


5-2018

Measuring Cogency in Argument in the Seventh-Grade English Classroom

Millie Gonzalez-Balsam

University of San Francisco, mgonzalezbalsam@dons.usfca.edu

Follow this and additional works at: <https://repository.usfca.edu/diss>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Methods Commons](#)

Recommended Citation

Gonzalez-Balsam, Millie, "Measuring Cogency in Argument in the Seventh-Grade English Classroom" (2018). *Doctoral Dissertations*. 436.

<https://repository.usfca.edu/diss/436>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, Capstones and Projects at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.

The University of San Francisco

MEASURING COGENCY IN ARGUMENT
IN THE SEVENTH-GRADE ENGLISH CLASSROOM

A Dissertation Presented
to
The Faculty of the School of Education
Learning and Instruction Department

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Millie Gonzalez-Balsam
San Francisco
May 2018

Copyright by Millie Gonzalez-Balsam. All rights reserved. Permission required for use of
the Teacher Designed Rubric to Measure Cogency (TDRMC).

THE UNIVERSITY OF SAN FRANCISCO

Dissertation Abstract

Measuring Cogency in Argument
in the Seventh-Grade English Classroom

Constructing a cogent argument that addresses real-world problems aids students in the development of critical thinking and requires students to present multiple perspectives in a credible manner. Yet, rubrics do not always measure students' reasoning. The purpose of this study was to create a valid and reliable instrument to measure cogency in argument. I created a Teacher Designed Rubric Measuring Cogency (TDRMC) based on Toulmin's model of argument for its emphasis on context-specific warrants, and I used Wilson's framework for assessment to operationalize the construct of cogency. I compared the TDRMC to the current standardized assessment rubric for the Common Core, the Smarter Balanced Assessment Consortium (SBAC). A mixed methods design with a convenience sample of 73 essays from seventh-grade students from a public middle school in northern California addressed three primary questions about the TDRMC: What is the reliability of the TDRMC, and how does it compare to the reliability of the SBAC? Are scores generated from the TDRMC more variable than scores generated from the SBAC? How does the TDRMC correlate to other established measures of writing and academic ability, such as grades, various SBAC scores, and grade point average (GPA), and how do these correlations

compare with those of SBAC Writing and established measures of writing and academic ability? Students essays were scored by a committee of four middle school English teachers. Both percent agreement and Cronbach's alpha showed that the TDRMC was more reliable than the SBAC. The TDRMC did not statistically capture more variability than the SBAC; however, visual inspection of the distribution suggests the TDRMC produced more variability in scores. The TDRMC significantly correlated with a range of external measures of academic ability. These correlations were comparable to those between SBAC and the established measures of writing and academic ability. Several nonsignificant trends suggest that the TDRMC was more highly correlated with sixth- and seventh-grade grade point average and SBAC English language arts than the SBAC. Overall results hint that the TDRMC is better at assessing the construct of cogency and is more highly correlated with academic performance measures than the SBAC.

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements for the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Millie Gonzalez-Balsam
Candidate

July 9, 2018
Date

Dissertation Committee

Dr. Nicola McClung
Chairperson

July 9, 2018
Date

Dr. Matthew Mitchell

July 12, 2018
Date

Dr. Helen Maniates

July 12, 2018
Date

ACKNOWLEDGMENTS

Many people have supported me in my journey of learning. Some have been important mentors who supported me as a teacher; others have supported me as a student in my research; and most importantly, I have received love and encouragement from those closest to me.

I could not have accomplished this research without the love and support of my family, Michael, Anita, and Alan. You all believed in me and supported me even when I was unsure of myself.

I will miss the many long conversations with Miles Myers. You pushed me to see the big picture and to ask the right questions, and you challenged me to continue to read, read, read everything, even when the resources appeared to be tangential to my work. You were truly a teacher's teacher, and I feel privileged that you were a part of my life, as a friend and an educator. Thank you, Sheridan Blau, for the conversations about argument, educational policy, and assessment.

In my teaching credential program, a professor once said to join the organization that supports your subject matter. I have been fortunate to be a member of the Curriculum Study Commission, where I have met many passionate and highly informed English teachers. It is the kind of organization where the question is not "what are you reading?" but "why haven't you read this?" The many conversations and conferences that I have been a part of with Commission members have pushed me and deepened my knowledge.

Assessment is and probably will continue to be a highly contested topic. Two people have provided me with a lens for how to engage in the conversation. For

providing me with a theoretical framework, I want to thank Diana Wilmot for strengthening my understanding of Wilson's framework by taking the time to answer my many queries and providing feedback to my work. For providing me with a philosophical lens that takes into account the social consequences of quantifying achievement, I want to thank Milton Reynolds. Thank you Tania Ivory for the extra pair of eyes.

Several colleagues from my cohort provided me with encouragement and food for thought over the years. I want to thank Sharon Radcliffe for the many conversations and arguments we had about argument and Blanca Pineda for her continual encouragement when I could not see the light at the end of the tunnel.

My committee members provided invaluable insight to different aspects of my research. Thank you, Dr. Helen Maniates, for providing the teacher voice on a subject that impacts us teachers and Dr. Mitchell for your attention to detail and support that aided me in my research. Both of you provided invaluable and constructive feedback.

I want to thank my committee chair, Dr. Nicola McClung, for her unwavering support in me and my research. She challenged me to hone in on clarifying my argument about the importance of assessment of argument and the educational implications for teachers and students.

Finally, I want to thank my husband, Gary Balsam, who has been on this roller coaster of a ride alongside me. You have been my sounding board and champion throughout this process. I could not have accomplished this without your love and support.

TABLE OF CONTENTS

ABSTRACT.....	ii
SIGNATURE PAGE	iv
ACKNOWLEDGMENTS	v
LIST OF TABLES.....	x
LIST OF FIGURES	xi
CHAPTER 1: INTRODUCTION.....	1
Statement of the Problem.....	1
Purpose of the Study.....	2
Significance of the Study.....	2
Theoretical Framework.....	3
Toulmin’s Model of Argumentation.....	3
The Importance of Argument Field in Toulmin	5
Arguments of Fact, Judgment, and Policy.....	6
The Berkeley Evaluation and Assessment Research Center Assessment System.....	6
Background and Need.....	9
Why Argument?.....	9
Moving Beyond the Primary Literary Text	10
Constructing Argument to Address Complex Real-World Problems.....	12
Six Components of Argument	14
Current Tools for Assessing Argument in the Seventh-Grade English Language Arts Classroom.....	18
College Career Readiness and the Importance of Argument.....	20
Accountability Policy.....	21
The Importance of Argument.....	22
Need for the Study	24
Research Questions.....	25
Summary.....	25
CHAPTER 2: LITERATURE REVIEW	26
Brief History of Assessment of Writing	26
Validity Applied to Writing Assessment	28
Rubrics	29
Emergence of Analytic Rubrics	29
Interlinear Exercises.....	31
The Importance of Construct Validity in Assessment of Writing	32
Toulmin’s Model of Argument: The Language of Academia	34
Argument and Reasoning.....	35
Argument as Applied to Literary Analysis.....	35

Empirical Research on Argument Using Toulmin’s Model	36
Role Warrants Play in the Validity of a Mathematical Proof	37
Scaffolding Argument in the Science Classroom	39
Connecting Selection of Evidence to the Claim	42
Summary	47
CHAPTER 3: METHODS	49
Research Questions	50
Research Design	50
Sample	51
Protection of Human Subjects	52
Instrument Generation	52
Building Block 1: Construct Map	52
Building Block 2: Item Design	53
Building Block 3: Outcome Space	57
Building Block 4: The Measurement Model	58
Construct Validity	58
Procedures	59
Phase 1: Instructional Procedures	60
Instruction of Toulmin’s Model of Argument	60
Administering the Prompt and Writing the Essay	60
Phase 2: Moderation Session	61
Phase 3: Calibration Protocol	61
Data Analysis	62
Summary	63
CHAPTER 4: RESULTS	65
Research Question 1	65
Moderation Sessions	65
Moderation Session for SBAC	65
Moderation Session for TDRMC	67
Establishing Interrater Reliability	69
Research Question 2	69
Research Question 3	71
CHAPTER 5: DISCUSSION	75
Overview of Argument	75
Summary of the Study	76
Limitations	77
Discussion of Research Questions	80
Research Question 1	80
Research Question 2	81
Research Question 3	82
Instructional Implications	83
Implications for Instruction of Toulmin’s Model of Argument	83
Revision of the TDRMC	83
Conclusion	84

Research Implications.....	86
Educational Implications	86
Summary.....	88
REFERENCES	91
APPENDICES	
APPENDIX A: University of San Francisco IRB Approval	102
APPENDIX B: Informed Consent.....	103
APPENDIX C: Rater Consent	105
APPENDIX D: Site Consent	107
APPENDIX E: Transcription of Moderation Session for SBAC	108
APPENDIX F: Transcription of Moderation Session for TDRMC.....	126

LIST OF TABLES

Table 1. Six Components of Argument With Examples From *The Outsiders* 16

Table 2. Iterative Process for Design of Instrument of Cogency..... 51

Table 3. Data Collection Procedures for the Study..... 59

Table 4. Descriptive Statistics..... 71

Table 5. Correlations Between SBAC, TDRMC, and External Measures 73

Table 6. Correlations Between SBAC and TDRMC With Sixth and Seventh ELA Grades
and Their Comparisons 73

LIST OF FIGURES

Figure 1. Aligning assessment with cognitive goals.....	7
Figure 2. The four building blocks of the BEAR system	8
Figure 3. Assessment triangle.....	9
Figure 4. Supplementary texts that complement and provide multiple perspectives	13
Figure 5. A schematic of Toulmin’s (1958/2003) theory of argument.....	15
Figure 6. Smarter Balanced rubric measuring elaboration of evidence.....	20
Figure 7. Construct map for cogency.....	54
Figure 8. TDRMC scoring guide measuring cogency and SBAC rubric measuring elaboration of evidence.....	55
Figure 9. TDRMC scoring guide measuring cogency and SBAC rubric measuring elaboration of evidence, as pertaining to <i>The Outsiders</i>	56
Figure 10. Distribution of SBAC scores from student’s essays.....	70
Figure 11. Distribution of TDRMC scores from student essays.....	70

CHAPTER 1: INTRODUCTION

Statement of the Problem

Argument is now the dominant mode of written discourse in the seventh-grade English language arts (ELA) classroom (National Governors Association Center for Best Practices Council of Chief State School Officers [NGA], 2010); that is, seventh-grade students must be able to construct argument, which includes demonstrating the ability to defend a claim that is based on evidence from a text and backed up with logical appeals (California Department of Education [CDE], 2015; Hillocks, 2011; NGA, 2010). This emphasis on moving beyond a primary text to incorporate supporting claims that are directly tied to evidence in literary and nonfiction texts is new and is required for argument as articulated in the Common Core State Standards (CCSS; NGA, 2010). The genre of argument includes informational or explanatory writing that defines, explains, or clarifies a concept, along with the mode of narrative writing used to express a personal or fictitious experience that explains or argues to persuade (NGA, 2010). All modes of discourse are concerned essentially with demonstrations of students' ability to construct argument. Despite the importance of argument in skilled writing, there is a dearth of valid and reliable measures of the construct, leaving educators ill equipped to assess student writing.

In the ELA classroom, argument has historically been constructed in response to literary texts. In elementary school, Grades K–5, students write persuasive arguments that are based on opinion about a literary text, the purpose of which is to convince the reader of the correctness of the opinion the student espouses. However, a shift from writing persuasive arguments that are supported by opinion to constructing argument that

is grounded in evidence takes place in middle school, comprising Grades 6–8. In addition, the concept of counterargument is introduced specifically in Grade 7, when students are expected to demonstrate their ability to oppose counterclaims (Smarter Balanced Assessment Consortium, 2015), because the strength of an argument is dependent on the degree to which it addresses counterarguments (Hillocks, 2011; Toulmin, 1958/2003). In addressing counterarguments, students write their own arguments, which differs from expressing an opinion in that the defense of a claim must cite evidence. Students use evidence from literary and nonfiction texts to present cogent reasoning in defense of their claims. Cogent reasoning requires students to engage in critically evaluating perspectives counter to their own and also requires writing that demonstrates the ability to present counterclaims that reflect a deeper analysis of the topic (Hillocks, 2011; NGA, 2010; Toulmin, 1958/2003).

Purpose of the Study

The purpose of this study is to create an instrument, the TDRMC, that is a valid and reliable measure of cogency as a component of the multidimensional construct of argument that addresses the gaps in earlier assessments. I chose cogency as the focus component because the reasoning is appropriate to the argument field (Hillocks, 2011).

Significance of the Study

This study adds to the research knowledge and administration of assessment by creating a valid instrument to assess cogency, a component of the multidimensional construct of argument. For the first 50 years of measurement's 100 years of history, assessment of writing ability focused more on reliability than on validity (Huot, O'Neill, & Moore, 2010), and validity was determined by reader consensus (Gere, 1980; Huot,

2002; Yancey, 1999). Operationalizing cogency will aid in constructing an instrument that validly measures the construct, enabling meaningful feedback to both student and teacher to aid in (a) understanding where the student's abilities lie and (b) informing instruction (Draney, 2009).

Theoretical Framework

This study employed two theoretical frameworks, Toulmin's (1958/2003) *model of argument* for operationalizing the construct of cogency and Wilson's (2005) *framework for assessment* articulated in the Berkeley Evaluation and Assessment Research Center (BEAR) assessment system, to design a valid and reliable instrument to assess the construct of cogency.

Toulmin's Model of Argumentation

In the past, writing in the seventh-grade classroom primarily drew from personal anecdotes or experiences. With the adoption of the CCSS (NGA, 2010), students are now required to construct argument with evidence that is grounded in literary texts and literary nonfiction.

The Toulmin model of argument is the most commonly used model in secondary schools and colleges (CDE, 2015; Graff, 2003; Kneupper, 1979; Lunsford, 2002). Toulmin's model lays out an argument in a procedural manner. For example, claim, data, and evidence provide a clear structure for students to aid them in clearly presenting an argument. A more sophisticated and controversial argument contains qualifiers, warrants, backing, and rebuttals, when necessary. The strength of the model, according to Kneupper (1979), is that it helps students develop arguments that are supported and shown to address counterclaims. It follows that writing is coherent, because each

element of the model is enumerated and the relationship between each point is made clear, which serves to delineate how the evidence supports the claim of the argument. Following Toulmin's model does not automatically result in students' ability to construct argument that is sound, but it helps them connect each element of the model and see how each point being made is connected.

