

Running head: ASSESSING A RUBRIC: EVALUATING A RUBRIC'S ABILITY TO
PROVIDE FUNCTIONAL DATA TO CRITIQUE CURRICULUM AND
IMPROVE INSTRUCTION

Assessing a Rubric: Evaluating a Rubric's Ability to Provide Functional Data to Critique
Curriculum and Improve Instruction

By:

John A. Gundrum

A Field Report Submitted
In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Education

at

University of Wisconsin Oshkosh

December 2007

COMMITTEE APPROVAL

 Advisor

12-8-07 Date Approved

 Advisor

12-8-07 Date Approved

TABLE OF CONTENTS

	Page
ABSTRACT	5
DEDICATION	6
ACKNOWLEDGEMENTS	7
CHAPTER 1: INTRODUCTION TO THE STUDY	8
Background, Justification, and Significance of the Study	10
General Purpose of the Study, Research Questions, and Definitions	12
Summary	13
CHAPTER 2: REVIEW OF THE RELATED LITERATURE	15
Significance/Importance	16
Current Practices/Methods	17
Collecting Assessment Data	18
Analyzing and Interpreting Assessment Data	19
Using Assessment Data	20
Challenges and Obstacles	21
Recommendations/Suggestions	24
Summary	26
CHAPTER 3: METHODOLOGY	27
Setting	27
Participants	28
Instrumentation	28
Procedures	29

Data Analysis	31
Summary	32
CHAPTER 4: RESULTS	33
Participant Demographics	33
Participants Perceptions of Assessment	35
Current Rubric “A” Data	37
Pilot Rubric “B” Data	38
Comparison of Current Rubric “A” and Pilot Rubric “B”	40
Summary	42
CHAPTER 5: DISCUSSION	43
Summation of the Purpose	43
Summation of the Literature	44
Summation of the Procedures	45
Interpretation of the Results	46
Limitations	49
Implications/Recommendations	50
Summary	51
REFERENCES	53
APPENDIX A: IRB APPROVAL	55
APPENDIX B: CURRENT RUBRIC (RUBRIC A)	57
APPENDIX C: PILOT RUBRIC (RUBRIC B)	59
APPENDIX D: SURVEY	61
APPENDIX E: WRITING ASSESSMENT PROMPT	65

APPENDIX F: STUDENT WRITING SAMPLE 67

APPENDIX G: PARTICIPANT DIRECTIONS 70

APPENDIX H: SUMMARY OF RESEARCH PROJECT SENT TO PARTICIPANTS 72

APPENDIX I: SHORT FORM CONSENT DOCUMENT 74

APPENDIX J: PARTICIPANT RESPONSES TO GENERAL SURVEY QUESTIONS 76

APPENDIX K: PARTICIPANT RESPONSES TO RUBRIC A SURVEY QUESTIONS 78

APPENDIX L: PARTICIPANT RESPONSES TO RUBRIC B SURVEY QUESTIONS 80

APPENDIX M: PARTICIPANT RESPONSES RE: STRENGTHS 82

APPENDIX N: PARTICIPANT RESPONSES RE: CONCERNS 84

APPENDIX O: PARTICIPANT RESPONSES RE: ADDRESSING CONCERNS 86

Abstract

This study was a response to the mandates of Public Law 107-110, commonly known as The No Child Left Behind Act of 2001, and its emphasis on the use of assessments to monitor student achievement, critique curriculum, and improve instruction. The study “assessed” the current rubric used to score a standardized district-wide student writing assessment in a Northeastern Wisconsin school district. Ten communication arts teachers from three high schools became participants. They were given an anonymous student writing sample, the current rubric in use by the district of the study, and a revised, pilot rubric. They were asked to assess the student writing sample using both rubrics and then respond to survey questions regarding the assessment process and each rubric’s ability to provide functional data to critique curriculum and improve instruction. Survey data from Likert Scale questions and participant responses to open-ended questions revealed they agreed that the use of assessment data to evaluate their teaching techniques is important. Echoing the existing literature, they also reported the most significant obstacle to accomplishing this task is lack of time. Participants also supported the pilot rubric over the current rubric in regard to its ability to provide more functional data for the purpose of critiquing curriculum and improving instruction.

Dedication

This field report and the compilation of graduate work that led to this Master's **D**egree

are dedicated to my parents,

Lawrence and Cordelia Gundrum.

They both passed away during my "graduate school journey."

This is for them.

Acknowledgements

I would like to acknowledge my wife, Tamie, whose patience, sacrifice, and support throughout the graduate school process made this achievement possible. I also thank the University of Wisconsin Oshkosh, the Special Education Department and its faculty, particularly Dr. Bertram Chiang and Dr. Wayne Swanger. I'm especially grateful to Dr. Swanger whose personal assistance as my advisor contributed a great deal to my ability to achieve this accomplishment.

Assessing a Rubric: Evaluating an Assessment Rubric's Ability to Provide Functional Data to Critique Curriculum and Improve Instruction

Chapter 1: Introduction to the Study

In the wake of the No Child Left Behind Act (N.C.L.B.) of 2001, the initiative to improve public education via higher standards and increased accountability (standards-based education reform) provoked public school administrators and educators to reevaluate curriculum and instruction. N.C.L.B. gave public schools “the flexibility to select the strategies that best meet their particular needs for improved teaching that will help them raise student achievement in the core academic subjects” (Executive Summary, 2004). The impetus for public schools became standards based instruction and assessment systems to test those standards. In response, administrators and educators created innovative performance based assessments with a dual purpose. Their primary purpose was to use the assessments as an instrument to measure students’ progress towards reaching standards (evidence of “student achievement”). Also, the assessments were designed as an instrument to critique curriculum and improve instruction (evidence of “improved teaching”). According to Murnane, Sharkey, and Boudett (2005, p. 269), “The student-assessment results that schools must report to satisfy No Child Left Behind requirements could be useful in pinpointing strengths and weaknesses in instructional programs . . .” The use of this data is critical to the process of educational improvement.

N.C.L.B. has made high stakes student assessment the major topic of interest in public education today. Or, as aptly stated by Stiggins and Chappuis, “No Child Left Behind has lit an assessment fire in our nation...” (2006, p. 13). The majority of the heat from this fire surrounds the use of assessments as an instrument to measure students’ progress towards reaching standards (evidence of “student achievement”). This has been the focus and priority of school districts, administrators, and teachers who strive to meet the mandates of N.C.L.B. and document

“adequate yearly progress” (A.Y.P.) of their students to avoid federal government intervention. What has been somewhat lost and ignored is the other advantage of administering assessments. Assessment results and data are invaluable to administrators and educators whose perpetual objective is to improve curriculum and instruction to better serve students.

If educators and administrators are to effectively use assessment results, a prerequisite skill is to be able to gather, interpret, and analyze the data. In referencing N.C.L.B., Sharkey and Murnane reflect, “Implicitly, the legislation assumes that teachers and administrators know how to learn from student assessment results and that they have the time and support that they need to do so” (2003, p. 78). These are the significant challenges this nation’s public educators have faced for the past six years as they strive to meet the mandates of N.C.L.B. The challenge has been sent by N.C.L.B to every public school district in this country. “Data-driven decision making requires educators not only to collect assessments but to be able to manage them. Getting ready for the challenge can be the hardest part” (Hall, 2003, p. 32). The essential issue is, given this set of challenges, what must educators and administrators do to ensure that their district complies with the requirements of this legislation and at the same time, improve not only their students scores, but more importantly, their students education. That is the ultimate goal of all educators.

Many school districts have risen to the challenge and responded with a number of new policies and procedures to evaluate students’ progress through assessments and use the data generated to evaluate curriculum and improve instruction. These procedures range from the creation of formative assessment instruments and the use of standardized state assessments to the collection of results and data to analyzing and interpreting the results in an attempt to evaluate instruction. One of these responses is the formation of committees to create new forms of

formative and summative assessments to gather a variety of data. Some studies (Boudett et. al., 2005) created committees of educators, administrators, and graduate students in a workshop setting to explore specific strategies of analyzing data from students' assessments. They then attempted to use the results to directly pinpoint weaknesses in curriculum, instruction, or methods of teaching. Other efforts (Murnane, Sharkey, and Boudett 2005) have focused on the use of technology, databases, and student information software to expedite the process of data entry and interpretation and to generate custom reports for the sole purpose of item analysis and evaluation.

This study is an effort to focus on the latter; how can student-assessment results assist administrators and educators evaluate and improve curriculum and instruction? Throughout this study, the issue of utilizing assessment data to evaluate curriculum and improve instruction is explored. Current practices in public education, challenges and obstacles facing administrators and educators, and recommendations and solutions to those challenges are topics that will be addressed. The remainder of chapter one contains the following in reference to this study: the background, justification, and significance, the problem, purpose, and objectives, the research questions, and the definitions of key terms.

Background, Justification, and Significance of the Study

As mentioned above, since .N.C.L.B., assessment has become the predominant topic of interest in public education today. Public school districts are challenged to create procedures to improve their delivery of instruction to all students through the use of assessments. One public schools response to the N.C.L.B. legislation was an initiative that set a goal to develop district-wide end-of-course assessments in communication arts that would eventually span all courses in grades seven through twelve. Beyond evaluating students' progress towards standards, the

primary purpose was to collect data on student achievement that would enable teaching staff to identify strengths and weaknesses in curriculum and methodology. Teachers from across the district formed committees by the grade level they taught to begin planning and developing performance based writing assessments. The assessments were created by teachers for the courses they currently taught and were based on the district standards and curriculum. Once created, these common assessments were to be administered by all communication arts teachers in grades seven through twelve.

The performance based writing assessments were administered within a specific window of time depending upon grade level (eg. all junior communication arts teachers would have their students complete the assessment in the month of December). The assessments were collected at a central location and then a team of teachers met for two full days to score each student's response given a standardized six point rubric. Each response was read and scored twice (once by two different teachers). If there was more than a one point discrepancy in any student's score, a third teacher would read it to determine which score was more appropriate and insure inter-reader reliability. When all the assessments were scored, they were entered into a district database and eventually returned to the students with feedback in the form of their scores and highlighted comments on the rubrics. Teachers were given access to the results and the data for their own students in the form of global percentages.

This study was created and conducted in response to the need for a more comprehensive and systematic method to utilize these assessments and students' scores on them to identify strengths and weaknesses in communication arts curriculum and instruction. This was the initial reason why these writing assessments were created. But, that objective has yet to be realized. The global scores and data provided by the current assessment rubric do not enable educators to

target specific deficit skill areas of student or their progress toward specific curriculum standards. The results also lack detailed data to inform and modify instruction. A new process and rubric that would enable assessment results to be quantified was needed. The new objective was to create an assessment rubric that would provide detailed data that is functional to educators and administrators. The ultimate goal is to create an assessment instrument that would generate data to pinpoint student strengths and weaknesses. Then, educators and administrators could have informed discussions and make critical decisions related to evaluating curriculum and improving instruction.

It was the goal of the researcher that this study of a curriculum based writing assessment would contribute to the field of existing research and knowledge about the use of student assessment results to assist educators in the evaluation of curriculum and the instruction. While others have performed similar research, it was hoped that the specific focus on writing assessments would make a unique contribution to the subject of communication arts instruction. With an understanding of the background, rationale, and significance of the study, the general purpose of the study becomes clearer.

General Purpose of the Study, Research Questions, and Definitions

The purpose of this project was to conduct a “case study” of an assessment instrument currently used to measure students’ writing ability. The assessment instrument (rubric) was evaluated specifically for its ability to provide detailed data for two purposes: (a) pinpointing strengths and weaknesses in student writing both individually and collectively and (b) evaluating curriculum standards and improve instruction.

The Research Question

The general question this study attempted to answer was, "How can an assessment instrument (rubric) better assist in the collection of data for the purpose of evaluating curriculum and improving instruction." This question subsumes other, more focused sub questions.

Sub questions.

1. What are the strengths and limitations of the current rubric in regard to providing that data?
2. Can the current rubric be modified and revised into a pilot rubric that would provide more functional data?

Definition of Key Terms

The following terms warrant definition as they specifically relate to this study:

1. Performance assessment: A measure of assessment based on authentic tasks such as activities, exercises, or problems that require students to show what they can do.
2. Rubric: An instrument or tool designed to measure in an objective, standardized manner.

