May 2016

A New Approach: Closing the Writing Gap By Using Reliable Assessment to Guide and Evaluate Cross-Curricular Argumentative Writing

Jody Bauer
University of Wisconsin-Milwaukee

A NEW APPROACH: CLOSING THE WRITING GAP BY USING RELIABLE ASSESSMENT TO GUIDE AND EVALUATE CROSS-CURRICULAR ARGUMENTATIVE WRITING

by

Jody A. Bauer

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Curriculum and Instruction

at

The University of Wisconsin, Milwaukee

May 2016

ABSTRACT

A NEW APPROACH: CLOSING THE WRITING GAP BY USING RELIABLE ASSESSMENT TO GUIDE AND EVALUATE CROSS-CURRICULAR ARGUMENTATIVE WRITING

by

Jody A. Bauer

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Donna Pasternak

Educational standards emphasize cross-curricular literacy and complex skills at the secondary level. These standards align to the heightened priority of argumentative writing across disciplines. With increasingly complex and shifting writing expectations, assessment practices need to be implemented to mirror these expectations. Two primary components to effective assessment are the ability to inform teaching and improve student learning. This study is designed to test the reliability of a process-oriented rubric as a tool to evaluate argumentative writing in the cross-curricular context, and as a tool to guide classroom practices based on student readiness.

The purpose of this study is to observe the reliability of a process-oriented argumentative writing rubric as an evaluation tool, and to measure the reliability of a process-oriented argumentative writing rubric that informs instruction. Teachers and their students from two inclusive, urban education classes participated in the study. One class represented a control group which received the first and last treatment conditions without experimental interventions. The second class represented an experimental group that engaged in a tiered intervention program guided by the assessment tool. Additionally, four reported educators participated in

analyzing the tool's reliability as a cross-curricular rubric for argumentative writing. The educators also completed surveys about their experiences using the rubric created for the study and traditional assessment practices.

Results suggest that tiered intervention guided by effective assessment makes significant gains in literacy achievement over a short period of time. Results also suggest that the rubric created for the study warranted more reliable ratings across cross-curricular educators than traditional assessment. Educators also reported a preference for the process-oriented approach to assessment over traditional methods.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank everybody who helped in the process of conducting the study, creating the report, and providing diligent guidance along this arduous journey.

I want to start by expressing my extreme gratitude to Jennifer Kohut. Without you, this research would not have been possible. Your collaboration and contribution to this work are immeasurable. You contributed your time as a fellow creator, investigator, researcher, and statistician. I am above all honored to call you my friend. Thank you for all that you have done, and I look forward to the heights of our future collaboration.

I would like to thank Jason Orozco, Amanda Vogel, Karen Lustig, and Jennifer Sagrillo for participating in the study and providing helpful and sincere feedback. I appreciate your time and input for this research.

I would also like to extend thanks to Donna Pasternak, Tania Habeck, and Candance Doerr-Stevens for the knowledge of research and advice given throughout this process. I learned much and benefitted greatly from your guidance and expertise.

Special thanks goes to Robin Harris and Ray Scolavino; I express extreme gratitude to the people who went out of their way to help and advise me during the creation of this study. I appreciate the time and dedication you provided to help me form the most credible and knowledgeable research possible. I truly appreciate your efforts.

CHAPTER 1
INTRODUCTION

As a secondary English teacher in an urban, public school, many of the students who enter my classroom are behind grade level expectations in their reading and writing skills. This observation is not limited to my classroom or school. According to the National Center for Education Statistics (2012), 74% of students entering high school score below literacy proficiency levels showing substantial deficits in their reading and writing ability when compared to national grade level expectations. The National Assessment of Educational Progress (2011) finds similar literacy statistics stating that only 40% of 12th grade students demonstrate proficiency in argumentative writing assessments. These deficits prove to be greater for minority students and students of lower socio-economic statues who score up to two and a half times lower than their suburban and private school peers, which is referred to as the achievement gap (Act, Inc., 2006).

Many high schools teachers have the responsibility to facilitate significant gains in student ability in order for students to read and write proficiently at grade level. Students may carry limited skills into the secondary setting due to a wide range of circumstances including family resources, socio-economic status, parental literacy rates, attitudes toward reading, educational resources, and myriad other possible factors (Chiu & Chow, 2015). Graham (2007) points out that writing competency is necessary for many post-secondary school and work opportunities. Given the importance of writing skills at every grade level and in future career opportunities, many educators and researchers have sought methods of increasing student literacy skills and closing the achievement gap.

When seeking to study teaching methodology, one must first look to how the results of instruction are assessed. Writing skills have been assessed using writing rubrics for years (Rudner and Schafer, 2002). De Leeuw (2016) finds that assessing writing performance using rubrics to provide feedback for multiple categories or standards is still a commonly used method of assessment in the United States. Writing rubrics have, on average, four to six categories of writing which are assessed all at once.

The factors assessed in student writing are determined by standards set by each state. Although the National Governors Association Center for Best Practices & Council of Chief State School Officers (2010) concludes that many states have adopted the Common Core State Standards (CCSS) over recent years, these new writing standard adaptations still correlate to Bloom's taxonomy and many of the former state standards. As a result, many writing rubrics used before the CCSS adaptations have been carried over and are still utilized. Examples and further analysis of these standards and procedures will be covered in-depth in the literature review.

Within each writing standard, the various levels at which students are assessed commonly correlate to a metric called "complexity," as determined by Bloom's Taxonomy (1976). Anderson and Krathwohl (2001) revised this taxonomy to meet more modern, out-come based educational objectives. Although assignments, projects, and evidence of learning are planned at varying levels, it may not be common practice to assess student results based on the level of complexity achieved in student products. Many assessments developed by teachers overly emphasize the memorization of procedures and facts (Porter et al., 1993). In regards to Bloom's Taxonomy, many teacher-made assessment highly emphasize memorization and recall skills, which represent lower levels of complexity (Huott, 2011). The act of planning and assigning

work at a high level of complexity does not guarantee that a student will be able to complete the work at the depth and quality at which the assignment intended. For example, asking a student to list factors is less complex than asking a student to judge factors, but the act of requesting that students judge or evaluate does not mean that students will be able to conduct an effective evaluation of factors. When assigning work with a high level of complexity, it is necessary to have a reliable method with which to measure the level of complexity actually achieved in student work.

Effective assessment of student work is just as significant as purposefully teaching skills to accommodate complex expectations set for the English Language Arts classroom. The National Research Council (2000) notes that effectively designed learning environments must be centered on assessment. There is an urgent call for authentic assessment that challenges students on an individual level, calls for higher-level thinking skills, and provides clear feedback for skill growth and understanding (Dolgin, Kelly, & Zelkha, 2010). Teachers and students alike may benefit from an assessment strategy that can effectively measure new, higher-level, and complex expectations while allowing both parties a chance to communicate and reflect on skills and deficits present in student work. Gardner (2008) suggests that designing such assessments is one of the most important jobs of an educator. Assessments should reflect the learning goals that define various environments; if the goal is to enhance higher-level skills such as analysis and evaluation, it is not sufficient to provide assessments that evaluate only memory, recall, comprehension, and other basic skills (National Research Council, 2000). If students are expected to demonstrate literacy proficiency with complex outcomes, then a tool to measure the complexity demonstrated would be helpful. The current study intends to create and evaluate a

tool which would measure complexity of student outcomes in the area of argumentative writing and be useful in the planning and instructional processes as well as the assessment process.

Experts agree that teaching a skill as complex as writing is difficult for teachers across the curriculum (Graham, 2007). The National Research Council (2000) describes one major role of assessment, when administered within the context of the classroom, as a source of feedback to improve both teaching and learning. An effective assessment tool can not only provide an outlet for teacher-student feedback, but can serve as a tool to improve and tier instruction. Appropriately designed assessments can help teachers rethink and organize their teaching practices (National Research Council, 2000). Teachers across the curriculum may benefit from an assessment strategy that will not only evaluate the complex skill demonstrated in student writing, but will also help to guide and reexamine teaching practices. As meta-analysis show, evaluating and monitoring student progress is not enough to create meaningful results in improving writing; in order to create meaningful change in writing ability, progress assessment must be utilized to inform students, planning, and teaching (Graham, Herbert, & Harris, 2015). Once instruction and skill strategies are guided by appropriately defined complex expectations and adjusted according to student performance, the probability of students achieving complexity in their work will be promoted.

In order meet the heightened and complex expectations of the Wisconsin state adopted Common Core State Standards, assessment and instruction should appropriately measure and tier complex literacy skills. The purpose of the current study is to:

1. Develop a tool, the Bauer & Kohut Argumentative Writing Rubric, based on the complexity levels detailed in Bloom's Taxonomy and on the expectations of the CCSS.

2. Observe the reliability of a process-oriented argumentative writing rubric created by the researchers for the current study as an evaluation tool.
3. Measure the consistency of the process-oriented argumentative writing rubric evaluation tool created for the current study in informing instruction.

Significance of Study

Both educators and students may benefit from instructional practices guided by reliable argumentative writing assessment at the secondary level. Reliability refers to the consistency of the scores produced by a measurement tool (Ross, 2006; Berkowitz, Wolkowitz, Fitch, & Kopriva, 2000; Huot, 1996). Complex writing expectations should be mirrored by reliable assessment practices that are feasible to implement with consistency across the curriculum.

The Common Core State Standards (CCSS) is an accepted form of standardized grade-level expectations implemented throughout Wisconsin. The CCSS promotes the concept of cross-curricular literacy. Literacy is defined as having the ability to read, write, listen, and think critically, and the ability to demonstrate these skills in different ways and for different purposes. Demonstrating literacy is a key foundational factor to building skills necessary for college and career readiness (Wisconsin Department of Public Instruction, 2011). According to the Department of Public Instruction (2011), elevating complex literacy ability is crucial for student secondary and post-secondary success. The department also notes that as students' progress through their academic experience, not only does their reading material become more challenging, but the need for independence in applying complex skills to difficult text become more pronounced. As students are expected to apply complex skills while reading and writing, teachers are expected to teach and assess performance of the desired skills.

Many educators hold that the primary purpose of classroom assessment is to inform teaching and to improve student learning (Fuchs & Fuchs, 2006; Mitchell & Neill, 1992). Educators and students alike require access to assessment tools that meet the demands of elevated complex literacy expectations and that reliably assess work and guide instruction according to literacy demands across curricular context. Educators use tools to guide and assess instruction and, as stated, students use these tools in order to make meaningful gains in literacy skills (Graham, Herbert, & Harris, 2015). Although current writing instruction may emphasize complex literary skills, there is not clear evidence that assessment of student writing in the United States focuses on this same criteria (Huot, 2002b). The significance of this study is to:

1. Provide educators with a process-oriented rubric that reliably evaluates argumentative writing in a cross-curricular context.
2. Provide educators with a process-oriented argumentative writing rubric that consistently guides complex literacy instruction.

Research Questions

For the purpose of the study, the following questions were addressed:

1. Is the Bauer & Kohut Argumentative Writing Rubric a reliable tool to evaluate argumentative writing in a cross-curricular context?
2. Is the Bauer & Kohut Argumentative Writing Rubric a consistent tool to guide instruction towards complex literacy expectations?

Definitions of Terms

The following terms will be used:

1. Assessment: The demonstration of student understanding and ability (Tileston, 2000). Assessment can refer to the process of evaluating this understanding and ability, or the assignment and rubric used to measure their achievement. The terms evaluation and assessment are used throughout this study interchangeably.
2. Traditional Assessment: Assessment that prioritizes correctness, grammar, and programmatic values in writing (Barrit, Stock, & Clark, 1986).
3. Rubric: An assessment tool that clearly dictates the skills and expectations of an assignment at multiple levels of achievement. (Brookhart, 1999).
4. Process Oriented Assessment: Assessments that seeks to gain insight into underlying processes and strategies demonstrated through student work (McTighe & Ferrara, 1998).
5. Strategy Instruction: Instruction that involves setting specific, reachable product goals and aligning processes or strategies in order to accomplish these goals (Graham, MacArthur, & Fitzgerald, 2013).
6. Formative Assessment: A framework that informs instruction through the assessment of student work (Risko & Walker-Dalhouse, 2010).
7. Tiered Instruction: The practice of modifying instruction to the purpose of tiering instruction to close the gap between learning goals and student readiness (Allen & Turville, 2010).
8. Literacy: The ability to read, write, listen, and think critically in different ways and for different purposes (Langer & Flihan, 2002).

9. Complexity: A synthesis of thought processes and strategies used to complete a task. (Davis & Sumara, 2008).
10. Difficulty: The amount of effort expended within a level of complexity (Tileston, 2000).
11. Intervention: Action to assist or improve a skill pertaining to an educational task (Methe & Riley-Tillman, 2008).
12. CCSS: Common Core State Standards. The CCSS is an accepted set of academic standards in Wisconsin that regulate what skills students should be able to demonstrate at the end of each grade level. For English Language Arts classrooms, the CCSS determine skill-based expectations for Reading Literature, Reading Informational Text, Writing, Language, and Speaking and Listening standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

Summary

The reading and writing deficits of a significant percentage of secondary students in the United States may affect post-secondary educational and career opportunities. The researchers outline a demonstrated benefit of assessment strategies that incorporate the process-oriented and complex expectations of literacy in the field of education. Cross-curricular educators may benefit from a reliable assessment tool applicable across subject areas that evaluates complex grade level literacy expectations represented by student work. Cross-curricular educators may also benefit from a consistent assessment tool that can inform and guide the organization and implementation of targeted instructional literacy strategies based on student outcomes.

CHAPTER 2
REVIEW OF LITERATURE

The Need for Modified Assessment

Swain and Mahieu (2012) argue that to design meaningful instruction, teachers must consider ways to meet student needs, reflect upon student results, and used this information as a means to improve future instruction. Just as instruction should be meaningful, so must be assessment. Meaningful assessment must be able to meet the needs of nationally standardized literacy expectations, the needs of instructional implementation, and the needs of the students. Assessment should be able to use used as an evaluation tool and a guide to inform instruction. If assessment is meaningful, then classroom instruction should reflect this value.

For almost fifty years, educators have grappled with strategies designed to support literacy goals in the cross-curricular context (Alvermann & Moore, 1991; Moore & Readence, 2001). Due to the flat or declining reading scores on national tests (e.g., Donahue, Daane, & Grigg, 2003), reforms have been implemented to increase literacy achievement in secondary classrooms. Many of these reforms have encouraged teachers to use literacy practices and instruction within their discipline (Moje, 2008). As literacy instruction develops across disciplines, it is necessary to reconsider its means of being meaningfully assessed. It is equally necessary to consider how meaningful assessment not only effects evaluation practices, but the entire act of literacy instruction (Huot, 2002b).

A complex approach to meaningful literacy assessment is not easily accommodated by traditional assessment. There is a gap between traditional assessment theories and the need for meaningful assessment that meets the needs of current standards and expectations, and that informs teaching practices (Huot & O'Neill, 2009; Huot, 2002b).

Traditional Assessment

English instruction has traditionally been strongly associated with grammar instruction (Wilson, 2011). Grammar instruction was one of the first widely accepted traits of the English curriculum (Applebee, 1974). Traditional English instruction emphasized the correction of errors in writing, the dissection of sentences, and the functions of parts of speech and words (Wilson, 2011). Traditional writing assessment reflects the priorities of traditional English instruction. Barrit, Stock, and Clark (1986) note that traditional writing assessment practices are usually based within specific disciplinary contexts, so that teachers evaluating student work are forced to make practical, pedagogical, programmatic, and interpretive judgments of the work without necessarily having to define abstract qualities, such as writing values. Traditional writing assessment has been characterized by stressing these practical, pedagogical, and programmatic components (for example, with the traditional emphasis of grammar or correction).

