rubrics.....	9
High-stakes Testing.....	11
Conclusion.....	13
Chapter 3 Design of the Study.....	15
General Description.....	15
Research Instruments.....	15
Sample and Sampling Techniques.....	16
Data Collection Approach.....	17

Data Analysis Plan.....	18
Chapter 4 Presentation of Research Findings.....	19
Introduction.....	19
Effectiveness of Authentic Assessments in Washington Township.....	19
Conclusion.....	25
Chapter 5 Conclusions, Implications, and Further Study.....	26
Introduction.....	26
Conclusions and Implications.....	26
Implications of Study on Leadership Skills.....	27
Further Study.....	28
References.....	31
Appendix A: Authentic Assessments.....	33
Appendix B: Teacher Survey.....	38
Biographical Data.....	41

## List of Figures

	Page
Figure 1 Comparison of Student Scores.....	22



## Chapter 1

### Introduction

Authentic assessment and scoring rubrics are catch phrases in education today. Everyone from politicians to parents has heard the terms. Teachers are being asked to test students knowledge using real world problems and assess student performance using scoring rubrics (Popham, 1997). As mandatory state benchmark exams employ authentic assessments and scoring rubrics, teachers feel compelled to use them in class on a more regular basis (Donlevy, 2000). Time often prevents teachers from developing assessments that they feel measure up to those students are presented with on state exams. As a result, teachers wrestle with the external pressure for their students to achieve high test scores on New Jersey state benchmark examinations and an internal belief in their ability to make sound academic choices rather than conforming their teaching to tests.

The intern conducted a case study of the effects of teacher developed assessments that coordinate with the science curriculum in Washington Township elementary schools upon student achievement on the state's Science assessment. As part of the study, a committee of teachers was convened to develop authentic assessments and measurement tools using rubrics matched to the school's curriculum and the state's science standards for grades one through five. Every effort was made to conform the products produced during the study to the recommendations of current educational research. The fourth grade teachers at Bells

## Statement of Purpose:

The purpose of this study was to discover the effectiveness of authentic science assessments on Washington Township elementary school students' understanding of science concepts as measured by the state of New Jersey's fourth grade science assessment. The outcome of the case study was a comprehensive report informing teachers and administrators of the study's conclusions. The results were used to make instructional and curricular decisions about the school's elementary science program.

A secondary purpose of the study was to develop the intern's ability to lead through support of the school's vision for student growth and academic achievement. By collaborating with faculty members to modify curriculum and improve student performance on assessments, the intern was able to encourage the professional growth of the staff. The project also developed the ability to use time effectively.

## Definitions

*Authentic assessment* – a test or project in which students are presented with a real-life problem they must solve. The problems can be in the form of an open-ended question or a presentation among other possibilities.

*Benchmark exams* – a term used to describe the high-stakes tests given at what are believed to be the most important transitional periods in a student's education. These are most often given in fourth grade, eighth grade, and eleventh grade.

*High-stakes testing* – State mandated exams required by policymakers in which students and schools face repercussions if an unsatisfactory percentage of students do not score at an acceptable level.

*Performance assessment* – a term used synonymously with authentic assessment.

*Proficiency/standards* – knowledge or skills students are expected to have mastered at a certain point in their schooling. They are often developed by national organizations for various subject matters and wholly or partially adopted by individual states.

*Scoring rubrics* – a list of skill criteria and qualitative descriptors used to assess student performance on a given task.

*Teaching to the test* – a phrase used to describe a situation in which educators choose to ignore otherwise valuable instructional material to focus their instruction only on the material that is tested on a high-stakes test.

#### Limitations

This study focused on the performance of students in grade four at Bells Elementary School in Washington Township. Five teachers participated in the study, including the intern. Each class was comprised of twenty-four students. The distinctive nature of the school sampled as well as the relatively short duration limit this study and the ability to generalize the findings to other educational institutions.

#### Setting of the Study

Approximately 20 miles from Philadelphia, Pennsylvania, Washington Township is a largely middle income community in Gloucester County, New Jersey. An elected mayor and a seven-member town council govern the township. It is a growing community of approximately 50,000 people largely newly transported individuals from Philadelphia. Most of the community's growth has taken place in the last fifteen years. A large percentage of the district's tax base is residential housing with a small percentage of taxes generated from commercial industry. This has placed a large tax burden on the

residents and has in large part lead to the defeat of six of the last seven school budgets (Comprehensive Annual Financial Report for Year Ended June 30, 2000).

The economic level of the families served by each school is largely middle income. Two high-density housing developments have been completed in the past two years modestly changing the social makeup of the district. These family's incomes qualify them for federally subsidized housing making them economically unique from most of the other families in the community. The growth of the community has lead to the recent expansion of school facilities. In 1995 voters approved a \$50 million bond referendum used for additions and renovations at every school building in the district. The district's students are educated in an early childhood center, six elementary schools, three middle schools, and one high school divided into two buildings---one for students in ninth and tenth grades and one for students in eleventh and twelfth grades---connected by a shared core building. The operating school budget for the year ended June 30, 2000 was \$87,683,394 (Comprehensive Annual Financial Report for Year Ended June 30, 2000).