The *claim* is the assertion. The *data* are the evidence, and the *warrant* represents the justificatory statement for the data constrained by the standards within the argument field, which includes the shared assumptions, values, ideas, and concepts for argument that are constrained by the norms within the specific disciplines (Lunsford, 2002; Rowland, 2008; Stygall, 1986). Often the warrant is defined as the hypothetical bridge that links the data to the claim. If warrants are not sufficient, then backing is provided. *Backing* refers to additional support for the warrants, which may require examples or definitions of abstract terms or philosophical concepts. If the warrants are accepted, then backing is not required.

Argument is analogous to jurisprudence because of Toulmin's emphasis on the use of justificatory statements (Toulmin, 1958/2003; Eemeren, Grootendorst, & Snoeck, 1996). In law, there is a procedural formulation of the shape and form of an argument. In a similar way, Toulmin's model lays out argument in a procedural format. This consistency is known as field invariant to indicate how the procedure for laying out an argument is similar regardless of the argument field. The emphasis is on the procedural manner in which an argument is laid out. The field-dependent component of Toulmin's model acknowledges that different criteria are applied to evaluating an argument because

constraints on an argument take into account the context in which an argument is made (Lunsford, 2002; Toulmin, 1958/2003).

The Importance of Argument Field in Toulmin

The evaluation of criteria is central in Toulmin's argument field because of its rejection that an absolutist model for evaluating criteria can be applied to any argument. The idea that warrants are context specific demonstrates to students the necessity of investigating the values and assumptions underlying the warrants (Hillocks, 2011; Lunsford, 2002; Stygall, 1986; Toulmin, 1958/2003). In this way, evidentiary reasoning is shown, because the data used to support a claim must be appropriate to the field (Eemeren et al., 1996; Hillocks, 2011; Lunsford, 2002).

The argument field is critical in understanding the role of warrants, because a warrant must do more than establish the criteria for accepting evidence to support a claim. The criteria for evaluating an argument in the argument field of geometry will be different from the criteria for evaluating an argument in the argument field of history. In this way, warrants explain the inferences derived from the reasoning and are seen to be context specific (Eemeren et al., 1996; Lunsford, 2002; Toulmin, 1958/2003). Thus the validity of an argument rests on the field-invariance and field-dependence aspects of argumentation in Toulmin's model, known as the *argument field* (Eemeren et al., 1996).

Modal terms are used throughout the procedural manner of laying out the argument to express the force of the claim within the argument field, because the warrants, or justificatory statements, depend on the field (Eemeren et al., 1996). The qualifiers point to the probabilistic nature of argument (Keith & Beard, 2008).

Arguments of Fact, Judgment, and Policy

Hillocks (2011) divided argument into three categories: *arguments of fact*, *arguments of judgment*, and *arguments of policy*. Regardless of what type of argument a student is making, a valid argument must include warrants and backing that are appropriate to the argument. Arguments of fact allow students to incorporate personal knowledge to aid in their use of justificatory statements, or warrants, that link their evidence to their claims. Arguments of judgment require an analysis of the values, or assumptions, because they cannot be supported by data. Arguments of policy focus on defining terms used in the warrants and backing that help establish the criteria for argument. Defense of all arguments requires students to defend the warrants and backing. The thinking involved in defense of warrants and backing is a complex act of reasoning.

The Berkeley Evaluation and Assessment Research Center Assessment System

For an instrument to be reliable and valid, the inferences drawn from the instrument should be representative of the construct being measured. The BEAR assessment system is a method of developing assessment aligned with instruction to provide a meaningful understanding of students' cognitive abilities relative to the curricular and cognitive goals and was developed by Dr. Mark Wilson. For the purposes of this study, I used an adapted version of the BEAR assessment system to create a valid and reliable instrument to measure cogency.

As illustrated in Figure 1, Wilson (2005) defined an instrument as a way of relating what is observed or manifest to what is being measured that is latent, unobserved, and theoretical.

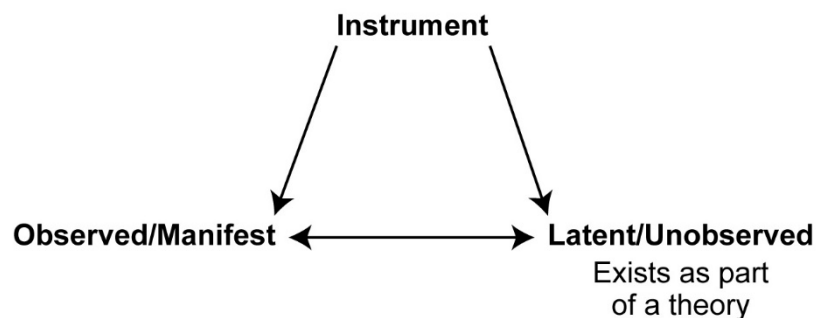


Figure 1. Aligning assessment with cognitive goals.

This model consists of four building blocks (Wilson, 2005): *construct map*, *item design*, *outcome space*, and *measurement model*. Each component of Wilson's model informs the next and should be seen as cyclical. As shown in Figure 2, the BEAR assessment system is a framework that makes assessment meaningful for several reasons: (a) It is tied to a cognitive theory of learning, (b) instruction and assessment are transparent, (c) the data inform the teacher and instructional practice, and (d) embedded in the design of the instruments are procedures to ensure reliability and validity.

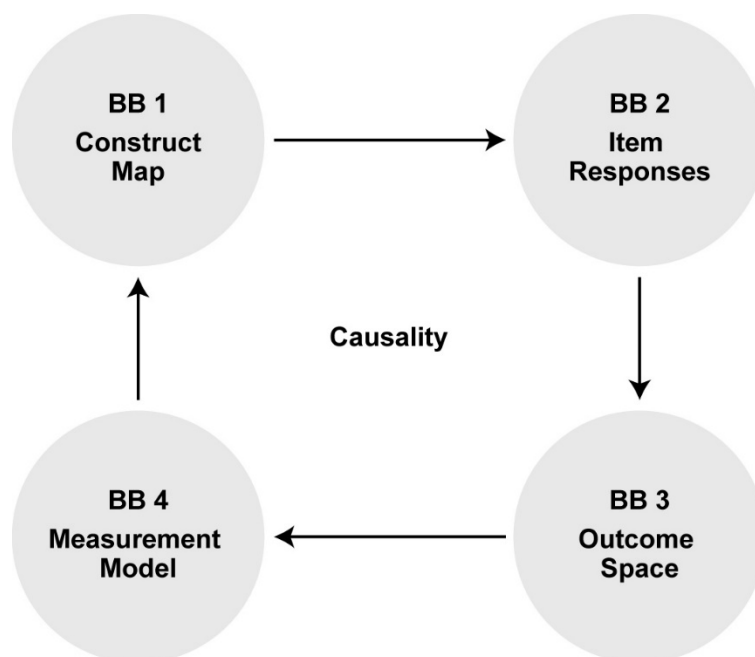
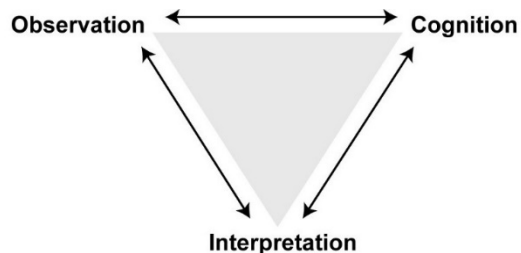


Figure 2. The four building blocks of the BEAR system. Adapted from *Constructing Measures: An Item Response Modeling Approach*, by M. Wilson, New York, NY: Taylor & Francis, 2005.

Wilson's framework for assessment is based on a cognitive theory known as the *assessment triangle* (National Research Council, 2001). The assessment triangle features three essential components, *cognition*, *observation*, and *interpretation*, as illustrated in Figure 3. Cognition refers to the theories each subject domain states a student must know to be competent (e.g., constructing a cogent argument), while observation refers to evidence of competence demonstrated in performance tasks (e.g., a written argument) as articulated by instructional practices reflective of the subject domain theories. Last, interpretation must reflect knowledge observed in performance tasks that are reflective of the cognitive theory of the subject domains (e.g., the scoring guide, also known as a rubric for this study). The three elements of this framework highlight the interconnectedness of the three components as a foundation of the measurement model. In this study, cognition is the construct map, observation refers to the prompt and reading

materials for the essay, while interpretation refers to whether the scores are reflective of the construct map.



*Figure 3. Assessment triangle. From *Knowing What Students Know: The Science and Design of Educational Assessment*, by National Research Council, Washington, DC: National Academy Press, 2001.*

Background and Need

Why Argument?

Acquiring skills to engage in argument is important to a student in the academic world and to a citizen participating in the political discourse of a democratic society. Academically, the skills involved in argument impact the manner in which people engage as citizens in political discourse. Argument requires taking a stance, presenting substantive evidence with appropriate underlying criteria pertinent to the argument, and critically evaluating the reasoning and underlying criteria of opposing viewpoints. In a multicultural society and democracy, these skills are essential tools that move people with different viewpoints toward acknowledgment of multiple perspectives (Emmel, Resch, & Tenney, 1996).

This type of reasoning is seen as an indicator of how well students may perform in college, because attending college means entering into a discussion of the world of

ideas. Establishing a culture of argument in schools is seen as necessary to give every student skills to enter into national political discourse (Graff, 2003). To be engaged is to be informed and open to the multiple perspectives of a diverse society. Reading and writing are seen as key to being informed and open. The abilities to interpret and make inferences are both important features of reading, especially the ability to make inferences, which demonstrates the ability to make reasoned judgments about the quality of a text. Inherent in this skill is the ability to critically evaluate other points of view. In this way, students are able to take positions on an issue and are able to construct argument.

Argument relies on appeals to logic, rather than pathos, and requires that students support their claims with evidence or facts and logical reasoning. Defense of ideas is the result of critically evaluating perspectives that run counter to the student's own. Critically evaluating the merit of counterarguments aids students in examining the fallacies in their own arguments. In this way, argument teaches critical thinking because it requires critical evaluation of multiple perspectives on an issue (CDE, 2015; Graff, 2003; Hillocks, 2011; Leeman, 1987). Yet writing argument is a difficult task for many students at the secondary level (McCann, 2010). Additionally, few empirical research studies have focused on reasoning skills in K–12 ELA classrooms (Fulkerson, 1996; Lunsford, 2002).

Moving Beyond the Primary Literary Text

The new standards for ELA require students to demonstrate a wider range of reading, writing, and complex reasoning skills that involve moving beyond the primary literary text. In the past, literary analysis of a novel was sufficient. Defense of a claim

usually did not require students to read a range of supplemental nonfiction literary texts. Evidence for claims was grounded in students' ability to provide evidence from the novel to support an analysis of the plot or character. Now, students need to provide supplemental evidence from content-rich nonfiction literary texts to defend claims on issues explicated in the primary literary text (NGA, 2010).

For example, in a seventh-grade classroom, students may read the novel *The Outsiders* by S. E. Hinton. Set in Oklahoma during the 1960s, the novel focuses on the rivalry between two gangs: the Greasers, who are the working-class boys from the east side of town, and the Socs, the wealthy socialites from the west side of town. The protagonist is Ponyboy, who at 14 years of age is the youngest of three Curtis brothers and lives on the east side. The eldest, Darry, is the head of the family since the death of their parents. Both Darry and the middle brother, Soda, work so Pony may continue his high school education. Their family also consists of the Greasers, boys who come from dysfunctional families—broken homes, in some cases headed by abusive or neglectful parents. Together the Curtis brothers and the Greasers make a family that provides unconditional love and support for each other, in contrast to the permissive, hands-off parenting of the families of the Socs. The Curtis brothers' definition of family is severely tested when Johnny, Ponyboy's best friend, kills a leader of the Socs in self-defense.

In the classroom, when reading *The Outsiders*, the issue of family is further explored with supplemental literary nonfiction texts, such as biographies, speeches, articles in journals, and memoirs (CDE, 2015), that address the themes explicated in the primary literary text. For example, during *The Outsiders* unit, students might read the article from the *New York Times* titled "The Changing American Family" by Natalie

Angier (2013). The article includes several stories that provide many perspectives to support and challenge the definition of family in *The Outsiders*. Reading a range of texts that are literary and literary nonfiction draws on, reinforces, and increases content knowledge about the themes raised in the novel. Ultimately, students should be able to critically evaluate, analyze, and address perspectives that may run counter to their claims.

As discussed, this increased demand for students to read within and across texts to construct written argument is new to the ELA standards (CDE, 2014). Now, students are required to build knowledge about the themes articulated in literary texts through reading and citing evidence from literary and supplemental literary nonfiction to support their claims. Although skilled argument always involves the use of evidence to support claims that are text based, now the texts also include literary nonfiction, placing additional reading, writing, and thinking demands on the student. Figure 4 shows how supplementary texts support the main text, which in this case is the novel *The Outsiders*. Supplementary texts serve different purposes, such as providing context for the novel, deepening conceptual understanding of themes, or providing multiple perspectives.

Constructing Argument to Address Complex Real-World Problems

The rationale for the new standards is to foster critical thinking skills by engaging students in making connections to the real world through a rigorous study of complex issues raised in the reading of literary texts, such as changing family structures, an increase in wealth disparities, and shifts in gender roles as explored in the unit on *The Outsiders*. Skilled defense of a claim comes from reading widely and deeply on an issue to demonstrate in writing the use of complex reasoning skills to critically evaluate the merit of positions that run counter to the claim. In doing so, students demonstrate their

ability to construct cogent arguments (CDE, 2015). *Cogency* is defined as the use of warrants and backing. For example, a student might construct an argument for broadening definitions of family after reading *The Outsiders* and then support his or her argument by drawing from the story titled “The Wedding Will Have to Wait,” included in the *New York Times* article “The Changing American Family,” to cite evidence for challenges to the nuclear family. The student might also acknowledge and address traditional views of the family by citing from “To Atlanta, by Way of Sri Lanka” in the same *New York Times* article but return to *The Outsiders* to cogently argue their position on broadening the definition of family.