Traditional approaches to assessment and grading have often overemphasized the importance of correctness, while at the same time minimizing the importance of higher-level thought. Writing instruction involving the explicit and systematic teaching of the parts of speech and structure of sentences has a negative effect on students, and is unlikely to help improve the quality of students' writing (Graham, MacArthur, & Fitzgerald, 2013). This is due to the idea that educators can often become distracted by focusing on surface-level errors instead of addressing the context and ideas of a piece of writing. Emphasizing corrective components has created a conflicted nature of writing assessment, and an overall lack of consensus about the role of assessment in writing and instruction (Huot, 2002b).

Traditional Assessment in Urban Education

Traditional assessment has become increasingly displaced in urban education. While urban schools are being populated by minority and immigrant children, the teachers in these classrooms are characterized by being predominantly middle-class European Americans, who are often educated in predominantly European-American institutions of higher education (Bireda & Chait, 2011; Ball, 1997; Gomez, 1993). These teachers often times have had little direct experience with cultural, ethnic, or linguistic diversity (Nieto, 1999). Concerns about the lack of diversity represented in urban teacher work-force are not new. For decades, the United States has sought to recruit minority teachers, but with little progress (Goldhaber, Theobald, & Tien, 2015). Cushner, McClelland, and Safford (2009) note that there is a considerable gap between the make-up of the urban student population and 21st century teacher workforce- data indicates that approximately 88%-90% of teachers in American are of European-American ancestry and from the middle class. They also note that changing demographic projections predict that by the year 2040, minority children will comprise more than half the children in classrooms.

A lack of diversity in the teacher work-force is significant because many educators don't recognize how their European-American upbringing and privileges may perpetuate social inequity in the classroom (Glimps & Ford, 2010). DeSherbinin (2004) recognizes a divide between European-American teachers and students who come from ethnic or socioeconomic backgrounds unfamiliar to middle-class white academia. A study by Ball (1991) revealed that European-American teachers scored assignments written in oral-based patterns, a more common pattern among minority students, lower than those written in mainstream academic literacy-based patterns. Traditional English instruction and forms of assessment tend to emphasize correctness, and teachers in Ball's study became more concerned with the technicalities of the

writing in their students' work. Similarly, Soares (2007) observed that in the mean scores on the SAT essay, minority students tended to score much lower than their European-American counterparts. These findings have been repeated by others (Gerald & Haycock, 2006), who found that the traditional assessment represented by the SAT essay scoring strongly correlated to socioeconomic and racial factors.

Huot (1996) stresses the idea that assessment practices need to be based upon the notion that educators are trying to assess the writer's ability to communicate in their writing. Students from low socioeconomic or diverse backgrounds receive mixed messages from teachers because teachers are often unaware of, unable, or unwilling to acknowledge the existence of conflicting cultures (that can emerge in writing)- this situation can contribute to students' failure in classrooms that require them to demonstrate their knowledge through writing (Ball, 1997). Students carry with them many negative, critical, correctness-centered ideas of traditional assessment and their writing (Huot, 2002b).

A lot more goes into the process of writing than the systematic skill of writing itself, such as higher-level thinking and processes. Huot (2002b) notes that what teachers assess ultimately determines what is valued. If this is true, then assessment needs to be modified to take other aspects of writing into consideration. The substance of writing must surpass the emphasis of its conventions. The focus of writing must be to determine the quality of writing itself (Swain & Mahieu, 2012).

The Rubric as a Tool for Assessment

Rubrics are tools to evaluate and provide guidance for student writing (Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz, 2009). Andrade (2005) notes that rubrics

strengthen learning by providing both students and teachers with a clear understanding of the criteria for writing and scoring.

The educational community began using rubrics in the 1970's in order to standardize ratings on writing exams, resulting in a holistic score based on a multidimensional set of criteria (Dirlam, 1980; Dirlam & Byrne, 1978). Rudner and Schafer (2002) note that rubrics are still used as a method to score writing. These types of scoring criteria guides provide the expectations of the scoring scale (Hunt & O'Neill, 2009). Brookhart (1999) further defines rubrics as a method in which to analyze the process or procedures used in the completion of the task.

Given that the two main functions of assessment are to inform teaching and to improve student learning (Fuchs & Fuchs, 2006; Mitchell & Neill, 1992), the use of rubrics as an assessment tool has multiple advantages in the classroom. Arter and McTigue (2001) note that rubrics are a beneficial instructional guide since teachers are better able to focus on specified learning targets as they choose instructional approaches that enable students to achieve desired learning outcomes. Rubrics also make the assessment process more accurate. Since the criteria of the learning targets are already established, teachers are more likely to be consistent in their evaluations of student work, and less likely to evaluate with individual skill bias (Wolf & Stevens, 2007). Peat (2006) suggests that rubrics, because of their distinct criteria, are able to increase objectivity in the evaluation of writing. This reinforces reliability of writing assessment across instructors and courses using a common rubric.

Specific evaluation criteria provide the qualities and traits needed for success (Graham, MacArthur, & Fitzgerald, 2013). Creating an effective rubric begins with the purpose of the assessment in mind so that the level to which the desired result is achieved can be aptly determined (Rudner & Schafer, 2002). The qualities identified will form the top level of the

grading scale (Brookhart, 1999). The bottom level of possible performance will then be determined, which will make it possible to find a mid-point between the top and bottom performance possibilities, making mid-level performance standards clear and appropriate (Rudner & Schafer, 2002). A well-developed rubric guides graders on focusing on specific criteria so that subjective opinions are minimized (Newell, Dahm, & Newell, 2002).

A common assumption among teachers is that providing a rubric will give students the tools with which to evaluate and revise their own writing in order to meet the standards. Students will not be able to self-assess to meet the expectations of a rubric if they do not understand its criteria. Essay scoring is notoriously subjective and varies even among well trained assessors- Providing clear expectations in the rubric makes the process of writing evaluation become more objective than grading without defined standards (Rudner & Schafer, 2002). This increase in the objectivity of scoring makes rubrics useful across all content areas (Schrock, 2000).

Holistic Grading

Rubrics can be characterized as either holistic or analytic (Moskal, 2000). Holistic rubrics provide a single score based on overall performance, while analytic rubrics give multiple scores over several dimensions (Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz, 2009). Holistic grading is one of the most commonly used forms of writing assessment and has been used in relation with many traditional forms of writing (Huot & O'Neill, 2009; Huot, 1996). To grade holistically is to see things as units or as wholes; when a teacher is grading holistically, they are evaluating the overall proficiency level of the work. Much like traditional assessment is considered "practical" and "programmatic," holistic grading can be considered a "scientific" and "objective" type of writing evaluation (O'Neill, 2003). Huot and O'Neill (2009) stress that

holistic scoring is only a means of ranking papers according to the criteria established in the rubric. Given that traditional assessment tends to overemphasize surface-level and systematic functions of writing, it is necessary to reconsider the type of criteria that influences our holistic assessment in order to best meet the needs of writing across contexts and content areas.

Cross Curricular Application

Literacy instruction is expected across content areas at the secondary level. Literacy is often characterized by generic specifications that do not take discipline-specific frameworks and characteristics into consideration. Because of this, literacy instruction has failed to be implemented in many cross-curricular secondary classrooms (Moje, 2008). English language arts, math, science, and social studies are examples of classes that hold distinct communities of practice, problem solving, and methods of practice (Siskin, 1994). Wilson (2011) describes that each of these disciplines holds different standards for addressing and assessing writing within their framework. Many cross-curricular educators do not feel appropriately trained to address writing instruction in their classes (Yore, 1991), nor do they feel equipped to address writing instruction in ways that meet the needs of their students in various contexts (Conley, Kerner, & Reynolds, 2005).

For the most part, traditional writing assessment has been constructed as a technical evaluation tool whose components are known only to those with specialized content knowledge (Huot, 1996). In regards to writing assessment, many traditional rubrics entail technical evaluations that are most effectively performed by teachers with strong English backgrounds. Many teachers feel discouraged by the need for specialized English knowledge in writing

assessment, and Huot and O'Neill (2009) consider this distrust understandable, given the highly technical aspects and discourse of traditional writing assessment.

Writing assessment should reflect the values of its various applications in a cross-curricular context. However, this need is currently not being met, and many non-English content area teachers are skeptical about how to contribute to current writing assessment practices that might not reflect the values and applications of literacy in their classes (Bacha, 2012; Huot, 1996). Since literacy is not limited to the content of specific texts, placing an equal value on the process of reading and writing is a central domain to the discipline (Wilson, 2011).

Assessment plays an important part of writing and its teaching, and it is significant to learn to use assessment in new ways as students learn to write in various and demanding contexts (Huot, 2002b). Writing assessment should be accessible to all educators across content areas, and provide evaluation strategies for the principles of writing that are valued within each context.

An Alternative to Traditional Assessment

Traditional writing assessment does not meet the needs of many urban, cross-curricular classrooms (Huot & O'Neill, 2009). The purpose of grading is not only to determine what a student can demonstrate at a particular point of study, but to help use this information to guide instruction based on the specific criteria of the assessment itself. However, much grading involves little (if any) learning or teaching (Huot, 2002a). Educators should reconsider what is valued as a measure of proficient writing, how this can be guided, and how this can be reflected in assessment. This entails addressing writing assessment as a social action, which means connecting assessment to teaching (Swain & Mahieu, 2012; White 1994; Lloyd-Jones, 1977). Huot (2002a) stresses the significance of using assessments to support the learning environment

for both teachers and students. If educators want to change or redefine the concept of writing assessment, contexts and influences that inform the assessment must be reconsidered (O'Neill, 2011). In order to make this change, educators must first identify what is valued in writing assessment, and determine how this value can be demonstrated through meaningful evaluation and instruction.

Process-Oriented Assessment

Process-oriented instruction is a framework that can shift the focus of programmatic writing assessment to a form of evaluation that concentrates more on the ideas of the writer. Delpit (1986) describes process-oriented instruction as one that centers itself on fluency and creative expression, rather than "correctness." Process-oriented assessment reflects the connection bridging assessment and instruction, straying from traditional writing assessment practices. This idea has been an ongoing educational theory for decades. McTighe and Ferrara (1998) mirror this idea, stating that process-oriented assessments seek to gain insight into underlying processes and strategies demonstrated by the student. This allows students the chance to put their thinking onto paper without the negative repercussions of overemphasized systematic writing conventions, which are evident in traditional assessment. Process-oriented instruction maintains high expectations, because the potential for the piece of writing to improve is not compromised by surface-level writing errors. Process-oriented instruction focuses on strengths, identifying "what is present" rather than "what is not present," which enables teachers to identify building blocks on which to guide instruction (Swain & Mahieu, 2012). McTighe and Ferrara (1998) also note that process-focused assessments provide information about students' learning

strategies and thinking processes, and it can provide this information to teachers while heightening students' own awareness of processes and worthwhile strategies.

Strategy Instruction

Not only does process-oriented instruction shift the focus of writing assessment to an evaluation of ideas in the work, but it provides educators opportunities to tier their instruction to meet writing goals in the students' work. Huot (2002b) notes that this process requires a different kind of classroom, in which all assessments are linked to helping writers improve. Graham, MacArthur, and Fitzgerald (2013) state that writing is a skill that draws on the use of many strategies to accomplish a variety of goals. This form of assessment must continuously gauge the proficiencies demonstrated in student work, and inform educators the direction in which to take their instruction accordingly. Strategy instruction is a technique that allows educators the opportunity to align their instruction with assessment by setting and measuring relevant goals. Strategy instruction involves setting specific, reachable product goals relating to the purpose of the assignment or characteristics of the final product, and teaching more generic processes or strategies for accomplishing these specific writing tasks (Graham, MacArthur, & Fitzgerald, 2013). Teaching adolescents strategies to improve their writing (generic, purpose-related, etc.) has a strong impact on the quality of their writing. Meta-analysis of the effects of 11 different writing interventions showed that strategy instruction was the most effective form of intervention (Graham, 2007). Writing strategy instruction has been found especially effective for students who have difficulty in writing, and is well supported by research in its positive effects on lower-achieving writers across a full range of ability (Graham, MacArthur, & Fitzgerald, 2013). A shift

must to occur in which ideas are the focus of value, and teachers can use their assessments as a measure of voice and a tool to guide instruction effectively.

Assessment as Instruction

There is a mutually-supportive relationship between assessment and instruction (McTighe & Ferrara, 1998). Huot (2002a) asserts that using writing assessment to promote teaching is one of the most crucial aspects of meaningful assessment. To use assessment practices as a form of supporting teaching, teachers must consider what impacts their assessment practices in the first place. Grades and assessments signify what is valued in instruction, therefore it is essential to connect how and what is valued to what is being taught (Huot, 2002b). Spandel (2006) notes that many rubrics created for the purpose of quick grading are ineffective. One existing collection of rubrics that are meant to be used as instructional tools are the 6+1 Trait writing rubrics. The collection contains seven rubrics, one in each of the following categories: ideas, organization, voice, word choice, sentence fluency, conventions, and presentation (Graham, McKeown, Kiuhara, & Harris, 2012). These rubrics were not deemed applicable to the current study as many of the seven categories listed are based on standard academic English conventions. Credible use of assessments relies on teachers' ability to design assessment based on instructional goals (Risko & Walker-Dalhouse, 2010). When creating meaningful assessment, instructors must approach their practice to determine exactly what needs to be assessed, what elements impact this assessment, and how this assessment can inform instruction.

The Need for Assessment as Instruction

Testing and assessment have become central to American education; standardized tests and assessments are being utilized to encourage the teaching of skills prescribed by state and local agencies (Rudner & Schafer, 2002). This form of standardized assessment can be seen from the skill-based expectations imposed on students at each grade level, to the format of progress and placement testing administered to students throughout their educational career.

In terms of argumentative writing assessment, standardized expectations have been established regarding the skills and knowledge students must be able to demonstrate. Argumentation in United States primary and secondary schools is most commonly taught using the Toulmin model (Ellis 2015). This model includes claim, evidence, warrant, backing, qualifier, and rebuttal. The most common writing tests that students take toward the end of their secondary schooling are the ACT and the SAT tests. The rubrics for both of these tests correlate to the Toulmin model of argumentation with variations in wording.

ACT Inc. (2015) notes that the ACT writing test scores on categories including:

- Clear, purposeful thesis (claim)
- Support of claims (evidence)
- Context for analysis (qualifier)
- Connecting ideas/transition (warrant)
- Reasoning (backing)
- Complexity for multiple perspectives (qualifier and rebuttal)
- Diction and conventions

The SAT writing rubric uses the Toulmin model (minus counter-argument or rebuttal) while emphasizing language and mechanics. According to The College Board (2015), the SAT writing rubric scores on categories including:

- Point of view (claim)
- Critical thinking (warrant)
- Appropriate examples and evidence (evidence/ grounds)
- Coherent progression of ideas (qualifier and/or backing)
- Diction and vocabulary
- Sentence variety
- Grammar, usage, and mechanics

Both tests, as well as many other forms of standardized assessment, expect students to know a variation of skills and strategies to apply to their final written product.