The district's six elementary schools have an approximate enrollment of 600 students at each school. The schools' teaching staffs range from 50 to 60, with class sizes averaging approximately 26 students. In addition, the student population is relatively stable with mobility rates in the schools ranging from 5% to 12%. While the majority of the student population is of Caucasian decent, students of nearly every culture are represented in the student population at each school. Student achievement is similar among the six schools, as demonstrated by the consistently similar standardized test scores of each school.

The curriculum of the district could be described as a more traditional curriculum with the infusion of many of more recent trends in education. All portions of the curriculum have been aligned New Jersey's core curriculum content standards, and each area of the curriculum is revised on a seven-year cycle. Most recently, the science curriculum was revised for the 2000-2001 school year. The size of the district allows it to have full-time gifted and talented teachers in each of the elementary schools providing additional human resources for the development of higher level thinking skills in the students. In addition, a full range of special services is available for students with special needs.

#### Organization of the Study

Chapter 1 introduced the purpose of the study---to determine the effectiveness of using authentic assessment to improve student achievement in science. Chapter 2 reviews current literature on the use of authentic assessment and scoring rubrics. The literature regarding high-stakes testing will also be presented. Chapter 3 presents a detailed analysis of the design of the study including the research design, the research instrumentation, the sampling techniques, the data collection approach, and the data analysis plan. Chapter 4 presents the research findings by answering each of the research questions and subquestions. The study concludes in chapter 5 with a presentation of the conclusions of the study as well as its implications and recommendations for further study.

## Chapter 2

### Review of Literature

#### Introduction

The purpose of this study was to discover the effects of authentic science assessments on Washington Township elementary school students' understanding of science concepts using a case study design. This study will result in a comprehensive report to inform teachers and administrators of its conclusions.

The standards and accountability movement has drastically changed the focus of schools in New Jersey and much of the rest of the country. The standards in New Jersey have been based largely on the recommendations of national councils in the various academic disciplines. Schools like Washington Township have developed and written their curriculum to align to the standards. New Jersey also has joined the many states that attach benchmark examinations to the standards. These benchmark examinations, also called high-stakes tests, are given in fourth, eighth, and eleventh grade to judge whether schools and students are making progress toward meeting the standards.

Along with the development of standards has come a recommendation to move away from the traditional objective assessments, composed primarily of multiple-choice responses, to what have been dubbed authentic assessments. Authentic assessments present students with real-world problems they must solve in some written format. Many of the items on state exams that measure student progress toward the standards are open-

ended. Students are asked to write an essay-like answer to the question using what they know about the appropriate concepts.

Washington Township schools have written their science curriculum in alignment with the State of New Jersey's standards for science instruction. However, many teachers have expressed concern about their ability to develop appropriate assessments to measure their students' progress toward the standards. They feel pressure for their students to do well on the state examination and they are fearful that their students will not be prepared for the format of the state's exam.

#### Authentic Assessment

There is almost uniform understanding among scholars as to what is meant by the term authentic assessment, often called performance assessment. Nearly every article used words similar to Terwilliger's (1997), assertion of the "real life" nature of the assessment. Students are presented with problems that they may face outside the classroom and are asked to present a solution to the problem using the knowledge and skills they have acquired.

Not all of those in educational circles define real-life in the same way. Some researchers suggest that a truly authentic assessment must present the student with a situation in which they do not already know the answer or how to find the answer (Herrington, Oliver, 2000). Only then will the student be challenged in a way they might be challenged in the real world. They are forced to think critically and creatively to find a solution. Others simply want to assess the students learning process over time using a tool such as a reflective journal (Prestidge, Williams-Glaser, 2000).

There really is no limit to the number of ways educators can interpret the term real-life in reference to authentic assessments. Collages, presentations, dramas, letters, debates, portfolios, journals, and open-ended responses are just a small sampling of the realistic assessment teachers and researchers have used. As Terwilliger (1997) stated, “What appears to be “realistic” to one individual often seems to be “unrealistic” to another.” Indeed, the term real-life is universally accepted, while its definition is up for debate. However, The underlying implication of authentic assessments is that they require deeper understanding of the concepts rather than a mere demonstration that the student has memorized the necessary information (Morris, 2001).

Many authentic assessments require the student to produce a speech or presentation to demonstrate a complex understanding of the concept (Wiggins, 1993). In addition to the project component of the assessment movement has been a push toward demand writing. Many including O’Hearn Downing (1995) have questioned the value of the more traditional objective testing. These individuals believe a timed open-ended question will reveal more about a student’s understanding of key concepts than the traditional objective tests. A parallel can be drawn between the real world of deadlines at work and the deadline of preparing a written cohesive response to one of these assessments. The theory has been implemented in the state of New Jersey’s Elementary School Proficiency Assessment---the state’s benchmark examination or high-stakes test for elementary students. Approximately one half of the exam consists of these types of open-ended questions.