Summary

Assessments can be a useful tool not to just gauge student achievement, but also as a means to evaluate curriculum and improve instruction. While there is a tremendous focus on the former through the influence of N.C.L.B., much is to be learned about the latter. If assessments are used solely as an instrument to measure students' progress towards reaching standards and their value as an instrument to evaluate curriculum and instruction is neglected, then administrators and teachers are missing an excellent opportunity to improve education. The ensuing chapter contains a review of the current literature relevant to this study. The review relates five of the predominant topics associated with using student assessment results to

improve curriculum and instruction: They are: the significance and importance of the process, current practices and methods of using assessment as an instrument of educational improvement, the challenges and obstacles faced by administrators and educators, and recommendations for school districts to meet those challenges and obstacles.

Chapter 2: Review of the Related Literature

As discussed in chapter one, the assessment of students serves two primary purposes in public education today: to appraise achievement and to improve instruction. Since the authorization of N.C.L.B. in 2001 and its mandates for accountability, choice, and flexibility, assessment has become paramount in public education. Truly, “No Child Left Behind has lit an assessment fire in our nation...” (Stiggins & Chappuis, 2006, p. 13)

Much has been written regarding the purpose and value of student assessments to measure progress and evaluate and improve curriculum and instruction. Scholars and studies relate how educators and administrators currently use assessment data to improve instruction, the obstacles and challenges they meet in the process, and recommendations and solutions to some of those obstacles. From standardized state examinations to district mandated open-ended writing responses, an immense amount of assessment data exists for administrators and educators to evaluate the quality of instruction they provide to students. The use of this data is critical to the process of educational improvement. The existing literature reveals administrators and educators are very adept at using assessments as a measure of student learning. However, the literature indicates there is a significant deficiency in administrators and educators ability to use assessment results as a tool for student learning.

The following is a review of the relevant literature regarding the use of assessment results to improve curriculum, instruction, and student learning. To ensure a focused summary of the literature, this review is organized in a topical or conceptual format. To narrow the scope, the focus of this review will address four of the predominant topics as they relate to using student assessment results to improve curriculum and instruction: significance/ importance, current practices/methods, the challenges/obstacles, and recommendations/solutions.

Significance/Importance

As stated earlier, the N.C.L.B. legislation of 2001 created a significant paradigm shift in public education. Six years later, school districts are still scrambling in a reactionary manner in the attempt to meet the mandates and comply with both the spirit and the letter of the law. As stated in the N.C.L.B Executive Summary, "School districts and schools that fail to make adequate yearly progress (A.Y.P) toward statewide proficiency goals will, over time, be subject to improvement, corrective action, and restructuring measures aimed at getting them back on course to meet state standards." With schools mandated to document adequate yearly progress of all students or face sanctions, the pressure to produce has become immense. As stated by Boudett et al., "State accountability systems that are based on test data and the N.C.L.B Act have put educators under great pressure to improve their students' scores on standardized tests" (2005, p. 700). Standardized assessment truly has become high stakes as schools failing to meet benchmarks by proving evidence of adequate yearly progress via assessment results face sanctions and remediation by the federal government.

In response, much has been written about the importance and advantages of using assessment data not just to measure "adequate yearly progress," but also as a means to improve instruction. This process had become critical as public school districts strive to satisfy the requirements of N.C.L.B. A common theme across the published literature emphasizes the critical need for educators to take full advantage of and utilize data from student assessments to improve curriculum and instruction. "If schools do not make constructive use of their test results, the net effect of the No Child Left Behind legislation will likely be a reduction in student learning" (Sharkey & Murnane, 2003, p. 79). Many scholars encourage administrators and educators not to view N.C.L.B. as an imposed mandate, but as an opportunity to create

procedures to assess their own curriculum and teaching methods. This is something sorely lacking in many districts and will be addressed later under “Challenges and Obstacles.” Shellard is one scholar who echoes this stance. “The key to school improvement in schools with standardized curriculums is the data obtained from frequent assessments of student performance. Principals and teachers can use this data not only to determine how well their students are learning but also to identify areas where intervention or changes in instruction are needed. Using data to drive improvement shows up as a characteristic of many successful schools” (2005, p. 30). The benefits of the process are clear. In their study, Murnane, Sharkey, and Boudett found schools that created formal procedures to use assessment results as a self-evaluation tool found “important implications for instructional improvement” (2005, p. 269). Or, as stated by Guskey, “Assessments can help improve education, but as long as we use them only as a means to rank schools and students, we will miss their most powerful benefits” (2003, p. 11).

In the past six years since the enactment of N.C.L.B., scholars have studied and explored public education’s attempt to meet the legislation’s mandates and expectations. A number of school districts have put in place procedures and practices with varying degrees of success. The next section is a brief summary of some specific strategies currently utilized by administrators and educators as found in the literature.

Current Practices/Methods

In response to N.C.L.B. mandates, many school districts have created a number of assessments and procedures to utilize data results to improve instruction. Although according to Murnane, Sharkey, and Boudett (2005), “Learning from student assessment results is an important activity in relatively few schools” (p. 270). Others employ various methods to collect, analyze/interpret and utilize student assessment data.

Collecting Assessment Data

Currently, the collection of data from students is achieved through two primary types of assessments: formative and summative. Stiggins and Chappuis further clarify the difference by describing summative assessments as assessments “of learning” and formative assessments as assessments “for learning” (2006). ●f the two, formative assessment results prove to be much more valuable for the purpose of evaluation and improving curriculum and instruction.

Summative assessments are “high-stakes” tests designed to provide information on the performance of districts and schools so resources and support can be well targeted (Using Classroom Assessment, 2006). These types of assessments are not particularly useful for classroom teachers. They provide information to teachers such as which students failed to master a particular skill, but they fail to tell what type of instruction a student needs to learn it. Specific information to assist educators in adapting instruction is obtained through classroom-based formative assessments.

Formative assessments are a form of summative assessment administered more often at regular intervals (often at the end of school calendar quarters). Their primary purpose is to determine which students are struggling and not progressing towards mastery of standards within the curriculum. It enables educators to identify students in need of assistance when that assistance is still timely and not too late. Stiggins and Chappuis refer to formative assessments as “assessment for learning.” They purpose ““assessment “for learning” happens in the classroom and involves students in every aspect of their own assessment to build their confidence and maximize their achievement. It rests on the understanding that students, not just adults, are data-driven instructional decision makers” (2006, Abstract). Whether summative or formative, once

collected, assessment data can be analyzed, interpreted, and then used to evaluate and improve curriculum and instruction.

Analyzing & Interpreting Assessment Data

The most efficient means of analyzing, interpreting, and distributing assessment data is through technology. The creation of district database systems and student information system software allows schools to access and analyze assessment data much more efficiently than in the past.

Murnane, Sharkey, and Boudett (2005) provide a specific example from Boston Public Schools (B.P.S.). B.P.S. uses an online database system referred to as “MyBPS.” MyBPS provides educators with basic student information such as class schedule and basic demographics. But, more importantly, this student information database allows administrators and educators to access assessment results with relative ease and efficiency. Administrators and educators can also generate multiple individualized reports in a variety of formats. Through these generated reports, educators can conduct detailed analysis including item analysis that enables them to determine which questions students did well on and which ones they struggled with. This allows educators to link test items to standards in the curriculum and enables them to review and revise instruction.

One of the more formal methods used to analyze assessment data to improve instruction has been through workshops. Boudett et al conducted a year-long workshop for teachers and administrators from Boston Public Schools and for students from the Harvard Graduate School of Education. They created school-based teams consisting of teachers, administrators, and graduate students in education. Its primary purpose was to help ten schools within B.P.S. use assessment data to evaluate and improve curriculum. The workshop explored different types of

data and data analyses and helped teams use data to analyze a school-specific problem and create and action plan. Its focus was to “place data analysis in the greater context of school improvement” (2005, p. 700). They spent the majority of their time working with real school data to solve problems. Their goals was “for participating schools to benefit from a structured and supported opportunity to make progress on work they needed to do and for Harvard graduate students to benefit from a truly authentic learning experience” (p. 701).

A common characteristic of many successful schools is their ability, as an entire staff, to collect, analyze, and interpret assessment data. Shellard reports that six schools districts, all of which had significantly increased student performance on state-mandated tests despite serving high percentages of at-risk students, experienced significant results by modifying instruction through assessment results. “All districts began their improvement after carefully reviewing assessment data” (2005, p. 31). It is only after student assessment data has been collected, analyzed, and interpreted that the most important part of the process can take place, putting assessment data to work to evaluate instruction.

Using Assessment Data

Murnane, Sharkey, and Boudett found that administrators and educators use student-assessment results and data in at least three ways: (a) to make decisions, (b) to justify a decision that has already been made or to support a predetermined stance, and (c) to develop discussion about what students know, what they are able to do, and how effective instruction has been (2005, p. 270-271). Clearly, in terms of improving instruction, the value lies with the latter. Current practices leave many educators and administrators with challenges and obstacles to overcome if they are to fulfill the requirements of N.C.L.B. and fully utilize assessment data to improve curriculum and instruction.

Challenges and Obstacles

Despite their best efforts through current practices and methods, public school administrators and educators face multiple significant challenges when they attempt to utilize student assessment results to improve curriculum and instruction. The major challenges and obstacles faced can be classified by the “four Ts”: time, technology, training, and tension.

One of the most significant challenges faced by educators in taking advantage of assessment results is time. With their daily responsibilities of lesson planning, teaching, grading, communicating with colleagues and parents, finding time to analyze assessment results to evaluate their curriculum and teaching techniques is a difficult task for most teachers. As stated by Shellard, “However, the multitude of demands placed on educators and the limited time available necessitates a system that is time-efficient as well as effective” (2005, p. 31). Sharkey & Murnane echo this sentiment saying, “Because time is the scarcest resource in most schools, teachers and administrators will learn from student assessment results only if they can obtain useful data quickly” (2003, p. 78).

The data obtained by teachers and administrators and their ability to use it quickly and effectively is dependent upon technology. Today, computers and software enable school districts to enter and access data with much more efficiency than in the past. According to Hall, “N.C.L.B. calls for a new type of data management and reporting system for schools and districts” (2004, p. 32). Larger districts are more likely to have student information systems or databases that enable educators to access and generate a variety of reports on both formative and summative assessments. The key to technology’s role in using assessment data to improve instruction lies in its availability and “user-friendly” format. District databases are ever-changing and evolving. With rapid advancement in bigger and better student information system software,

school districts often update or entirely change the software used to store and access student data with little or no transition for teaching staff. In order for educators to find the technology “user-friendly,” sufficient training is necessary.

Perhaps the biggest inhibiting factor facing educators in their attempt to use assessments to improve instruction is a lack of training, skill, and expertise. According to Hall, “The accountability measures associated with N.C.L.B are more numerous and more difficult to collect and understand” (2004, p. 32). Sharkey & Murnane found training that schools did provide for their teachers and administrators in using data tends to focus on how to use technology and software (2003, p. 79). However, utilizing and learning from student assessment results lacks priority in a majority of schools. Murnane, Sharkey, and Boudett report many schools’ staff lack the expertise to learn from assessment results. They also found that educators “varied markedly in their knowledge about analysis of student-assessment results” and “many participants had never looked at the questions asked on a standardized state test” (2005, p. 269 & 273). Sharkey and Murnane found that “teachers and administrators also need to learn a more difficult set of skills: how to ask instructionally relevant questions of data and how to answer such questions. Most educators have not learned these skills in pre-service training” (2003, p. 79). Some educators lack of knowledge or ignorance regarding the usefulness of student assessment data results in their failure to understand or appreciate the significant benefits of such a process. Other schools assessment experience was simply limited to analyzing scores on standardized assessments strictly because of accountability requirements. They failed to see the process as a means for evaluating and improving instruction. The fact that administrators and educators struggle with issues of time, technology, and lack of training, often leads to the last significant obstacle in their attempt to utilize assessment data to improve instruction, tension.