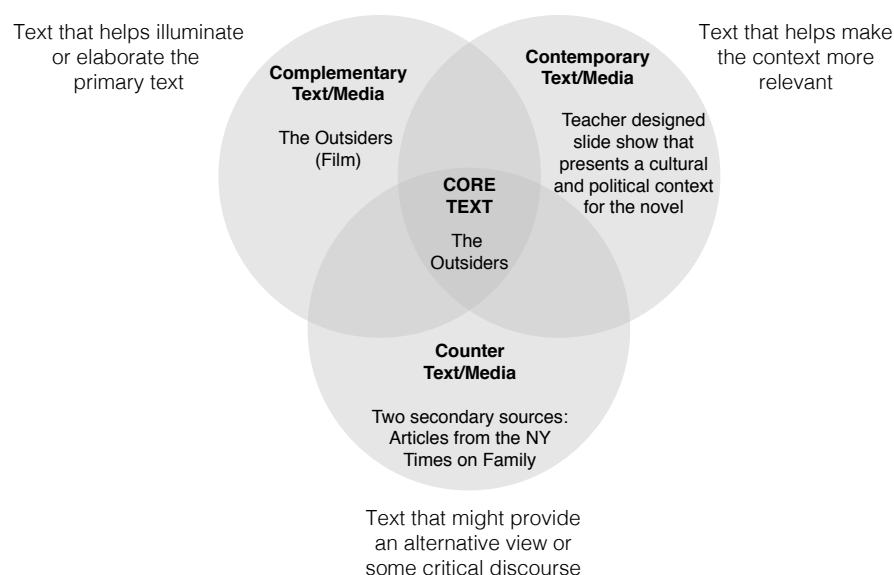


Figure 4. Supplementary texts that complement and provide multiple perspectives.

In the course of constructing argument, the dominant focus is on using appeals to logic, which is more than simply writing to persuade. While both forms of argument have at their core the outcome of changing the reader’s mind, the emphasis on appeals to

logic requires students to support claims backed up with evidence that is credible and logically reasoned (Hillocks, 2011; Toulmin, 1958/2003).

Toulmin (1958/2003) proposed a theory of argument that is multifaceted and that can also be understood in the context of measurement as “multidimensional” (Wilson, 2005). As illustrated in Figure 5, Toulmin’s model of argumentation includes assertion of a claim that is both articulated in the thesis statement and supported with evidence or data. Warrants are the justificatory statements to explain the evidentiary reasoning or support for the claim and use appeals to logic. The backing of warrants must be appropriate by providing definitions of the values and assumptions in the warrant that are acceptable and integral to the topic—to the argument field.

The warrant is the most important component in Toulmin’s model, and yet it is the most misunderstood. A warrant must be evaluated within the context of the argument field, also known as the subject domain. In this way, the conclusion may be deduced from the evidence in a logical and reasoned way in context. Backing supports the warrants and is based on the criteria that further provide a definition for the more intangible aspects of the warrant or specific examples for arguments of fact (Hillocks, 2011).

Six Components of Argument

For example, if in a seventh-grade unit on *The Outsiders* a student challenges the concept of the nuclear family with the thesis, also known as the claim, that “What matters most is what the family contains, not what it looks like,” then backing must support the established criteria for the definition of family. To provide backing, the student might draw on the stories from the aforementioned *New York Times* article.

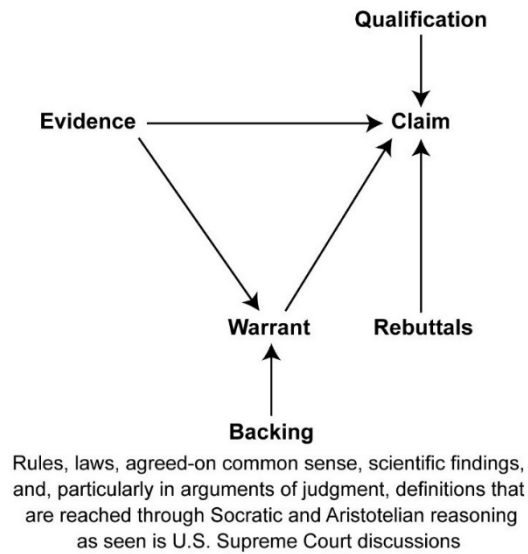


Figure 5. A schematic of Toulmin's (1958/2003) theory of argument. From *Teaching Argument Writing: Grades 6–12*, by G. Hillocks, Portsmouth, NH: Heinemann, 2009.

Table 1 presents the procedural manner in which a student would lay out an argument using Toulmin's model of argumentation and examples from *The Outsiders*. The manner in which Toulmin's model of argumentation ensures the cogency of an argument is demonstrated in the examples of the warrant and the backing. Justificatory statements explain the logical reasoning behind the selection of the evidence used to defend a claim and demonstrate how a warrant is context specific (Lunsford, 2002; Stygall, 1986). This process requires a student to critically evaluate perspectives that run counter to his or her claim, because citing evidence from texts is not enough: The reasoning must be evidentiary to support analysis.

Table 1

Six Components of Argument With Examples From The Outsiders

Dimension	Description/characteristics	Examples from <i>The Outsiders</i>
Claim	The argument put forth by the student	“What matters most is what the family contains, not what it looks like.”
Data/evidence	Support from the <i>New York Times</i> articles to support the claim that the Curtis brothers and their extended family, the Greasers, are the best representation of a family	“Relationships like these— independent of biology but closer and more enduring than friendship—have been documented in various cultures throughout history.”
Warrant	Underlying values and assumptions that are mainly implicit	Relationships based upon unconditional love and support are what constitute a family
Backing	Additional definitions of values and assumptions to strengthen the warrant	The definition of family is not constrained by biology. “Anthropologists have traditionally used the term ‘fictive kin’ to separate such relationships from ‘true’ kinship based on blood or law, but many researchers have recently pushed back against that distinction, arguing that self-constructed families are no less real or meaningful than conventional ones.”
Qualifier	Acknowledgment of the possibility of alternative views	“It’s the backbone of how we live,” said David Anderson, 52, an insurance claims adjuster from Chicago. “It means everything,” said Linda McAdam, 28, who is in human resources on Long Island.

Table 1 (*continued*)

Dimension	Description/characteristics	Examples from <i>The Outsiders</i>
Counterargument	Presentation of weakness in the argument	We're sappy family romantics. When an informal sample of 52 Americans of different ages, professions, and hometowns were asked the first thought that came to mind on hearing the word "family," the answers varied hardly at all. Love! Kids! Mom! Dinner!
Rebuttal	Evidence for the counterclaim	Families, they say, are becoming more socially egalitarian over all, even as economic disparities widen. Families are more ethnically, racially, religiously, and stylistically diverse than half a generation ago—than even half a year ago.

Note. Six components of argument adapted from Toulmin (1958/2003).

Since Toulmin's model of argumentation asserts that the argument is an argument of probability, then it must be qualified. The dimension of *qualifiers*, which is necessary in pointing out the possibility of alternative views, involves attention to grammatical structures. These qualifiers include the use of modals, which aid the writer in expressing the degree of certainty in the construction of an argument (Toulmin, 1958/2003). Modals explain how strong the evidence is in support of the claim and typically include words such as *sometimes*, *most*, or *usually*. Acknowledgment of the counterclaim is expressed with qualifiers, while *rebuttals* address the counterclaim.

To sum up, the purpose of constructing argument is to demonstrate the validity of a claim with support of logical reasoning. Claims must be well supported with evidence that considers counterarguments that are substantive and that reflect critical thinking and logical reasoning (Leeman, 1987; Kneupper, 1979; Graff, 2003; McCann, 2010; Warren, 2010; Hillocks, 2011). Important to note is that while Toulmin's model of argument contains similar components of argument structure, such as claim, data, evidence, counterclaim, and rebuttal, its emphasis on a context-specific warrant distinguishes it from a generic argument model. Toulmin's focus on the context-specific nature of the warrant establishes the criteria for the underlying values and assumptions implicit in the warrant.

Current Tools for Assessing Argument in the Seventh-Grade English Language

Arts Classroom

Although seventh-grade students are expected to provide evidence to support claims made in an argument, tools for the assessment of argument are limited. As Figure 6 shows, the descriptor for the highest score of 4 in the SBAC rubric, for example, states that a student must provide evidence to back up a claim, but the language of the instrument does not state that the response must demonstrate the ability to construct cogent arguments that use evidentiary reasoning. In Chapter 1 of the English Language Arts/English Language Development Framework for California Public Schools (CDE, 2015), the reference of the expectation that students will be able to demonstrate cogent reasoning with the application of the new CCSS is stated:

They actively seek the wide, deep, and thoughtful engagement with high-quality literary and informational texts that builds knowledge, enlarges experience, and

broadens worldviews. They reflexively demonstrate the cogent reasoning and use of evidence that is essential to both private deliberation and responsible citizenship in a democratic republic.

In other words, the language of the descriptor regarding relevant evidence does not explicitly state the cogency of the evidence, that is, whether the conclusion is logically deduced based on the criteria of the values and assumptions of the particular topic or argument field. The rubric uses subjective language when it states that the evidence must be thorough and convincing, but what is missing is an assessment of whether evidentiary reasoning is present, or, in other words, whether the reasoning is appropriate to the evidence and argument field (Hillocks, 2011; Toulmin, 1958/2003). Therefore, in addition to assessing evidence, an instrument must assess a student's complex reasoning skills and, in particular, whether the evidentiary reasoning is appropriate to the argument field.

The SBAC descriptors provide information regarding a student's strengths and weaknesses but lack an evaluation of communicative intent, that is, how well language expresses meaning (Gere, 1980). Additionally, current assessments, such as the SBAC, fail to adequately capture the variance in the complexity of cognitive abilities, especially when measuring the complex reasoning needed to construct cogent arguments. This gap in assessment makes it difficult to draw valid inferences from student scores. Current assessments also do not measure proficiency on a continuum (Pellegrino, Chudowsky, & Glaser, 2001), owing in large part to the lack of criteria that clearly delineate the specific construct being measured (Haladyna & Rodriguez, 2013; Nodoushan, 2014).

 4-Point Argumentative Performance Task—Writing Rubric (Grades 6–11)

4 POINTS

The response provides thorough and convincing support/evidence for the argument(s) and claim that includes the effective use of sources (facts and details). The response clearly and effectively expresses ideas, using precise language:

- comprehensive evidence from sources is integrated; references are relevant and specific
- effective use of a variety of elaborative techniques*
- vocabulary is clearly appropriate for the audience and purpose
- effective, appropriate style enhances content

* Elaborative techniques may include the use of personal experiences that support the argument(s).

3 POINTS

The response provides adequate support/evidence for the argument(s) and claim that includes the use of sources (facts and details). The response adequately expresses ideas, employing a mix of precise with more general language:

- adequate evidence from sources is integrated; some references may be general
 - adequate use of some elaborative techniques
 - vocabulary is generally appropriate for the audience and purpose
 - generally appropriate style is evident
-

2 POINTS

The response provides uneven, cursory support/evidence for the argument(s) and claim that includes partial or uneven use of sources (facts and details). The response expresses ideas unevenly, using simplistic language:

- some evidence from sources may be weakly integrated, imprecise, or repetitive; references may be vague
 - weak or uneven use of elaborative techniques; development may consist primarily of source summary or may rely on emotional appeal
 - vocabulary use is uneven or somewhat ineffective for the audience and purpose, with inconsistent or weak attempt to create appropriate style
-

1 POINT

The response provides minimal support/evidence for the argument(s) and claim that includes little or no use of sources (facts and details). The response's expression of ideas is vague, lacks clarity, or is confusing:

- evidence from the source material is minimal or irrelevant; references may be absent or incorrectly used
 - minimal, if any, use of elaborative techniques; emotional appeal may dominate
 - vocabulary is limited or ineffective for the audience and purpose, with little or no evidence of appropriate style
-

0 POINTS

- Unintelligible
 - In a language other than English
 - Off-topic
 - Copied text
 - Off-purpose
-

Figure 6. Smarter Balanced rubric measuring elaboration of evidence.

College Career Readiness and the Importance of Argument

College readiness refers to the content knowledge and cognitive strategies students need to keep up with the academic rigor of college. In the context of argument, college readiness is defined as the ability of students to show the logical reasoning behind support of claims and to demonstrate their ability to identify credible sources as they construct argument. To demonstrate proficiency in the cognitive strategies necessary for writing argument in seventh grade, students must use skills that include analysis,

interpretation, precision and accuracy, problem solving, and reasoning, all skills that are demonstrated in college-level academic writing (Conley, 2007; Conley, Drummond, de Gonzalez, Rooseboom, & Stout, 2011; Deane, 2011). The 2011 American College Testing (ACT) results showed that only one-fourth of the students who took the test demonstrated college readiness (Conley, 2007; Rothman, 2012).

In a survey conducted by Peter D. Hart Research Associates and Achieve Group Inc. (2005), 300 college professors and 400 employers were asked to give their opinions regarding college readiness and adequate preparation for the workforce. Approximately 82% of college professors noted that incoming freshmen were not prepared for the rigors of college. More than one-third of high school graduates were not prepared for entry-level jobs, while close to half of college graduates were inadequately prepared to demonstrate competence at positions beyond entry-level work (Hart Research Associates, 2005).

Accountability Policy

A Nation at Risk (U.S. National Commission on Excellence in Education, 1983) helped to usher in the national standards movement in 1994. The agenda of the movement focused on establishing standards for each core subject to help boost student achievement, accompanied by tests to measure if students were meeting the standards (Clinchy, 1998). The standards movement was propelled in 2002 with the No Child Left Behind Act of 2002 (NCLB), which stipulated that schools with low scores on standardized tests would incur consequences (Fritzberg, 2004).

The passage of NCLB required all subgroups (racial, English language learner [ELL], and low socioeconomic status) to make progress on standardized tests to show

that schools were meeting adequate yearly progress (AYP). Schools that failed to make AYP were restructured, closed down, or even taken over by the state (Fritzberg, 2004). The NCLB act elevated standardized tests from a measurement tool to the focus for schools and districts to improve student achievement (Clark, 2011). In addition, NCLB stipulated that all students must achieve proficiency by 2014, a condition that penalized schools in urban school districts with students from low socioeconomic backgrounds and communities with linguistic diversity.

The legacy of NCLB has been a new face to the accountability movement. Race to the Top was predicated on the idea that choice in the form of charter schools and standardized testing that measured the quality of teachers, achievement of students, and progress of schools were the best ways of reforming schools (Ravitch, 2013). States had to compete for funds from the federal government and, in doing so, had also to agree to the above. And as Ravitch pointed out in her book *Reign of Error*, this competition for funds was the first time the federal government did not operate from the position of addressing inequities when it provided funds for education but instead relied on competition between states.

The Importance of Argument

The CCSS (CDE, 2015) framework stated that argument must be taught, and it is currently elevated above the mode of writing persuasive essays. This shift is in response to data that show that incoming college freshmen are not able to construct argument. According to the National Center for Education Statistics (2004), approximately 40% of incoming college freshmen at 4-year institutions were enrolled in a remedial course, while at California state universities, 46% of incoming college freshmen were required to

enroll in remedial English classes (Ali & Jenkins, 2002). The statistics were worse for students from low socioeconomic backgrounds; approximately 60% graduate from high school, while one in seven will successfully complete his or her education at a 4-year institution.