Another popular aspect of authentic assessment in the literature is the self-assessment opportunities it can provide (Prestidge, Williams-Glaser, 2000). Students can



evaluate their own participation in a group project and look for ways that they might improve their own performance. The multiple ways in which students can be assessed also provides the opportunity for students to help teachers determine how the quality of their work will be assessed (Ferrell, Skillings, 2000). Engaging students in these ways is believed to stimulate their interest and thinking which is a goal of this process.

While nearly all see the benefits of authentic assessment some researchers have sounded cautionary alarms. Terwilliger (1997) agrees with much of what is behind the authentic assessment movement, yet he believes that, “the rhetoric of the reformers is misleading and largely unsupported by data.” He does not believe that all other forms of assessment should be set aside, but should be included along with authentic assessments. Furthermore, he says the term authentic implies that these assessments are superior to other assessments. He worries that by de-emphasizing knowledge in favor of only higher order thinking tasks students will lack the knowledge base necessary in so many fields. For example, historians must have a tremendous knowledge base before excelling in their field. Terwilliger believes that both knowledge and higher order thinking skills need to be assessed and neither should be set aside in favor of the other.

### Rubrics

Authentic assessments present a unique dilemma for the instructor. Unlike objective tests, which lend themselves to quantitative analysis, authentic assessments are much more subjective and consistent standards are difficult to maintain. In an effort to standardize the scoring of these assessments rubrics are used. A rubric is defined as “an assessment tool that uses clearly specified evaluation criteria and proficiency levels to gauge student achievement of those criteria” (Montgomery, 2000). In other words, the

students are presented with clear goals they can attempt to meet in order to demonstrate their competency level. Thus, the guesswork is removed from their preparation for assessment.

Besides the obvious benefit of rubrics in helping teachers evaluate student work accurately and consistently, rubrics are said to aid in instructing students and promoting self-assessment. Educators such as Skillings and Ferrell (2000) suggest that rubrics should be used to instruct students in the skills and proficiencies they need to write successfully. Further, they feel students should have input in the development of the rubric. This enables the students to analyze their own writing performance and the standards that can best measure it. Higher level thinking is the goal of authentic assessment and rubrics are believed to be one more tool in helping students exercise this skill.

Some, including W. James Popham (1997) are critical of rubrics as they are currently used. Popham charges that “the vast majority of rubrics are masquerading as contributors to instruction when, in reality, they have no educational impact at all.” He does not suggest that rubrics have no purpose in education, but merely that most contain flaws that keep them from being used to their full potential. Popham feels that many rubrics are too task specific; they focus on the task rather than the more general skill students need to have. Conversely, other rubrics are excessively general. Overly general rubrics give students and teachers no cues as to what skills students need to demonstrate. Popham says another fault of many rubrics is their overly detailed nature. Their size makes them burdensome to teachers and students alike. Finally, Popham believes that many teachers confuse the test of the skill with the skill itself. This focus during

instruction prevents students from mastering a skill that they can transfer to many different types of problems. These problems must be addressed before rubrics will reach their full potential.

Much of what is expected of well-designed rubrics is just the opposite of the aforementioned problems. Teachers should use three to five specific evaluative criteria, avoiding subjective and hard to measure, general, evaluative terms (Montgomery, 2000). Popham (1997) suggests that the evaluative criteria must represent the focal attributes of the skills that are being assessed. By focusing on skills rather than tasks, teachers can successfully instruct students using the criteria of a rubric.

#### High-Stakes Testing

Not to be dismissed from this discussion is the influence of high-stakes testing, or benchmark examinations, on the authentic assessment movement. Many of the examinations require students to complete authentic assessments. High-stakes testing is used because these tests have become the centerpiece of the standards-based reform movement. This movement uses tests to track schools and students, promote students, and set minimum standards for graduation (Hoffman, Assaf, & Paris, 2001). These pressure for these tests come largely from business and political leaders who are concerned about the future economic competitiveness of the United States (Donlevy, 2001). More skilled workers will be needed and there is fear that the present school system will not produce those workers. Mathematics and science are most often a focus of the testing and standards (Donlevy, 2000). Much of the public demands for accountability and the well-entrenched fear for the future lead many to believe that these tests are here to stay.

Most educators resent how these tests have been thrust upon them with little or no input from teachers. The political nature and diagnoses of the achievement problem and subsequent testing further compounds the negative feelings of educators (Hoffman, Assaf, & Paris, 2001). Teachers often fear that the tests themselves will so drive the curriculum that other equally valuable skills and knowledge will be forgotten or ignored by educators (Kubow & DeBard, 2000).