Standardized skill expectations also impact what students must be taught at every grade level. Many states around the country have curriculum standards, and state-developed assessments to observe the practice of those standards. Most state standards define expected outcomes- what skills and content students must be able to demonstrate proficiently- but do not advise specific strategies or pedagogy to achieve these proficiencies (Rudner & Schafer, 2002). Educators must recognize the necessity of assessing students in a way that guides their learning in light of standardized expectations. Rudner and Shafer (2002) echo this idea, expressing their impression that testing is more than accountability in the classroom; it is also a means to improve education itself. Since standardized expectations take such priority in the American education system, teachers must express that priority in their assessment techniques. Standards are

pointless until you define how to assess them; every teacher cannot be expected to teach the required skills if the expectations are not explicit and tangible (Bambrick-Santoyo, 2007).

Standardized testing and expectations play a huge role in the evaluation and skill-based monitoring that students must undergo throughout their educational career. In order for assessments to be best-practice and meaningful, they must then directly reflect the expectations of these standards. Assessments can improve instruction by providing information that indicates students' strengths, needs, and specific instructional necessities (Risko & Walker-Dalhouse, 2010). By clearly establishing how standardization connects with classroom assessment, teachers can more effectively create meaningful assessment based off of student needs, and that best informs instruction in consideration to these expectations.

Formative Assessment

Assessment itself can be effective tool for planning (Peverini, 2009). In a balanced assessment system, educators utilize both formative and summative assessment practices. Garrison and Ehringhaus (2013) identify summative assessments as assessments that are given periodically to determine what a student knows or can do. These forms of assessments can be tests, end-of-term exams, or standardized assessments. Formative assessment differs from summative assessment in that, instead of determining what a student can or cannot do, it is part of the instructional process. Formative assessment is a framework that informs instruction through the assessment of student work. Assessing students' work to guide instruction is a longstanding practice of teachers (Risko & Walker-Dalhouse, 2010). Scriven (1967) first made the distinction between summative and formative assessment roles in the context of program evaluation. Bloom (1969) was the first researcher to use Scriven's formative assessment

terminology to make distinctions with respect to students (Wiliam and Thompson, 2008; Black & William, 2003). According to Bennet (2011) Bloom's idea of formative evaluation was to provide feedback and commentary of improvement at each point of the teaching and learning process.

In current-day classrooms, formative assessment is used during instruction to provide feedback to adjust teaching and learning to improve the achievement of intended instructional outcomes (McManus, 2008). Part of creating credible formative assessments requires teachers to carefully plan and determine what should be measured, and how it should be measured (Popham, 2008). Traditional rubrics contain numerous factors designed to be a holistic determination of the writer's proficiency in multiple categories of skill. These rubrics do not effectively inform instruction, as there are too many factors present at once to realistically inform an instructor of the instructional tasks at hand. When students miss a factor in learning, instructors must able to identify the misunderstanding and use that knowledge to plan further instruction (Bambrick-Santoyo, 2007).

Considering the priority of standardized expectations in the classroom, formative assessment strategies must be able to align themselves with correlated and specific tasks. Once this assessment criteria is specified, teachers can then make direct applications to instruction adjustments when necessary, to best accommodate the needs of their students (Risko & Walker-Dalhouse, 2010). Tomlinson (2001) recommends that continual adaptations are made to instruction in order to maintain an appropriate level of challenge dependent on students' demonstration of these tasks. She advocates a framework that uses specific tasks to achieve proficiency of outlined concepts.

There have been recorded positive impacts of the formative assessment process, and researchers claim that a large body of evidence suggests that formative assessment is a vital feature of classroom work, and its development can raise standards (Popham, 2008; Rudner & Schafer, 2000). If state and local agencies expect their standardized expectations to be elevated in schools, then clear expectations and criteria must be represented in the classroom assessment itself in order to more effectively promote achievement and guide instruction. One method of establishing clear expectations and criteria is through the use of tiered instruction.

Tiered Instruction

According to Allen and Turville (2010), the purpose of tiering instruction is to close the gap between what students are actually able to do and what they are expected to do. Wixson and Valencia (2014) suggest that instead of teachers considering which tasks are "best" when planning, they should consider which tasks are appropriate within the context of the learning goal and student ability. Essentially, teachers begin with a standard and then utilize student needs and current levels to determine what to teach in order for the student to be able to reach that standard. With the support of considerate planning and tiered instruction, students may become more engaged and increase their success in meeting the required learning goals Tiered instruction supports formative assessment through clear goals, indicators of student readiness, and open-endedness to support and engage all levels of learners.

In order to determine the effectiveness of assessment, educators must clearly define the learning goals inherent in the assessment (Bennett, 2011). Although writing is a recursive, complex process, writing skills are generally not learned all at once. In order to take students from lower-level to higher-level learning goals, one skill needs to be presented at a time. After

individual skills are taught, they can be combined in order to achieve higher-level learning goals; this scaffolded process, or instructional framework, includes the skill steps necessary to perform each part of the whole process (Tomlinson, 2001).

The Common Core State Standards Connection

The Common Core State Standards (CCSS) is an accepted form of standardized, grade-level expectations implemented throughout Wisconsin and various other states (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). The CCSS promotes the concept of cross-curricular literacy. Cross-curricular teachers are expected to support literacy by implementing critical reading and writing work in their classes (Calkins, Ehrenworth, & Lehman, 2012). The CCSS emphasizes writing as heavily as reading, and the potential of reading being assessed through writing creates an even more critical significance on writing (Calkins, Ehrenworth, & Lehman, 2012).

Additionally, the CCSS explicitly requires higher-order thinking skills as a means to prepare students for college readiness (Magner, Soulé, & Wesolowski, 2011). These higher-order, or complex, skill requirements are integrated throughout the standards in every discipline. These complex expectations are also evident in the cross-disciplinary literacy expectations of the CCSS, which are established in history/social studies, science, and technical subject courses (Magner, Soulé, & Wesolowski, 2011). Standards in English language arts at the secondary level include five sub-standards: reading literature, reading informational text, writing, speaking and listening, and language. The focus of the CCSS in the current study is the connection between cross-disciplinary literacy expectations, complex skills, and the assessment of argumentative writing standards.

Literacy

The last five years have given a significant emphasis on literacy achievement on secondary students in the United States (Moje, 2008). Increasingly, educators observe the correlated relationship between learning to write and learning to read (Pearson, 2002). Graham and Herbert (2011) conducted a meta-analysis on the ways writing instruction impacts reading skills and found that the reading and writing processes are interactive. Increases in either reading or writing practices correlate to increases in the other. Many of the processes and strategies for writing involve an interaction with text, and therefore involve the processes and strategies of reading (Carter, 1990; Valencia, McGinley, & Pearson, 1990; Freedman, Dyson, Flower, & Chafe, 1987; Johnston, 1987; Bruffee, 1986; Faigley, 1986; Smith, 1982; Macrorie, 1976; Moffett, 1968). The writing process has a direct connection to reading process, and researchers know that these skills often draw from the same pool of background knowledge (Graham, MacArthur, & Fitzgerald, 2013). Deford (1981) also advocates the supportive and interactive relationship between the reading and writing processes.

Langer and Flihan's (2002) research has shown that writers tend to combine what they have learned about language, structure, and style from the texts they have encountered as readers; they also reflect on their knowledge of texts and experiences as a way of producing and synthesizing ideas for writing. When approached as similar and related processes rather than isolated skills, reading and writing can promote the development of reading, writing, and thinking.

More than only skill processes, research examines the relationships between reading and writing as cognitive and social processes as well (Langer & Flihan, 2002). The combination of these processes is referred to as literacy. The Wisconsin Department of Public Instruction

identifies literacy as "the ability to read, write, listen, speak, think critically and perform in different ways and for different purposes" (2011, p. 18). Literacy involves more than simply reading and writing skills, but context and application. The necessity to promote literacy has been established by the CCSS. "Reading, writing, speaking, listening and critical thinking must be integrated into each discipline across all grades so that all students gradually build knowledge and skills toward college and career readiness" (Wisconsin Department of Public Instruction, 2011, p. 23). Not only is literacy achievement a predictor for academic success, it is a basic requirement for participation in civic life and in the global economy (Graham, MacArthur, & Fitzgerald, 2013). Knowing this significance, demonstration of literacy skills is an expectation of all content-area teachers.

Literacy is as significant in mathematics, engineering and art courses as it is in science, social studies, and English (Wisconsin Department of Public Instruction, 2011). The positive impact of literacy processes can be seen at a cross-curricular level. Students' understanding of science, social studies, and language arts texts is improved when they write about what they read (Graham & Hebert, 2010). As the reading and writing processes are interactive, increases in either reading or writing correlate to the other. (Graham, 2011). Across content areas, the benefits and applications of literacy can be observed.

Argumentative Writing

The educational emphasis on argumentative writing skills has increased as the CCSS has taken a focus on literacy across the curriculum. Calkins, Ehrenworth, and Lehman (2012) note that the two biggest literacy cornerstones across disciplines is providing students with extensive reading and meaningful opportunities for writing a range of informational, argumentative, and

narrative texts. They emphasize that prioritizing argumentative and informational writing best suits the needs of implementing cross-curricular literacy according to the CCSS. As argumentation was already of considerable interest to content-area educators, this increased emphasis has put teaching of argumentative techniques at the forefront of education (Bell, 2000).

The CCSS cover three types of writing: argumentative, informational/explanatory, and narrative (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). The style of argumentative writing most used in the United States is the Toulmin Model which consists of grounds, claim, evidence, warrant, backing, qualifier, and rebuttal (Ellis, 2015). The Toulmin Model closely aligns with the wording of the CCSS argumentative writing standard and sub-standards which includes using analysis, evidence, and reasoning to support claims (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

Argumentative writing has been prioritized in classrooms aligned with the CCSS (Calkins, Ehrenworth, & Lehman, 2012). Not only is argumentative writing trending with national standardized assessments, but it provides students an outlet to interact with their reading and writing in an original and structured fashion. It is argued that the skill of argumentation is necessary in order to participate in a democratic society and is also a major factor in academic, career, and personal success (Chaffee et al., 1999). The Wisconsin Department of Public Instruction (2011) mirrors these sentiments, emphasizing that argumentative writing and literacy skills are critical to college and career readiness. Teachers must have the tools to implement and assess argumentative in order to meet heightened literacy expectations.

Complexity

The CCSS places a much greater emphasis on higher-level skills than previous educational reforms, with a focus on complexity and independence in skills across disciplines (Calkins, Ehrenworth, & Lehman, 2012). There is a distinct difference between complexity and difficulty. When a task is difficult, one will put forth more effort than they would when completing a task with less difficulty. Effort is in no way synonymous with critical thinking. Sousa (1995) stresses that there is a significant difference between complexity and difficulty. Tileston (2000) notes that complexity refers to the thought processes that the brain uses to deal with information, while difficulty refers to the amount of effort used within a level of complexity. A learner might use a large amount of energy on difficulty while working at a low level of complexity.

Complexity refers to the compilation of skills required to address a task with a relevant and effective response. Davis and Sumara (2008) describe complexity as the study of learning and learning systems that encompasses specified knowledge and context. They argue that complexity points towards some sort of system that demonstrates learning. Skill and knowledge-based collectives can become broader, more refined, and more capable of diverse possibilities.

At times, texts and assignments represent an elevated level of difficulty, but student response may not require a high level of complexity. Text does not automatically have a certain level of complexity; complexity is the way in which the reader interacts with the text and the task factors given (RAND Reading Study Group, 2002). Writing complexity reflects the ideals of critical engagement, analysis, evaluation, and perspective.

Complexity in the current study is defined as a synthesis of thought processes and strategies used to complete a task. Students may be able to engage with text or in writing while

thinking critically about the topical concept, yet struggle with English language conventions. As Rezaei and Lovorn (2010) found that teachers are prone to be significantly influenced by mechanics and convention errors over content, even if mechanics and conventions aren't the focus of the writing task. In order to gauge the complexity of student writing, it is imperative that instruction and evaluation narrow solely to content and purposefully overlook mechanics and conventions.

Bloom's Taxonomy

Swain and Mahiew (2012) are clear that in order to embed expected results into a curriculum, teachers must be aware of all stages in the process, and utilize those stages in planned activities. They argue that many designed assessment systems do not get used long enough for researchers to determine their real long-term results. According to Tileston (2000), Bloom's Taxonomy (1976) is a system created in 1976 and is still widely utilized by teachers today to measure task complexity. Each level of Bloom's Taxonomy "…represents a different level of complexity" (Tileston, 2000, p. 37). However, Bloom's does not measure the complexity achieved in students' responses to the assignment, only the depth of the assignment given. A shift must occur where this Taxonomy can be represented not only in the assignment, but in the assessment of student work. Such a framework could better support the assessment of process-oriented goals of writing in the cross-curricular context.

The levels of Bloom's Taxonomy outline various degrees of complexity as they pertain to the demonstration of skill and knowledge acquisition. According to Crebert, et. al. (2011), and Tileston (2000), the six levels of Bloom's Taxonomy that assess the level of complexity

represented by an assigned task include: knowledge, comprehension, application, analysis, evaluation, and synthesis.

Bloom's level one, *knowledge*, is the ability to recall and recite information. When asked to choose, find, define, label, match, name, or recall, the subsequent thought needed to answer the question is considered a low-complexity process (Crebert, et. al., 2011; Tileston, 2000).

Figure 1: Bloom's Taxonomy Level 1- Knowledge

Bloom's Taxonomy Level 1	Essential Questions
Knowledge	• What is? • Where is? • When did? • Can you list? • Why did? • How would you explain?

Bloom's level two, *comprehension*, raises complexity to the next level by asking one to explain something that one understands. When asked to compare, contrast, demonstrate, interpret, infer, rephrase, summarize, or classify, the critical thinking required is increased significantly from the previous level of the taxonomy (Crebert, et. al., 2011; Tileston, 2000).

Figure 2: Bloom's Taxonomy Level 2- Comprehension

Bloom's Taxonomy Level 2	Essential Questions
Comprehension	• How would you classify/ compare…? • Will you state in your own words… • What is the main idea of…? • Which statements support…? • What is meant by…? • How would you summarize…?

Bloom's level three, *application*, requires one to apply their knowledge and ideas to a task or situation. At this level, one is required to take or decide on an action in response to an input or condition. When students are asked to build, construct, develop, interview, organize, experiment, plan, or model their work, knowledge and comprehension must be transferred and utilized, not simply explained (Crebert, et. al., 2011; Tileston, 2000).

Figure 3: Bloom's Taxonomy Level 3- Application

Bloom's Taxonomy Level 3	Essential Questions
Application	• How would you use…? • How would you solve…? • How would you organize…? • What approach would you use to…? • What would result if…? • What elements would you choose to change?

Bloom's level four, *analysis*, is similar to application, but the difference is that while application is more of a physical or spatial process, analysis is the cerebral process of breaking something down into manageable parts. The ability to categorize, discover, examine, simplify, or determine function or motive are thought processes that require a high level of conceptual ability or critical thinking (Crebert, et. al., 2011; Tileston, 2000).

Figure 4: Bloom's Taxonomy Level 4- Analysis

Bloom's Taxonomy Level 4	Essential Questions
Analysis	• How _ is related to _? • Why do you think…? • What is the motive? • How would you classify…? • What evidence can you find? • What is the relationship between…? • What is the function of…? • What ideas justify…?