Tension associated with assessment data is a considerable barrier for both administrators and educators. For educators, the primary source of tension results from fear that students' assessment results will be used as a point of comparison among other teachers by administration. Others report concerns that assessment results may be used for evaluation purposes. Due to these factors, many teachers become reluctant to participate in the process. As indicated by Sharkey and Murnane, "teachers' feelings of vulnerability can jeopardize the activity. Comparison of results are inevitable and potentially threatening" to many educators (2003, p. 80). What ensues is a lack of cooperation and the direct result is a stalemate in the process. When teachers become reluctant to share and discuss the assessment results of their students, the opportunity for the entire school to learn larger lessons becomes limited. Teacher tension relative to assessments is also found in their view of administrative commitment and support. Teachers who do not feel administration is "in it for the long haul" and are not committed to the process also feel reluctant to initiate or participate in the process. Others question whether their analyses of the data and the subsequent changes it may dictate will be supported. Tension felt by administrators lies not in questions of comparisons, evaluation, or support, but issues of access and privacy or confidentiality and liability. Once again, Sharkey and Murnane reveal, "Many administrators worry about liability in giving teachers access to potentially sensitive information on students whom they do not teach (2003, p. 79).

The N.C.L.B. legislation mandates public schools to use assessment data to ensure all students are progressing toward standards and experience success. However, it also assumes that administrators and educators are already trained and have the time and support to do so. The existing research indicates this is simply not the case. Despite a number of significant obstacles and challenges faced by administrators and educators in their quest to efficiently and effectively

utilize student assessment data to improve curriculum and instruction, the literature does reveal some solid recommendations to overcome them.

Recommendations/Suggestions

The literature shares recommendations for administrators and educators that will assist in efforts to improve curriculum and instruction through assessment results. These recommendations are meant to assist schools in addressing some of the obstacles and challenges they face in their efforts to meet the mandates of N.C.L.B. More importantly, they provide strategies to connect assessment to school improvement. There are several suggestions for administrators and educators to address the “four Ts” discussed previously in “Challenges and Obstacles”: time, technology, tension, and training.

In an attempt to address the “four T” challenges, Sharkey and Murnane offer the following specific suggestions for learning from student assessment results (2003, p. 78).

Time/Opportunity.

1. Build time into the schedule for examining student assessment results.
2. Provide processes for teachers to talk about their students’ work.

Technology/Tension.

1. Update information frequently to reflect student mobility.
2. Enable staff to “define instructionally relevant groups of students in the databases and examine assessment results for these groups.”
3. Give teachers and administrators software that is easy to use and that addresses the questions that teachers frequently ask, both about performances

of their current students relative to benchmarks and about the performances of the students they taught in the previous year.

4. Design the database system so that school principals can use discretion in resolving the tension between access and privacy.

Training/Expertise.

1. Teacher training on formative assessments would contribute to more reliable scoring.
2. Prepare teachers and administrators to use the software and to frame and address instructionally relevant questions.
3. Invite teams from schools with expertise in analyzing assessment results to demonstrate to educators from other schools what they have done and how it contributed to instructional improvement

They also recommend that districts acquire student information databases and software that stores data centrally but enables administrators and educators to access assessment results for specific groups of students in their schools. Effective databases make it simple for educators to access end-of-year summative assessment data to incorporate with their own formative assessment results.

Shellard (2005) suggests that school faculty meet to discuss and create a list of the twelve most important standards for each subject. They should then design short formative assessments that over the course of the school year, address all twelve standards. Administrators should then provide time in faculty meetings or scheduled group time to grade the assessments to determine, based on the results, what changes may be needed in the schedule, curriculum, and teaching strategies in the coming weeks, months, or following year. Love (2001) affirms, assessments

should be an “integral part of instruction... The standards talk about curriculum, instruction, and assessment together as a single vehicle... What unites them is a common focus on student learning.” Shellard (2005) believes that for schools to meet the challenges of N.C.L.B., they must integrate all three: curriculum, instruction, and assessment.

Summary

Assessment of student learning has become and continues to be one of the most significant issues in public education today. N.C.L.B truly “has lit an assessment fire in our nation...” (Stiggins and Chappuis, 2006, p. 13). Administrators and educators attempting to integrate curriculum, instruction, and assessment to meet the N.C.L.B. mandates clearly have a formidable challenge. From current practices and methods, obstacles faced by schools, to suggestions and recommendations of scholars, the literature reveals the complexities of using student assessment data to evaluate and improve curriculum and instruction. The existing literature reveals much has been learned in the past six years since the authorization of N.C.L.B. legislation, but, there is still much to be learned. There are significant deficiencies in the existing literature detailing specific, tangible techniques teachers can implement immediately to evaluate their curriculum and instruction techniques. As aptly stated by Sharkey & Murnane, “The increased emphasis on student testing will improve the quality of education only if educators use student assessment results effectively. If we want all schools to use test results to inform meaningful school-wide instructional improvement, we must find ways to overcome the technology, knowledge, and opportunity challenges inherent in this work” (2003, p. 81)

The following chapter will detail the methodology used in this “case study” of an existing assessment’s ability to generate data for the purpose of evaluating curriculum and improving instruction.

Chapter 3: Methodology

The two primary purposes of student assessments are a means to measure student progress and achievement, and to evaluate curriculum and improve instruction. This study was created in response to one school districts' need for the latter. The study involves creating, evaluating, and piloting a new assessment rubric in a high school communication arts course. The goal is to attain data that could be quantified to measure the effectiveness of curriculum standards and instructional strategies. This information would be shared with administrators and educators with the intent of evaluating the current curriculum and improving instructional strategies. Although an assessment process was in place, the present instrument provides global data that does not enable educators to draw specific conclusions about students' strengths and weaknesses. These global scores did not target specific deficit skill areas. The rubric results did not provide administrators and educators with the means to make specific comparisons between the assessment results and curriculum standards. This is a prerequisite to using an assessment to improve instruction. Thus, a new assessment instrument was needed to acquire quantitative data that could be used to evaluate specific standards within the curriculum. These results would also enable educators to evaluate their instructional strategies to improve their teaching techniques. This chapter relates the methods used in executing this case study of a new assessment instrument. The research setting, participants, instruments, procedures, and methods of data analysis are described.

Setting

This study was conducted between September and December of the 2007-2008 school year in a public school district in Northeastern Wisconsin. The school district serves students in a city of approximately 77,000 residents. Ranked by size, it is one of the state's ten largest school

districts with an enrollment of approximately 15,000 students in 2007. The district contains forty schools categorized as: traditional, alternative, specialized charter, virtual academy, and online. Of those, nine are high schools (three traditional, two alternative, three specialized charter, and one online) serving approximately 5,000 students. This study was conducted with the assistance of faculty from the three traditional high schools.

Participants

The participants in this study were ten teachers from the school district's three traditional high schools. All of the participants were certified in and currently teaching communication arts (English) in grades nine through twelve. All were district colleagues or fellow faculty with the researcher. The participants were selected based on their knowledge, skill, and expertise in using standardized rubrics to score performance based writing assessments as evidenced by the following criteria. All participants had: (a) Five or more years teaching experience in communication arts including instruction and assessment of student writing, (b) Formal training by the school district in the process of administering and scoring performance based writing assessments with a standardized rubric, and (c) Participated in the formal process of reading and scoring district-wide standards-based student writing assessments with a standardized rubric.

Instrumentation

Several specific instruments were reviewed and utilized in the data collection process for this study. Two assessment instruments (rubrics) used in scoring student essays were used. The assessment rubric currently in use by the district in this study (Appendix B) was revised and modified into a new pilot rubric (Appendix C). A survey was developed seeking teacher responses to comparisons between rubrics and their input on the use of the rubrics to assess

student writing and evaluate curriculum standards. (Appendix D). A cover page accompanied the survey with specific directions to be followed by the participant.

Procedures

The researcher began the study by analyzing and evaluating the district's current procedures for assessing eleventh grade students' writing skills through a performance based assessment. The current writing prompt and directions were reviewed as well as the corresponding rubric used to score student responses (Appendix E & B). Data reports of the scores from students who took the assessment during the 2006-2007 school year were analyzed. The data was reported in simple statistics including averages, mean, median, and mode. Student scores on the six-point scale were reported in percentages. These reports enabled teachers to compare each of their classes with one another, with the school average, and the district average. Comparisons could be made between classes taught by the teacher, other classes at the same school, and other classes across the district. After the review of the instrument and data reports, it became clear that the rubric provided scores and data that lacked functionality. Although students benefited from the experience and the individual feedback of their writing skills, there was little value for educators in terms of evaluating specific curriculum standards they taught or instructional methods they used. The assessment did not provide educators with data they could use to evaluate specific target deficit skill areas of students. This prompted the researcher to modify the rubric to produce data that would be functional and assist educators in evaluating specific curriculum standards and instructional methods for the broader objective of improving instruction.

The researcher modified the current rubric by converting general comments under the six score descriptions into six specific skill categories: (a) Introduction/Conclusion, (b) Task,

(c) Organization/ Transitions, (d) Development/Focus, (e) Language/Vocabulary, and (f) Sentence Structure (Appendix C). This new pilot rubric would assess the same skills, but would enable the results to be reported as strengths or weaknesses in a specific skill. It was hoped that this would enable educators to then use the data results to make more meaningful comparisons. This new pilot rubric would generate data that could be used to evaluate specific curriculum standards and teaching techniques.

Once created, face validity was established through the review of the pilot rubric and the survey instrument by three individuals: an Associate Professor in the Department of Special Education at the University of Wisconsin Oshkosh, the Director of Humanities for the school district of the researcher who coordinates communication arts curriculum, instruction, and assessment at the secondary level, and a high school communication arts teacher who also served as the Curriculum Support Specialist in the high school of the researcher. With suggestions and feedback from these evaluators, the rubric and survey were revised into their final form (Appendix C & D). Prior to the collection of data, approval for the study including instruments used to gather data was granted by the Institutional Review Board (I.R.B.) of the University of Wisconsin Oshkosh and the Director of Humanities of the school district of the researcher.

After approval, the participants were sent a letter explaining the rationale of the study and the procedures involved (Appendix H). Participants also received a consent form for the investigation requesting their participation (Appendix I). Participants then received the instrumentation documents for the study: (a) a sample of a student response to the assessment prompt (Appendix F), (b) a copy of the initial rubric (Appendix B), (c) a copy of the new, revised pilot rubric (Appendix C), and (d) detailed directions with the survey (Appendix G).

The participants were asked to read the sample student response, score it using the original rubric, score it using the revised rubric, and then complete the survey. Data was collected through teacher responses to survey questions. Each survey item had a corresponding Likert rating scale. Other items on the survey included questions of direct comparison between the two rubrics and open ended response questions. The participants were asked not to discuss the process with anyone else until they had completed the rubrics and the survey. The participants were given a two week window to complete the scoring and the survey. Participants failing to return the survey after two weeks were sent an email reminder. The results were returned to the researcher via interdepartmental mail from all three high schools.

Data Analysis

The resulting data from the survey instrument and procedures described above were organized and analyzed both qualitatively and quantitatively. Likert scale responses were recorded, averaged, and analyzed. Open ended questions were recorded, compared, and organized into categories of similarities and differences. Reliability was measured in two ways. Inter-rater reliability was measured through analysis and comparison of participant responses on both rubrics. Reliability was also assessed through specific questions on the survey. Validity was also assessed through participant responses to specific survey questions. Concurrent validity was measured through a comparison of global scores between both rubrics. Treatment fidelity was addressed through analysis of participant survey responses and evaluating their ability to complete the survey thoroughly, correctly, and effectively. These results were reported and interpreted in narrative text and displayed in bar graphs and charts within tables for visual interpretation.

Summary

This chapter explained the methodology used in this study of a pilot assessment rubric. The “subject” of this case study is the rubric used in a standardized writing assessment among three high schools in a northeastern Wisconsin school district. After an evaluation of the current rubric and the data results from the 2007-2008 school year, the researcher decided to modify and revise the original rubric into a new, pilot rubric. With feedback from an associate professor, an administrator, and an educator, the initial pilot rubric and survey were critiqued and modified into a final survey and pilot rubric. Then, both rubrics were given to ten communication arts teachers with the survey seeking their professional input on the new rubric’s functionality and ability to provide data results from student writing that can more usefully be used by administrators and educators in evaluating curriculum standards and improving instruction. The subsequent chapter presents the results obtained from these methods.

Chapter 4: Results

As stated in Chapter 1, the No Child Left Behind Act of 2001 has created a marked difference in the way in which public education views and uses student assessment results. Curriculum based assessments have dual purposes. Their primary purpose is to assess students' progress towards reaching curriculum standards. Also, assessments are to be used as the instrument to critique curriculum and improve instruction.

The purpose of this survey study was to explore the latter of the two. Ten teachers of high school communication arts volunteered to become participants. They were given an anonymous student essay and were asked to score it using the current rubric used by the district of their employment and a new, pilot rubric. Then, they responded to survey questions relative to each rubric and that rubric's ability to provide functional data to critique curriculum and improve instruction.