The CCSS (CDE, 2015) framework explicitly stated that evidence must exhibit content knowledge and demonstrate the academic discourse of the particular discipline as they construct meaning. Because college readiness is a major rationale behind the formation of the CCSS and its accompanying assessment, the SBAC, the importance of student proficiency in the ability to construct argument is emphasized because it is the dominant mode of discourse in college. The CCSS adopted the National Governor's Association anchor standards, which express the foundational skills necessary for a student to demonstrate college readiness.

The adoption of the CCSS and their accompanying assessments—the SBAC and the Partnership for the Assessment of College and Career Readiness (PARCC)—is the latest iteration of standardized assessment to measure college readiness. With the adoption of the CCSS, a new paradigm of assessment has been created that includes formative, interim, and summative assessment. The CCSS and accompanying assessments support college readiness and have elevated academic writing as a demonstration of college readiness. The ability to construct argument is what will be measured at the seventh-grade level. If the SBAC does not properly measure the construct of argument, and because assessment drives instruction, is instruction ensuring that students will be college ready?

Need for the Study

The SBAC is a summative assessment designed for accountability; therefore it is not to be used as formative assessment or to inform instruction. However, we are expecting students to graduate with the skill of constructing arguments that demonstrate cogent reasoning, yet we are currently lacking in an assessment tool that targets student understanding of cogency that informs instruction. Therefore the use of the TDRMC will aid teachers in the instruction of argument that meets the standard of cogency articulated in the English Language Arts/English Language Development Framework for California Public Schools (CDE, 2015).

Beyond the new standards and requirements of the Common Core, argument is an important concept for developing citizens to participate in a democracy who have the skills to consider and act on important issues in our society. Yet, some scholars have argued that the language used to frame argument is often presented in a charged manner in which the end goal is to smash the opposition rather than generate understanding of the issue (Lakoff & Johnson, 2008). Entrenched positions that lead to polarization on real-world issues, such as climate change, income inequality, and race relations, are just a few with which current students will have to grapple. Understanding involves acknowledgment of the underlying values and assumptions of the counterargument as well as the underlying values and assumptions of our own arguments. When this is accomplished, the nuances and complexities of issues can be acknowledged, and we can adjudicate important real-world issues. Therefore research on assessment of argument is necessary.

Research Questions

The research questions for this study follow:

1. What is the reliability of the TDRMC, and how does it compare to the reliability of the SBAC?
2. Are scores generated from the TDRMC more variable than scores generated from the SBAC?
3. How does the TDRMC correlate to other established measures of writing and academic ability, such as grades, various SBAC scores, and grade point average (GPA), and how do these correlations compare with those of SBAC Writing and established measures of writing and academic ability?

Summary

My intent in conducting this study was to create an instrument (TDRMC) to measure cogency, a component of the multidimensional construct of argument. Students essays were scored using both the TDRMC and the SBAC to determine which instrument was more reliable and valid in assessing the construct of cogency. Additionally, the TDRMC was compared to the SBAC to determine which instrument produced more variance of scores of student essays. Finally, both instruments were correlated with real-world measures to determine which had better external validity.

CHAPTER 2: LITERATURE REVIEW

The purpose of this study is to create an instrument that measures cogency. The ability to use cogent reasoning is a key cognitive strategy in the construction of argument. Current instruments assess dimensions of argument but do not assess cogency. The intent of the instrument in the present study is to provide valid inferences regarding assessment of students' ability to use complex reasoning as they construct cogent arguments. Additionally, this instrument is to be used to capture the variability in students' performance.

The review of the literature is presented in three sections. The first section contains a brief historical overview of the assessment of writing, in particular, how to quantify assessment. In addition, this section examines how the tension between reliability and validity swings back and forth due to the interpretation of a rubric's use in assessing writing. The second section looks at how and why analytic and holistic rubrics were created. The creation of these rubrics was an attempt to address writing assessment. The third section provides a rationale for creating a writing instrument that assesses cogency in argument. This section also presents an in-depth analysis of Toulmin's model of argument, the framework of which informs my instrument on cogency. Additionally, this section attests to the importance warrants play in constructing cogent arguments, which is the thesis of this study. Included are studies that incorporate Toulmin's model of argument in science and mathematics classrooms.

Brief History of Assessment of Writing

The question of what it means to effectively measure writing is one that has historical roots. For much of its history, the assessment of writing has swung back and

forth between the concepts of reliability and validity of instruments (Yancey, 1999). The degree of agreement between raters in writing assessment determined the reliability of an instrument, while in the field of educational measurement, it referred to the degree to which an instrument could yield consistent measures after repeated administrations of the instrument (Huot et al., 2010). For the first 50 years of education and psychological measurement's 100-year history, reliability was elevated above the concept of validity, particularly in the assessment of writing (Greenberg, 1992; Huot et al., 2010; Yancey, 1999). According to Yancey's historical overview of writing, reliability dominated from 1950 to 1970. Reliability referred to interrater reliability (see also Huot, 2002).

During this time period, writing was assessed through administering tests that focused on students' abilities in vocabulary and grammar. These tests were seen as objective measures of writing that enabled teachers and administrators to place students into the proper writing classes (Huot et al., 2010; Yancey, 1999).

Until 1970, there was no unified theory of validity. Measurement researchers either held the view that the types of evidence used to make predictive inferences were integral to the validation process of a test or advocated a unified theory of validation. The unified theory of construct validity became the dominant theory accepted by the measurement community (Anastasi, 1986; Embretson, 2007; Guion, 1977; Messick, 1980). The construct model of validity stressed the idea that the test does not get validated; instead, the interpretations and evidence of the test get validated. What emerged from this model were methodological principles that called attention to transparency and explicitness regarding the proposed interpretations of the test, theory, and alternate interpretations.

From 1970 to 1986, the emergence of rubrics dominated, particularly the practice of using the holistic rubric to score essays. From 1986 to 1999, portfolio assessment was used. Portfolio assessment focused on samples of student writing, which in turn revealed that objective measures of writing did not focus on what instruction of writing was. Grammar and vocabulary did not constitute writing. As a result of a combination of the emergence of rubrics and portfolio assessments, the concept of validity became important in the assessment of writing.

Validity Applied to Writing Assessment

In his 2002 book *(Re) Articulating Writing Assessment*, Huot discussed how the concept of validity in the 1980s went beyond the idea that the validity of a test depended on whether it measured what it was trying to assess. The social consequences of tests created a division in the measurement community. For Messick (1975), social consequences of interpretations of tests were seen as integral and important components of validity. In addition to the inclusion of consequential validity, Shepard (1997) and Linn (1997) advocated that validity must involve an examination and analysis of the test scores. However, incorporating the social consequences of an interpretation of test scores as part of the validity process was not embraced by all (Mehrens, 1997; Popham, 1997a), because it was seen as compromising the integrity of the validation process.

As applied to writing assessment, Huot (2002) noted that a test's validity must rest on its theoretical rationale and from evidence in the form of student essays that aid in providing support for the use and interpretation of test scores. In educational measurement, validity came to be understood as a unitary concept under the the idea that in addressing the social consequences of a test, threats to its validity would result if the

test were not used as an assessment of writing that informs instruction and informs students of their proficiency.

Rubrics

Rubrics were seen as a means by which English teachers could assess writing. The holistic rubric in particular was most commonly used from 1970 to 1986 (Yancey, 1999). The term *rubrics* is synonymous with *scoring guides*. Rubrics are most commonly used to assess writing in an English classroom for Grades K–12. Rubrics provide students with the academic expectations for the writing task by explicitly explaining the academic standards and objectively communicating a student’s academic progress, and as a result, they are thought to promote student learning (Allen & Tanner, 2006; Andrade, 2005; Hafner & Hafner, 2003; Howell, 2014; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). For the purposes of this study, analytic and holistic rubrics were examined because they are the most commonly used in K–12 classrooms (Andrade, 2005; Hafner & Hafner, 2003; Reddy & Andrade, 2010).

Emergence of Analytic Rubrics

Diederich, French, and Carlton (1961) were commissioned by the College Board to conduct a study to determine best practices for producing interrater reliability. This study was instrumental in establishing research in writing assessment, along with the development of essential components of what we now know as the analytic rubric. The study consisted of 53 readers who included college professors, writers, editors, lawyers, and business executives who served as readers of 330 college freshman essays.

Agreement on essay scores was reflected in a low interrater reliability score of .31. The

low scores for interrater reliability underscored the necessity of creating a reliable instrument to measure writing (Greve, Morris, & Huot, 2018; Huot et al., 2010).

An unexpected finding of the study led to the creation of what are currently known as *analytic rubrics*. Factor analysis of the scores revealed that agreement of comments were sorted into five categories: ideas, mechanics, wording, flavor, and form. According to Hampel (as cited in Haswell, 2014), Diederich's intent was for the categories to be used as a common vocabulary to discuss student work. However, a decade later, the five categories became components of the analytic rubric. The holistic rubric soon followed in the 1970s. In addition, this study went beyond examining the lack of consensus among raters toward examining the criteria raters used to score writing. This study was instrumental, because it took into account the idea that writing assessment comprised more than just establishing interrater reliability.

Analytic rubrics place the criteria of the assignment and levels of achievement in a grid. The criteria for the specific writing task are expressed in rows on the left-hand side of the instrument, while the levels of achievement are expressed in columns. In this way, the different characteristics of a writing assignment are individually graded. Students receive specific information regarding their strengths and weaknesses (Allen & Tanner, 2006; Andrade, 2005; Hafner & Hafner, 2003). The negative aspect of this type of scoring is the danger in confirmation bias; that is, the score of one scale may influence a rater to give the writing a similar score on another scale, because the criteria for each category may be vague (Haladyna & Rodriguez, 2013; Nodoushan, 2014).

Interlinear Exercises

Godshalk, Swineford, and Coffman (1966) investigated different ways of establishing validity for the measurement of writing ability. The authors of the study were interested in determining if writing a 20-min essay was a more reliable and valid measurement of a student's writing ability than the practice of answering objective multiple-choice questions, along with the completion of interlinear exercises that consisted of editing passages of writing that contained errors. More than 600 11th- and 12th-grade students participated in the study over a 3-week period. Criteria for the objective portion and interlinear exercises were established by the total scores of 25 readers of the 5 essays, who assigned a score of 3, 2, or 1 to each of the essays. To determine if the fifth essay topic was predictive of a student's writing ability, the scores of the other four essay topics were added up. The next year, 145 teachers were asked to reread two of the essays to compare the ratings teachers assigned under field conditions as opposed to the ratings assigned under the test conditions of the 25 teachers.

One of the most important results of the study was the establishment of reliability for rating essays, which was accredited to several factors: (a) require students to produce more than one essay and (b) require a different reader for each written essay. Yet the most critical finding was the high correlation between the scores of the objective multiple-choice questions of writing and the interlinear exercises with the scores on the writing samples, the qualifier being that the objective multiple-choice questions were only valid when evaluated against the criteria established by the reliability of the written essays. Most importantly, this study established a protocol for scoring writing using a holistic rubric.

Holistic rubrics are one-dimensional, meaning that all the criteria for a specific writing assignment are described in one scale, with levels of achievement typically scaled from 1 to 6. In contrast to analytic rubrics, only one score is assigned. Holistic rubrics present an overall picture of what the writer can do. The feedback communicated to the teacher and student is a picture of what the student can do as opposed to where the student is on the performance continuum. In contrast to analytic rubrics, dimensions of writing are not scored separately; instead, they are collapsed into a single category, and as a result, a clear diagnosis of a student's competency is not presented (Nodoushan, 2014). Lack of specific feedback impedes students' progress.

What Diederich et al.'s (1961) and Godshalk et al.'s (1966) studies both failed to take into account was that a measurement of writing should assess actual writing that goes beyond the ability to correct errors, which was required of students in the completion of interlinear exercises. In addition, criticism came from the National Council of Teachers of English (Braddock, Lloyd-Jones, & Schoer, 1963), which pointed out the necessity of valuing the assessment of ideas and analysis expressed in writing.

The Importance of Construct Validity in Assessment of Writing

Composition teachers believe that for a test to be considered valid, it must contain three pieces of evidence: content validity, construct validity, and predictive validity (Greenberg, 1992). Yet at the same time, the validity of an instrument was for the most part determined by the test's creator (Huot et al., 2010). By the mid-1960s, construct validity had become more relevant as a theory for writing assessment. Construct validity emphasized the need for a theoretical framework for a measure, so that inferences derived from the operationalization of the construct validated the interpretation of scores (Huot et

al., 2010). Thus theories must be well developed to support construct models. The theory must show how it is representative of the construct it is measuring, then it must describe the observable traits linked to the construct (Cronbach & Meehl, 1955). In other words, Cronbach and Meehl's starting point in determining the validity of a test was determining if the construct was supported by the theory. Cronbach and Meehl argued that construct validity should be part of the validation process of tests because it provides a deeper understanding of the underlying theoretical attributes of the construct being measured. For Cronbach, the inferences deduced from a test must be explained by the construct. In this way, Cronbach elevated construct validity above content and criterion (Brennan, 2006).

Yet seminal research on writing assessment (Diederich et al., 1961; Godshalk et al., 1966) has primarily focused on criterion validity. The issue of validity kept rearing its head throughout the many iterations of writing assessment.

Validity assures that the inferences derived from the scores can be used to implement educational policy. To address the most constructive way a rubric may be used, Turley and Gallagher (2008) suggested that educators, administrators, and others in education who formulate policy, along with large-scale assessment vendors, should use four questions to guide them in their use of a rubric. The questions deal with what the rubric will be measuring; whether it will be used in the classroom, by administrators, or for large-scale assessment; who interprets the results; and for what purposes the results will be used. Turley and Gallagher echoed Messick's theory of validity (Huot, 2002; Messick, 1995).

While there are a variety of rubrics, research regarding how they contribute to informing instruction and assessing writing has been mixed, often because the instruments have not shared the same theoretical framework of what constitutes writing as a starting point (Gere, 1980; Haladyna & Rodriguez, 2013). Rubrics that are specific to the performance task are most effective in providing data that inform instruction (Arter, 2000). Criteria must be specifically defined to identify the construct and to present a continuum that describes where the students' proficiency in the construct lies (Haladyna & Rodriguez, 2013; Nodoushan, 2014).

At the same time, there is a lack of theory associated with writing assessment (Gere, 1980). The assessment of writing has placed more emphasis on developing instruments while ignoring the necessity of developing theory on how writing is evaluated. Although analytic and holistic rubrics derive their validity from interrater reliability, neither measures the communicative intention of the writer—the ability to communicate meaning (Gere, 1980).