While many have criticized the non-educator influence of high-stakes tests, some positive effects have been recognized. Many schools that have been slow to change are now aligning themselves to the standards adopted by policymakers (Kubow & DeBard, 2000). These results are considered positive, even among the most vocal critics of the benchmark examinations. In addition, other national organizations such as the American Educational Research Association (AERA, 2000) have given limited endorsement to this type of testing by developing standards they feel if followed would ensure the avoidance of the potential pitfalls of these examinations. They assert that sanction decisions about students and schools should not center on scores from a single exam. Many opportunities must be given and alternatives should be provided for students whose scores do not seem to represent their achievement level. Another important and pertinent recommendation is the suggestion that multiple test forms should routinely be introduced to avoid a narrowing in curriculum and “teaching to the test.” An overly focused curriculum with the aim of higher test scores illegitimizes the exam and is of little educational benefit to students. High-stakes testing has definitely changed the educational landscape and while it presents dilemmas to educators and students it can produce positive accountability if implemented in a sensitive and thoughtful manner.

## Conclusion

The consensus in the literature suggests that the educational community strongly supports authentic assessments and the use of rubrics to score those assessments. These real-world problems require students to use higher level thinking skills and force them to demonstrate a deeper understanding of the underlying concepts. Most researchers believe these assessments are superior to the more traditional, objective assessments. Yet, some caution that the educational community should not use the promise of authentic assessments to completely forsake traditional evaluation but should merge authentic assessments with other more time-tested measurement tools.

Similarly, the use of rubrics to systematize scoring into a list of criteria with more objective standards has received almost universal praise. In addition, many laud the value of rubrics instructional and self-reflective properties. Students are able to begin a project knowing what the teacher will assess and how they will assess it. This will give them an opportunity to look back when they have completed the assignment to determine areas in need of improvement. Teachers can use them to aid in instruction and help reduce the subjective nature of many projects and assessments. However, Popham cautions that instructors should guard against falling into the trap of developing overly specific or overly general rubrics.

The authentic assessment movement has occurred at the same time that high-stakes testing has come to national prominence because of the use of authentic assessments within these tests. High-stakes testing put pressure on educators and students to perform up to standards or face some type of sanction. This pressure has led to positive results such as the updating of the curriculum in many schools. However,

educators are leery of the political nature of the exams. In addition, the narrowing of curriculum by teaching to the test has been exposed as a negative side effect of the exams. It is clear that schools should modernize their curriculum. However, they should avoid narrowing it to a point where the quest for higher test scores takes precedence over a thorough curriculum.

## Chapter 3

### Design of the Study

#### General Description

This case study sought to determine the effectiveness of teacher-made, open-ended, authentic assessments on student achievement in elementary science. This action research study utilized both qualitative and quantitative data gathering methods. Fourth grade students and their teachers at Bells Elementary School were the subjects of the study. Surveys and interviews of fourth grade teachers answered the research questions, validated and clarified data generated from student scores on open-ended authentic assessments modeled after the State of New Jersey's Elementary School Proficiency Assessment, the state's benchmark examination for elementary students.

#### Research Instruments

Two open-ended authentic assessments with similar prompts were given to the 4<sup>th</sup> grade students involved in the study. Students were asked to use their knowledge of science concepts, contained in the Washington Township Elementary Science Curriculum, to explain the thinking of a character placed in an everyday science situation. Each of the questions had an accompanying scoring rubric modeled after the State of New Jersey's generic rubric for open-ended science assessments which is used for assessing student performance on the Elementary School Proficiency Assessment. The first authentic assessment asked students to explain in a written essay how a student could

predict the weather based on the clouds in the sky. The second authentic assessment asked students to explain how a student could determine how a particular rock was formed based on where it was found and what it looked like. Each authentic assessment was given at the time the concepts were presented in the school's curriculum. The assessments were given six to eight weeks apart, depending upon the pacing of the individual teacher (see Appendix A).

In addition, a survey was developed to gather information on teacher perceptions of the impact of the assessments on student performance, how often the teachers used open-ended questions, how they affected instructional choices, and any other evidence of the impact the assessment had on student learning (see Appendix B). The surveys were distributed after the students had completed both of the open-ended questions. In addition, teacher observations and opinions collaborated and buttressed the data collected from the students' scores on the assessments. When necessary, individual interviews were conducted to clarify responses to the survey. These interviews focused on incomplete or contradictory responses by teachers to the survey, and were not used to gain information above and beyond the data that the survey was intended to gather.

#### Sample and Sampling Techniques

The subjects of this study were one hundred and twenty fourth grade students and five fourth grade teachers at Bells Elementary School in Washington Township, New Jersey. The students comprised the entire fourth grade class of Bells School, including special needs students. Two of the classrooms contained special education students who were assisted by the in-class support of a special education teacher. The make-up of the classes was comparable to the make-up of the other fourth grade classes in the district.



There were four female teachers and one male teacher in the study. Their years of teaching experience varied from five years to twenty-eight years. All of the teachers had a bachelor's degree in elementary education, and one had completed a master's program in elementary education.

The sampling techniques used were both convenient and purposeful. The use of students and teachers in the intern's own school insured participation by the subjects of the study. Because the purpose of the study was to determine the affect of authentic assessments on student achievement in Washington Township, this nonprobability sample was appropriate, as the results of this study did not need to be generalizable to the population at large.