Bloom's level five, *evaluation*, reflects a student's ability to judge. In order to evaluate, one must have an expert knowledge and comprehension of the topic and be able to analyze and compare the efficacy of various outcomes. One is not able to effectively judge an action that one does not understand. Before one is able to defend, criticize, dispute, justify, recommend, assess, or deduct, one will have to critically engage with and analyze multiple variables. The process of evaluation is the second highest level of functional complexity in the taxonomy (Crebert, et. al., 2011; Tileston, 2000).

Figure 5: Bloom's Taxonomy Level 5- Evaluation

Bloom's Taxonomy Level 5	Essential Questions
Evaluation	• Do you agree with…? • How would you prove…? • Assess the value of… • What would you recommend? • What choices would you have made? • How would you prioritize…? • What information would you use to support/justify?

Bloom's level six, *synthesis*, represents the highest level of task complexity in the taxonomy. Level six reflects the ability to create something new or to take something apart and put it back together in a different way. This level of Bloom's Taxonomy involves building, combining, composing, designing, developing, imagining, inventing, improving, and/or testing. These are the skills that require the most complex thought processes needed for task completion (Crebert, et. al., 2011; Tileston, 2000).

Figure 6: Bloom's Taxonomy Level 6- Synthesis

Bloom's Taxonomy Level 6	Essential Questions
Synthesis	• What changes would you make to solve…? • How would you improve…? • What would happen if…? • Can you propose an alternative? • How would you design…? • How would you test…? • Predict the outcome of… • Construct a model that would change…

Complexity and the Process-Oriented Approach

The CCSS addresses text complexity, and uses three measures to determine text complexity. The first measure looks at the meaning, language, and knowledge demands of a text, which correlates to Bloom's first level of complexity: knowledge. The second measure relies on readability measures, which correlate to Bloom's second level of complexity: comprehension. The third measure includes reader and task issues such as motivation, experience, and task complexity, which include the remaining levels of complexity and rely on student engagement with a text (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

However, the CCSS does not mention complexity as it aligns with writing standards. Given the significance of literacy in the CCSS, and the correlation between reading and writing skills (Carter, 1990; Valencia, McGinley, & Pearson, 1990; Freedman, Dyson, Flower, & Chafe,

1987; Johnston, 1987; Bruffee, 1986; Faigley, 1986; Smith, 1982; Macrorie, 1976; Moffett, 1968), assessment practices must reflect the complexity expectations related to writing as well. A study conducted by ACT, Inc. (2006) found that the ability to utilize the information read in analysis, evaluation, or synthesis processes such as writing is just as significant for college and career readiness as the ability to read complex texts. This concept stresses the equal value of engaging in complex materials through performing complex literacy tasks.

Rudner and Schafer (2002) note that delineation, or breaking down all aspects of content, level, and complexity of a subject is imperative in utilization of standards for effective planning. Once this is done, alignment, or utilizing this delineation, will be structured to create links between information. This allows for effective scaffolding of the learning activities. Davis and Sumara (2008) note that complexity is emergent, in that its collective components arise through the interactions of many sub-components. They claim that in order to be emergent in complexity, one must "level-jump," or be able to look at a circumstance holistically, while being able to pay attention to the conditions of its emergence. Bloom's Taxonomy is one possible tool to provide framework for the delineation and alignment processes, as the levels of complexity specified in Bloom's Taxonomy align with CCSS expectations for reading and writing standards.

Teachers historically feel tension about their roles as the teacher, reader, and rater in a final assessment (Huot & O'Neill, 2009; Conley, Kerner, & Reynolds, 2005; Barrit, Stock, & Clark, 1986). In order to assuage this tension, assessment practices can be performed within specific, or more narrowed contexts. This may allow the rater to make more pedagogical assessments without conflicting influences or multiple values that may be too abstract to judge objectively. Assessment should be performed within a particular context as a clearly defined event (Huot, 1996).

Rezaei and Lovorn (2010) investigated the reliability of rubrics in assessment among multiple raters. The research focused on one essay that demonstrated great conventions of English language, while not fully addressing the prompt. The research also focused on another essay with content completely addressing the prompt, but lacking in academic English conventions. They found that the essay raters were significantly influenced by mechanical characteristics or conventions over content, even though conventions comprised a minor amount of the rubric. Teachers tend to get stuck on conventions when grading writing. Assessing only content messages within writing, while specifically ignoring conventions and mechanics, is necessary to form the foundation of effective writing on which conventions can later be built.

Creating and Supporting Modified Assessment

In order to address the needs of argumentative writing assessment across disciplines, the researchers for this study created the Bauer & Kohut Argumentative Writing Rubric (Appendix H). The components included in the rubric are detailed below:

Origin of Content

A rubric should link to the instructional objectives to which the school adheres (Risko & Walker-Dalhouse, 2010). Considering that many schools throughout the United States have adapted the CCSS, rubrics could be used to assess the criteria outlined by the standards. Unfortunately, The National Research Council (2001) found that it is difficult to assess deep comprehension, informational organization, and strategy usage when using classroom assessment based on standards. One possibility for this lack of depth in assessment is that alignment with state assessment criteria tends to focus more on literal comprehension

(Applegate, Applegate, McGeehan, Pinto, & Kong, 2009). In order to create more meaningful and well-rounded assessment, the components of an assessment- Bauer & Kohut Argumentative Writing Rubric- must first be identified (Bennett, 2010).

Lanning (2013) points out that the Common Core State Standards were written in a way in which many separate skills are embedded in one standard: this intentionally encourages integrated skill instruction from lower-level skills into more rigorous and complex competencies. Using the CCSS as an end point should be one small part of the planning process. The Common Core State Standards Initiative (2010) recognizes the necessity of linear instruction that first builds foundational skills and concepts within each standard so that "as students advance through the grades and master the standards in reading, writing, speaking, listening, and language, they are able to exhibit with increasing fullness and regularity these capacities of the literate individual" (p. 7). The Bauer & Kohut Argumentative Writing Rubric created for the current study was designed as a linear, process-oriented tool intended to guide instruction and assessment. Each necessary component of the desired end goal is present within the steps of the rubric, in order to clearly establish learning goals as they progress from lower-level to higher-level processes. The Bauer & Kohut Argumentative Writing Rubric lends itself to clear, differentiated instruction, as students are able to work at multiple levels within one each step of the framework.

The Bauer & Kohut Argumentative Writing Rubric specifies an incremental increase in complexity aligned to the expectations of argumentative writing detailed by the CCSS. These expectations are clearly specified and tiered to help students arrive at argumentative writing goals using the assessment as a guide to learning and instruction.

Instructional Strategies Aligned to the Rubric

The reliability of the Bauer & Kohut Argumentative Writing Rubric to guide instruction is dependent on formative assessment and tiered instruction. To accommodate these components within the framework of modified assessment for this study, the researchers also aligned supported instructional strategies to each level of the rubric. Considering the prevalence of complexity in the literacy expectations of the CCSS, the researchers for this study correlated instructional strategies for argumentative writing as they align with Bloom's Taxonomy. Since research supports the use of assessment as a guide to inform instruction (Risko & Walker-Dalhouse, 2010), it is also important to consider strategies that might align to the criteria of the assessment. The Bauer & Kohut Argumentative Writing Rubric combines the expectations of the CCSS and Bloom's Taxonomy to address the assessment of argumentative writing across disciplines. The strategies in this section (categorized by the corresponding level of Bloom's Taxonomy) reflect the expectations of the Bauer & Kohut Argumentative Writing Rubric.

Bloom's level one, *knowledge*, specifies a basic and surface-level understanding: to remember. According to Marzano (2012), students should be able to demonstrate an understanding of the content. When working at this level, educators should assess strictly for this comprehension while deemphasizing irrelevant skill demonstrations that may be present in the student's work. In this level, there should be no emphasis on punctuation, spelling, or grammar (Marzano, 2012). Strategies utilized at level one may include factual questions intended to determine general information by using words such as who, what, when, where, and how (Chaffee et al., 1999). The information gained from these types of questions will be objective data used to assess surface-level understanding of a topic. If basic knowledge of a topic is not acquired, one will be unable to move through the remaining stages of complexity.

Bloom's level two, *to understand*, starts to gauge comprehension as it relates to the individual. When the reader is able to interact with the text, comprehension occurs (Kucer, 2001). This level promotes the idea of reading and writing as an interaction, and assesses the student's ability to engage in the resulting interaction personally. Pardo (2004) describes the instructor responsibility at this level to create learning opportunities that encourage the kind of interaction which creates meaning. Within this level of taxonomy, students should work on building their comprehension through context, connection, and inquiry.

The comprehension of text can be evaluated across all content areas by putting content into one's own words, or interpreting knowledge. Strategies that implement learning at the comprehension level include activating background knowledge, utilizing concept maps or charts, sketching through the text, monitoring comprehension, and questioning interpretation. According to Pardo (2004), concept maps and charts show background knowledge, connection to the text, and connection across texts. Students are able to support the unknown with the known of their own personal knowledge index or experiences. Sketching through the text is an annotation method that uses pictures instead of words. According to Daniels and Steineke (2011), students who stop reading periodically to create a visual representation of the text experience a low level of difficulty with a high level of student engagement with a text. Sketching encourages interactions with text that are more likely to emerge in later writing than textual annotation. Pardo (2004) defines monitoring comprehension as finding out what students understand and do not understand in order to fix the gaps in comprehension. The more a student can advocate for their understanding, the more effectively they can demonstrate this understanding through assessment. Questions of interpretation allow students to demonstrate interpretation skills such as

chronological relationships, process relationships, cause and effect relationships, or casual relationships (Chaffee et al., 1999).

Bloom's level three, *apply*, encompasses students' ability to apply logical and appropriate structure to writing to fit the needs of a task. This level requires an ability to effectively pre-plan or to revise writing, both of which can be problematic to novice writers. Application may be assessed by looking for the presence of a clear thesis and a logical order of reasons and warrants following a claim, including appropriate breaks in writing. Strategies used to improve application ability include surface revisions such as conventions or morpheme changes and non-surface revisions such as adding, moving, or omitting information (De La Paz & Sherman, 2013).

Bloom's level four, *analyze*, requires students to express multiple perspectives or address relationships. Ferretti et al. (2009) claim that students do not attend to multiple viewpoints or opposing positions enough. In order to assess analysis, an instructor looks for the description of more than one perspective and the purposeful analytical understanding of each. Examples of teaching strategies to improve analysis include asking critical questions and the S.C.A.N. technique. Song and Ferretti (2012) give the following examples of critical questions: "How sure are you that said consequences will actually happen?", "What evidence do you have to support your assertion that the consequences will happen?", and "What are other possible outcomes?" The S.C.A.N. technique asks the questions: "How much sense does it make?" "What connections can I make?" "Can anything be added?" "Can I note any errors?"

Bloom's level five, *evaluate,* requires students to present a well-developed claim and counter-claim which include possible strengths and limitations. Although students may evaluate to what degree they have met the expectations at any level of complexity, once a student is able to perform a task at all lower levels of complexity, self-assessment at a holistic level becomes

more meaningful. Through self-assessment, learners are able to think about their progress and find more ways to improve and revise their work (Kavaliauskiene, 2004). Work may be revised according to the strengths or weaknesses in the writing. According to Sommers (1980), revising is an important step which can help students to create meaningful texts and also is a means through which they can analyze their papers in terms of content, organization, communicative purpose, and genre. This level of the taxonomy focuses not only on the comprehension and analysis demonstrated by the student, but on holistic evaluation to meet the needs of the argumentative writing assessment. Revision correlates more closely with improved writing than does almost any other form of writing instruction (Beyer, 1979). Individual reflective evaluation of work can support clearer, more concise, and more purposeful writing related to content and criteria.

Strategies that may be used to learn and improve evaluation techniques include self-monitoring, checklist revision, peer revision, and evaluating arguments. Self-monitoring requires students to conduct ongoing interaction with their writing through observation and recording (Menzies, Lane, & Lee, 2009). It is a reflective strategy that promotes independence in recognizing the strengths and weaknesses of writing. Kavaliauskiene (2004) suggests that with self-assessment, learners get an opportunity to consider their own progress and find ways in which to change, adapt, or improve. Self-monitoring also builds independence in writing. Checklist revision is a more scripted alternative to self-monitoring. Raimes (1983) specifies that this is a useful starting point in analysis of specific elements of writing because teachers can devise checklists that focus on specific tasks, which can then be adjusted for students' needs. The strategy of evaluating the strengths and weaknesses of an argument (or counterargument) is a crucial element to evaluating an argumentative writing assessment as a whole. Chaffee et al.

(1999) point out that one must be able to *evaluate* the effectiveness of an argument in order to *create* an effective argument. Examples of evaluation questions may include: How true are the reasons being used to support the claim? To what extent do the reasons support the conclusions? To what extent does the conclusion follow logically from the reasons given?

Bloom's level six, *create*, essentially combines knowledge and ability of all of the previous levels of complexity in order to utilize holistic argumentative skills and critical thinking, and to compose with specific needs, audience, and situation in mind. Achieving at this level requires students to compile information together in new ways or patterns to find different solutions (Crebert, et. al., 2011). In order to see students jump into the process of creating an effective and purposeful piece of writing to meet the criteria of an assessment, teachers must be aware of the levels of process complexity required to complete an assigned task. Teachers must then present strategies that tier learning and guide students to the end goal by recognizing where each student stands on the proverbial ladder. By connecting the evaluation of literacy to Bloom's taxonomy, a final assessment should demonstrate a naturally heightened correlation to which both instructor and student see a clear path. Along the way up this ladder, students must be assessed for process skills and guided to higher-level thinking instead of being corrected (and possibly held back) from higher levels of complexity due to surface level mistakes in their writing.

Strategies aligned to assessment should be implemented according to student readiness and learning goals. Formative assessment and tiered instruction help bridge the gap between student demonstration and targeted outcomes. In order to purposefully guide instruction according to assessment, instructors should be able to align their instructional practices with the desired learning goals detailed within the assessment.

Conclusions on Assessment

Traditional assessment can limit teachers' ability to emphasize complex, higher-level processes in their evaluation of student writing. This trend is even more prevalent in classes populated by minority students or students of varying socioeconomic status, who tend to create written expression that is more true to spoken vernacular than academic English (Whitney, 2005; Ball, 1996). Curriculums that have adapted the CCSS have cross-disciplinary literacy expectations. Alignment to the CCSS and standardized tests such as the ACT and SAT has created a high importance on the skill of argumentative writing. Teachers across the curriculum are expected to implement reading and writing instruction. Traditional assessment tends not to accommodate diverse contexts of writing, which often times leaves cross-curricular educators frustrated or differed from writing instruction and evaluation (Huot & O'Neill, 2009).

Assessment needs to reflect the contexts and purposes to which it applies. Writing assessment should be accessible for teachers across the curriculum, and accessible as a measure of the expression and ideas on behalf of urban students. In order to achieve these considerations, the emphasis of assessment should shift from evaluating surface-level correctness to complex, higher-level ideas and processes present in the writing. Process-oriented frameworks can assist this shift in assessment focus, especially when supported with tiered strategy instruction and formative assessment strategies.

The researchers in this study created an argumentative writing rubric with the intention of accommodating the assessment of complex, higher-level processes in student writing in a cross-curricular context. The researchers aligned the objectives on the rubric to the CCSS argumentative writing standards and Bloom's Taxonomy. Considering that the reliability of the

rubric to guide instruction is dependent on formative assessment and tiered instruction, the researchers also aligned supported instructional strategies to each level of the rubric.