Toulmin's Model of Argument: The Language of Academia

Argument is the language of academia (Graff, 2003), and writing argument involves evaluating the merit of multiple perspectives to address counterclaims. Toulmin's model of argumentation is the most common model taught at the secondary level (CDE, 2015; Graff, 2003; Kneupper, 1979; Lunsford, 2002) and involves assertion of a claim (what is being argued), evidence, warrants, qualifiers, and rebuttals, if necessary. The purpose of argument is to persuade the reader that the claim the writer has asserted on a topic is valid and therefore must be accepted.

Argument and Reasoning

Argument is viewed as academic writing and uses “key cognitive strategies” that include analysis, interpretation, precision and accuracy, problem solving, and reasoning (Conley, 2007). Prior to writing, the student must engage in research to select evidence in support of the claim. Evidence that is credible includes a variety of sources that are synthesized instead of summarized (Graff, 2003; Hillocks, 2011; Warren, 2010).

Constructing argument develops critical thinking skills because it requires problem solving and critical evaluation of counterargument, which results in new knowledge.

Reasoning that is cogent and evidentiary must support the process (CDE, 2015).

Argument as Applied to Literary Analysis

In a seventh-grade English classroom, students construct argument in response to literary texts and nonfiction literary texts. The merit of a literary text is judged by the importance of the theme the student believes the writer is communicating. Defense of the interpretation of the text requires students to establish criteria to defend the inferences drawn from the texts. For example, in the novel *The Outsiders*, the motif of family is recurring. If a student writes the thesis statement “In the novel *The Outsiders*, S. E. Hinton uses the motif of family to assert that ‘what matters most is what a family contains, and not whether it is a nuclear family,’” then the student must establish the criteria for the conceptual definition of family. The inferences the student draws from the novel and additional literary nonfiction texts must be defended with warrants and backing, appropriate to the argument, and in this way, students write complex arguments that demonstrate reasoning that is evidentiary, to support claims (Hillocks, 2011; Lunsford, 2002; Stygall, 1986; Toulmin, 1958/2003).

Until the adoption of the CCSS (NGA, 2010), writing in an English classroom has traditionally focused on analysis of plot or theme in literary texts in which a student's opinion was supported by evidence from the primary literary text. Now evidence must come from multiple texts that support the construction of an argument, not an opinion. This type of argument can be considered empirical (Wolfe, 2011), which refers back to the idea of evaluating or refuting, in this case, multiple pieces of data in the form of literary fiction, nonliterary fiction, and nonfiction on an issue explicated in a literary text.

Empirical Research on Argument Using Toulmin's Model

Although many valuable articles reflect teacher action research conducted in the classroom on implementing Toulmin's model of argument or teaching argument in general in an English classroom, none have been empirical research studies. However, what emerges from the literature on action research in an English classroom and empirical studies in science and math classrooms is a consensus regarding the importance of the warrant in Toulmin's model—and the difficulty in teaching it (Alcock & Weber, 2005; Keith & Beard, 2008; McCann, 2010; Nardi, Biza, & Zachariades, 2012; Rex, Thomas & Engel, 2010; Stygall, 1986; Warren, 2010; Wilson, 2005). Much of the research on argument concerns studies that have examined how students construct argument in science and math classes, because reasoning and thinking are necessary skills for evaluating or refuting alternate theories (Simon, 2008).

This section discusses several studies that used Toulmin's model to evaluate the reasoning articulated in the warrants that students used to support their justification for selecting evidence used to support their claims. Hillocks (2011) claimed that Toulmin's

framework would provide a better assessment of the reasoning students used to construct cogent arguments.

Role Warrants Play in the Validity of a Mathematical Proof

An important skill required of mathematics students is the ability to validate mathematical arguments to demonstrate that a theorem is true, yet little research has been conducted examining how students go about validating a proof by examining the warrants for their claim.

Alcock and Weber (2005) conducted an exploratory study to examine whether students took the validity of a warrant into account when determining the proof of a flawed mathematical argument. Students were asked to prove that (\sqrt{n}) is valid in presenting a sequence that diverges to infinity. The researchers were interested in (a) understanding how students determined the logic of an argument and (b) how they determined the content used to legitimate the proof as valid in a line-by-line analysis.

In the study, Toulmin's model of argument was used by mapping warrants onto the line-by-line analysis of a mathematical argument. The similarity in both is the idea that every claim supported by evidence must show how the warrants connect the evidence to the claim. A series of assertions make up a mathematical proof, which, as in composing argument for expository essays, requires evidence/data and warrants to support its assertions. The warrants in this study served to illustrate the logical reasoning students used to demonstrate the validity of each assertion. The mathematical proof contained minor errors in the first three lines, while the last line contained a major flaw, because inherent in the first three lines was an invalid warrant. The conclusion that follows an invalid warrant is an invalid argument.

The theorem and mathematical proof were provided to the students so they could determine if the proof needed modifications. Pairs of students were encouraged to work together while the interviewer asked the students questions to clarify their thinking processes as they made modifications. The data from the interviews were coded into three groups. The first group of students identified that the warrants used in the first three lines of the proof were invalid and therefore the proof could not be accepted. In their reasoning, the students noted that the warrant in the proof could not stand up to their counterexamples. The second group of three students noted that the proof was invalid because the definitions for divergence were incorrect, and therefore they could not accept the data. Yet this group of students did not find errors in the last line. The third group of five students accepted the proof, but when the interviewer asked them to examine the proof again, by closely reevaluating the last two lines, students rejected the proof. The reasoning one of the students used was a counterexample, which showed that the warrant was invalid. The last group of two students stated that the proof was valid but could not come up with reasons to support their claims.

Whereas most of the student responses focused on the truth of the proofs, only three students considered whether the warrants were valid in substantiating the mathematical argument. Students were able to complete the process of verifying the proof of an argument but needed prompting to go through the steps to ascertain whether the warrants were valid.

The implications of this study called attention to the importance of teaching students how to identify warrants and to take into consideration how they substantiate the proof of a mathematical argument. To differentiate between the implicit logic of

assertions and the legitimacy of an argument, a warrant must be substantiated to determine its validity.

Toulmin's model of argumentation stresses the important role a warrant plays in providing support for the evidence, so the claim is accepted. The importance of the warrant requires students to evaluate the logic and complex reasoning as a means of determining whether a claim, or a proof, in this case, can be seen as valid.

Scaffolding Argument in the Science Classroom

Previous research (Driver, Newton, & Osborne, 2000) regarding the function of argument and its impact on scientific literacy asserted the necessity for establishing norms for argument in the science classroom. Argument helps to frame knowledge as socially constructed, as opposed to how it has commonly been presented, which is through a positivist approach. Within a positivist framework, the emphasis is on the procedural manner of conducting research that confirms existing theories, while current research embraces construction of knowledge that emphasizes discursive practices. Within this framework, students evaluate the validity of evidence and examine alternate theories, and implicit in this framework is the importance of the reasoning used throughout the process. In this way, students develop an understanding of how they come to know what they know, because they engage in examining the reasoning used to support evidence for a claim. Students develop epistemic beliefs because they use the language and practices of scientific inquiry, which aids them in determining whether the warrants for the claim are valid.

Sandoval and Reiser (2004) developed a framework of scientific inquiry that emphasized argumentation as a central focus of instructional practice to foster scientific

inquiry in students. This framework was used to create Explanation/Constructor, a software tool that would aid students in understanding the explicit linking of explanations to questions posed.

Explanation/Constructor provided domain-specific templates to help students explain how their evidence answered the questions. The domain-specific templates were explanation guides that consisted of scientific theories relevant to the particular questions posed. The authors contended that inquiry-based practices alone do not result in improved conceptual understanding of content. A critical aspect to aid students in constructing argument when engaging in inquiry-based practices is providing learning environments that help students develop these epistemic practices. Argumentation in a science classroom is best at developing epistemic practices. These practices focus students' attention on developing justificatory statements that explain the warrant for their answers. In this way, students develop conceptual understanding of the scientific principles that constitute what is warranted. The study was conducted over a 2-year period in an upper-class, predominantly Caucasian high school located in a suburb of Chicago. The participants for Year 1 ($n = 69$) came from two levels of biology classes: one honors class and two regular classes. For Year 2, the participants ($n = 87$) came from two honors classes and two regular classes of biology.

The software tool Explanation/Constructor presented questions on three investigations: (a) the natural selection process of finches from the Galápagos Islands (GF), (b) the history of how bacteria adapt to antibiotics (TB), and (c) the study of panthers in their natural habitat in North America. For the investigation of the GF, students were presented with the following information on a computer screen: questions

that presented explanations of a causal nature, explanation guides, and a box where students provided rationale for their chosen explanations. The students' explanations had to be linked to one of the explanation guides. Results of the case studies reveal that students constructed epistemic conversations because of the use of domain-specific templates.

In Year 1, over a 4-week period, students studied Darwin's natural selection theory then were responsible for completing lab activities that demonstrated their understanding of the theory.

For Year 2, the focus of the study was to investigate how students evaluated their explanations and those of their peers by considering whether claims made were clear and supported. Revisions to the software tool Explanation/Constructor included providing evidence more easily on the computer screen, which the authors hoped would result in students selecting and citing data into the field labeled "Explanations Text" on the computer screen. To prompt reflection during the process of constructing an explanation, a detailed rubric for assessing students' and peers' explanations was provided. In addition, a feature was added that enabled students to facilitate reviewing the data used as evidence for explanations.

Students' statements were coded and analyzed in the following categories: lack of data ($n = 26$, 23.5%), lack of mechanism ($n = 22$, 20.18%), affirmation ($n = 15$, 13.76%), counterclaim ($n = 11$, 10.09%), lack of information ($n = 11$, 10.09%), say more ($n = 9$, 8.26%), alternative claim ($n = 7$, 6.42%), objection ($n = 3$, 2.75%), suggestion ($n = 3$, 2.75%), and need for warrant ($n = 2$, 1.83%). The results indicate that students used the

rubrics to assess the criteria for peers' explanations of evidence, because the expectations were clearly articulated in the rubrics.

Assessment of the limitations of students' own explanations were coded and analyzed. Students' statements were coded into the following categories: verisimilitude ($n = 34$, 30.91%), alternative claim ($n = 28$, 25.45%), limited extent ($n = 17$, 15.45%), restatement ($n = 17$, 15.45%), ambiguity in data ($n = 6$, 5.45%), question ($n = 3$, 2.73%), assumption ($n = 2$, 1.82%), lack of time ($n = 2$, 1.82%), and absurd alternative ($n = 1$, 0.91%). The results indicate that students did not look at their claims and evidence as a possible reason for the limitations of their explanations but instead looked at the limitations of explanations provided by Explanation/Constructor.

Overall, the results indicate that peer critiques of explanations were most effective in focusing student attention on the validity of evidence selected to support a claim. In this way, reference to Toulmin's model is made in the study's emphasis of the importance of the role of reasoning—that it must be evidentiary to demonstrate how the claim is supported.

Connecting Selection of Evidence to the Claim

Sandoval and Millwood (2005) noted the importance of epistemic tools in aiding the development of reasoning as a means of fostering conceptual understanding of material. When students are required to explain what constitutes a warranted claim, it encourages the use of knowledge grounded in theory of the specific domain. With the use of epistemic tools, students are provided an opportunity to engage in discourse that encourages justification and citing of evidence to support claims. At the same time, the authors noted that the software tool Explanation/Constructor was contextualized to the

specifics of the curriculum to support their reasoning. As a result, epistemic beliefs were developed regarding what constitutes scientific knowledge.

Sandoval and Millwood (2005) expanded on the research conducted by Sandoval and Reiser (2004). In this study, the authors assessed students' warrants for their claims. The authors contended that when Toulmin's model of argumentation is only used to analyze the components of a student's argument, the context-specific nature of the warrant is not taken into account. Yet, it is this aspect of the warrant, as Toulmin noted, that provides validity for an argument. The particular focus of this study was an investigation of how Explanation/Constructor was used to aid students in connecting their selection of evidence for their claims. The goal of this additional feature to the software tool was the idea that the explicit linking of the evidence to the claim would prompt students to provide warrants that justified their selection of evidence that was grounded in the theory of natural selection. The questions that framed students' explanations in the study were as follows: (a) Was there evidence to support claims? and (b) How are the data linked to the claims in students' explanations?

A qualitative research design was conducted over a 4-week period. The participants ($n = 87$) included 43 girls and 44 boys from high school biology classes. Two classes ($n = 52$) were taught at the honors level, while the other two classes ($n = 35$) were taught at the regular track. As in the study conducted by Sandoval and Reiser (2004), students worked on an evolution unit concerning Darwin's natural selection theory. For the first 2 weeks, the study focused on (a) students' explanations for the phenomenon of why a minority of GF survived a catastrophe and (b) why antibiotics became resistant to TB.

Structural analysis was used after students' explanations were coded.

Explanations were placed into the following four categories representative of the four components of the theory: (a) selective pressure, (b) impact of selective pressure, (c) identification of the differential trait in beak size that explains which GFs survived, and (d) the effect of the selective nature of the trait. The goal of this section of the analysis was to examine the quality of arguments and the inclusion of warrants. Next, the qualities of the warrants were analyzed to determine whether the data provided were sufficient and relevant in justifying students' claims. Finally, rhetorical references to cited data were analyzed to examine how students used the data to assert their claims, for example, did they make specific reference to graphs or tables in their explanations?

The findings for conceptual quality of explanations show that while the majority of students were able to provide explanations for the theory of natural selection for the GF investigation, only half were able to explain a component of the theory of selective advantage for the TB investigation. Both groups demonstrated an understanding of selective pressure and its impact, yet for the components of differential traits and the effect of the selective nature of the trait, students did not provide many warrants. The overall scores for the GF investigation, on a scale from 0 to 8, had a mean of 6.83 with a standard deviation of 0.75, while the overall scores for the TB investigation had a mean of 6.90 with a standard deviation of 1.72.

For sufficiency of data, citation of data was very low (73% of $n = 22$ groups) for the GF investigation. The overall score for this investigation had a mean of 1.61 with a standard deviation of 0.52. For the TB investigation, most students cited data to justify their claims ($M = 11.95$; $SD = 0.70$). For both problems, there was a strong correlation

between the sufficiency of data and the warrant. For the GF investigation, Spearman's $\rho = .677, p < .001$, while for the TB investigation, Spearman's $\rho = .772, p < .001$. The scores for the data show that when students understood the data, warranted claims were supported with relevant data.