#### Data Collection Approach

Student scores were gathered from two open-ended authentic assessments in December of 2001 and in January of 2002. Both assessments asked students to make an inference based on their knowledge of science concepts that they were taught as a part of the district's science curriculum. The teacher and the intern scored the students' responses using a three-point rubric designed by the district's elementary science committee. These scores were combined for a total raw score. This process is based on the State of New Jersey's model for scoring student responses on the Elementary School Proficiency Assessment.

The participating teachers were also surveyed to gain more data concerning the effectiveness of the assessments and the rubrics as instruction aids. The survey consisted of questions requiring likert scale responses, as well as open-ended questions. The questions were designed to help the intern determine whether the teachers viewed the

assessments as authentic, how the authentic assessments affected the instructional choices of the teachers, what further training the teachers needed, and how effective the authentic assessment strategies were in improving student achievement. Any ambiguous or incomplete responses were followed up with a one on one interview with the participant to clarify their views.

#### Data Analysis Plan

After the student responses had been scored, the results of the two open-ended authentic assessments were compared in an effort to determine any improvement in student performance. The time limits of the study prevented the use and analysis of student scores on the Elementary School Proficiency Assessment.

The data analyzed in this study was gathered predominately from the survey instrument completed by the teacher participants in this study. The intern compared the responses of the teachers in an attempt to determine if they felt the assessments had any consistent impact on instruction and student learning. The survey responses were also used to determine the validity of the assessments as truly authentic in nature. In addition, the results helped direct the intern's decisions regarding the need for further staff development in the area of authentic assessment and instruction.

## Chapter 4

### Presentation of Research Findings

#### Introduction

This study attempted to determine the effectiveness of authentic assessments in developing fourth grade students' understanding of science concepts at Bells Elementary School in Washington Township, New Jersey. The case study focused on the reaction and performance of teachers and students on open-ended assessments developed by the district elementary school science committee. These measurement tools were designed to test student understanding of concepts in the district's curriculum in a manner consistent with current research and the format of the State of New Jersey's Elementary School Proficiency Assessment. The intern also attempted to determine how the assessments would effect the instructional choices of teachers. In addition to the effect on instruction, the intern sought to determine what characteristics the teachers looked for in authentic questions.

#### Effectiveness of Authentic Assessments in Washington Township

The inclusion of open-ended authentic assessments on the Elementary School Proficiency Assessment has brought about significant change in the teaching practices of fourth grade teachers in Washington Township. As a result of the shift in emphasis from multiple choice standardized tests to a combination of multiple choice and open-ended authentic assessment the fourth grade teachers at Bells Elementary School said they have made instructional choices for their students that focus more on writing. They discussed

the use of science journals and writing explanations for questions discussed in class. The teachers felt they had continued to improve their questioning during classroom discussion by focusing on questions that require higher-level thinking skills. An example of this is the formation of generalizations developed from the concepts of a particular portion of the science curriculum.

The surveyed teachers also acknowledged that their student assessment practices have changed as result of the state's standards and assessment. All of the teachers noted that they include open-ended questions on every student test. Some of those teachers expressed that they had used open-ended questions prior to the development of the Elementary School Proficiency Assessment, but they now use them more frequently. In addition, the teachers said that they have attempted to use open-ended questioning more during instruction, followed by testing. As a result, they have de-emphasized factual recall during discussions and testing.

The teachers are very positive about the use of the new authentic assessments. They are appreciative of the district science committee's preparation of questions. They feel the questions prepare students for the state's fourth grade exams, while relating to and enhancing the district's curriculum. The teachers also expressed that the timing of the committee's project was good. The development of the assessments came after the adoption of the curriculum, making the utilization of the project more manageable and not nearly as overwhelming.

All of those surveyed were of the opinion that the students in their classroom enjoyed the newer assessments. They attributed the comfort level of the students to their growing familiarity with this type of assessment. The use of these types of questions at

lower grade levels was cited as evidence of this observation. Several teachers noted that the students were aware that they needed to effectively answer questions in this format on the Elementary School Proficiency Assessment, so they were receptive to the practice of demonstrating their mastery of concepts in this manner. Another respondent stated that this type of assessment more accurately demonstrates the level at which a student is performing. This improves the teacher's ability to monitor student progress. She felt that most students enjoyed the freedom of being able to explain themselves in this format, while some struggled with their ability to express the concepts. A more traditional objective test might not bring a student's true understanding of concepts to the surface as readily as these open-ended authentic assessments.

Student scores on the Elementary School Proficiency Assessment were not available at the writing of this thesis. As a result, students in fourth grade at Bells Elementary School were given two open-ended assessments nearly two months apart. The scores demonstrated no change over a two-month period using the new authentic assessments developed by the science committee. Figure I shows the average score for the pretest was 2.19, while the average score for the posttest was 2.18. The scores were out of a possible score of three. Scores rose for thirty-nine students, fell for forty-one students, and remained unchanged for twenty-nine students.