CHAPTER 3
RESEARCH METHODOLOGY

Research Procedures

The purpose of this chapter is to provide an explanation of the procedures, instrumentation, and research design used in the current study to investigate the reliability of a rubric created to guide the instruction and assessment of argumentative writing in the cross-curricular context.

Students in two inclusive, urban education classes participated in the study. Students from both classes participated in ten 50-minute instructional courses over a period of two weeks. The control group received traditional instructional practices during the study. The experimental group engaged in brief literacy intervention guided by the Bauer & Kohut Argumentative Writing Rubric created for this study. Data will be collected and analyzed to determine significance of achievement rate between each group. Additionally, four reported cross-curricular educators participated in determining the reliability of the rubric as an evaluation tool for student argumentative writing. Each educator was asked to rate 28 student samples of argumentative writing between 1 (low achievement) to 5 (advanced achievement) using a traditional rubric in addition to the rubric created for the study. Investigators will analyze the data to determine the interrater reliability of each rubric and the parallel forms reliability comparing the results of each rubric.

Population and Sample

Both the control group and experimental group were determined based on the placement of ninth graders in the researchers' classes. This type of sampling is referred to as convenience

sampling, and McMillian (1996) notes that it is one of the most common and widely used sampling strategies in educational research. Because the nature of convenience sampling may create bias, both classes were asked to complete a baseline on-demand argumentative writing assessment to determine homogeneity between groups. A coin was flipped to determine which class was the control group and which was the experimental group.

The racial demographics of the school as of 2015 are 69% black/non-Hispanic, 14% Hispanic, 9% Asian, 8% white/non-Hispanic, .4% Native American, and .06% multiracial. 13% of students tested at the proficient level in reading in the 10th grade.

Students in two standard (non-advanced) level urban public high school English classes participated in the study. A parent consent and child assent form was given to students in two classes taught by the same team of two instructors. 47 forms were sent home and 47 were returned by the students with the appropriate signatures. Five of 47 students received special education services according to an individualized education plan and were included in the regular education class. There were no exclusions by the instructors; all students in the classes participated in the treatment of their particular group and data was collected for the study only from students who returned the parent consent and student assent form. If a participant was absent for either the first or last condition, the closest result to the missing result was used in its place.

Regular cross-curricular educators also participated in the study. Teachers were recruited on a voluntary basis. The researchers in this study asked colleagues across disciplines to participate voluntarily and without compensation. The demographics of the teachers who participated in the study were 75% European-American, and 25% Hispanic. One educator was a first year teacher, one educator had 32 years of experience, one educator had 22 years of

experience, and the last had 13 years of experience. The educators taught regular education classes in the disciplines of social studies, English language arts, math, and health and wellness education.

Five educators originally participated in establishing initial interrater reliability for the use of the traditional rubric and the Bauer & Kohut Argumentative Writing Rubric during the study. While four of the educators were able to establish common ratings between their applications of each rubric, one educator felt the application of both rubrics wasn't manageable. This educator reported a strong disciplinary difference in writing assessment between the rubrics in the study and those used in her own science class, and therefore didn't continue the study. The four reporting educators received and signed a consent form. All data and survey questions collected from the educators were used and analyzed in the study. Survey questions (listed on page 60 in the "Results for Qualitative Survey" section) were distributed to the educators instead of a face-to-face or over-the-phone interview because in-person methods could have created bias through the participants' acquaintances with the researchers for this study.

All data collected in this study was given voluntarily and all scores and responses were kept confidential. No participants were offered monetary incentives nor other incentives to participate in the study.

Research Instruments

The rubric in the current study was designed as a linear process intended to guide instruction. Each necessary component of the desired end goal is presented incrementally within the steps of the rubric. The Bauer & Kohut Argumentative Writing Rubric provides clear, differentiated instruction, as students are able to work at multiple levels within each step of the

rubric framework. The criteria of the rubric has been established through its alignment with Bloom's Taxonomy and the Common Core State Standards. The reading and writing expectations detailed by the Common Core State Standards were necessary to consider when creating the rubric due to their widespread adoption across the United States, including in Wisconsin. The necessity to promote literacy has been emphatically established by the CCSS. The Common Core State Standards also emphasize the application of literacy skills as they apply to informational text, as "Reading Informational Text" is one of the main umbrella standards in which to assess secondary students in the English Language Arts classroom. Students are expected to engage in complex literacy skills in a cross-curricular context, and the Bauer & Kohut Argumentative Writing Rubric aligns itself with the relevant content and theoretical application to accommodate this need.

Informational articles were used in alignment with argumentative writing prompts throughout the study with the experimental and control groups. Each article was acquired from Newsela (www.newsela.com). Newsela provides leveled reading materials. Davis (2016) notes that Newsela levels their articles according to Lexile bands, with consideration to the background knowledge and subject maturity of each text. Each informational article used during the study had the same level of difficulty in order to eliminate any variables regarding text difficulty. Each article was levelled at the sixth grade reading level. This reflected the baseline data collected from the control group and experimental group. Students in each group had an average score of 1.64, which placed them among the first and second lower-level tier of the Bauer & Kohut Argumentative Writing Rubric. The researchers for this study felt that maintaining a sixth grade reading level throughout the study would eliminate variables that could be attributed to fluctuating text difficulty.

Additionally, each article was edited to approximately 1.5 pages in length. The researchers for this study did not change the wording or difficulty of the text, but aimed to maintain consistency in the difficulty and length of the text by omitting excess details. This was to help eliminate any variables that could exist if the text difficulty or length were to vary. While the articles had homogenous reading level and length, the content of each article was different. Each article consisted of opinion pieces on two sides of an issue. The decision to use articles containing different content throughout the study was made to reduce the effects of repeated exposure, which could potentially affect student work outcomes.

One argumentative writing prompt was provided for each article. While the content of the articles changed, the directions for the writing prompt were always the same:

"Read the article provided. In your writing, take one of the positions given on the following prompt or present a different point of view on this question. Create an argumentative response that you believe effectively and thoroughly supports your opinion."

The researchers for this study intentionally excluded any specific writing goals in the argumentative prompt (such as including multiple perspectives, a clear thesis, or counter argument) to help eliminate instructional bias with the experimental group. This was done to help determine the effectiveness of tiered strategies aligned with the rubric and student readiness versus teaching that is not guided by the rubric and student readiness. If the rubric proves to be a reliable guide for instruction, improvement in student work should be evident based on the sequencing and responsiveness of rubric-guided literacy instruction alone.

All of the participant writing samples in this study were assessed using the Bauer & Kohut Argumentative Writing Rubric. In the experimental group, instruction aligned to the Bauer & Kohut Argumentative Writing Rubric would build on the readiness of students, as determined by assessing their writing in response to informational text. Student writing was scored based on the indicators of proficiency within each level of the Bauer & Kohut Argumentative Writing Rubric. Instructors assessed student work by analyzing its content, and determining its proficiency based on the objectives that determine each level of the Bauer & Kohut Argumentative Writing Rubric. Subsequent lessons were guided by the rubric, and were continuously informed by student data throughout the period of the study. Data was maintained in a student growth chart to track the level of written responses to each informational article.

Targeted brief literacy intervention mini-lessons were planned to teach each increasing level of complexity as efficiently as possible. The literacy intervention strategies (outlined in the "Adapting Bloom's Taxonomy to Argumentative Writing Strategies" section in Chapter 2) were different for each level of the Bauer & Kohut Argumentative Writing Rubric. The strategy of intervention was determined by the readiness demonstrated by a majority of students in the class based on their proficiency in certain levels of the Bauer & Kohut Argumentative Writing Rubric. Different strategies were implemented throughout the study as students demonstrated progress in their proficiencies. Each strategy implemented during this study has been researched according to its application to the instructional goals detailed in each level of the research rubric. The strategies were guided by the expectations set in place by the Bauer & Kohut Argumentative Writing Rubric. Student work and readiness informed the practice of moving to the next set of strategies and skills. Instruments required to implement each strategy such as worksheets or post-it notes were provided. The researchers do not want to focus on the effectiveness of any one

strategy, but instead on the reliability of using the research rubric to guide instruction. If the rubric is a reliable guide, student results should show statistically significant progress. The authors are not prescribing literacy strategies for teachers to use; instead, they are presenting a tool which can be implemented for strategy planning, instruction, and assessment by any teacher at any level and across the curriculum.

In the control group, instruction was based on the district's literacy plan developed at the secondary school of the study. This literacy plan was based on teaching students to use a structured reading annotation system before responding in writing. The annotation strategy implemented with the control group was predetermined by the school learning team. The school of the study implemented the district literacy plan through a cross-curricular initiative of reading informational text with argumentative response. Students in all subject areas were expected to be able to analyze informational text and create a coherent argumentative writing response that meets the expectations of the Common Core State Standards. This initiative parallels the objectives of the current study. The school of the study emphasized the skill of text annotation, a skill that could align with multiple levels represented on the experimental rubric. The students in the control group received the same assessments as the students in the experimental group, but did not have instruction that was guided by the results of their work. The instruction in the control group instead used the annotative skills and strategies prescribed by the school's literacy initiative according to the prescribed pacing guide.

In addition to testing the Bauer & Kohut Argumentative Writing Rubric's consistency in guiding instruction, the researchers for this study asked four participating educators to rate student writing samples using the Bauer & Kohut Argumentative Writing Rubric and a traditional rubric in order to test the reliability of each rubric as an evaluation tool. The

traditional rubric, the ACT Writing Test Scoring Rubric (Appendix B), is a holistic six point rubric that aligns itself to the argumentative writing expectations of the ACT Test. The ACT Writing Test Scoring Rubric was chosen because it was created through the ACT organization as a means in which to set expectations and assess the argumentative writing that aligns with the test. Both the ACT Writing Test Scoring Rubric and the Bauer & Kohut Argumentative Writing Rubric utilize expectations of the Toulmin Model of argumentative writing (Ellis, 2015) and function on a six-point scale.

Data Collection

Students in two inclusive, urban education classes participated in the study. Students from one class represented the control group, and students from the other class represented the experimental group. Students from both groups completed an initial argumentative writing sample in response to informational text. The researchers for this study analyzed this data using a t-test to determine homogeneity between the two groups. The study was conducted in ten 50-minute class periods over the course of two-weeks. Throughout the period of the study, both the control group and experimental group received identical assessment pieces, and these pieces were evaluated by the researchers for this study per the Bauer & Kohut Argumentative Writing Rubric. Daily procedures included providing a mini-lesson on a brief literacy intervention strategy, handing out a new article for written response to a prompt utilizing the new strategy, and directing participants to apply the strategy throughout reading and written response. Each participant had a file folder in which all artifacts were collected. All participant products were collected and scored each day according to the six point scale represented on the Bauer & Kohut Argumentative Writing Rubric. This rubric was not provided to participants to eliminate bias that

could arise through recognition of assessment criteria. At the end of the study, data of the final assessment of each group was analyzed using an unpaired t-test to determine if statistical significance exists between the levels of improvement shown by the control group and experimental group. Any negative gains were adjusted to zero. The researchers for this study anticipated that any statistically significant increases found on behalf of the experimental group could be attributed to the consistency of the Bauer & Kohut Argumentative Writing Rubric in guiding instruction to meet elevated literacy goals.

Additionally, four reported cross-curricular educators participated in determining the reliability of the Bauer & Kohut Argumentative Writing Rubric as an evaluation tool for student argumentative writing. Each educator was asked to rate 28 student samples of argumentative writing between 1 (low achievement) to 5 (advanced achievement) using a traditional rubric, the ACT Writing Test Scoring Rubric, in addition to the Bauer & Kohut Argumentative Writing Rubric. Interrater reliability was established among the educators for the use of both rubrics. To determine initial interrater reliability, instructors coached the educators in the use of each rubric. They then asked each educator to rate the one sample between 1 to 5 using the ACT Writing Test Scoring Rubric, and to rate a separate sample between 1 to 5 according to the Bauer & Kohut Argumentative Writing Rubric. Investigators calculated the number of adjacent ratings, then totaled the number of matched ratings, and converted the fraction to a percentage. After initial interrater reliability was established, the teachers rated the 28 student samples of argumentative writing. Data was analyzed to determine the interrater reliability of each rubric using an ANOVA test. The researchers for this study also determined parallel forms reliability comparing the results of each rubric using a paired t-test to open the discussion of the application of the Bauer & Kohut Argumentative Writing Rubric as a consistent tool for argumentative

writing assessment across subjects. Teacher survey questions were also taken into consideration for this portion of the research.

Participant samples were collected in file folders and analyzed over the course of the study. Assessment scores for each writing sample were entered by hand on a spreadsheet which was later entered into an electronic spreadsheet on which participants were identified by number. Ratings of student work were collected from educators in charts and entered into an electronic spreadsheet. Survey questions from teachers were written and submitted on paper. All data was reported anonymously.

Data Analysis

Results for Rubric as a Tool to Guide to Instruction

An independent-samples t-test was conducted to determine the homogeneity between the control group and the experimental group. There was not a statistically significant difference between the scores of the control group (M=1.64, SD=0.64) and the experimental group (M=1.77, SD=0.61) conditions (Table 1); t (45)=0.7255, p=0.4719. The results indicate that the control group and experimental group are homogeneous in initial literacy ability.

Table 1. *t*- test Results Assessing Control and Experimental Group Homogeneity

Group	n	Mean	t	df	sig 2-tailed	Std. error difference
Control	25	1.64	0.7255	45	0.4719	0.183
Experimental	22	1.77				

An independent-samples t-test was conducted to compare the difference in level of improvement from pre to post between the control group and experimental group. There was an extremely statistically significant difference between the scores of the control group (M=0.85, SD=0.73) and the experimental group (M=2.48, SD=1.12) conditions (Table 2); t (45)=5.9969, p < 0.0001. The average increase in score of the control group was 0.85 and the average increase in score of the experimental group was 2.48. The results indicate that while both methods of instruction are effective in achieving student growth, using a process-oriented rubric to guide planning and instruction has an increased effect on literacy achievement.

Table 2. *t*- test Results Comparing Control and Experimental Group Increases

Group	n	Mean	t	df	sig 2-tailed	Std. error difference
Control	26	0.85	5.9969	45	<0.0001	0.272
Experimental	21	2.48				

Discussion on Rubric as a Tool to Guide to Instruction

The present study predicted that using a linear, focused rubric to guide literacy planning, instruction, and assessment would result in increased levels of growth on student writing samples. This hypothesis was supported in the study as the students who were presented with the experimental, or rubric method of planning and instruction achieved higher rates of skill improvement than students presented with the control, or traditional method of planning and instruction. The control group which was taught according to a traditional pacing guide and assessed using a holistic method increased their scores by 49% while the experimental increased their scores by 130%.

This result is consistent with the literature which says pacing guides, which are traditionally used to plan instruction, are not as effective as using assessment to plan instruction (Peverini, 2009). The result of the current study also confirms the importance of assessment type as the literature suggests that traditional, holistic rubrics contain too many factors to realistically and effectively guide instruction (Bambrick-Santoyo, 2007).

There were various possible factors influencing the significantly greater results of the experimental group and future research may attempt to determine the level of effect of these variables. Possible interactions include the process-oriented planning method, different instructional techniques, specific strategies utilized at each level, the addition of revision in the

writing process, continuing to use the same piece of writing for revision at each level, and the instructional guidance of a linear rubric.