The findings and implications of this study echo Toulmin's framework in confirming the necessity of analyzing the quality of a student's argument as well as its structure. In so doing, the importance of context-specific warrants in the argument is highlighted. As a result, this provides stronger support for the claim.

Understanding of how evidence supports a claim requires that students determine the reliability and validity of the evidence. Examining the evidence in this way helps develop reasoning skills that aid students in evaluating the evidence used to support their arguments.

Roberts and Gott (2010) implemented a module of instruction that explicitly taught concepts of evidence. The design of the study was exploratory and descriptive. Participants included 65 students who ranged in age ($n = 19-45, M = 21.06$ years). Over a 9-week period, the evidence module used in this study emphasized a procedural understanding of the concepts of evidence and sought to answer whether this module, along with the addition of open-ended investigations, would promote the transfer of the procedural concepts of evidence toward questions that focused on the validity of the claim. The procedural concepts included Toulmin's model of argument, which focused primarily on warrants, backing, qualifiers, and rebuttals. Inherent to procedural understanding are the reliability and validity of the procedures used to obtain and verify the evidence.

The concepts of evidence were grouped into two subsets or layers. The first subset or layer (Layers A–C) concerned data collection and included measurement and repeated measurement of a single variable to establish the validity of the claim made in the investigation. The validity of the design of the instrument must support the relationship of the variables, because this relationship will impact reliability and in turn any interpretations deduced from the relationship. Layers A–C represent what the authors noted as core in developing a procedural understanding of the concepts of evidence. The core concepts of evidence are noted as being influenced by previous research and were delineated in Layers D–F.

Pre- and posttests were administered, which presented socioscientific scenarios. Socioscientific scenarios are real-world issues that deal with moral and ethical consequences of a decision. The example used in the pre- and posttests was the same. Students were provided with a presentation that concerned a local cement company's request to switch to a fuel that burned questionable chemicals. Next, students were provided with a summary of the issue. The writing prompt asked students to write a letter that addressed concerns regarding the safety issues of the request made by the cement company. The letter should keep in mind what types of questions residents might raise concerning the request to switch fuel. The time frame to complete the written performance was 10 min.

Questions about the claim of safety were coded using Toulmin's model. Student responses were coded into the following categories: all questions, broader socioscientific debate, substantive backings, all procedural questions, data, procedural warrant, procedural backings, qualifiers, and rebuttals. The categories with the three largest effect

sizes were ‘all procedural questions,’ ‘procedural backings,’ and ‘broader socioscientific debate.’

The findings of the study show that evidence-based questions helped students raise questions about the claim using Toulmin’s model of argumentation. As a result, the importance of the warrant was more clearly demonstrated. The implications of this study refer back to Toulmin’s model, because the cogency of an argument includes evidence and reasoning that are appropriate to the argument field, otherwise known as the subject domain. Therefore we see that a teacher’s direct instruction in selecting evidence would aid students in constructing cogent arguments.

Summary

In general, Toulmin’s model of argumentation has been popular in science and math classrooms as a theoretical framework for analyzing how evidence supports a claim because of the model’s focus on the warrant (Alcock & Weber, 2005; Bottcher & Meisert, 2011; Sandoval & Millwood, 2005; Simon, 2008; von Aufschnaiter, Erduran, Osborne, & Simon, 2008). Therefore, in this study, Toulmin’s model was used to establish criteria for each descriptor level of cogency and was articulated in a new instrument for writing assessment called the TDRMC. As discussed in the historical overview, validity of the assessment is diminished when context is not included, which in this case refers to the context and theory of writing assessment. The discussion of rubrics referred to the goal of this study in addressing vague descriptors.

Criteria of the construct must be delineated at each descriptor level to inform students of their proficiency levels on a continuum of the construct. In addition, when criteria are specific at each descriptor, feedback on students’ progress helps to inform

teacher instruction. Finally, as discussed in empirical studies of Toulmin's model, the importance of constructing a cogent argument was demonstrated in the use of warrants to provide support for the selection of evidence, which in turn strengthened the claim.

CHAPTER 3: METHODS

The purpose of this study is to create an instrument that validly and reliably measures the construct of cogency, which is one dimension of the multidimensional construct of argument. The instrument includes a construct map, a prompt, a scoring guide, and student responses. The overall construct of cogency, a dimension of the multidimensional construct of argument used in the study, is based on Toulmin's model of argumentation, which is identified by the CCSS (NGA, 2010) and guides the teaching of argument in high schools and universities in the United States (CDE, 2015; Graff, 2003; Lunsford, 2002; Kneupper, 1979).

This study employed the BEAR assessment system (Wilson & Sloane, 2000) to design an instrument to measure cogency. It is a framework rooted in a theory of cognition that views student proficiency within a particular subject domain as a progression of learning along a continuum (Brown & Wilson, 2011). The BEAR assessment system (Wilson & Sloane, 2000) integrates formative assessment into the curriculum to provide meaningful feedback to students as they are learning rather than providing only a single summative assessment (Wilson & Scalise, 2006). Feedback that is explicitly related to the learning goals of a task is most effective in aiding student learning and promotes growth by moving the student forward (Cheesebro, 2003; Civikly, 1992; Shepard, 2005; Timperley & Parr, 2009; Trauth-Nare & Buck, 2011). Feedback that contains a clear idea of the learning goal, gives an understanding of where the student is on the continuum, and closes the gap between misconception and action is the most useful (Sadler, 1989). Therefore a reliable and valid measure will be informative for both the student and the teacher. Students will obtain a clearer picture of their level of

proficiency and will additionally have an understanding of what they need to do to move forward. Teachers will also have a better understanding of instruction to aid students in moving them forward.

Research Questions

1. What is the reliability of the TDRMC, and how does it compare to the reliability of the SBAC?
2. Are scores generated from the TDRMC more variable than scores generated from the SBAC?
3. How does the TDRMC correlate to other established measures of writing and academic ability, such as grades, various SBAC scores, and GPA, and how do these correlations compare with those of SBAC Writing and established measures of writing and academic ability?

Research Design

This study employed a mixed methods approach guided by the BEAR assessment system to develop an instrument to measure cogency. The method was an iterative process used to design the TDRMC, and data were collected in four phases. Each phase of the design of the instrument informed the next and supported the reliability and validity of the instrument, as Table 2 illustrates (Wilson, 2005). The four phases are known as the *four building blocks*, as outlined in Wilson's framework for assessment. Table 2 provides an overview of the procedures for this study according to the building blocks of Wilson's design.

Table 2

Iterative Process for Design of Instrument of Cogency

Building block	Iterative step
1. Construct map	1. Review of the literature of argument 2. Creation of a construct map of cogency
2. Item design	3. Essay prompt administered to 73 Grade 7 students in an English classroom
3. Outcome space	4. Rubric/scoring guide to identify qualitative levels of student proficiency in writing argument (cogency)
4. Measurement model	5. Determine if student responses can be mapped back to the construct map for cogency

Sample

Approximately 73 students from a larger sample of students I taught ($n = 125$) in a seventh-grade English class in a public middle school located in northern California took part in this study. This convenience sample captured a range of academic proficiency, diversity of students according to gender, and socioeconomic status representative of the school district.

The demographics of the school district were as follows: Latino, 10%; African American, 2%; Asian American, 34%; and White: 46%. The ELL population is 12%, 11% are students with disabilities, and 9% are socioeconomically disadvantaged. The percentage of students receiving free and reduced-price lunch combined is 9.4%, which accounts for a total of 1,163 at my school site out of the total district enrollment of 12,399 students.

Protection of Human Subjects

I submitted an application for this dissertation study to the University of San Francisco Institutional Review Board for the Protection of Human Subjects before I implemented the instrument used in this study (Appendix A). Individuals' identities were concealed along with any other identifying information. Parent approval consent forms, which also provided the opportunity for parents to opt out of having their children participate, were sent home (Appendix B). I provided contact information in case parents had any questions about the study. Rater approval consent forms (Appendix D) were obtained from my colleagues who participated in the moderation session. Individuals' identities were concealed. Additionally, I obtained site approval (Appendix C).

Instrument Generation

Building Block 1: Construct Map

The first step in measuring cogency involves creating a construct map, which in this study is based on Toulmin's model of argumentation. The most relevant aspect of the model with regard to the current study is the emphasis on students' ability to use reasoning to support the evidence for their claims. This skill is demonstrated in their use of warrants and backing, which I have defined as cogency, and is presented in the construct maps. I have developed the construct map used in this study based on Toulmin's model of argumentation because this model serves as the best possible framework for building a measure for cogency, a dimension of argument because of its emphasis on the context-specific nature of the warrants (Toulmin, 1958/2003).

The construct map guides the operationalization of the construct; that is, the idea behind a construct map is to create a visual representation of the construct that describes

the expectations of reasoning for each level of student performance. The cognitive theory being measured should be represented in the construct map (Wilson, 2005). A full understanding of the theory behind the variable measured informs the construct map's construction. A complete construct map indicates the qualitative levels of the trait. The levels of proficiency are ordered from high to low such that the construct map fully delineates a continuum of proficiency. According to Wilson, this type of measurement looks at student learning as falling into levels of proficiency, in contrast to assessment, which gives a fixed and limited characterization resulting in a simple understanding of whether or not students "get it." Wilson's framework instead informs a teacher of the level of proficiency. The construct map is subject to clarification and input from experts in the field and is revised according to the experts' feedback. The process of refining the instrument is continual, extending beyond the initial design of the construct map. My construct map described the level of difficulty in the construction of cogency (on the left side) and the kind of observed responses associated with each level (on the right) along the continuum of the cogency construct (Wilson, 2005). After I created my construct map, I solicited feedback from my expert panel to determine if the map was representative of the construct.

Figure 7 describes the levels of the construct of cogency. The levels present the learning trajectory of the construct and provided the basis for the scoring guide for measuring cogency.

Building Block 2: Item Design

The scoring guide was aligned with the qualitative levels of the responses delineated in the construct map. On the left side of Figure 8 is the scoring guide, which

was designed directly from the construct map; on the right side is the SBAC rubric, measuring the dimension of “Elaboration of Evidence,” which is aligned to the dimension of cogency on the TDRMC scoring guide. The construct of cogency was defined as a combination of warrants and backing from Toulmin’s model of argumentation. Figure 8 is a generic scoring guide, whereas Figure 9 is an example of the scoring guide as it pertains to the unit on the novel *The Outsiders*, in response to the theme of “what constitutes a family.”

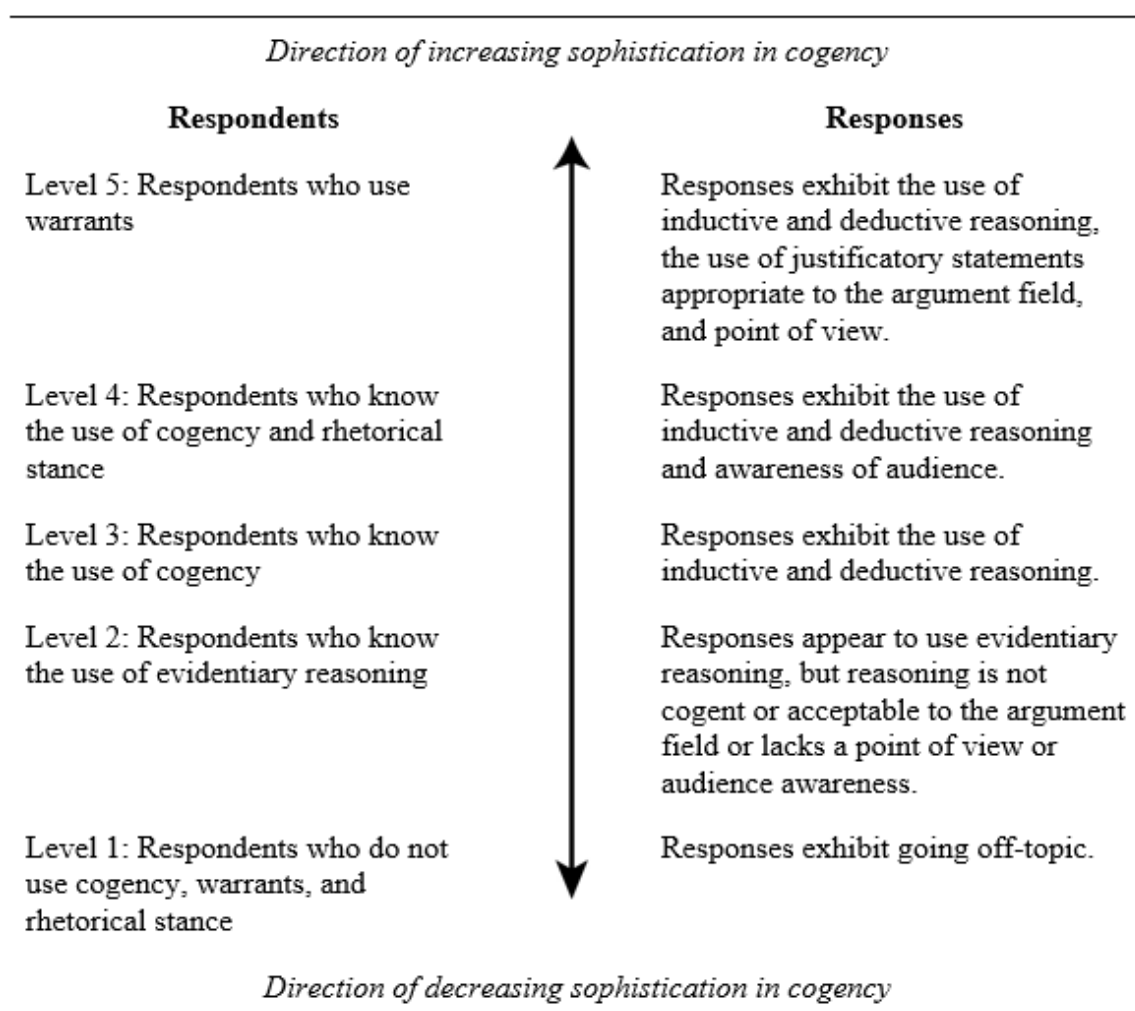


Figure 7. Construct map for cogency.