All of the teachers at Bells Elementary School feel that their students have benefited from the use of open-ended questions and rubrics. The teachers stated that they have seen children demonstrate a deeper understanding of concepts on these written assessments than on other forms of testing. They also noted a continual improvement over time by students on open-ended assessments. The increasing success of their

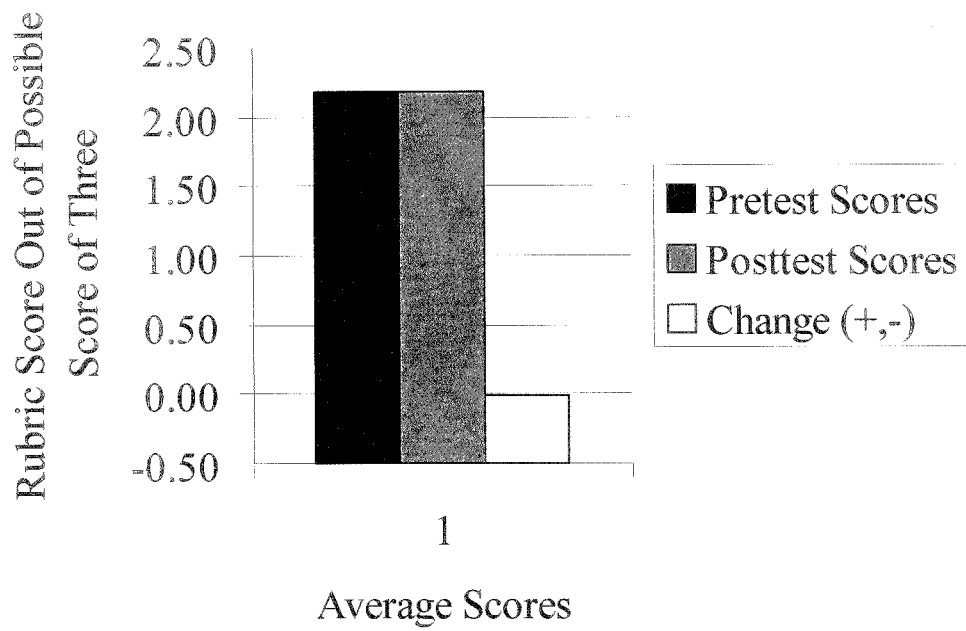


Figure I. Comparison of average student scores on two open-ended questions developed by the district science committee given approximately two months apart.

students encouraged them. Finally, teachers cited the benefits of well-rounded responses by students that create a better picture of their students' mastery of the concepts. This provided them with more information to use when making instructional decisions for their classes.

The surveyed teachers agreed that the assessments developed by the science committee were authentic. The most important positive characteristic for the teachers was the local development of these testing tools in contrast to assessments created by a distant publishing company. They felt items were realistic in that they put students in situations to which students could relate and might face in their everyday lives. They appreciated the fact that students must apply the concepts learned in real-world, common, sense situations to successfully answer a prompt. The respondents also mentioned the fact that the assessments incorporated a concrete element as a positive characteristic.

At this point the teachers are happy with the level of staff development related to authentic assessments. None of them saw a need for any further training. They felt comfortable after several years of administering the Elementary School Proficiency Assessment and receiving district training. One teacher did express interest in learning to write assessments for subjects other than science.

The data collected suggests that the 4<sup>th</sup> grade teachers at Bells Elementary School are very excited and positive about the open-ended authentic assessments and rubrics developed by the district's elementary science committee. The teachers recognized the importance of using this type of measurement tool to enhance their own instruction and student assessment. In addition, the teachers felt the questions addressed the pragmatic goal of developing skills necessary to achieve at a high level on the Elementary School

Proficiency Assessment. Of greater importance was the ability to develop better thinking and communicating skills in their students as they apply basic science concepts taught in school to hypothetical situations they may face in real life. The teachers believe strongly that their student's performance has improved using new authentic assessments, however, at this time there is no empirical data to support that belief. It is likely that more time is needed before quantitative data could be gathered to buttress the teachers' opinions. However, quantitative data would not be privy to the daily interaction the instructors share with their students. The data collected to this point suggests that the students at Bells Elementary School are benefiting from the open-ended authentic assessments developed by the science committee.

In the opinion of their teachers the students also seem to appreciate and value the new assessments. Some teachers believe the students enjoy the freedom of expressing their ideas in a more open format. This supports the school of thought put forward by Wiggins and others that these types of assessments are better than the more traditional forms of assessment. Additionally, teachers have made it clear to the students that they will need the skills necessary to successfully answer these questions when they take the state's mandatory fourth grade exam. If for no other reason the majority of students accept the benefits of these types of questions.

The authentic assessment practices of the Elementary School Proficiency Assessment have affected the pedagogical practices of teachers at Bells Elementary School. Teachers not only include more open-ended questions on their regular classroom exams but have also incorporated more writing in their routine classroom instruction. In addition, the teachers find themselves stressing the higher-level thinking questions during



class discussions. However, current research and professional standards encourage this practice making it difficult to determine whether this change in instructional behavior would have occurred to this extent naturally, without the pressures of the Elementary School Proficiency Assessment. None the less, it is clear the high-stakes test has played a significant role in the changes in classroom instruction at Bells Elementary School.