Results for Initial Interrater Reliability

In order to determine reliability of scoring procedure between participants before the rating of study samples, two initial pieces of argumentative writing were scored by each of the five initial participants; one sample scored with each rubric. Reliability scoring proceeded independently without participant knowledge of which rubric was created for the current study and without the reliability observer present. Thus, reliability estimates are conservative compared to real-time observations when two observers code simultaneously as did the researchers for this study. While four of the participants were able to establish common ratings between their applications of each rubric, one participant felt the application of both rubrics wasn't manageable. This participant reported a strong disciplinary difference in writing assessment between the rubrics in the study and those used in her own practice, and therefore didn't continue the study. For the remaining four reporting participants, the average kappa was 1, with all participants rating samples within one point when compared to all other raters for each writing sample. There were no indications of systematic observer differences. These results indicate that interrater reliability is consistent between raters for each rubric condition as there was 100% agreement.

Results for Rubric as a Reliable Cross-Curricular Evaluation Tool

A one-way between subjects ANOVA was calculated to compare the effect of rubric usage on grading consistency between raters in both traditional holistic rubric and focused linear

rubric conditions. There was a significant difference at the p<.05 level (Figure 1) in grading consistency between raters using the traditional rubric [F(3, 108) = 0.75, p = 0.525]. There was no significant difference at the p<.05 level (Figure 2) in grading consistency between raters using the experimental rubric [F(3,107) = 1.56, p= 0.204]. These results indicate a higher level of consistency between raters using the experimental Bauer & Kohut Argumentative Writing Rubric when compared to the consistency between raters using the traditional writing rubric.

Figure 7: Interval Plot Comparing Traditional Rubric Ratings

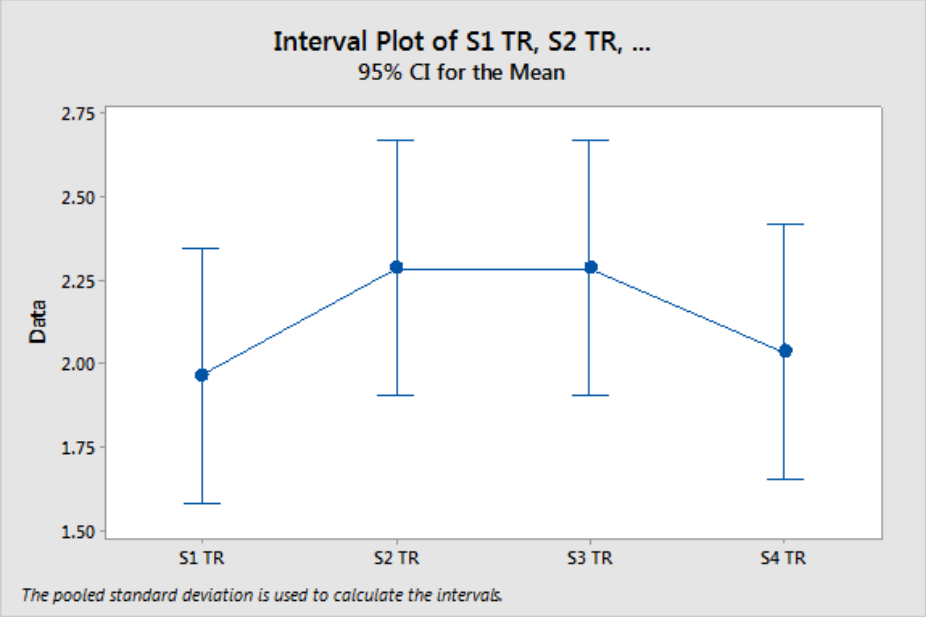
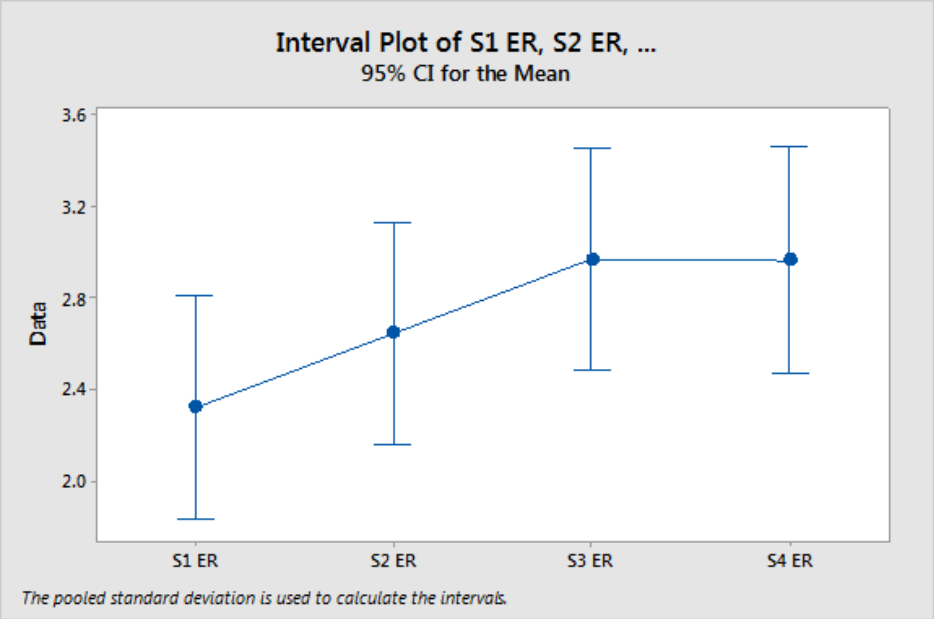


Figure 8: Interval Plot Comparing Experimental Rubric Ratings



Discussion on Rubric as a Reliable Cross-Curricular Evaluation Tool

The present study predicted that using a focused, linear rubric to guide instruction would result in increased consistency of ratings between raters. This hypothesis was supported in the study as the ratings of writing samples were less consistent between raters when using the traditional rubric and more consistent between raters when using the experimental rubric. The significance threshold was set at .05 and the data concludes that the results of the traditional rubric rating variance may have been due to chance and the results of the experimental rubric rating consistency were not due to chance, but to the use of the Bauer & Kohut Argumentative Writing Rubric. These results indicate that the Bauer & Kohut Argumentative Writing Rubric is a consistent and reliable tool for use among cross-curricular teachers.

Parallel Forms Reliability

A paired-samples t-test was conducted to compare the traditional rubric and experimental rubric ratings for each writing sample (Table 2). There was an extremely statistically significant

difference between the traditional rubric rating of a writing sample (M=2.14, SD=1.02) and the experimental rubric rating of that same sample (M=2.74, SD=1.31) for each writing sample; t(111)=8.03, p = <0.0001. These results indicate that raters scored a writing sample differently for each rubric used as a scoring guide. They also indicate that the average scores of each writing sample using the traditional rubric were lower than the average scores of the matched sample using the experimental writing rubric.

Table 3. t-test Results Comparing Control and Experimental Rubric Ratings

Group	n	Mean	t	df	sig 2-tailed	Std. error difference
Control	112	2.14	8.0344	111	<0.0001	0.074
Experimental	21	2.74				

Discussion on Parallel Forms Reliability

The present study predicted that using a focused, linear rubric to assess a writing sample would result in different scores than the rating of the same writing sample scored using a traditional, holistic rubric. This hypothesis was supported in the study as the raters, when using the traditional rubric for scoring, rated the writing at a lower level than they rated the same piece of writing using a focused, linear rubric for scoring. The difference between the mean scores for the traditional rubric and experimental rubric was .70, and raters using the control rubric scored writing samples lower on average than raters using the experimental rubric.

Raters scored student samples lower and less reliably using traditional assessment methods. This finding supports the research saying that teachers may be limited in their assessment of the writing samples due to an overemphasis on conventions and other holistic grading measures (Huot, 2002b). Focusing on conventions has a negative effect and is unlikely to help improve the quality of students' writing (Graham, MacArthur, & Fitzgerald, 2013). The

current study suggests that this type of grading which looks for what is missing rather than what is present tends to result in lower scores.

Raters scored student samples higher and more reliably using the experimental rubric. This suggests that teachers were able to look at the student writing as a more complete entity, and evaluate the ideas in the samples through the process-oriented criteria established in the experimental rubric. A process-oriented approach maintains high expectations, because the potential for the piece of writing to improve is not compromised by surface-level writing errors. Process-oriented instruction focuses on strengths, identifying "what is present" rather than "what is not present" (Swain & Mahieu, 2012).

Results for Qualitative Survey

Following participation in rating using both rubrics, each participant was asked the same set of questions about their experience. While the participants were not informed about the sources of each rubric so as to not create bias in the results of their ratings, the source of each rubric was revealed before the survey questions were completed. For the purposes of the survey, the ACT Writing Test Scoring Rubric was labelled "Rubric #1," and the Bauer & Kohut Argumentative Writing Rubric was labelled "Rubric #2." Rater 1 is represented by a Hispanic English language arts teacher with 13 years of experience, Rater 2 is represented by a European-American math teacher with 22 years of experience, Rater 3 is represented by a European-American health and wellness education teacher with 32 years of experience, and Rater 4 is represented by a European-American first year social studies teacher.

Survey Question 1: Describe your experience evaluating student work with Rubric #2:

- Rater 1 Answer: "It felt 'easier'... less to focus on."
- Rater 2 Answer: "Overall, I found it (mostly) easy to understand, and I referred to the "look fors" several times to help me.
- Rater 3 Answer: n/a
- Rater 4 Answer: "Rubric #2 was a lot easier to use. The use of Bloom's Taxonomy was great- it let me as the evaluator to see the key elements of writing that I am grading."

Survey Question 2: Describe any advantages of using this tool to assess argumentative writing (Rubric #2):

- Rater 1 Answer: "This rubric really puts the emphasis on key components of argumentative essays."
- Rater 2 Answer: "Very easy to understand! This went much faster than Rubric #1. It seemed to scale up from 1 to 6 as to what was expected."
- Rater 3 Answer: "It's simpler than the other tool, but in some cases, 'breaks' weren't apparent."
- Rater 4 Answer: "Clear and concise"

Survey Question 3: Describe any disadvantages of using this tool to assess argumentative writing (Rubric #2):

- Rater 1 Answer: "Complete lack of grammar."
- Rater 2 Answer: "I had a hard time with some of the student samples that fit more than one level of the rubric."

- Rater 3 Answer: "There should be additional levels for students who are in between skills."
- Rater 4 Answer: "This rubric doesn't provide a skill focus on grammar (like Rubric #1), although one can argue that grammar falls under organization."

Survey Question 4: How does using this tool compare to traditional assessment strategies (Rubric #1):

- Rater 1 Answer: n/a
- Rater 2 Answer: "I found Rubric #2 much easier to use and much clearer. I had to refer to Rubric 1 multiple times while looking at each sample."
- Rater 3 Answer: "It's more complex, but it may be more precise when it comes to distinguishing one level from another."
- Rater 4 Answer: "Instead of looking at every little thing, the instructor can grade (argumentative) writing as a whole."

Survey Question 5: Do you feel you graded the work differently between Rubric #1 and Rubric #2? Is so, explain:

- Rater 1 Answer: "I felt that Rubric #2 afforded more leniency in grading."
- Rater 2 Answer: "Yes. I think more samples were scored higher on Rubric 2, but I don't know if I can say why."
- Rater 3 Answer: "Rubric #2 was easier and faster to use than Rubric #1."
- Rater 4 Answer: "Yes. Based on the terms from Bloom's Taxonomy, I graded the work based on whether or not the student met that standard/term."

Survey Question 6: Overall opinion/ concluding thoughts:

- Rater 1 Answer: "I would use rubric #1 for an assignment that students had significant time to complete. Rubric #2 seems very handy for on demand writing."
- Rater 2 Answer: n/a
- Rater 3 Answer: "We need to help students present other people's points of view and 'cite' them properly. Students, in some cases, are very passionate about one perspective, but we need to give them some guides to see all sides of an issue, so they become informed in their decision-making."
- Rater 4 Answer: "I much preferred Rubric #2."

Discussion on Qualitative Survey

A majority of the raters reported a preference for the Bauer & Kohut Argumentative Writing Rubric. While it was noted that a potential disadvantage of the rubric was its lack of emphasis on mechanics like grammar, the raters overwhelmingly found the Bauer & Kohut Argumentative Writing Rubric easy and accessible. Raters noted that the rubric was clear and easy to understand. Raters also noticed a difference in their approaches to holistic assessment between the emphases of each rubric. These findings support the research because the raters' preference for the experimental rubric aligns with its cross-curricular application. Given the highly technical aspects and discourse of traditional writing assessments, many teachers feel discouraged by traditional writing assessment (Huot & O'Neill, 2009). By eliminating the emphasis on convention, cross-curricular educators felt that the experimental rubric was the more accessible tool in this study for grading student writing.

Summary

In summary, Chapter 3 has presented the findings and results of the data analysis of the research for this study. The data collected tested the consistency of the Bauer & Kohut Argumentative Writing Rubric as a tool to guide instruction. The data collected also tested the reliability of the Bauer & Kohut Argumentative Writing Rubric as a means to evaluate argumentative writing at a cross-curricular level. The survey questions addressed the perspectives of cross-curricular educators in terms of their experiences using the Bauer & Kohut Argumentative Writing Rubric compared to a traditional assessment rubric.

CHAPTER 4
FINDINGS

With increasingly complex and shifting literacy expectations, assessment practices need to be implemented that mirror these expectations in a reliable manner. The function of a rubric is broken down into two components: The ability to consistently assess work, and the ability to consistently inform student learning (Fuchs & Fuchs, 2006; Mitchell & Neill, 1992). Educators and students alike need access to assessment tools that meet the demands of elevated complex literacy expectations, and that reliably assess work and guide instruction according to these demands across curricular context. The purpose of this chapter is to discuss the objectives of the study and how the findings reflect the intention of the research.

Conclusions Based on Research Findings

Reliability as a Tool to Guide Instruction

An extreme statistical significance in literacy improvement was established between the control group using traditional instruction, and the experimental group guided by the Bauer & Kohut Argumentative Writing Rubric. While both groups made gains in literacy ability, the experimental group exhibited a much greater increase in growth by the end of the study. This research suggests that utilizing assessment that details process-oriented goals and scaffolds strategies can greatly impact the achievement of learners. Based on the findings of this study, educators should have access to assessment tools that provide the opportunity to guide and inform instruction based on student readiness.

Reliability as a Cross-Curricular Tool to Evaluate Argumentative Writing

The Bauer & Kohut Argumentative Writing Rubric proved not only to be a reliable evaluation tool, but a more reliable tool than the traditional method of assessment used in this study. Teachers were able to more reliably rate student samples using the Bauer & Kohut Argumentative Writing Rubric. Teachers reported that they found this rubric easier to use than the traditional rubric. They also noted that they felt the expectations on the Bauer & Kohut Argumentative Writing Rubric were clear and easier to understand than the traditional rubric. While the participant who taught English language arts noticed the lack of grammar and convention on the rubric as a potential weakness, the other participants appreciated the simplicity of learning goals represented by each level of the rubric. This indicates that a process-oriented assessment tool is more accessible to educators in a cross-curricular context, and therefore warrants more reliable results in evaluating work.

A gap between how the teachers rated the samples per each rubric was also evident. There was a statistically significant difference in ratings between the two rubrics. Teachers tended to give lower ratings when using the traditional rubric, and higher ratings when using the Bauer & Kohut Argumentative Writing Rubric. Some teachers speculated on their survey questions that this change in grading could have been due to the lack of emphasis on conventions, and the opportunity to view the writing holistically per its components. This indicates that the Bauer & Kohut Argumentative Writing Rubric shifted the manner in which teachers rated the student samples based on the holistic emphasis of each rubric.