Teacher Designed Rubric Measuring Cogency	Smarter Balanced Rubric Measuring Elaboration of Evidence
4-Point Argumentative Performance Task—Writing Rubric (Grades 6–11)	
<p>LEVEL 5 Deductive and Inductive Reasoning The response demonstrates the ability to define theme as it is defined traditionally (<i>deductive</i>) and the ability to re-define (deductive) the theme that utilizes an analysis of examples used in the text. (<i>inductive</i>) Justificatory Statements: Selection of evidence and use of warrants The selection of evidence is relevant, and appropriate to argument field, and the warrants provide evidentiary reasoning in support of the claim. Awareness of Audience The backing includes underlying values and assumptions for the criteria of the warrants.</p>	<p>4 POINTS The response provides thorough and convincing support/evidence for the argument(s) and claim that includes the effective use of sources (facts and details). The response clearly and effectively expresses ideas, using precise language:</p> <ul style="list-style-type: none"> • comprehensive evidence from sources is integrated; references are relevant and specific • effective use of a variety of elaborative techniques* • vocabulary is clearly appropriate for the audience and purpose • effective, appropriate style enhances content <p>* Elaborative techniques may include the use of personal experiences that support the argument(s).</p>
<p>LEVEL 4 Justificatory Statements: Selection of evidence and use of warrants The selection of evidence is relevant, and appropriate to argument field, but the warrant does not clearly provide evidentiary reasoning in support of the claim. Awareness of Audience The backing includes some underlying values and assumptions for the criteria of the warrants.</p>	<p>3 POINTS The response provides adequate support/evidence for the argument(s) and claim that includes the use of sources (facts and details). The response adequately expresses ideas, employing a mix of precise with more general language:</p> <ul style="list-style-type: none"> • adequate evidence from sources is integrated; some references may be general • adequate use of some elaborative techniques • vocabulary is generally appropriate for the audience and purpose • generally appropriate style is evident
<p>LEVEL 3 Justificatory Statements: Selection of evidence and use of warrants Some but not all of the evidence selected is relevant, or appropriate to the argument field. The warrants provide incomplete evidentiary reasoning in support of the claim. Awareness of Audience The backing does not include underlying values and assumptions for the criteria of the warrants.</p>	<p>2 POINTS The response provides uneven, cursory support/evidence for the argument(s) and claim that includes partial or uneven use of sources: (facts and details). The response expresses ideas unevenly, using simplistic language: The response provides minimal support/evidence for the argument(s) and claim that includes little or no use of sources (facts and details). The response’s expression of ideas is vague, lacks clarity, or is confusing:</p> <ul style="list-style-type: none"> • some evidence from sources may be weakly integrated, imprecise, or repetitive; references may be vague • weak or uneven use of elaborative techniques; development may consist primarily of source summary or may rely on emotional appeal • vocabulary use is uneven or somewhat ineffective for the audience and purpose • inconsistent or weak attempt to create appropriate style
<p>LEVEL 2 Justificatory Statements: Selection of evidence and use of warrants The selection of evidence is not connected to the argument field and the warrants do not provide evidentiary reasoning in support of the claim Awareness of Audience The backing does not include underlying values and assumption for the criteria of the warrants.</p>	<p>1 POINT The response provides minimal support/evidence for the argument(s) and claim that includes little or no use of sources: (facts and details). The response’s expression of ideas is vague, lacks clarity, or is confusing:</p> <ul style="list-style-type: none"> • evidence from the source material is minimal or irrelevant; references may be absent or incorrectly used • minimal, if any, use of elaborative techniques; emotional appeal may dominate • vocabulary is limited or ineffective for the audience and purpose • little or no evidence of appropriate style
<p>LEVEL 1 Off-topic</p>	<p>0 POINTS</p> <ul style="list-style-type: none"> • Unintelligible • In a language other than English • Off-topic • Copied text • Off-purpose

Figure 8. TDRMC scoring guide measuring cogency and SBAC rubric measuring elaboration of evidence.

Teacher Designed Rubric Measuring Cogency	Smarter Balanced Rubric Measuring Elaboration of Evidence
4-Point Argumentative Performance Task—Writing Rubric (Grades 6–11)	
<p>LEVEL 5 <i>Deductive and Inductive Reasoning</i> The response demonstrates the ability to define family as it is defined traditionally (<i>deductive</i>) and the ability to re-define family that utilizes an analysis of examples used in the text (<i>inductive</i>). <i>Justificatory Statements: Selection of evidence and use of warrants</i> The selection of evidence is relevant, such as actions the characters take that describe what a family is, or thoughts described that reflect family bonds, and is appropriate to the argument field. The warrants provide evidentiary reasoning in support of the claim. <i>Awareness of Audience</i> The backing includes underlying values and assumptions for the criteria of family so the warrants are understandable to the audience, by providing any background information that is necessary.</p>	<p>4 POINTS The response provides thorough and convincing support/evidence for the argument(s) and claim that includes the effective use of sources (facts and details). The response clearly and effectively expresses ideas, using precise language:</p> <ul style="list-style-type: none"> • comprehensive evidence from sources is integrated; references are relevant and specific • effective use of a variety of elaborative techniques* • vocabulary is clearly appropriate for the audience and purpose • effective, appropriate style enhances content <p>* Elaborative techniques may include the use of personal experiences that support the argument(s).</p>
<p>LEVEL 4 <i>Justificatory Statements: Selection of evidence and use of warrants</i> The selection of evidence is relevant, such as actions the characters take that describe what a family is, or thoughts described that reflect family bonds. As a result, the selection of evidence is appropriate to the argument field. The warrant does not clearly provide evidentiary reasoning in support of the claim because background information is not provided. <i>Awareness of Audience</i> The backing includes some underlying values and assumptions for the criteria of the warrants.</p>	<p>3 POINTS The response provides adequate support/evidence for the argument(s) and claim that includes the use of sources (facts and details). The response adequately expresses ideas, employing a mix of precise with more general language:</p> <ul style="list-style-type: none"> • adequate evidence from sources is integrated; some references may be general • adequate use of some elaborative techniques • vocabulary is generally appropriate for the audience and purpose • generally appropriate style is evident
<p>LEVEL 3 <i>Justificatory Statements: Selection of evidence and use of warrants</i> Some but not all of the evidence selected is relevant, and appropriate to the argument field, such as actions the characters take that describe what a family is, or thoughts described that reflect family bonds. The warrants provide incomplete evidentiary reasoning in support of the claim. <i>Awareness of Audience</i> The backing does not include underlying values and assumptions for the criteria of the warrants.</p>	<p>2 POINTS The response provides uneven, cursory support/evidence for the argument(s) and claim that includes partial or uneven use of sources: (facts and details). The response expresses ideas unevenly, using simplistic language: The response provides minimal support/evidence for the argument(s) and claim that includes little or no use of sources (facts and details). The response’s expression of ideas is vague, lacks clarity, or is confusing:</p> <ul style="list-style-type: none"> • some evidence from sources may be weakly integrated, imprecise, or repetitive; references may be vague • weak or uneven use of elaborative techniques; development may consist primarily of source summary or may rely on emotional appeal • vocabulary use is uneven or somewhat ineffective for the audience and purpose • inconsistent or weak attempt to create appropriate style

Figure 9. TDRMC scoring guide measuring cogency and SBAC rubric measuring elaboration of evidence, as pertaining to *The Outsiders*.

LEVEL 2**Justificatory Statements: Selection of evidence and use of warrants**

The selection of evidence is **not relevant** such as the actions the characters take that **describe what a family is, or thoughts described that reflect family bonds, and is not appropriate to the argument field**. As a result, the warrants **do not** provide evidentiary reasoning in support of the claim

Awareness of Audience

The backing **does not** include underlying values and assumption for the criteria of the warrants.

LEVEL 1

Off-topic

1 POINT

The response provides minimal support/evidence for the argument(s) and claim that includes little or no use of sources: (facts and details). The response's expression of ideas is vague, lacks clarity, or is confusing:

- evidence from the source material is minimal or irrelevant; references may be absent or incorrectly used
- minimal, if any, use of elaborative techniques; emotional appeal may dominate
- vocabulary is limited or ineffective for the audience and purpose
- little or no evidence of appropriate style

0 POINTS

- Unintelligible
- In a language other than English
- Off-topic
- Copied text
- Off-purpose

Figure 9 (continued)

Building Block 3: Outcome Space

The outcome space is the scoring guide, which delineates qualitative levels of cogency evidenced in the writing performance task. To address interrater reliability, Wilson (2005) recommended training raters along with a monitoring system that would provide feedback to the raters to ensure calibration. Training and monitoring consist of providing raters with a conceptual understanding of the construct, having them grade the writing samples that demonstrate a wide range of ability, providing opportunities for raters to discuss and justify their scores, and providing feedback to the raters regarding their efforts to calibrate. This is known as *moderation*. My role was to facilitate the discussion with the raters. Sample student responses were added as exemplars to the scoring guide at each of the qualitative levels of the feature of cogency categorized on a hierarchical level to guide scoring. Providing an opportunity for the raters to discuss the essays helped them to calibrate because they were provided an opportunity to discuss their rationale for scoring student papers at a particular level, and in this way, they more

carefully chose writing samples that were reflective of the construct delineated in the scoring guide (Kennedy, 2005).

Building Block 4: The Measurement Model

The measurement model in the traditional BEAR assessment system uses item response theory (IRT) to identify if the item difficulties are accurately mapped according to the underlying theory. In this study, I was unable to use IRT because I only have two item scores per student (one for the SBAC rubric and one for my scoring guide). However, I needed to look at the distribution of scores from my sample to see if they represented the range of the levels across the construct map. This provided support for the levels created on the construct map. If I did not have any scores at some level, one could have questioned if the level was necessary. Using correlations allowed me to check for external validity.

Construct Validity

In this study, I conducted a review of the literature on cogency, and this review formed the basis of the construct map and scoring guide. An expert panel provided feedback on the construct map and scoring guide to ensure construct validity. The expert panel consisted of two former executive directors of the National Council of Teachers of English. One of the two is currently the English education program coordinator at Teachers College, Columbia University. The second member of the expert panel was a cofounder of the Bay Area Writing Project, which was a precursor to the National Writing Project. This panelist was also a consultant with the Institute for Research on Teaching and Learning and a senior research fellow at the Institute for Standards, Curricula, and Assessment.

Procedures

From approximately November 7 to December 14, this study took place. Students read the novel *The Outsiders* from October 10 to November 4. In addition, students read the *New York Times* article “The Changing American Family.” Furthermore, students had an opportunity to view the film version of the novel. The procedures for this study involved three timelines, as outlined in Table 3.

Table 3

Data Collection Procedures for the Study

Date range	Tasks
Nov. 7–Dec. 5, 2016	<p>Nov. 7–9: Read <i>New York Times</i> article and annotate materials</p> <p>Nov. 10–15: In-class instruction of Toulmin’s model of argument</p> <p>Nov. 16–21: Students provided with the prompt, then write their rough drafts in class</p> <p>Nov. 22: Students submit rough draft</p> <p>Dec. 5: Students submit final draft.</p>
Dec. 6–16, 2016	<p>First 10 essays scored by the author of this study, first with the SBAC rubric and then with the TDRMC scoring guide</p>
Feb. 12, 2017	<p>Trained three raters on SBAC and TDRMC and facilitated moderating session for raters to calibrate and identify exemplars for each level (~7 hours)</p>

Phase 1: Instructional Procedures

Over the course of 3 days, and following the reading of the primary text, *The Outsiders*, students read the secondary text from the *New York Times*, titled “The Changing American Family.” The *New York Times* article was read as a whole class with discussion to clarify any concepts or material in the article. The purpose of whole-class discussion was to aid students in annotating essential concepts.

Prior to presenting the prompt, I devoted one class period of 55 min to establishing what students believed constitutes a family, because the theme of family was the focus of the unit.

Instruction of Toulmin’s Model of Argument

The following instruction and class discussion took place in three 55-min class periods. A slide show was presented to students that explained each component of Toulmin’s model. Particular attention was devoted to the role warrants play in the model for each set of examples with which students had an opportunity to practice. This activity took place in whole-class discussion. For example, I would display a claim supported with evidence and several examples of warrants on the SmartBoard. In groups of four, students would come up with the warrant that was most appropriate to the argument field for the claim. Once students demonstrated their understanding of warrants, I presented the prompt for the essay.

Administering the Prompt and Writing the Essay

The prompt I administered to students stated,

In the novel *The Outsiders*, S. E. Hinton asserts that what matters most is what a family contains, not what it looks like. Do you agree with her depiction of

family? Take a position and support it with the following materials: the novel *The Outsiders*—S. E. Hinton, and one essay from the *New York Times* titled “The Changing American Family.”

The *New York Times* article presented an overview of the changing structure of the American family.

Students wrote their rough drafts in class over five class periods of 55 min. In this way, I was able to provide assistance in their construction of argument to address the prompt. After students submitted their rough drafts for my feedback, students made revisions and submitted final drafts. I graded and posted the essays.

After I posted the grades, I read a letter aloud to each class asking for parental consent to use student essays in this study. Students were told that their grades would not be impacted whether their parents agreed to participate or opted out of the study.

Phase 2: Moderation Session

In preparation for the moderation session, I randomly selected a sample of 30 essays, then I scored five essays first with the SBAC rubric, then the same set of five essays with the TDRMC scoring guide as a way of identifying exemplars for use during the calibration protocol session.

Phase 3: Calibration Protocol

This session was convened to establish interrater reliability. The training sessions for the SBAC rubric and the TDRMC scoring guide, along with calibration of exemplars, were audiotaped and transcribed (see Appendices E and F). A total of four raters were present. Three of the raters, including me, were middle school English teachers, and the

fourth rater was a teacher on special assignment who provided instructional guidance to secondary English teachers in the school district.

For approximately 2.5 hours, the raters were trained on how to use the SBAC rubric in the category “Elaboration of Evidence” with the identified exemplars. During this time period, the raters scored five exemplars with the SBAC, then we compared scores and had a discussion when there was widespread disagreement on the scores. We followed the same procedure for the TDRMC scoring guide.

After we reached agreement and clarity regarding how to use the SBAC rubric, we silently scored 10 essays at a time. We compared our scores and discussed any widespread discrepancies. To determine reliability, percentage agreement was calculated for the first set of 10 essays. This process of scoring, discussing, and calculating percentage agreement was then repeated two more times. We scored a total of 30 essays in this manner using the SBAC rubric.

After lunch, the raters were trained on how to use the TDRMC using identified exemplars. During this time, we made revisions to the wording of the TDRMC. This session took approximately 1.5 hours, then we had to stop for the day. Each rater had the same set of 30 papers to grade using the TDRMC scoring guide. At the end of the day, I emailed a revised version of the TDRMC on which we had all agreed. The grades and essays were handed back to me within a week of the moderation session.

Data Analysis

SBAC and TDRMC scores for the full sample of 73 essays were assigned by the rater with the highest interrater reliability score during the moderation session. These data were used for the following analyses.