The teachers accepted the developed questions as authentic. The teachers credited this to the fact that teachers in Washington Township who have taught the subject matter developed the assessments for teachers in Washington Township. It is clear that these teachers are reticent to accept the pre-made assessments produced by publishing companies. The curriculum in Washington Township is not textbook driven, making the tests developed by the publishers less attractive to the teachers, and less applicable or realistic to the students. The teachers feel that the questions developed by the committee do indeed put the students in situations that could be described as real life. This provides the students with opportunities to apply their learning in a more authentic fashion.

### Conclusion

The data collected for this study does suggest that open-ended, or authentic, assessments developed by the Washington Township Elementary Science Committee have had positive effects on the students and teachers at Bells Elementary School. The teachers cited an improvement in student performance and demonstrated understanding of science concepts. Students were observed accepting the need for and value of the questions. Additionally, teachers have made what they feel are positive changes in their instructional and student assessment choices. The evaluative tools have been accepted as authentic by the fourth grade staff at the school.

## Chapter 5

### Conclusions, Implications, and Further Study

#### Introduction

In this study the intern attempted to determine the effectiveness of using authentic assessments to improve student understanding of elementary science concepts. The data collected demonstrates that student understanding does improve as a result of the use of authentic assessment related to real world situations. Teachers have changed their instructional practices to better prepare students for authentic assessment and to strengthen their thinking skills. Through the completion of this project the intern was able to lead educators by developing a vision with committee members and working with them to make the vision a reality. In spite of the apparent success, further study is needed to ascertain the long-term effects of the assessments on the performance and the understanding of students at Bells Elementary School. The committee must continue to revisit the authentic assessment package and search for ways in which they can improve it.

#### Conclusions and Implications

The use of authentic assessments has had a positive impact on student understanding of science concepts at Bells Elementary School. The teachers have changed their instructional choices. They have focused their questioning of students in such a way as to dig deeper into the concepts they are teaching. The tests they use to measure student achievement include more open-ended authentic assessments than they

did before the science committee created the assessment packet. In addition, the teachers have accepted the assessments as authentic due to the life-like situations in which they place students.

The 4<sup>th</sup> grade teachers have also observed that students at Bells Elementary School appear to enjoy and accept the authentic assessments. The students are very aware of their need to perform well on the Elementary School Proficiency Assessment and welcome the opportunity to practice the skills they will need to demonstrate on the examination. They have also become more familiar with these authentic assessments, and the teachers feel their performance has improved with the increased exposure to this type of assessment. At this time there is no objective data to support this belief. The time frame of the project made it impossible to use scores from the state's exam. More time and study is needed before a definitive conclusion can be drawn.

As a result of the preliminary effectiveness of the authentic assessments in improving the science program at Bells School, the intern recommends that teachers continue to use the authentic assessments as part of their efforts to improve their instruction. This will foster a deeper understanding of science concepts. The assessments appear to have a positive effect. Further use and study of their impact should help the teachers better develop their students understanding of science concepts.

#### Implications of Study on Leadership Skills

The intern's ability to lead educators grew as a result of this study. This study required an articulation of a vision for science instruction and assessment. The intern worked with the committee to develop the vision for greater use of life-like, or authentic assessments. The committee worked toward the goal while the intern used their input to

formulate a plan and share responsibility with the members. The plan required a reasonable timetable for completion. As a result, the intern had to find a strategy that would allow the committee members to complete their tasks without having a negative impact on their other teaching duties. It was also important to avoid frustrating the staff and hurting the morale of the committee members. The intern successfully balanced the needs of the project and the needs of the staff involved.

The intern also demonstrated an ability to refine curriculum in a manner supported by research. A study was conducted of current trends and understandings concerning student assessment. The research supported and encouraged the use of authentic open-ended assessments. The information was shared with committee members and used to plan the creation of the open-ended authentic assessments. The research provided crucial support for the decisions the committee made about student assessment. Through this process the intern modeled life-long learning for the staff and helped them develop a curiosity and reliance on current research.

#### Further Study

While the results of this study have suggested that the use of authentic assessments has had a positive impact on instruction and student performance at Bells Elementary School, there are more questions that need to be answered. How will student understanding of science concepts change over time? The relatively short duration of this study made it difficult to use the rubrics designed by the committee to determine if student scores actually improved. Elementary School Proficiency Assessment scores were unavailable. Tracking these scores over a period of years would give valuable information that could be used to support or weaken the teachers belief that these

assessments have had a positive effect on student understanding of elementary science concepts.

Some editing and evaluation was conducted on the authentic assessments created by the committee. However, this was limited in scope. It is important that the committee reassess the authentic assessments after they have been used more extensively to determine if any changes are needed. This is likely to happen naturally. Students will expose the weaknesses of seemingly well thought out test questions. The committee should check back on a regular basis with teachers to ascertain any poorly worded or designed authentic assessments. The test items should be reworked to maximize the level to which they stimulate greater thinking in students.