This chapter will relate the results of the participants' assessment of the sample student essay and their responses to the survey questions. The organization of this chapter reflects the structure of the survey. Part one focuses on the demographics of the participants. Part two relates participant's perceptions of assessments as a tool of curriculum improvement. Part three reveals the survey results relevant to the current rubric used by the district of the participants. Part four relates the survey results from the pilot rubric. And, part five focuses on a direct comparison of survey data of both rubrics.

Participant Demographics

The participants were selected through an email invitation to all high school communication arts teachers employed in the district of the researcher. As described in Chapter 3, there are three public high schools within a district that services approximately 77,000

students in grades kindergarten through twelve. For the purpose of this study, these three high schools are referred to as schools “X,” “Y,” and “Z.” Thirty-eight high school communication arts teachers were emailed. Ten teachers from the three schools volunteered to participate in the study. For the sake of identification, they were randomly assigned with the letters “A” through “J.” There were representatives from all three high schools. Even though participants were selected randomly from a convenience sample, there was a relatively even distribution of participants from all three schools. Refer to Table 1.

Table 1: Participants by School

School	X	Y	Z
Total CA Staff	14	11	13
Participants	3	4	3
Percentage of Staff Participating	21%	36%	23%

Personal demographic information surveyed from participants included years of teaching experience and current teaching position by grade level. The average number of years of experience teaching communication arts for all ten participants was 13.5 years with a range of 29 years between the teachers with the least to most experience. Current grade levels taught ranged among all four grades in high school. Refer to Table 2.

Table 2: Teaching Experience of Participants

Participant	Years Teaching Experience	Current Grade Level(s) Teaching	High School
A	9	9th & 10th	X
B	5	9th, 11th, & 12th	Z
C	21	11th	Z
D	10	12th	Y
E	11	9th	Y
F	15	11th	Z
G	8	10th	Y
H	17	11th & 12th	X
I	34	12th	X
J	5	9 & 11th	Y
	Mean 13.5 Years		
	Range 29 Years		

Participants were sent the following via interdepartmental mail: (a) a direction sheet, (b) a summary of the research project, (c) the IRB Short Form Consent Document, (d) the current junior level writing prompt used by the district of the study, (e) an anonymous student essay response to the writing prompt, (f) the current rubric used by the district to assess the junior level writing assessment, (g) a pilot rubric, and (h) a survey. After reading the directions, the writing prompt, and the student essay, participants scored it using the current rubric (A), the pilot rubric (B), and then responded to questions on the survey instrument. The survey included questions with a five point Likert Scale and open ended response opportunities. The next section reveals participant’s perceptions of the use of assessment data to critique curriculum and improve instruction.

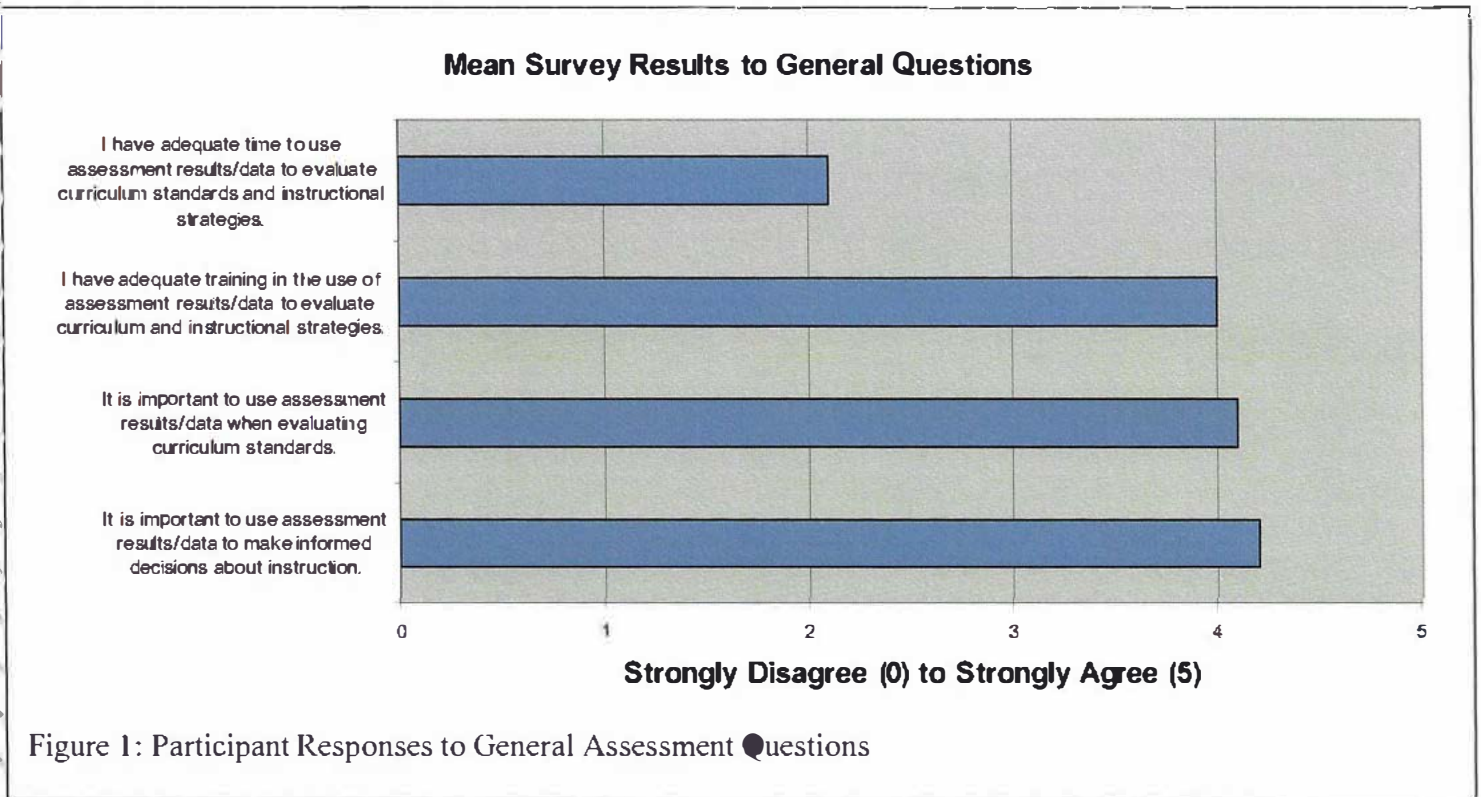
Participants Perceptions of Assessment

On the survey, participants were asked to respond to four general questions soliciting their opinion regarding the following: (a) the use of assessments to make informed decisions about instruction, (b) the use of assessments to evaluate curriculum, (c) their training in the use of assessments to evaluate curriculum, and (d) time constraints on their ability to use assessment data to evaluate curriculum. The itemized results appear in Table 3.

Table 3: Itemized Participant Responses to General Assessment Questions

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
General					
It is important to use assessment results/data to make informed decisions about instruction.	3	6	1		
It is important to use assessment results/data when evaluating curriculum standards.	2	7	1		
I have adequate training in the use of assessment results/data to evaluate curriculum and instructional strategies.		10			
I have adequate time to use assessment results/data to evaluate curriculum standards and instructional strategies.			3	5	2

The responses of all ten participants were averaged. Results are seen in Figure 1.



The average responses of the ten participants revealed the following. The average response to the statement “I have adequate time to use assessment results/data to evaluate curriculum standards and instructional strategies” was a 2.1, slightly above “Disagree.” The average response to the statement “I have adequate training in the use of assessment results/data to evaluate curriculum and instructional strategies” was a 4.0, “Agree.” The average responses to the statements “It is important to use assessment results/data when evaluating curriculum standards” and “It is important to use assessment results/data to make informed decisions about instruction” were 4.1 and 4.2 respectively. Both falling slightly above the “Agree” category on the Likert Scale. After surveying participants about the relationship between assessments and

curriculum improvement, they were surveyed about their perceptions of the current rubric's ability to provide meaningful data for that purpose.

Current Rubric "A" Data

Participants were asked to score the anonymous student essay using the current rubric in use by the district of the researcher, hereafter referred to as "Rubric A." Participants were asked to calculate the time it took to score the essay using Rubric A and then they were asked to give the essay a global score as directed on Rubric A. Participant responses to these two tasks are illustrated in Table 4.

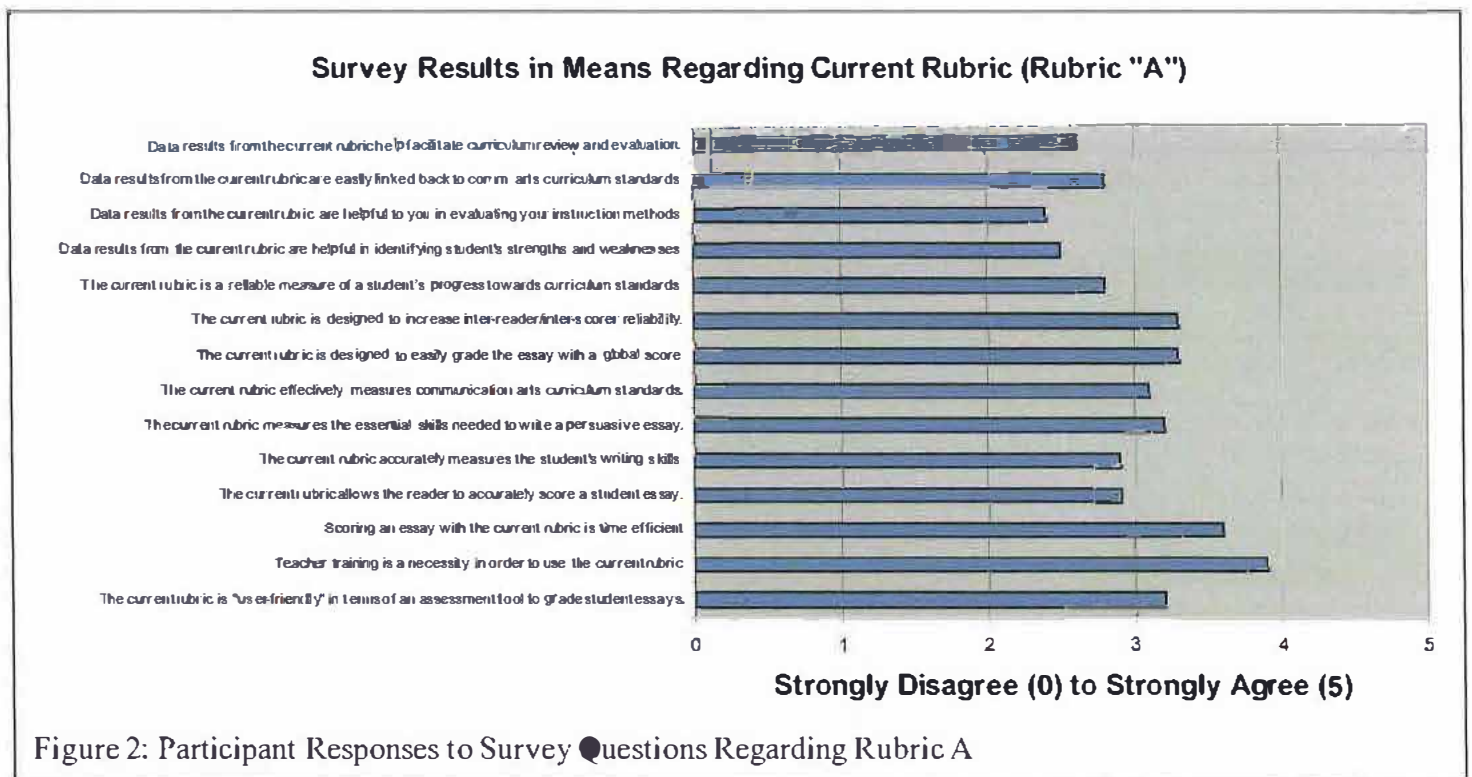
Table 4: Duration to Score Essay Using Rubric A and Holistic Score of Essay Using Rubric A

Participant	Time to Score Essay Using Rubric A	Global Score of Essay using Rubric A
A	12 Minutes	4
B	3 Minutes	5
C	2.5 Minutes	4
D	7 Minutes	4
E	5 Minutes	4
F	10 Minutes	3
G	2.5 Minutes	3
H	12 Minutes	3
I	4 Minutes	4
J	4 Minutes	4
Mean	6.2 Minutes	3.8

The average time to score the essay of all ten participants using Rubric A was 6.2 minutes. The average global score of the student essay using Rubric A was a 3.8 on a 6 point scale.