This research suggests that a process-oriented assessment tool is more accessible to cross-curricular educators as a tool to evaluate argumentative writing. This research also suggests that such a tool should be made available to educators to evaluate argumentative writing

reliably in the cross-curricular context. Based on the findings of this study, using a holistic, process-oriented assessment tool is more accessible and more reliable as a means to evaluate argumentative writing across the curriculum.

Limitations of the Study

Participation in the study was voluntary. Because of the voluntary basis of participation, students may not have remained consistent throughout the duration of the study. Lack of motivation or absence during the study may have impacted student results.

Additionally, the use of a rubric in general could impact the ability to evaluation student work on behalf of the investigators. Since rubrics tend to describe general and synthesized criteria, unique characteristics in student writing may not be noticed or accounted for (Delandshere & Petrosky, 1998; Haswell & Wyche-Smith, 1994). Students may make improvements, but if they are not specified by the Bauer & Kohut Argumentative Writing Rubric, they may not be scored or considered as improvement.

The study is limited to a sample size of 47 pupils of one teacher in one urban school district. This number is unable to produce a wide variant of results. Additionally, only one English Language Arts class was used as an experimental group to test the consistency of the Bauer & Kohut Argumentative Writing Rubric as a guide to inform instruction. Since the rubric was created with the intention of evaluating and guiding instruction in a cross-curricular context, cross-curricular teachers and classes should have the opportunity to contribute data about its reliability in various settings.

While the content of the rubric aligns itself to relevant educational constructs (Bloom's Taxonomy and the Common Core State Standards), the rubric has not been tested for its validity.

Gay (1987) notes that determining reliability is a prerequisite for determining validity. Validity should be established in the future to assure the utmost effective use of the rubric.

Recommendations for Future Research

On the basis of the study, the following recommendations for future research are encouraged:

Future research should aim to establish the validity of the Bauer & Kohut Argumentative Writing Rubric and its alignment to the Common Core State Standards and Bloom's Taxonomy. Future research should also measure the reliability of the rubric to guide instruction across curriculum and context. Secondary education holds a high emphasis on argumentative writing in the cross-curricular context. This research highlighted the achievement of the Bauer & Kohut Argumentative Writing Rubric in one English Language Arts class. Other classes must be represented in this work to fortify the applicability of the rubric in the cross-curricular setting.

The duration of the study was a two-week period. Generally, the more time students have to practice skills, the more effective the intervention can be. However, Sharif, Ozuah, Dinkevich, and Mulvihill (2003) express concern that there is very little published data about the efficacy of brief literacy intervention in modern literacy research. Because this brand of research is so new, there are not many studies that provide data about the overall effectiveness or ineffectiveness of the duration of time spent on brief literacy intervention. While some researchers (Nugent, 2010; Venville & Dawson, 2010; Brooks, 2007; Vaughn et. al., 2000) have had significant gains with brief literacy intervention, Singleton (2009) and Truch (2003) warn educators that the rate of gain may decelerate after the first 12 hours of intervention. To continue testing the effectiveness and consistency of the Bauer & Kohut Argumentative Writing Rubric as a guide to instruction,

the research could be adapted in two ways. First, the study may be conducted with the addition of one or more data samples from the control group and experimental group collected beyond the time period of guided brief literacy instruction in order to gauge skill retention of short term guided intervention. Second, the guided brief literacy instruction could extend beyond a two-week time period to test the consistency of the rubric as a guide for instruction over long periods of time.

References

ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.

ACT, Inc. (2015). *The ACT writing test scoring rubric.* Retrieved from http://www.act.org/aap/pdf/Writing-Test-Scoring-Rubric.pdf.

Allen, L., & Turville, J. (2010). *Differentiating by Readiness: Strategies and Lesson Plans for Tiered Instruction, Grades K-8.* New York, NY: Taylor & Francis.

Alvermann, D.E., & Moore, D.W. (1991). Secondary school reading. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 951-983). White Plains, NY: Longman.

Anderson, L. W., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Andrade, H.G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53,* 27-30.

Applebee, A.N. (1974). *Tradition and reform in the teaching of English: A history.* Urbana, IL: National Council of Teachers of English.

Applegate, A.J., Applegate, M.D., McGeehan, C.M., Pinto, C.M., & Kong, A. (2009). The assessment of thoughtful literacy in NAEP: Why the states aren't measuring up. *The Reading Teacher, 62(5),* 372-381. doi:10.1598/RT.62.5.1

Arter, J., & McTighe, J. (2001). *Scoring Rubrics in the Classroom*. Thousand Oaks, CA: Corwin Press.

Bacha, N.N. (2012). Disciplinary writing in an EFL context from teachers' and students' perspectives. *International Journal of Business and Social Sicence, 3(2),* 233-256.

Ball, A. F. (1997). Expanding the dialogue on culture as a critical component when assessing writing. *Assessing Writing, 4(2),* 169-202.

Ball, A.F. (1996). Expository writing patterns of African American students. *The National Council of Teachers of English,* 27-36.

Ball, A. F. (1991). *Organizational patterns in the oral and written expository writing of African-American Adolescents.* Unpublished doctoral dissertation, Stanford University.

Bambrick-Santoyo, P. (2007). Data in the driver's seat. *Educational Leadership, 65*(4), 43-46.

Barritt, L., Stock, P.L., & Clark, F. (1986). Researching practice: Evaluationg assessment essays. *College Composition and Communication, 38,* 315-327.

Bell, P. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education, 22(8),* 797-817. Doi:10.1080/095006900412284

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, and Practice, 18(1),* 5-25.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8(2-3),* 70-91.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Erlbaum.

Berkowitz, D., Wolkowitz, B., Fitch, R., & Kopriva, R. (2000). *The use of tests as part of high-stakes decision-making for students: A resource guide for educators and policy-makers.* Washington, DC: U.S. Department of Education.

Beyer, B. (1979). Writing to learn in social studies: Pre-writing and rewriting to learn. *Social Education, 44,* 187-189.

Birda, S., & Chait, R. (2011). Increasing teacher diversity: Strategies to improve the teacher workforce. *Center for American Progress.* Retrieved from https://cdn.americanprogress.org/wp-content/uploads/issues/2011/11/pdf/chait_diversity.pdf

Black, P., & William, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29(1),* 7-74.

Bloom, B. S. (1976). *Human Characteristics and school learning.* New York, NY: McGraw-Hill.

Brookhart, S.M. (1999). The art and science of classroom assessment: The missing part of pedagogy. *ASHE-ERIC Higher Education Report, 21(1).*

Brooks, G. (2007). *What works for children with literacy difficulties? The effectiveness of intervention scheme*. London: The Department for Children, Schools, and Families.

Bruffee, K.A. (1986). Social construction, language, and the authority of knowledge: A bibliographical essay. *College English, 48,* 773-790.

Calkins, L., Ehrenworth, M., Lehman, C. (2012). *Pathwasy to the common core: Accelerating achievement*. Portsmouth, NH: Heinemann.

Carter, M. (1990). The idea of expertise: An exploration of cognitive and social dimensions of writing. *College Composition and Communication, 41,* 265-286.

Chaffee, J., McMahon, C., & Stout, B. (1999). *Critical thinking, thoughtful writing: A rhetoric with readings.* Boston, NY: Houghton Mifflin Company.

Chiu, M. M., & Chow, B. W. (2015). Classmate characteristics and student achievement in 33 countries: Classmates' past achievement, family socioeconomic status, educational resources, and attitudes toward reading. *Journal of Educational Psychology, 107(1),* 152-169.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English Language Arts.* Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

College Board (2015). *SAT: Essay scoring.* Retrieved from https://sat.collegeboard.org/scores/sat-essay-scoring-guide.

Crebert, et al., (2011). *Critical evaluation skills toolkit*. Retrieved from http://www.griffith.edu.au/gihe/resources-support/graduate-attributes

Culham, R. (2003). *6+1 Traits of Writing.* New York, NY: Scholastic Professional Books.

Cushner, K., McClelland, A., & Safford, P. (2009). *Human diversity in education: An integrative approach.* Boston, MA: McGraw Hill Higher Education.

Daniels, H., & Steineke, N. (2011). *Texts and lessons for content-area reading.* Portsmouth, NH: Heinemann Publishers.

Davis, B., & Sumara, D. (2008). Complexity as a theory of education. *Transnational Curriculum Inquiry, 5(2),* 33-44.

Davis, N. (2016). Grade to Lexile conversion table. *Newsela.* Retrieved from http://support.newsela.com/hc/en-us/articles/201218865-Grade-to-Lexile-Conversion-Table

De La Paz, S., & Sherman, C. K. (2013). Revising strategy instruction in inclusive settings: Effects for English learners and novice writers. *Learning Disabilities Research & Practice, 28(3),* 129-141.

De Leeuw, J. (2016). *Leadership of Assessment, Inclusion, and Learning.* Switzerland: Springer International Publishing.

Delandshere, G., & Petrosky, A. (1998). Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher, 27(2),* 14-25.

Deford, D. E. (1981). Literacy: Reading, writing, and other essentials. *Language Arts, 58(6),* 652-658.

Delpit, L. (1986). Skills and other dilemmas of a progressive Black educator. *Harvard Educational Review*, 56(4), 379-385.

Desherbinin, J. (2004). White professors can help uproot racism. *The Chronical of Higher Education*. Retrieved from http://chronicle.com/article/White-Professors-Can-Help/23743/

Dirlam, David. & Byrne, Maureen. (1978). *Standardized Developmental Ratings*. [Washington, D.C.]: Distributed by ERIC Clearinghouse

Dirlam, D. K. (1980). Classifiers and cognitive development. In S. & C. Modgil (Eds.), *Toward a Theory of Psychological Development*. Windsor, England: NFER Publishing, 465-498

Dolgin, J., Kelly, K., & Sarvenaz, Z. (2010). *Authentic Assessment for the English classroom*. Urbana, IL: National Council of Teachers of English.

Donahue, P., Daane, M., & Grigg, W. (2003). *The nation's report card: Reading highlights 2003* (No. NCES 2004452). Washington DC: National Center for Education Statistics.

EMCC General Education Abilities Matrix. Retrieved from http://www.estrellamountain.edu/sites/default/files/docs/saac/emcc-gen-ed-abilities-march2011.pdf

Ellis, L. M. (2015). A critique of the ubiquity of the Toulmin model in argumentative writing instruction in the U.S.A. In Eemeren, F. H., & Garssen, B. (Eds.), *Scrutinizing Argumentation in Practice* (pp. 201-214). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Faigley, L. (1986). Competing theories of process: A critique and a proposal. *College English, 48*, 527-542.

Ferretti, R.P., Lewis, W.E., & Andrews-Wekcerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101,* 577-589.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35,* 39-50.

Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L.W. Gregg & E.R. Steinberg (Eds.), *Cognitive process in writing* (pp. 31-50). Hillsdale, NJ: Erlbaum.

Freedman, S. W., Dyson, A. H., Flower, L. & Chafe, W. (1987). *Research in writing: Past, present, and future* (Tech. Rep. No. 1). Berkeley, CA: University of California, Center for the Study of Writing.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: Why, why, and how valid is it? *Reading Research Quarterly, 41(1),* 93-99.

Gardner, T. (2008). Designing writing assignments. Urbana, IL: National Council of Teachers of English.

Garrison, C., & Ehringhaus, M. (2013). *Formative and summative assessments in the classroom*. Dover, NH: Measured Progress. Retrieved from https://www.amle.org/portals/0/pdf/articles/Formative_Assessment_Article_Aug2013.pdf

Gay, L.R. (1987). Selection of measurement instruments. In *Educational Research: Competencies for Analysis and Application* (3rd ed.). New York, NY: Macmillan.

Gerald, D., & Haycock, K. (2006). *Engines of inequality: Diminishing equity in the nation's premier public universities*. Washington, DC: Education Trust.

Glaus, M. (2013). Text complexity and young adult literature: establishing its place. *Journal of Adolescent & Adult Literacy, 57(5),* 407-416.

Glimps, B.J., Ford, T.N. (2010). White power and privilege: Barriers to culturally responsive teaching. *International Journal of Educational Policies, 4(1),* 39-52.

Goldhaber, D., Theobald, R., & Tien, C. (2015). *The Theoretical and Empirical Arguments for Diversifying the Teacher Workforce: A Review of the Evidence*. Center for Education Data & Research: University of Washington Bothell. Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202015-9.pdf

Gomez, M. L. (1993). Prospective teachers' perspectives on teaching diverse children: A review with implications for teacher education and practice. *Journal of Negro Education,* 62(4), 459-474.

Graham, S., & Hebert, M., (2011). Writing to read: a meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710-744.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading*. Washington, DC: Alliance for Excellent Education.

Graham, S., Herbert, M., & Harris K. R., (2015). Formative assessment and writing: a meta-analysis. *The Elementary School Journal, 115(4),* 523-547.

Graham, S., MacArthur, C.A., & Fitzgerald, J. (2013). *Best practices in writing instruction* (2nd ed.). New York City, NY: Guilford Press.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R., (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology 104(4),* 879-896.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445-476.

Haswell, R., & Wyche-Smith, S. (1994). Adventuring into writing assessment. *College Composition and Communication, 45,* 220.236.

Huitt, W. (2011). Bloom et al.'s taxonomy of the cognitive domain. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved [date], from http://www.edpsycinteractive.org/topics/cognition/bloom.html

Huot, B. (2002a). *(Re)articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.

Huot, B. (2002b). Toward a new discourse of assessment for the college writing classroom. *College English, 65(2),* 163-179.

Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication, 47(4),* 549-566.

Huot, B., & O'Neill, P. (2009). *Assessing writing: A critical sourcebook*. Boston, NY: Bedford/ St. Martin's.

Johnston, P. (1987). Assessing the process and the process of assessment in the language arts. In J. Squire (Ed.), *The dynamics of language learning: Research in reading and English* (pp. 335-357). Urbana, IL: National Council of Teachers of English/ ERIC.

Langer, J. A., & Flihan, S. (2002). Writing and reading relationships: Constructive tasks. *Writing: Research/ Theory/ Practice*, 1.

Lanning, L. A. (2013). *Designing a concept-based curriculum for English language arts.* Thousand Oaks, CA: Corwin Press.

Lloyd-Jones, R. (1977). Primary trait scoring. In Cooper, C.R. & Odell, L. (Eds.), *Writing: Describing, measuring, and judging*. Urbana, IL: NCTE.

Kavaliauskiene, G. (2004). Quality assessment in teaching English for specific purposes. *ESP World*. Retrieved from http://esp-world.info/Articles

Kucer, S.B. (2001). *Dimensions of literacy: A conceptual base of teaching reading and writing in school settings*. Mahwah, NJ: Erlbaum.

Macrorie, K. (1976). *Telling writing*. Rochelle Park, NJ: Hayden.

Magner, T., Soulé, H., & Wesolowski, K. (2011). *P21 Common core toolkit: A guide to aligning the common core state standards with the framework for 21st century skills.* Washington, DC: Partnership for 21st Century Skills. Retrieved from http://www.p21.org/storage/documents/P21CommonCoreToolkit.pdf

Marzano, R.J. (2012). Art & science of teaching. *Educational Leadership*, 82-83.