The authentic assessments created by the science committee followed the pattern of open-ended questions used by the State of New Jersey on the Elementary School Proficiency Assessment. This is a good starting point, but there are other assessment options that are equally valuable and authentic. Further research should focus on the possibilities for life-like assessments other than the open-ended authentic assessments developed by the committee. These activities should align well to the district's curriculum. Many teachers have already developed their own projects that require students to demonstrate a deep understanding of science concepts. The committee should consider collecting these ideas as well as creating other authentic assessments supported by current research. They can study the quality of the assessments and add them to the open-ended authentic assessments already in use by the teachers in the district.

This study has determined that the work completed by the science committee has already demonstrated some improvement in student understanding of science concepts.

By studying these ideas further, and continually revisiting the work already completed, the committee can progressively work toward maximizing the quality of the science program at Bells Elementary School.

## References

- Comprehensive Annual Financial Report for the Fiscal Year Ended June 30, 2000. Board of Education of the Township of Washington School District. Washington Township, New Jersey.
- Donlevy, J.G. (2000). The dilemma of high-stakes testing: what is school for? International Journal of Instructional Media, 27, (4), 331-337.
- Donlevy, J.G. (2001). High-stakes environments and effective student-teacher relationships: some lessons from special education. International Journal of Instructional Media, 28, (1) 1-8.
- Herrington, J. & Oliver, R. (2000). An instructional design framework for authentic learning environments. Educational Technology Research and Development, 48, (3), 23-48.
- Hoffman, J.V., Assaf, L.C. & Paris, S.G. (2001). High-stakes testing in reading: today in Texas, tomorrow? The Reading Teacher, 54, (5), 482-492.
- Kubow, P.K. & DeBard, R. (2000). Teacher perceptions of proficiency testing: a winning Ohio suburban school district expresses itself. American Secondary Education, 29, (2), 16-25.
- Montgomery, K. (2000). Classroom rubrics: systematizing what teachers do naturally. The Clearing House, 73, (6), 324-328.
- Morris, R.V. (2001). Drama and authentic assessment in a social studies classroom. The Social Studies, 92, (1), 41-44.

O'Hearn Downing, S. (1995). Teaching writing for today's demands. Language Arts, 72, (3), 200-204.

Ongoing Information for the Community (2001). [www.wtps.org/info/index.html](http://www.wtps.org/info/index.html)

Popham, W.J. (1997). What's wrong-and what's right-with rubrics. Educational Leadership, 54, (1), 72-75.

Position statement of the american educational research association concerning high-stakes testing in prek-12 education (2000). Educational Researcher, 29, (8), 24-25.

Prestidge, L.K. & Glaser, C.W. (2000). Authentic assessment: employing appropriate tools for evaluating students' work in 21<sup>st</sup> century classrooms. Intervention-in School-and-Clinic, 35, (3), 178-182.

Skillings, M.J. & Ferrell, R. (2000). Student-generated rubrics: bringing students into the assessment process. Teaching Reading, 53, (6), 452-455.

Terwilliger, J. (1997). Semantics, psychometrics, and assessment reform: a close look at "authentic" assessments. Educational Researcher, 26, (8), 24-27.

Wiggins, G.P., (1993). Assessing Student Performance. San Francisco: Jossey-Bass Publishers.



Appendix A  
Authentic Assessments



**The Method of Formation Determines the Classification of a Rock  
(Tim found a rock near a stream...) Open-Ended Question  
Rubric**

- 3 points**      The explanation should demonstrate a clear understanding of how sedimentary rocks are formed, and the action of water on stone formation.
- 2 points**      Response demonstrates some understanding of how a sedimentary rock is formed, but not a clear understanding of why the rock was found near a stream.
- 1 point**        Some correct information but does not explain what sedimentary rock is or how it is formed.
- OR**
- Some correct information and some irrelevant information.
- 0 points**        Some attempt. Totally irrelevant response.

Additionally, students' papers may be unscorable for the following reasons:

<u>Code</u>	<u>Reason</u>
NR	No Response
OT	The paper was off-topic
NE	The paper was in a language other than English
WF	The paper was written in the wrong format, e.g., the student responded with a poem instead of an essay.
FR	The paper contained only a fragment of a response.

Unscorable student papers receive a score of 0.





Appendix B

Teacher Survey

# Survey

## Impact of Open-Ended Science Assessments on Instruction and Student Performance

Name: \_\_\_\_\_

1. What are your feelings about using the assessments developed by the science committee?
  
  
  
  
  
  
  
  
  
  
2. How would you describe the reaction of your students to the assessments developed by the science committee?
  
  
  
  
  
  
  
  
  
  
3. Do you feel that your students have benefited from your use of open-ended questions and rubrics?

If yes, what have you observed that leads you to believe this?

If no, what would make the assessments more effective?





## Biographical Data

Name	Timothy L. Trow
High School	Evansville High School Evansville, WI
Undergraduate	Bachelor of Arts Elementary Education Cedarville University Cedarville, OH
Graduate	Master of Arts School Administration Rowan University Glassboro, NJ
Present Occupation	4 <sup>th</sup> Grade Teacher Bells Elementary School Washington Township Public School Sewell, NJ