Participants responded to fourteen survey questions related to their experience in assessing the student sample using Rubric A. Responses were recorded on a five category Likert Scale that ranged from "Strongly Agree" to "Strongly Disagree." Likert Scale categories were given numerical values ranging from 5, "Strongly Agree, to a 1, "Strongly Disagree." Responses of all

ten participants to all fourteen questions were tabulated and averaged. The results appear in Figure 2.



After responding to questions regarding Rubric A, participants scored the same student sample using the pilot rubric. They then responded to the same fourteen questions, however, this time, they were related to their experience in assessing the student sample using the pilot rubric.

Pilot Rubric "B" Data

Again, participants were asked to score the anonymous student essay using the pilot rubric, hereafter referred to as "Rubric B." Participants were asked to calculate the time it took to score the essay using Rubric B and then they were asked to give the essay a global score as directed on Rubric B. Participant responses to these two tasks are illustrated in Table 5.

Table 5: Duration to Score Essay Using Rubric B and Holistic Score of Essay Using Rubric B

Participant	Time to Score Essay Using Rubric B	Global Score of Essay using Rubric B
A	4.5 Minutes	4
B	5 Minutes	4.5
C	3.5 Minutes	4
D	4.5 Minutes	4.5
E	5 Minutes	4.16
F	12 Minutes	4
G	4.5 Minutes	3.3
H	15 Minutes	3.33
I	5 Minutes	3.8
J	4 Minutes	4.33
Mean	6.3 Minutes	3.99

The average time to score the essay of all ten participants using Rubric B was 6.3 minutes. The average score of the student essay using Rubric B was a 3.99 on a 6 point scale. Participants again responded to the same fourteen survey questions related to their experience in assessing the student sample using Rubric B. Responses were recorded on the same five category Likert Scale ranging from “Strongly Agree” to “Strongly Disagree.” Likert Scale categories were again given numerical values ranging from 5, “Strongly Agree, to a 1, “Strongly Disagree.” Responses of all ten participants to all fourteen questions were tabulated and averaged. The results appear in Figure 3.

Survey Results in Means Regarding Pilot Rubric (Rubric "B")

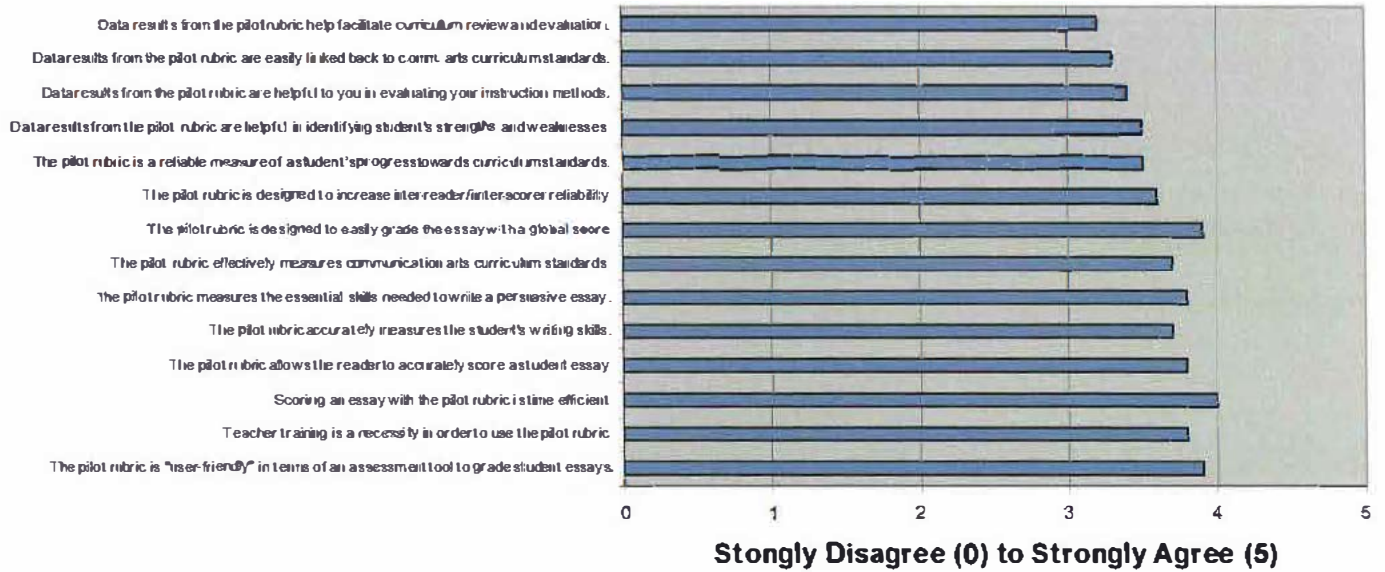


Figure 3 Participant Responses to Survey Questions Regarding Rubric B

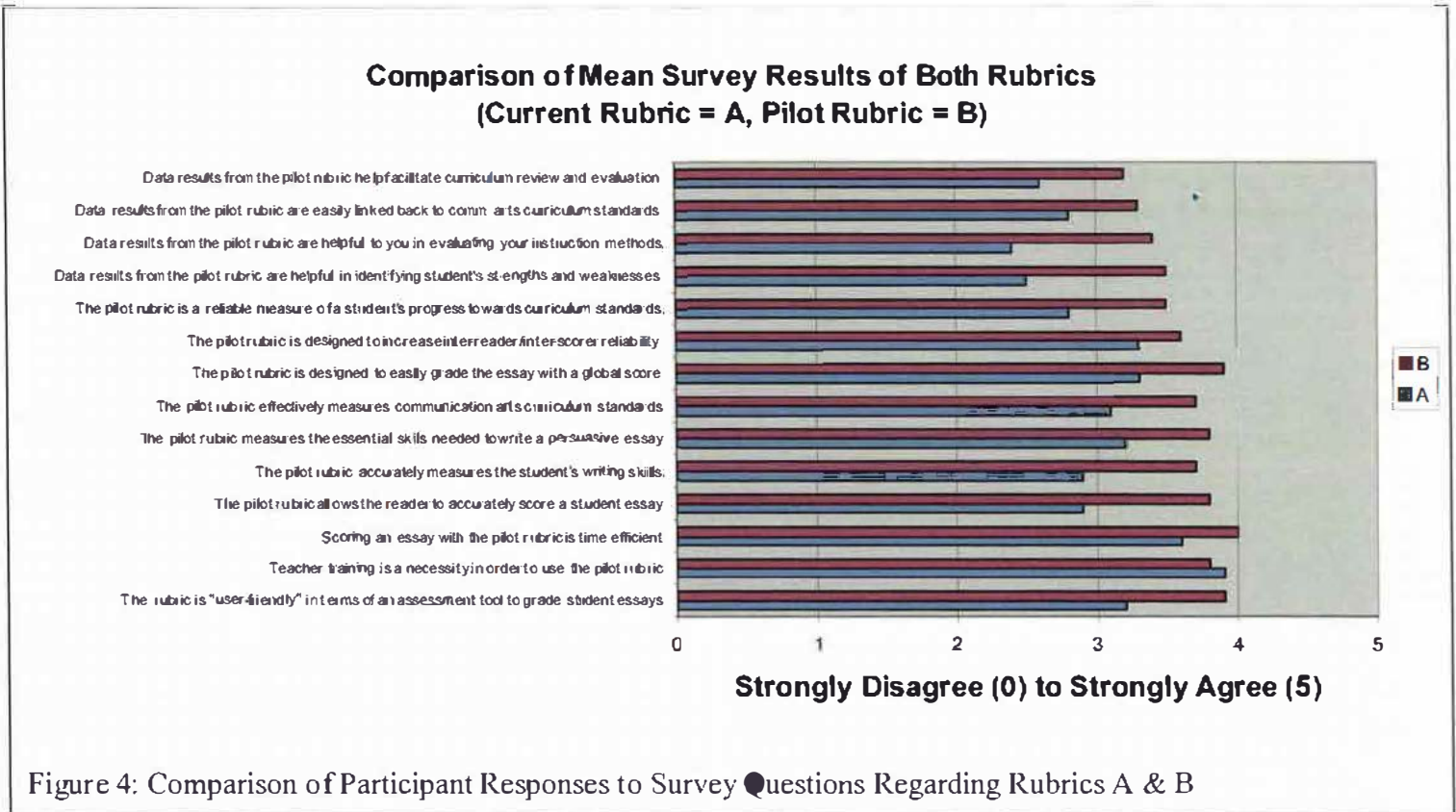
Comparison of Current Rubric "A" and Pilot Rubric "B"

As depicted in Table 6, a direct comparison of the time needed to assess the same student sample using Rubric A and Rubric B reveals a one-tenth of one minute difference. A direct comparison between the global scores of the sample essay between Rubrics A and B reveals a .19 point differential on a six point scale.

Table 6: Comparison of Mean Time and Global Scores of Rubric A and Rubric B

	Mean Time to Assess Student Sample	Mean Global Score of Student Assessment
Current Rubric (Rubric A)	6.2 Minutes	3.8
Pilot Rubric (Rubric B)	6.3 Minutes	3.99

When Figures 2 and 3 are merged, a much clearer comparison can be made regarding participant responses to the fourteen questions from the survey regarding Rubric A and Rubric B. The results can be seen in Figure 4.



The final three survey questions were answered after participants completed the assessment of the student sample using both rubrics and after they had completed the fourteen survey questions for each. The final three questions were: (a) Which rubric provides more functional data to be used by teachers to evaluate students' writing skills?, (b) Which rubric provides more functional data to be used by teachers to evaluate and modify instructions and lesson plans?, and (c) Which rubric provides more functional data to be used by teachers to evaluate and modify curriculum? The results of participant responses appear in Table 7.

Table 7: Itemized Participant Responses to Rubric Comparison Questions

	Current Rubric (Rubric A)	Pilot Rubric (Rubric B)	NA
Which rubric provides more functional data to be used by teachers to evaluate students' writing skills?	0	8	2
Which rubric provides more functional data to be used by teachers to evaluate and modify instructions and lesson plans?	0	8	2
Which rubric provides more functional data to be used by teachers to evaluate and modify curriculum?	0	7	3

Summary

Ten communication arts teachers employed in the district of the researcher were surveyed regarding the use of assessment results to critique curriculum and improve instruction. They participated in the scoring of an anonymous student essay using the rubric currently in use in the district and a new, pilot rubric. Their responses to survey questions relating their experience and opinion of each rubric's ability to provide data for the use of curriculum improvement is detailed above. The results presented above as well as narrative responses of participants to open ended questions will be more thoroughly summarized, discussed, and evaluated in terms of their significance in the following chapter.

Chapter 5: Discussion

Since the authorization of the No Child Left Behind Act (N.C.L.B.) of 2001 assessment of student learning has become and continues to be one of the most significant issues in public education today. In this high stakes assessment environment, it is a critical challenge for educators to efficiently utilize student assessment results to critique the curriculum they teach and improve the day-to-day instruction in their classroom. This study, *Assessing a Rubric: Evaluating a Rubric's Ability to Provide Functional Data to Critique Curriculum and Improve Instruction*, was this researcher's attempt to explore an option to address this challenge in the district of his employment. The results of this study were presented in the previous chapter. This final chapter will briefly summarize the following: (a) the purpose of the study, (b) the relevant literature reviewed for this study, and (c) the procedures used in the data collection process. It will also analyze, interpret, and discuss these results and attempt draw conclusions from the data. The limitations inherent in the methodology will be discussed and the implications of the results will be explored. The chapter concludes with recommendations for further research and suggestions for public school educators and administrators regarding the use of assessments and rubrics to improve the education of the students they serve.

Summation of the Purpose

The purpose of this project was to conduct a "case study" of an assessment instrument (rubric) currently used in the district of the researcher. The rubric was evaluated specifically for its ability to provide detailed data for two purposes: (a) pinpointing strengths and weaknesses in student writing both individually and collectively and (b) evaluating curriculum standards and improve instruction. The general research question which drove this study was "How can an assessment instrument (rubric) better assist in the collection of data for the purpose of evaluating

curriculum and improving instruction.” Two secondary questions were created in an attempt to narrow and focus: “What are the strengths and limitations of the current rubric used in the district of the study in regard to providing that data?” and “Can the current rubric be modified and revised into a pilot rubric that would provide more functional data?”