McManus, S. (2008). *Attributes of effective formative assessment*. Washington, DC: Council for Chief State School Officers. Retrieved from http://www.ccsso.org/publications/details.cfm?PublicationID=362

McMillan, J. H. (1996). *Educational research fundamentals for the consumer (2nd ed.).* New York, NY: Harpercollins Publishers, Inc.

McTighe, J., & Ferrara, S. (1998). Assessing learning in the classroom. *National Education Association,* 1-37.

Menzies, H.M., Lane, K.L., & Lee, J.M. (2009). Self-monitoring strategies for use in the classroom: A promising practice to support productive behavior for students with emotional or behavioral disorders. *Beyond Behavior,* 27-35.

Methe, S.A., & Riley-Tillman, T.C. (2008). An informed approach to selecting and designing early mathematics interventions. *School Psychology Forum: Research into Practice, 2,* 29-41.

Mitchell, R., & Neill, M. (1992). *Criteria for evaluation of student assessment systems.* Washington, DC: National Forum on Assessment.

Moffett, J. (1968). *Teaching the universe of discourse*. Boston: Houghton Mifflin.

Moje, E. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy 52* (2), 96-107.

Moore, D.W. (1996). Contexts for literacy in secondary schools. In D.J. Leu, C.K. Kinzer, & K.A. Hinchman (Eds.), *Literacies for the 21st century: Research and practice* (45th yearbook of the National Reading Conference, pp. 15-46). Chicago: National Reading Conference.

Moskal, B.M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation, 7(3).* Retrieved from http://PAREonline.net/getvn.asp?v=7&n=3

National Assessment Governing Board (2011). *Writing framework for the 2011 National Assessment of Educational Progress, pre-publication edition.* Iowa City, IA: ACT, Inc.

National Assessment Governing Board (2009). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (2012). *The Nation's Report Card: Writing 2011* (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors. Retrieved from http://www.corestandards.org

National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Press Academy.

Newell, J.A., Dahm, K.D., & Newell, H.L. (2002). Rubric development and interrater reliability issues in assessing learning outcomes. *Chemical Engineering Education, 36,* 212-215.

Nieto, S. (1999). What does it mean to affirm diversity in our nation's schools? *The School Administrator, 56(5),* 32-34.

Nugent, M. (2010). Teaching our traveler children to read; An action research project. Support for Learning, 25(2), 55-62.

O'Neill, P. (2003). Moving beyond holistic scoring through validity inquiry. The *Journal of Writing Assessment, 1(1),* 47-65.

O'Neill, P. (2011). Reframing reliability for writing assessment. *The Journal of Writing Assessment, 4(1).* Retrieved from http://journalofwritingassessment.org/article.php?article=54

Pardo, L. S. (2004). What every teacher needs to know about comprehension. *International Reading Association,* 272-280.

Pearson, D.P. (2002). Thinking about the reading and writing connection. *The Voice, 7(2),* 8-9.

Peat, B. (2006). Integrating writing and research skills: Development and testing of a rubric to measure student outcomes. *Journal of Public Affairs Education, 12,* 295-311.

Peverini, S. (2009). The value of teacher expertise and good judgement: Recent inspiring reading about assessment. *Language Arts, 86(5),* 398-402.

Porter, A.C., Kirst, M.W., Osthoff, E.J., Smithson, J.S., & Schneider, S.A. (1993). *Reform up close: A classroom analysis*. Draft final report to the National Science Foundation on Grant No. SPA-8953446 to the Consortium for Policy Research in Education. Wisconsin Center for Education Research, University of Wisconsin-Madison.

Popham, W.J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Raimes, A. (1983). *Techniques in teaching writing*. Oxford: Oxford University Press.

RAND Reading Study Group (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND.

Rezaei, R.A., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing, 15(1),* 18-39.

Risko, V. J., & Walker-Dalhouse, D. (2010). Making the most of assessments to inform instruction. *The Reading Teacher, 63(5),* 420-422.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation, 11(10),* 1-13.

Rudner, L., & Schafer, W. (2002). *What teachers need to know about assessment*. Washington, DC: National Education Association.

Scriven, M. (1967). The methodology of evaluation. In T.W. Tyler, R.M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally.

Shrock, K. (2000). *Kathy Shrock's Guide for Educators.* Retrieved from http://www.mines.edu/Academic/assess/Resource.htm

Singleton, C. (2009). *Intervention for dyslexia. A review of published evidence on the impact of specialist dyslexia teaching*. Downloaded from http://www.thedyslexia-spldtrust.org.uk/medial/downloads/articles/13-intervention_for_dyslxia_research_report.pdf

Siskin, L.S. (1994). *Realms of knowledge: Academic departments in secondary schools.* London: Falmer.

Smith, F. (1982). *Writing and the writer*. New York: Holt, Rinehart & Winston.

Soares, J.A. (2007). *The power of privilege: Yale and America's elite colleges.* Stanford, CA: Stanford University Press.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication, 31,* 378-388.

Song, Y., & Ferretti, R. P. (2012). Teaching critical questions about argumentation through the revising process: effects of strategy instruction on college students' argumentative essay. *Springer Science=Business Media B.V.,* 67-90.

Sousa, D. (1995). *How the brain learns.* Reston, VA: National Association of Secondary School Principals.

Spandel, V. (2006). Speaking my mind: In defense of rubrics. *The English Journal, 96(1),* 19-22.

Stallard, C. (1974). An analysis of the writing behavior of good student writers. *Research in the Teaching of English, 8,* 206-218.

Stellmack, M.A., Konheim-Kalkstein, Y.L., Manor, J.E., Massey, A.R., & Schmitz, J.P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology, 36,* 102-107.

Swain, S., & LeMahieu, P.G. (2012). Assessment in a culture of inquiry: The story of the National Writing Project's analytic writing continuum. In N. Elliot & L. Perelman (Eds.), *Writing Assessment in the 21st Century: Essays in Honor of Edward M. White*. New York: Hampton Press.

Thompson, E. H. (1985). Self-assessment and the mastery of writing. In J. Beard & S. McNaff (Eds.), *Testing in the English language arts: Uses and abuses.* Rochester, MI: Michigan Council of Teachers of English.

Tileston, D.W. (2000). *10 best teaching practices: How research, learning styles, and standards define teaching competencies.* Thousand Oaks, CA: Corwin Press, Inc.

Tomlinson, C.A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Alexandra, VA: Association for Supervision and Curriculum Development.

Truch, S. (2003). Comparing remedial outcomes using LIPS and Phono-Graphix: An in-depth look from a clinical perspective. Unpublished manuscript. The Reading Foundation, Calgary, Alberta, Canada.

Valencia, S., McGinley, W., & Pearson, P.D. (1990). Assessing literacy in the middle school. In G. Duffy (Ed.), *Reading in the middle school* (2nd ed., pp. 124-141). Newark, DE: International Reading Association.

Vaughn, S., Gerten, R. & Chard, D.J. (2000). The underlying message in learning disabilities intervention research: Findings from research synthesis. *Exceptional Children, 67(1),* 99-114.

Venville, G. J., & Dawson, V. M. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *The Journal of Research and Science Teaching, 47(8),* 952-977.

White, E. (1994). Issues and problems in writing assessment. *Assessing Writing, 1,* 11-27.

Whitney, J. (2005). Five easy pieces: Steps toward integrating AAVE into the classroom. *The National Council of Teachers of English, 94(5),* 64-69.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning.* New York: Erlbaum.

Wilson, A. A. (2011). A social semiotics framework for conceptualizing content area literacies. *Journal of Adolescent & Adult Literacy, 54*, 435-444. doi:10.1598/JAAL.54.6.5

Wisconsin Department of Public Instruction (2011). *Common Core State Standards for Literacy in All Subjects.* Retrieved from http://standards.dpi.wi.gov/sites/default/files/imce/cal/pdf/las-stds.pdf

Wixson, K. K., & Valencia, S. W. (2014). CCSS-ELA: Suggestions and cautions for addressing text complexity. *The Reading Teacher, 67(6),* 430-434.

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching, 7(1),* 3-14.

Yore, L.D. (1991). Secondary science teachers' attitudes toward and beliefs about science reading and science textbooks. *Journal of Research in Science Teaching, 28(1),* 55-72. Doi:10.1002/tea.3660280106

Appendix A

Bauer & Kohut Argumentative Writing Rubric

The Bauer & Kohut Argumentative Writing Rubric

BLOOMS TAXONOMY LEVEL	1 Remember MI	2 Understand BA	3 Apply BA+	4 Analyze PR	5 Evaluate PR+	6 Create AD
PROFICIENT / SKILLED	A list of basic facts and information are given.	Summary and interpretation of ideas from the text; extending or explaining how the ideas relate or compare.	A clear thesis or central idea that relates to the scenario is presented. Reasons follow claim. Structure of writing moves the reader through text in logical order with appropriate breaks.	Multiple perspectives or relationships are discussed and/or considered. All views are clearly expressed.	Argument is clearly made evaluating the strength and weaknesses of chosen position. Relevant evidence is used to support claims. Sources are properly cited.	Writing is clear; is focused on central idea; is organized logically; is appropriate to audience & purpose. Relevant details, evidence, anecdotes, etc. are interpreted, warranted, and cited. A complex understanding of the topic is evident.
LOOK FOR	• Facts listed • Interpretation of facts missing	• Summary and/or paraphrase of textual information • Interpretation of ideas from the text	• Argumentative organizational structure • Thesis *followed by* claim *followed by* reasons / evidence • Paragraph breaks present as needed	• Writer perspective clarified and supported • Other perspective(s) clarified and refuted	• Textual evidence supports each claim/ perspective • Sources of evidence are cited	• Proficient combination of levels 1-5 • All evidence is relevant and warranted • Topic complexities are made clear
NOTES						

Appendix B

ACT Writing Test Scoring Rubric

	Ideas and Analysis	*Development and Support*	*Organization*	*Language Use*
Score 6: **Responses at this scorepoint demonstrate effective skill in writing an argumentative essay.**	The writer generates an argument that critically engages with multiple perspectives on the given issue. The argument's thesis reflects nuance and precision in thought and purpose. The argument establishes and employs an insightful context for analysis of the issue and its perspectives. The analysis examines implications, complexities and tensions, and/or underlying values and assumptions.	Development of ideas and support for claims deepen insight and broaden context. An integrated line of skillful reasoning and illustration effectively conveys the significance of the argument. Qualifications and complications enrich and bolster ideas and analysis.	The response exhibits a skillful organizational strategy. The response is unified by a controlling idea or purpose, and a logical progression of ideas increases the effectiveness of the writer's argument. Transitions between and within paragraphs strengthen the relationships among ideas.	The use of language enhances the argument. Word choice is skillful and precise. Sentence structures are consistently varied and clear. Stylistic and register choices, including voice and tone, are strategic and effective. While a few minor errors in grammar, usage, and mechanics may be present, they do not impede understanding.
Score 5: **Responses at this scorepoint demonstrate well-developed skill in writing an argumentative essay.**	The writer generates an argument that productively engages with multiple perspectives on the given issue. The argument's thesis reflects precision in thought and purpose. The argument establishes and employs a thoughtful context for analysis of the issue and its perspectives. The analysis addresses implications, complexities and tensions, and/or underlying values and assumptions.	Development of ideas and support for claims deepen understanding. A mostly integrated line of purposeful reasoning and illustration capably conveys the significance of the argument. Qualifications and complications enrich ideas and analysis.	The response exhibits a productive organizational strategy. The response is mostly unified by a controlling idea or purpose, and a logical sequencing of ideas contributes to the effectiveness of the argument. Transitions between and within paragraphs consistently clarify the relationships among ideas.	The use of language works in service of the argument. Word choice is precise. Sentence structures are clear and varied often. Stylistic and register choices, including voice and tone, are purposeful and productive. While minor errors in grammar, usage, and mechanics may be present, they do not impede understanding.
Score 4: **Responses at this scorepoint demonstrate adequate skill in writing an argumentative essay.**	The writer generates an argument that engages with multiple perspectives on the given issue. The argument's thesis reflects clarity in thought and purpose. The argument establishes and employs a relevant context for analysis of the issue and its perspectives. The analysis recognizes implications, complexities and tensions, and/or underlying values and assumptions.	Development of ideas and support for claims clarify meaning and purpose. Lines of clear reasoning and illustration adequately convey the significance of the argument. Qualifications and complications extend ideas and analysis.	The response exhibits a clear organizational strategy. The overall shape of the response reflects an emergent controlling idea or purpose. Ideas are logically grouped and sequenced. Transitions between and within paragraphs clarify the relationships among ideas.	The use of language conveys the argument with clarity. Word choice is adequate and sometimes precise. Sentence structures are clear and demonstrate some variety. Stylistic and register choices, including voice and tone, are appropriate for the rhetorical purpose. While errors in grammar, usage, and mechanics are present, they rarely impede understanding.
Score 3: **Responses at this scorepoint demonstrate some developing skill in writing an argumentative essay.**	The writer generates an argument that responds to multiple perspectives on the given issue. The argument's thesis reflects some clarity in thought and purpose. The argument establishes a limited or tangential context for analysis of the issue and its perspectives. Analysis is simplistic or somewhat unclear.	Development of ideas and support for claims are mostly relevant but are overly general or simplistic. Reasoning and illustration largely clarify the argument but may be somewhat repetitious or imprecise.	The response exhibits a basic organizational structure. The response largely coheres, with most ideas logically grouped. Transitions between and within paragraphs sometimes clarify the relationships among ideas.	The use of language is basic and only somewhat clear. Word choice is general and occasionally imprecise. Sentence structures are usually clear but show little variety. Stylistic and register choices, including voice and tone, are not always appropriate for the rhetorical purpose. Distracting errors in grammar, usage, and mechanics may be present, but they generally do not impede understanding.

	Ideas and Analysis	*Development and Support*	*Organization*	*Language Use*
Score 2: **Responses at this scorepoint demonstrate weak or inconsistent skill in writing an argumentative essay.**	The writer generates an argument that weakly responds to multiple perspectives on the given issue. The argument's thesis, if evident, reflects little clarity in thought and purpose. Attempts at analysis are incomplete, largely irrelevant, or consist primarily of restatement of the issue and its perspectives.	Development of ideas and support for claims are weak, confused, or disjointed. Reasoning and illustration are inadequate, illogical, or circular, and fail to fully clarify the argument.	The response exhibits a rudimentary organizational structure. Grouping of ideas is inconsistent and often unclear. Transitions between and within paragraphs are misleading or poorly formed.	The use of language is inconsistent and often unclear. Word choice is rudimentary and frequently imprecise. Sentence structures are sometimes unclear. Stylistic and register choices, including voice and tone, are inconsistent and always appropria rhetorical purpose. Distracting errors in grammar, usage, and mechanics are present, and they sometimes impede understanding.
Score 1: **Responses at this scorepoint demonstrate little or no skill in writing an argumentative essay.**	The writer fails to generate an argument that responds intelligibly to the task. The writer's intentions are difficult to discern. Attempts at analysis are unclear or irrelevant.	Ideas lack development, and claims lack support. Reasoning and illustration are unclear, incoherent, or largely absent.	The response does not exhibit an organizational structure. There is little grouping of ideas. When present, transitional devices fail to connect ideas.	The use of language fails to demonstrate skill in responding to the task. Word choice is imprecise and often difficult to comprehend. Sentence structures are often unclear. Stylistic and register choices are difficult to identify. Errors in grammar, usage, and mechanics are pervasive and often impede understanding.