Summation of the Literature

The review of the relevant literature in chapter two was organized in a topical format narrowing the focus to four predominant subtopics: (a) significance/ importance, (b) current practices and methods, (c) challenges and obstacles, and (c) recommendations and solutions. The existing literature confirmed the importance of using assessment results to improve instruction and student performance and the challenges it poses on educators. As stated by Boudett et al., “State accountability systems that are based on test data and the N.C.L.B Act have put educators under great pressure to improve their students’ scores on standardized tests” (2005, p. 700). Current methods created by educators to meet the mandates of N.C.L.B. include both formative and summative assessments. Stiggins and Chappuis clarify the difference by describing summative assessments as assessments “of learning” and formative assessments as assessments “for learning” (2006). Of the two, formative assessment results prove to be much more valuable for the purpose of evaluation and improving curriculum and instruction. Despite their best efforts, educators face tremendous challenges in their attempts to utilize formative assessment results to critique curriculum and improve instruction. These challenges can be summarized in the “four Ts.” Time, technology, training, and tension are the major obstacles inhibiting educators from fulfilling the mandates of N.C.L.B. As administrators and educators grapple with these challenges, several schools have created solutions to the challenges of the “four Ts.” Suggestions include building time into the school schedule to examine assessment results,

creating and purchasing student information software systems to streamline the process, and collaborating with schools with expertise in analyzing assessment results. From current practices and methods, obstacles faced by schools, to suggestions and recommendations of scholars, the literature reveals the complexities of using student assessment data to evaluate and improve curriculum and instruction. Or, as stated by Sharkey & Murnane, "If we want all schools to use test results to inform meaningful school-wide instructional improvement, we must find ways to overcome the technology, knowledge, and opportunity challenges inherent in this work" (2003, p. 81). The existing literature reveals much has been learned in the past six years since the authorization of N.C.L.B., but, there is still much to be learned. That was the motivation for this researcher. The research procedures described in the methodology of this field report was an attempt by this researcher to learn how to better utilize assessment results in the district of his employment for the purpose of improving instruction.

Summation of the Procedures

The "subject" of this case study was the rubric used in a standardized writing assessment among three high schools in a northeastern Wisconsin school district. The current rubric was evaluated and modified by the researcher into a new, pilot rubric. A survey instrument was developed to solicit feedback from fellow communication arts teachers about their views on the effectiveness of each rubric's ability to provide functional data for the purpose of evaluating curriculum. With feedback from an associate professor, an administrator (Director of Humanities), and a fellow educator of communication arts, the initial pilot rubric and survey were critiqued and modified into a final survey and pilot rubric. Then, thirty-eight communication arts teachers from all three high schools were petitioned as participants. Ten volunteers became the participants in the study. Participants were sent: (a) a direction sheet, (b) a

summary of the research project, (c) the IRB Short Form Consent Document, (d) the current junior level writing prompt used by the district of the study, (e) an anonymous student essay response to the writing prompt, (f) the current rubric used by the district to assess the junior level writing assessment, (g) a pilot rubric, and (h) a survey. After reading the directions, the writing prompt, and the sample student essay, participants scored it using the current rubric (A), the pilot rubric (B), and then responded to questions on the survey instrument. All materials were returned to the researcher and the results were tabulated for analysis. The methods employed in this study produced some rather intriguing results

Interpretation of the Results

General Survey Questions

Participants responded to general survey questions regarding their ability to utilize assessment results to improve instruction. Responses of all ten participants, reflected in means, reveal they “agree” to “strongly agree” that it is important to use assessment data to make informed decisions about instruction and evaluate curriculum standards (mean response 4.2 and 4.1 respectively). They also “agree” that they have adequate training in using assessment results for this purpose (mean response 4.0). However, reflecting what was revealed in the review of the current literature in chapter 2, one of the four Ts, time, was related as a significant obstacle. The mean response to the question “I have adequate time to use assessment results to evaluate curriculum standards and instructional strategies” was 2.1, slightly above “disagree.” Clearly, time continues to be the major obstacle to effectively utilizing assessment data as outlined in N.C.L.B.

Time and Global Score Comparison of Both Rubrics

In a direct comparison of the student writing sample's global score and the time necessary to complete the scoring between Rubric A and Rubric B, participant ratings of Rubric A and Rubric B were very similar. The mean global score of the student sample using Rubric A of all ten participants was 3.8 from the 6 point scale. The mean global score of Rubric B was 3.99 from the 6 point scale. The difference in global scores was 0.19. The mean time to score the student sample of all ten participants using Rubric A was 6.2 minutes. The mean time for Rubric B was 6.3 minutes. The time differential (0.10 minutes) was nearly negligible. Neither rubric had a distinct advantage in terms of the time it took the participants to score it. Also, the pilot rubric appears to be a reliable measure of the student sample when compared to the rubric currently in use by the district of the study to assess student essays.

Survey Question Comparison of Both Rubrics

One of the sub-questions that drove this study was "can the current rubric be modified and revised into a pilot rubric that would provide more functional data?" Based on survey results and open-ended participant responses, the answer appears to be a resounding "yes." As seen in chapter 4 via tables, figures, and descriptive narrative, the resulting data from participant survey responses suggests the pilot rubric (Rubric B) has significant advantages over the current rubric (Rubric A) in terms of providing meaningful and functional data for the purpose of critiquing curriculum and improving instruction. When the participants were asked to directly compare both rubrics "head-to-head," in questions displayed in Table 7, none of the ten participants selected the current rubric over the pilot rubric in terms its ability to provide functional data to evaluate students writing skills, to evaluate instruction and lesson plans, and to evaluate and modify curriculum. As seen in Figure 4 in chapter 4, participant responses to the fourteen survey

questions assessing each rubric's ability to provide meaningful data, Rubric B "outranked" Rubric A on all fourteen questions. Open-ended participant responses support this conclusion.

In regard to the current rubric, Rubric A, Participant "D" remarked, the "layout is difficult to read. Score is too generalized and doesn't provide much opportunity for specific comments on strengths/weaknesses." Participant "E" responded, the "current rubric took time to read and decipher the differences between the scores. A score of "4" or "3" (I struggled a bit based on the rubric) doesn't give the student much feedback."

Participants also had specific comments regarding Rubric B's ability to provide meaningful data. Participant "C" stated, "Although Rubric A takes less time to use, the formatting of Rubric B is easier to explain to students. It also relates more clearly to the standards/instruction methods." Participant "G's" comments in reference to the pilot rubric were, "this rubric is better than the other because it separates writing elements. Students get the whole picture (global score) and see their strengths and weaknesses." And, Participant "F" said, "I prefer this rubric, though it still needs work. I don't think student failure to recognize a counter argument shows a weak understanding of persuasion, though it is certainly not adequate. The scores for "Task" still need work, but I believe this is a better rubric." Participant "E" stated, "Could layout criteria a bit better in terms of organization, but overall, very "user friendly." I liked this method much better." Participant "D" may have said it most succinctly by responding simply, "Use the pilot rubric..." In the head-to-head comparison, the participants clearly favored Rubric B, the pilot rubric, over Rubric A, the current rubric. And, given the negligible difference between global scores of both rubrics and time needed to assess the student sample, Rubric B clearly "outperformed" Rubric A.

Limitations

Prior to making recommendations based on the research results discussed above, limitations inherent in this study should be addressed. The original pool of possible participants for this study was thirty-eight communication arts teachers from three high schools within the same district. Having participants from three different schools versus selecting participants from the same school adds credibility to the results in terms of limiting building bias. The sample size of participants could have been larger. Ten teachers from a pool of thirty-eight meant 26% of the possible pool participated in this study. Ideally, this percentage could have been higher. Since all ten participants are district colleagues and acquaintances with the researcher, there is the potential for an unknown or “hidden” bias. Participant’s personal relationship with the researcher may have influenced their decision to participate or not to participate in this study and could have potentially influenced their responses to survey questions. Another potential limitation to the study was within the methodology and the scoring process of the student essay sample. All participants had prior experience in assessing student essays using the current rubric, Rubric A, which is still currently in use in the district of the study. None of the participants were familiar with the pilot rubric, Rubric B. The familiarity with Rubric A could have created bias. Participant’s use of Rubric A may have affected the time to score the student essay sample since participants were already familiar with it. In open ended comments, two participants reflected this possibility. Participant “A” stated, “I think “B” can provide better data eventually, but “A” is easier for me to use simply because I have scored with it so many times.” Participant “I” stated the following in regard to Rubric B, “It took me longer than the other only because I was less familiar, despite previewing...” Lastly, one must question whether the results attained from the pilot rubric will directly impact student achievement in a positive manner. This could potentially

become the next step in the process or become the basis for a future study. Despite these potential limitations, the findings of this study have multiple implications and enable this researcher to make specific recommendations to administration in the district of the study and broader, more general recommendations to public educators in general.

Implications/Recommendations

The results described above have created direct implications for the school district described within this study and broader implications for public school educators and administrators in general who strive to improve their procedures and policies to meet the mandates of N.C.L.B. From these implications, some clear recommendations can be made. First, specifically for the district described in this study, there would appear to be a significant advantage in adopting the pilot rubric to replace the current rubric used annually to assess student writing. If assessments are truly to be “designed as an instrument to critique curriculum and improve instruction and provide “evidence of improved teaching,” this must become a serious consideration. It is the intention of this researcher to share the results of this study and these recommendations with the administration of this school district.

The participants in this study offered some suggestions and recommendations of their own in response to the question, “What should be done to address the concerns and challenges of the current District-Wide Standards-Based writing assessments?” Participant “A” stated, “Discussion with administrators about what we would do with our time to discuss results. There also needs to be less complaining among our teachers about the process. It’s in place, it has driven individual change in the classroom (at least mine), it’s not going away we need to work with it.” Participant “B” echoed one of the four Ts presented in the review of literature, teacher tension and educators concern about assessments results used by administrators as a way to

evaluate teachers. "The data should be used for something meaningful but NOT teacher achievement because no special treatment is give to students with learning disabilities (like extra time, read aloud, proofread) and many score poorly because they are used to having more time/proof read during a regular class day." Participant "H" echoes another one of the four Ts, expressing concerns of time and compensation, "More time and money given to scoring and analysis of data." Beyond these suggestions and recommendations from the participants, the research supports other reforms for public educators and their pursuit to improve the entire student assessment process and comply with N.C.L.B.

Another, more general recommendation for other public educators and administrators would be to question and reevaluate all of their assessments and their ability to provide truly meaningful data. A similar process used in the methodology of this study to "assess an assessment" or a rubric and its ability to provide truly functional data that goes beyond testing students for testing sake should be a continuous process throughout all of public education. It may be advisable to put all of our assessments to the test to the same degree that we test our students.

Summary

According to the No Child Left Behind Executive Summary of 2004, N.C.L.B. gave public schools "the flexibility to select the strategies that best meet their particular needs for improved teaching that will help them raise student achievement in the core academic subjects." A mandate and challenge was sent to public education. The challenge to public educators everywhere was to use assessments as an instrument to measure students' progress towards reaching standards (evidence of "student achievement") and design and create assessments as an instrument to critique curriculum and improve instruction (evidence of "improved teaching").

We must continually question the assessment tools and rubrics we use to measure students' progress. There must be the same focus and emphasis on the quality of the rubrics used to measure student achievement as there is on the data results obtained from these rubrics. We must give equal importance to the "measure" as we do on the results we obtain from that "measure." If educators and administrators are to get beyond assessing students for the sake of assessment only, then more functional assessment scoring tools or rubrics must be developed. Then, and only then, will public education be able to meet the mandates of No Child Left Behind.

References

- Boudett, K. P., Murnane, R.J., City, E., & Moody, L. (2005). Teaching educators: How to use student assessment data to improve instruction. [Electronic version] *Phi Delta Kappan*, 86, 700-706.
- Executive Summary of the No Child Left Behind Act of 2001. (2004). Retrieved June 21, 2007, from U.S. Department of Education Web site:
<http://www.ed.gov/nclb/overview/intro/execsumm.html>
- Guskey, T.R. (2003). How classroom assessments improve learning. [Electronic version] *Educational Leadership*, 60, 6-11.
- Hall, D. (2004). Preparing for the data deluge. [Electronic version] *Learning and Leading with Technology*. 32, 33-35.
- Karge, B. (1998). Knowing what to teach: Using authentic assessment to improve classroom instruction. [Electronic version]. *Reading & Writing Quarterly*. 14, 319-330.
- Love, N. (2001) The heart of systematic reform. [Electronic version] *ENC Focus*. 8, 14-15.
- Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using student-assessment results to improve instruction: Lessons from a workshop [Electronic version]. *Journal of Education for Students Placed at Risk*. 10, 269-280.
- Sharkey, N. S., & Murnane, R. J. (2003). Learning from student assessment results: A necessary, if difficult, response to NCLB. [Electronic version] *Educational Leadership*, 61(3), 77-82.
- Shellard, E. (2005). How assessment data can improve instruction. [Electronic version] *Principal*. 84, 30-32.

Stiggins, R.J. (2002). Assessment crisis: The absence of assessment for learning. [Electronic version] *Phi Delta Kappan*. 83, 758-765.

Stiggins, R. (2004). New assessment beliefs for a new school mission. [Electronic version] *Phi Delta Kappan*. 86, 22-27.

Stiggins, R., & Chappuis, J. (2006). What a difference a word makes: Assessment "for" learning rather than assessment "of" learning helps students succeed. [Electronic version] *Journal of Staff Development*. 27, 10-14.

Using classroom assessment to improve instruction. (2006). *The Center for Comprehensive School Reform and Improvement: Newsletter*, Retrieved July 6, 2007, from www.centerforscri.org

Appendix A: Institutional Review Board Approval



September 5, 2007

[REDACTED] n.
[REDACTED]

Dear Mr. Gundrum:

On behalf of the UW Oshkosh Institutional Review Board for Protection of Human Participants (IRB), I am pleased to inform you that your application has been approved for the following research: Assessing a Rubric: Evaluating a Rubrics Ability to Provide Functional Data to Critique Curriculum and Improve Instruction.

Your research protocol has been classified as EXEMPT. This means you will not be required to obtain signed consent. However, unless your research involves **only** the collection or study of existing data, documents, or records, you must provide each participant with a summary of your research that contains all of the elements of an Informed Consent document, as described in the IRB application material. Permitting the participant, or parent/legal representative, to make a fully informed decision to participate in a research activity avoids potentially inequitable or coercive conditions of human participation and assures the voluntary nature of participant involvement.

Please note that it is the principal investigator's responsibility to promptly report to the IRB Committee any changes in the research project, whether these changes occur prior to undertaking, or during the research. In addition, if harm or discomfort to anyone becomes apparent during the research, the principal investigator must contact the IRB Committee Chairperson. Harm or discomfort includes, but is not limited to, adverse reactions to psychology experiments, biologics, radioisotopes, labeled drugs, or to medical or other devices used. Please contact me if you have any questions (PH# 920/424-7172 or e-mail: rauscher@uwosh.edu).

Sincerely,

Dr. Frances Rauscher
cm

Dr. Frances Rauscher
IRB Chair

cc: Wayne Swanger
1214

Appendix B: Current Rubric (Rubric A)

Score = 6 Essays within this score range demonstrate effective skill in responding to the task.

The essay shows a clear understanding of the task; takes a position on the issue; may offer a critical context for discussion. It addresses complexity by examining different perspectives on the issue, or by evaluating the implications and/or complications of the issue, or by fully responding to counter arguments to the writer's position. Development of ideas is ample, specific, and logical. Most ideas are fully elaborated. A clear focus on the specific issue in the prompt is maintained. The organization is clear: the organization may be somewhat predictable or it may grow from the writer's purpose. Ideas are logically sequenced. Most transitions reflect the writer's logic and are usually integrated into the essay. The introduction and conclusion are effective, clear, and well developed. The essay shows a good command of language. Sentences are varied and word choice is varied and precise. There are few, if any, errors to distract the reader.

Score = 5 Essays within this score range demonstrate competent skill in responding to the task.

The essay shows a clear understanding of the task; takes a position on the issue; may offer a broad context for discussion. It shows recognition of complexity by partially evaluating the implications and/or complications of the issue, or by responding to counter arguments to the writer's position. Development of ideas is specific and logical. Most ideas are elaborated, with clear movement between general statements and specific reasons, examples, and details. Focus on the specific issue in the prompt is maintained. The organization is clear, although it may be predictable. Ideas are logically sequenced, although simple and obvious transitions may be used. The introduction and conclusion are clear and generally well developed. Language is competent. Sentences are somewhat varied and word choice is sometimes varied and precise. There may be a few errors, but they are rarely distracting.

Score = 4 Essays within this score range demonstrate adequate skill in responding to the task.

The essay shows an understanding of the task; takes a position on the issue; may offer some context for discussion. It may show some recognition of complexity by providing some response to counter arguments to the writer's position. Development of ideas is adequate, with some movement between general statements and specific reasons, examples, and details. Focus on the specific issue in the prompt is maintained throughout most of the essay. The organization is apparent but predictable. Some evidence of logical sequencing of ideas is apparent, although most transitions are simple and obvious. The introduction and conclusion are clear and somewhat developed. Language is adequate, with some sentence variety and appropriate word choice. There may be some distracting errors, but they do not impede understanding.

Score = 3 Essays within this score range demonstrate some developing skill in responding to the task.

The essay shows some understanding of the task; takes a position on the issue but does not offer a context for discussion. It may acknowledge a counter argument to the writer's position, but its development is brief or unclear. Development of ideas is limited and may be repetitious, with little, if any, movement between general statements and specific reasons, examples, and details. Focus on the general topic is maintained, but focus on the specific issue in the prompt may not be maintained. The organization is simple. Ideas are logically grouped within parts of the essay, but there is little or no evidence of logical sequencing of ideas. Transitions, if used, are simple and obvious. An introduction and conclusion are clearly discernible but underdeveloped. Language shows a basic control. Sentences show a little variety and word choice is appropriate. Errors may be distracting and may occasionally impede understanding.

Score = 2 Essays within this score range demonstrate inconsistent or weak skill in responding to the task.

The essay shows a weak understanding of the task; may not take a position on the issue, or the essay may take a position but fail to convey reasons to support that position, or may take a position but fail to maintain a stance. There is little or no recognition of a counter argument to the writer's position. The essay is thinly developed. If examples are given, they are general and may not be clearly relevant. It may include extensive repetition of the writer's ideas or of ideas in the prompt. Focus on the general topic is maintained, but focus on the specific issue in the prompt may not be maintained. There is some indication of an organizational structure, and some logical grouping of ideas within parts of the essay is apparent. Transitions, if used, are simple and obvious, and they may be inappropriate or misleading. An introduction and conclusion are discernible but minimal. Sentence structure and word choice are usually simple. Errors may be frequently distracting and may sometimes impede understanding.

Score = 1 Essays within this score range show little or no skill in responding to the task.

The essay shows little or no understanding of the task. If the essay takes a position, it fails to convey reasons to support that position. It is minimally developed. It may include excessive repetition of the writer's ideas or of ideas in the prompt. Focus on the general topic is usually maintained, but focus on the specific issue in the prompt may not be maintained. There is little or no evidence of an organizational structure or of the logical grouping of ideas. Transitions are rarely used. If present, an introduction and conclusion are minimal. Sentence structure and word choice are simple. Errors may be frequently distracting and may significantly impede understanding.

No Score Blank, Off-Topic, or Illegible

Appendix C: Pilot Rubric (Rubric B)

PILOT STUDY RUBRIC

AASD GRADE 11 COMMUNICATION ARTS WRITING ASSESSMENT

	Introduction/Conclusion	Task	Organization/Transitions	Development/Focus	Language/Vocabulary	Sentence Structure
6 Effective	The introduction and conclusion are clear, and effectively written.	The essay shows a clear understanding of the task; takes a position on the issue; may offer a critical context for discussion. It addresses complexity by examining different perspectives on the issue, or by evaluating the implications and/or complications of the issue, or by fully responding to counter-arguments to the writer's position.	The organization is clear: the organization may be somewhat predictable or it may grow from the writer's purpose. Ideas are logically sequenced. Most transitions reflect the writer's logic and are usually integrated into the essay.	Development of ideas is ample, specific, and logical. Most ideas are fully elaborated. A clear focus on the specific issue in the prompt is maintained.	The essay shows a good command of language.	Sentences are varied and word choice is varied and precise. There are few, if any, errors to distract the reader.
5 Competent	The introduction and conclusion are clear and competently written.	The essay shows a clear understanding of the task; takes a position on the issue; may offer a broad context for discussion. It shows recognition of complexity by partially evaluating the implications and/or complications of the issue, or by responding to counter-arguments to the writer's position.	The organization is clear, although it may be predictable. Ideas are logically sequenced, although simple and obvious transitions may be used.	Development of ideas is specific and logical. Most ideas are elaborated, with clear movement between general statements and specific reasons, examples, and details. Focus on the specific issue in the prompt is maintained.	Language is competent.	Sentences are somewhat varied and word choice is sometimes varied and precise. There may be a few errors, but they are rarely distracting.
4 Adequate	The introduction and conclusion are somewhat clear and adequately written.	The essay shows an understanding of the task; takes a position on the issue; may offer some context for discussion. It may show some recognition of complexity by providing some response to counter-arguments to the writer's position.	The organization is apparent but predictable. Some evidence of logical sequencing of ideas is apparent, although most transitions are simple and obvious.	Development of ideas is adequate, with some movement between general statements and specific reasons, examples, and details. Focus on the specific issue in the prompt is maintained throughout: most of the essay.	Language is adequate, with some sentence variety and appropriate word choice.	There may be some distracting errors, but they do not impede understanding.
3 Developing	An introduction and conclusion are discernible but underdeveloped.	The essay shows some understanding of the task; takes a position on the issue but does not offer a context for discussion. It may acknowledge a counter-argument to the writer's position, but its development is brief or unclear.	The organization is simple. Ideas are logically grouped within parts of the essay, but there is little or no evidence of logical sequencing of ideas. Transitions, if used, are simple and obvious.	Development of ideas is limited and may be repetitious, with little, if any, movement between general statements and specific reasons, examples, and details. Focus on the general topic is maintained, but focus on the specific issue in the prompt may not be maintained.	Language shows a basic control.	Sentences show a little variety and word choice is appropriate. Errors may be distracting and may occasionally impede understanding.
2 Inconsistent	An introduction and conclusion are discernible but are reflect inconsistent skill.	The essay shows a weak understanding of the task; may not take a position on the issue, or the essay may take a position but fail to convey reasons to support that position, or may take a position but fail to maintain a stance. There is little or no recognition of a counter-argument to the writer's position.	There is some indication of an organizational structure, and some logical grouping of ideas within parts of the essay is apparent. Transitions, if used, are simple and obvious, and they may be inappropriate or misleading.	The essay is thinly developed. If examples are given, they are general and may not be clearly relevant. It may include extensive repetition of the writer's ideas or of ideas in the prompt. Focus on the general topic is maintained, but focus on the specific issue in the prompt may not be maintained.		Sentence structure and word choice are usually simple. Errors may be frequently distracting and may sometimes impede understanding.
1 Minimal	If present, an introduction and conclusion are minimal.	The essay shows little or no understanding of the task. If the essay takes a position, it fails to convey reasons to support that position.	There is little or no evidence of an organizational structure or of the logical grouping of ideas. Transitions are rarely used.	It is minimally developed. It may include excessive repetition of the writer's ideas or of ideas in the prompt. Focus on the general topic is usually maintained, but focus on the specific issue in the prompt may not be maintained.		Sentence structure and word choice are simple. Errors may be frequently distracting and may significantly impede understanding.

Appendix D: Survey

Years of teaching experience in communication arts? _____

Current grade level(s) you are teaching?

General

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1					
2					
3					
4					

Additional comments: (use the back of sheet if necessary)

Current Rubric (Rubric A)

1	How long did it take you to score the writing sample with the current rubric _____ Minutes					
2	The current rubric is "user-friendly" in terms of an assessment tool to grade student essays.					
3	Teacher training is a necessity in order to use the current rubric .					
4	Scoring an essay with the current rubric is time efficient.					
5	The current rubric allows the reader to accurately score a student essay.					
6	The current rubric accurately measures the student's writing skills.					
7	The current rubric measures the essential skills needed to write a persuasive essay.					
8	The current rubric effectively measures communication arts curriculum standards.					
9	The current rubric is designed to easily grade the essay with a global score.					
10	The current rubric is designed to increase inter-reader/inter-scorer reliability.					
11	The current rubric is a reliable measure of a student's progress towards curriculum standards					
12	Data results from the current rubric are helpful in identifying student's strengths and weaknesses.					
13	Data results from the current rubric are helpful to you in evaluating your instruction methods.					
14	Data results from the current rubric are easily linked back to comm. arts curriculum standards.					
15	Data results from the current rubric help facilitate curriculum review and evaluation.					

Additional comments: (use the back of sheet if necessary)

		Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Pilot Rubric (Rubric B)						
1	How long did it take you to score the writing sample with the pilot rubric Minutes					
2	The pilot rubric is "user-friendly" in terms of an assessment tool to grade student essays.					
3	Teacher training is a necessity in order to use the pilot rubric .					
4	Scoring an essay with the pilot rubric is time efficient.					
5	The pilot rubric allows the reader to accurately score a student essay.					
6	The pilot rubric accurately measures the student's writing skills.					
7	The pilot rubric measures the essential skills needed to write a persuasive essay.					
8	The pilot rubric effectively measures communication arts curriculum standards.					
9	The pilot rubric is designed to easily grade the essay with a global score.					
10	The pilot rubric is designed to increase interrater/inter-scorer reliability.					
11	The pilot rubric is a reliable measure of a student's progress towards curriculum standards.					
12	Data results from the pilot rubric are helpful in identifying student's strengths and weaknesses.					
13	Data results from the pilot rubric are helpful to you in evaluating your instruction methods.					
14	Data results from the pilot rubric are easily linked back to comm. arts curriculum standards.					
15	Data results from the pilot rubric help facilitate curriculum review and evaluation.					
Additional comments: (use the back of sheet if necessary)						

		Current Rubric (Rubric A)	Pilot Rubric (Rubric B)	NA
1	Which rubric provides more functional data to be used by teachers to evaluate students' writing skills?			
2	Which rubric provides more functional data to be used by teachers to evaluate and modify instructions and lesson plans?			
3	Which rubric provides more functional data to be used by teachers to evaluate and modify curriculum?			
Open Ended Response Questions: (use the back of sheet if necessary)				
1	What are the current strengths in the District-Wide Standards-Based writing assessment process?			
2	What are the current concerns with the District-Wide StandardsBased writing assessment process?			
3	What should be done to address the concerns and challenges of the current District-Wide Standards-Based writing assessments?			



Appendix E: Writing Assessment Prompt

**11th Grade
(December)**

ID # _____
School Code _____
Teacher Code _____

Writing Assessment

In timed or tested writing, a writer responds to a prompt within a set period of time. The prompt gives the writer the topic.

The Task:

Using the writing process, complete a well-developed response to the writing prompt below. You may use the back of this page to plan your writing (This pre-writing/planning space might include notes, lists, webs, outlines or anything else that might help you plan your writing). Upon completion of your planning, use your time to write a “final copy” of your response.

Be sure to carefully analyze the prompt so that you fully understand the writing form, the audience and the subject matter of this writing task.

The Prompt:

Should high school students hold jobs during the school year?

Write an essay that takes a position on whether or not high school students should hold jobs during the school year.

Your essay should:

- take a clear position on the issue
- express your reasons for having that specific opinion
- include supporting detail and strong reasoning
- acknowledge the counter-argument

Appendix F: Student Writing Sample

Kids in high school have many different activities going on. Holding a part time job is one of them. Some may think it puts too much pressure on kids, but it is important that these youth hold them during the school year because of the many lessons they learn from the experience.

Holding a job during the school year teaches organization skills. It helps to show kids where their priorities should be.

Planning around work schedules helps kids to plan other things and keep track of when school projects and assignments are due. Organization is important once kids go into the real world and it needs to be learned early so that it can be applied later on in life.

Another life lesson that holding a job during the school year teaches, is responsibility. Showing up for work on time, and still taking the initiative to finish homework will later on help kids to take things seriously. Again, priorities are set and kids must learn to follow through on them.

These responsibilities will also help kids make better decisions about what is really important.

Earning money during the school year is also important. The summer doesn't last long enough in order ^{to} save up a sufficient amount of money to help you get a start in life. Money needs to be earned throughout the year. This life skill will also teach kids that they need to get out and work in order to get money, despite whatever else is going on. This determination will also help them feel good about themselves, because they are pulling their own weight.

Kids should hold jobs during the school year. Jobs teach organization, responsibility, and how to earn money. These experiences will help kids later in life and thus give them a better chance in the real world.

Appendix G: Participant Directions

Please find the following paper clipped with this coverpage:

1. Summary of research project (one page)
2. Participant consent form from UW (one page)
3. The standardized prompt and directions to the district junior writing assessment
4. An anonymous student response to the district junior writing assessment (two pages)
5. The current rubric used by the district (RUBRIC "A") (one page)
6. The pilot rubric (RUBRIC "B") (one page)
7. Survey (three pages)

Directions:

1. Read the "Summary of Research Project" (This is a UWO requirement)
If you have any questions or concerns, please contact me
2. Read the "UWO Short Form Consent Document" (This is a UWO requirement) please date and sign
3. Read the standardized prompt and directions to the district junior writing assessment
(If you are already familiar with the junior level writing assessment, you can skip this)
4. Familiarize yourself with both rubrics, current rubric used for the junior writing assessment (RUBRIC A) and the pilot rubric (RUBRIC B)
5. Read the sample student response to the writing prompt
6. Score it using the current rubric used by the district (RUBRIC "A")
Give the response a final global score of between a "No Score" through a Six
Keep track of the time it takes to score it
7. Score it using the pilot rubric (RUBRIC "B")
Score each subtopic with a 1 through 6
Add all six "subscores" and divide by 6. This is the Global Score
Record Global Score on top of rubric
Keep track of the time it takes to score it
8. Complete the three page survey as accurately and honestly as possible
(There are questions specific to both rubrics and general questions about the process)
9. Enclose all materials in envelope and return to me at North

Notes: These directions are not intended to insult anyone's intelligence. I am striving to make the process as standardized and "antiseptic" as possible.

Although this may seem like a lengthy process, it should take approximately 30 minutes.

Once again, your participation is sincerely appreciated. THANK YOU!

If you have any questions whatsoever, please contact me at North.



Appendix H: Summary of Research Project Sent to Participants

SUMMARY OF RESEARCH PROJECT

The purpose of this summary is to describe the research study, "Assessing a Rubric: Evaluating an Assessment Rubric's Ability to Provide Functional Data to Critique Curriculum and Improve Instruction," and to explain the study's scope, aims, and purpose.

The reasonably expected benefits of the project are the collection of information and data to evaluate a current assessment instrument used in the district for the purpose of evaluating curriculum and improving instruction. The project is expected to be beneficial to educators and administrators by the insights it may provide about current practices in assessing student writing through standardized rubrics.

The procedures that will be used involve administering an anonymous student essay sample with two rubrics, the current rubric used within the district and a new pilot rubric. Participants will be asked to assess the sample using both rubrics and then respond to a series of survey questions related to the rubrics and their assessment experience. Upon completion, participants will be asked to return the writing sample, the rubrics, and the survey to the researcher. The researcher will analyze the results of the rubrics and surveys and document the data in a thorough report with informed recommendations regarding the use of assessments to evaluate curriculum and improve instruction.

It is reasonably foreseeable that you will experience minimal discomfort. The risk of harm that could result from your participation in the project is negligible. The alternative procedures that could have been used in this study include interviewing or observing participants. The expected duration of your participation is approximately thirty minutes to read the sample student essay, score it with both rubrics, and complete the corresponding survey.

Your participation in the study is completely voluntary--you do not have to participate and you can stop at any time. If you refuse to participate now, or withdraw from the study later, it will have no effect on any regular services or benefits available to you at the University of Wisconsin Oshkosh.

Any personal information used in this study will be treated confidentially. Information which identifies you as an individual will not be released, without your consent, to anyone for purposes which are not directly related to this research study.

If you have any question about this study, or your rights, you may call or write:



You will be given a copy of this statement, which serves to acknowledge the fact that you have been informed about the project and that you have voluntarily agreed to participate.

Appendix I: Short Form Consent Document

UNIVERSITY OF WISCONSIN OSHKOSH
SHORT FORM CONSENT DOCUMENT

I, the undersigned, have had presented to me a description of the research study, "Assessing a Rubric: Evaluating an Assessment Rubric's Ability to Provide Functional Data to Critique Curriculum and Improve Instruction," which explained the study's scope, aims, and purpose; the expected duration of my participation; the procedures which will be used (including experimental procedures); the reasonably expected benefits to myself or others; the reasonably foreseeable discomfort and risk of harm which could result from my participation in this project; and other information required under procedures for informed consent.

I have been provided with a copy of the summary that was presented to me.

I agree to participate in this study.

Date

Signature of Research Participant

(Print) Name of Research Participant

I, the undersigned, have witnessed the presentation of the description of the research study to the above named research participant (or participant's legally authorized representative).

Signature of Witness

(Print) Name of Witness

Appendix J: Participant Open Ended Comments/Responses to General Survey Questions

Participant E

“Teachers need time to look at student scores and use the information to aid classroom instruction.”

Participant G

“Assessment is a general term. Certainly I place more emphasis on my classroom assessment than I do on a district or state assessments. All, however, have their place in determining curriculum and instructional strategies.”

Participant H

“Our district will not pay or give time to #4.”

Appendix K: Participant Open Ended Comments/Responses to Rubric A

Participant A

“I have never seen any data results.”

Participant D

“Layout is difficult to read. Score is too generalized and doesn’t provide much opportunity for specific comments on strengths/weaknesses.”

Participant E

“Current rubric took time to read and decipher the differences between the scores. A score of “4” or “3” (I struggled a bit based on the rubric) doesn’t give the student much feedback.”

Participant F

“The current rubric/assessment is only one snap shot of the curriculum. The issue is not as much the rubric as it is the fact that we are only assessing a minute part of the curriculum.”

Participant G

“This rubric links too many skills together. It’s definitely user-friendly (the user is the teacher), but it doesn’t breakdown skill assessments for the student. Plus, I imagine they would simply look at their scores and toss the sheet without reading the entire description, which may be only partially true anyway.”

Participant H

“The current rubric is simply that, a rubric. You use it. I learn more about what to work on in class by simply reading what my students write.”

Participant I

“To me, it’s more user friendly because I’ve been on the 11th grade scoring team for three years. English teachers would not need much training. I doubt teachers from other disciplines could score reliably in general.”

Participant J

“I don’t feel these assessments are usually an accurate picture of a student’s writing ability.”

Appendix L: Participant Open Ended Comments/Responses to Rubric B

Participant A

“I thought it was interesting that the pilot rubric was not divided up by the six traits. I think “B” can provide better data eventually, but “A” is easier for me to use simply because I have scored with it so many times.”

Regarding questions 12-15

“Sorry, but I would have to see how all the data is tabulated, displayed, and how much time I am given as an instructor to analyze it.”

Participant C

“Although Rubric A takes less time to use, the formatting of Rubric B is easier to explain to students. It also relates more clearly to the standards/instruction methods.”

Participant E

“Could layout criteria a bit better in terms of organization, but overall, very “user friendly.” I liked this method much better.”

Participant F

“I prefer this rubric, though it still needs work. I don’t think student failure to recognize a counter argument shows a weak understanding of persuasion, though it is certainly not adequate. The scores for “Task” still need work, but I believe this is a better rubric.”

Participant G

“This rubric is better than the other because it separates writing elements. Students get the whole picture (global score) and see their strengths and weaknesses.”

Participant H

“The graphic design or presentation of Rubric B was more enjoyable, “friendly,” to me. The last ones, 12,13,14,15, I would have to wait and see.”

Participant I

“It took me longer than the other only because I was less familiar, despite previewing
Same comments as for Rubric B English teachers would generally be familiar with it. The concepts and terminology. 12 & 13 = “helpful” is the key word. I see little difference other than organization. That does make it more user-friendly. Still have the problem like on the Task column. Some descriptors were 4, some 3. Then it becomes and individuals judgment call on which is more important. But, of course, of the scoring process is (informed) individuals judgment calls.”

Participant J

“I think assessments are a good idea, but I think the process, in general is flawed.”

Regarding questions 11-15

“For group/district scoring sessions. For my own purposes, this rubric would be useful (all disagrees would move to agrees)”