To analyze the first research question (What is the reliability of the TDRMC, and how does it compare to the reliability of the SBAC?), an account of what happened in the calibration scoring process was described via an audiotape of the calibration protocol process and then transcribed. During the calibration protocol, the TDRMC was revised to ensure fidelity of the construct. In addition, percentage agreement from the calibration scoring session was obtained by first scoring 30 papers with the SBAC, then scoring the same 30 papers with the TDRMC. Interrater reliability was measured using Cronbach's alpha.

To answer the second research question (Are scores generated from the TDRMC more variable than scores generated from the SBAC?), I calculated the standard deviations of SBAC and TDRMC scores and then compared the variance from the dependent samples using the Morgan–Pittman test.

To analyze the third research question (How does the TDRMC correlate to other established measures of writing and academic ability, such as grades, various SBAC scores, and GPA, and how do these correlations compare with those of SBAC Writing and established measures of writing and academic ability?), I correlated SBAC and TDRMC scores with real-world measures of academic ability. These measures included grades, various SBAC scores, and GPA. To determine which instrument had a higher correlation with the real-world measures, I conducted a test for the significance of the difference between dependent correlations.

Summary

The purpose of this study was to design an instrument to measure the cogency of an argument. A cogent argument was defined as using context-specific warrants as

articulated in Toulmin's model of argument. The study used a mixed methods design with a convenience sample of middle school students. Students were instructed on how to use Toulmin's model of argument as they constructed essays in response to a prompt for a unit on the novel *The Outsiders* by S. E. Hinton. Student essays were scored with the SBAC rubric and the TDRMC scoring guide. The TDRMC was revised during the calibration protocol of the moderation session. Interrater reliability was established during the moderation session. The rater with the highest item-total correlation was chosen to score all essays once with the SBAC and again with the TDRMC. The subset of 30 essays was put back into the larger sample of essays to total 73 essays. Variance of scores was determined by comparing the standard deviations of SBAC and TDRMC. Correlating the scores of the TDRMC and SBAC determined which instrument had better external validity.

CHAPTER 4: RESULTS

The purpose of this study was to create a valid and reliable instrument (TDRMC) to measure cogency in students' writing. This measure's reliability, variability, and validity were compared to those of an established measure, the SBAC. The moderation session established interrater reliability for the TDRMC and SBAC instruments and informed modifications to the TDRMC instrument. In addition, I explored in this study whether the TDRMC scoring guide measuring cogency generated more variability in scores than the SBAC. Finally, I examined the correlations between the TDRMC and other established measures of writing and academic achievement.

Research Question 1

Research Question 1 asked, What is the reliability of the TDRMC, and how does it compare to the reliability of the SBAC? Establishing the TDRMC's reliability involved several steps: (a) holding a moderation session to establish interrater reliability for both instruments, (b) revising the TDRMC during the moderation session to ensure the instrument was reflective of the construct of cogency, (c) calculating the percentage agreement on scores, and (d) calculating the measure's reliability using Cronbach's alpha.

Moderation Sessions

Moderation Session for SBAC

The first step in the moderation session was a calibration protocol ensuring that all raters agreed on what constituted Levels 2–4 for the SBAC instrument. Three teachers and I scored 30 student essays, which were a subset of the larger sample of

student essays ($N = 73$) for this study. The focus was on agreement about what constituted writing performance on each of the levels, on a scale of 1 to 4.

The beginning of the calibration session involved a review of the SBAC instrument. We all agreed that a Level 2 represented uneven cursory support for an argument and claim. In particular, the integration of sources (evidence) was weak and consisted primarily of a summary of the evidence (Appendix E).

We discussed the SBAC's description of what constitutes a Level 3 using an exemplar identified as a 3. We all agreed that the response provided adequate as opposed to comprehensive support. While evidence for the argument and claim included the use of sources, such as facts and details, it employed a mixture of precise with more general language. We all agreed that qualifiers such as adequate versus comprehensive elaborative techniques differentiated a 3 from a 4 (Appendix E).

We discussed the SBAC's description of what constitutes a Level 4 using an exemplar identified as a 4. To score a 4, the response demonstrated the use of precise language, evidence from sources was integrated, and references were relevant and specific (Appendix E).

To determine the percentage agreement among the raters for the SBAC, we scored the 30 student essays. The SBAC instrument yielded a percentage agreement of .50 for the first set of 10 essays, .52 for the second set of 10 essays, and .52 for the third set of 10 essays. Cronbach's alpha for the SBAC was .741.

To determine which rater had the highest agreement with other raters, I correlated the raters' scores for each instrument. The raters' item-total correlations with each

other's scores ranged from .526 to .601 for the SBAC. In later analysis, I used the scores from the rater with the highest item-total correlation (Rater 4).

Moderation Session for TDRMC

On the same day as the session to calibrate the SBAC, a calibration protocol session was conducted with the same three teachers and me for the TDRMC instrument. Because this was a new measure, the calibration session included discussions and changes to the wording of the TDRMC instrument to clarify the instrument for Level 5.

Revisions to Level 5. For a Level 5 paper, the raters agreed that the section of the TDRMC titled "Awareness of Audience" primarily focused on establishing the criteria for the warrant in constructing an argument (Appendix F). The language changed from "Backing includes underlying values and assumptions for the criteria of the warrants" to "Backing includes some underlying values and assumptions for the criteria of the warrants."

Revisions to Level 4. The language for a Level 4 paper changed for the section titled "Awareness of Audience" from "The backing includes some underlying values and assumptions for the criteria of the warrants" to "Backing includes mostly clear and consistent underlying values and assumptions for the criteria of the warrants" (see Appendix F).

Revisions to Level 3. In the language for a Level 3, changes were made to the sections titled "Justificatory Statements" and "Awareness of Audience" (see Appendix F). For the section titled "Justificatory Statements," the language changed from "Some but not all of the evidence selected is relevant or appropriate to the argument field. The warrants provide incomplete evidentiary reasoning in support of the claim" to "Some but

not all selection of the evidence is relevant and is thereby not appropriate to the argument field. The warrants provide incomplete evidentiary reasoning in support of the claim.”

We qualified the language to describe if the selection of evidence was not relevant, in which case, it cannot be appropriate to the argument field. For the section titled “Awareness of Audience,” the language changed from “The backing does not include underlying values and assumptions for the criteria of the warrants” to “Backing includes unclear and inconsistent underlying values and assumptions for the criteria of the warrants.” A Level 3 demonstrated incomplete evidentiary reasoning.

Revisions to Level 2. The description for a Level 2 changed for the sections titled “Deductive and Inductive Reasoning” and “Awareness of Audience” (see Appendix F). For the section titled “Deductive and Inductive Reasoning,” the language changed from “The writer demonstrates the ability to define the theme as it is defined traditionally (*deductive*) and the ability to re-define the theme that utilizes an analysis of examples used in the text (*inductive*)” to “The writer does not demonstrate the ability to define the theme as it is defined traditionally (*deductive*) and does not demonstrate the ability to re-define the theme that utilizes an analysis of examples used in the text (*inductive*).” The language for the section titled “Awareness of Audience” changed from “Backing does not include underlying values and assumptions for the criteria of the warrant” to “Backing may or may not include underlying values and assumptions for the criteria of the warrants.” The language regarding a student’s ability to define and redefine theme was impacted by the students’ inability to include underlying values and assumptions for the criteria of the warrants in the student’s construction of argument.

Establishing Interrater Reliability

The percentage agreement among the raters for the TDRMC was .62 for the first set of 10 papers, .57 for the second set of 10 papers, and .72 for the third set of 10 papers. Cronbach's alpha for the TDRMC was .816. The results show more agreement among the raters with the TDRMC instrument. This level of agreement is higher than the same group of raters had with the SBAC, $\alpha = .741$.

The raters' item-total correlations with each other's scores ranged from .398 to .787. The results for establishing an analysis of each rater's reliability show that Rater 4 was the most reliable, which allowed me to enter the scores from the subset of 30 papers from Rater 4 back into the larger pool of students' scores.

Research Question 2

Research Question 2 asked, Are scores generated from the TDRMC more variable than scores generated from the SBAC? Figures 10 and 11 show the distribution of SBAC and TDRMC scores. The results from both SBAC and the TDRMC show students scored most frequently at a Level 3; however, the TDRMC appears to have produced more variability in scores.

To determine if the TDRMC did yield more variability in scores than the SBAC, I first compared the measures by comparing the standard deviations. Table 4 presents the mean, standard deviation, and variance from both instruments. The scores from the SBAC and TDRMC instruments were from the same students, which made them dependent samples. To compare the variance from the dependent samples, I used the Morgan–Pittman test. The results show that my instrument does not statistically capture more variability than the SBAC.

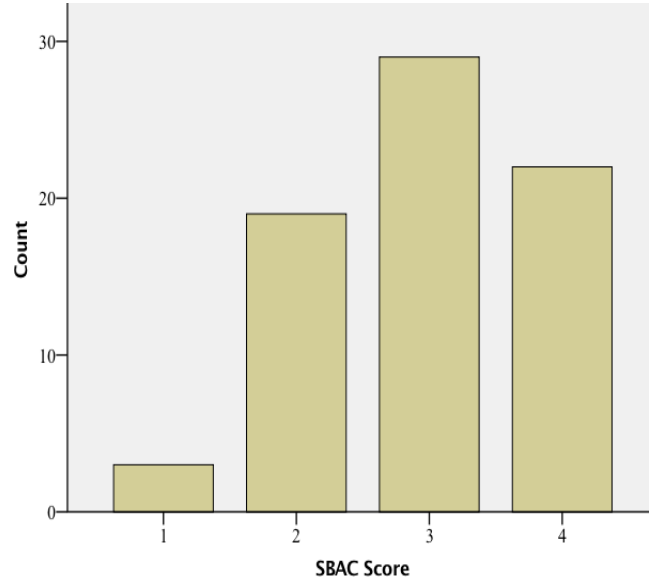


Figure 10. Distribution of SBAC scores from student's essays.

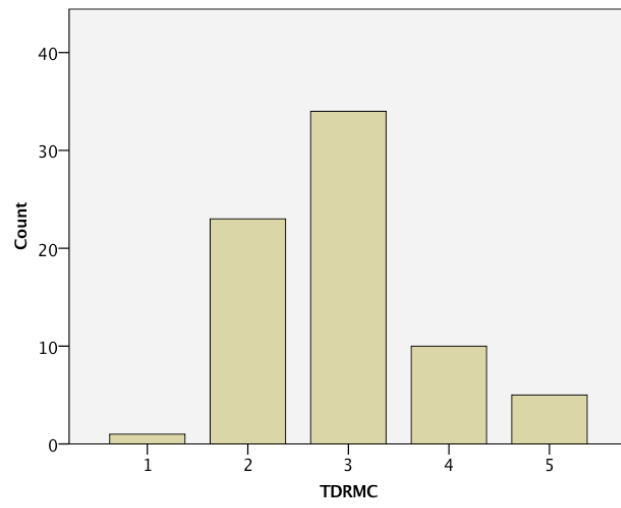


Figure 11. Distribution of TDRMC scores from student essays.

Table 4

Descriptive Statistics

	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>	Variance
SBAC	73	1.00	4.00	2.97	.88	.78
TDRMC	73	1.00	5.00	2.96	.89	.79

Note. SBAC = Smarter Balanced Assessment Consortium. TDRMC = Teacher Designed Rubric Measuring Cogency.

Research Question 3

Research Question 3 asked, How does the TDRMC correlate to other established measures of writing and academic ability, such as grades, various SBAC scores, and GPA, and how do these correlations compare with those of SBAC Writing and established measures of writing and academic ability? To determine if the TDRMC has better external validity than the SBAC, I calculated correlations between SBAC and TDRMC and real-world measures of students' performance: sixth-grade GPA, first-semester seventh-grade GPA, raw scores for sixth-grade SBAC ELA, and the scores for sixth-grade SBAC Writing. The SBAC ELA scores are the cumulative scores for the following categories: reading, writing, listening, and research/inquiry. The scores for writing were pulled out of the overall cumulative score. The results are shown in Table 5.

The correlations show that both the TDRMC and the SBAC are significantly correlated with all the variables. I then conducted several tests for the significance of the difference between dependent correlations to determine if the TDRMC had a higher correlation with the real-world measures.

The results for the test of the significance of the difference between dependent correlations show that the correlation between the SBAC and sixth-grade GPA and the correlation between the TDRMC and sixth-grade GPA were not statistically different, $p = .422$. The results for the test of the significance of the difference between dependent correlations showed that the correlation between the SBAC and seventh-grade GPA and the correlation between the TDRMC and seventh-grade GPA were also not statistically different, $p = .632$. The results for the test of the difference between sixth-grade SBAC ELA and TDRMC and sixth-grade SBAC ELA with SBAC were not statistically significant, $p = .592$. The results for the test show that the correlation between sixth-grade SBAC Writing and SBAC and sixth-grade SBAC Writing and TDRMC were not statistically different, $p = .960$. These results show that the TDRMC's and SBAC's correlations with the external measures did not statistically differ from each other. However, while findings suggest that there were no statistically significant differences across the two instruments, there was a trend suggesting that TDRMC was more highly correlated with various performance measures (sixth- and seventh-grade GPA and SBAC ELA) than SBAC (see Table 5).

To further test the validity of the TDRMC, I used Kendall's tau to correlate the measures of SBAC and TDRMC with sixth- and seventh-grade ELA grades. I used Kendall's tau because sixth- and seventh-grade ELA grades are an ordinal variable. The results for the tests are shown in Table 6.

Table 5

Correlations Between SBAC, TDRMC, and External Measures

	SBAC	TDRMC	Sixth GPA	Seventh GPA ^a	Sixth SBAC ELA	Sixth SBAC Writing
SBAC	–					
TDRMC	.601**	–				
Sixth GPA	.461**	.535**	–			
Seventh GPA ^a	.499**	.541**	.889**	–		
SBAC ELA	.478**	.529**	.681**	.562**	–	
SBAC Writing	.419**	.414**	.482**	.562**	.816**	–

Note. $N = 73$. ELA = English language arts. GPA = grade point average. SBAC = Smarter Balanced Assessment Consortium. TDRMC = Teacher Designed Rubric Measuring Cogency.

^aFirst semester.

** $p < .01$ (2-tailed).

Table 6

Correlations Between SBAC and TDRMC With Sixth and Seventh ELA Grades and Their Comparisons

	SBAC	TDRMC	Sixth ELA	Seventh ELA	p -Values from SBAC/TDRMC correlation comparison
SBAC	–				
TDRMC	.557**	–			
Sixth ELA	.379**	.455**	–		.119
Seventh ELA	.432**	.546**	.445**	–	.017

Note. ELA = English language arts. SBAC = Smarter Balanced Assessment Consortium. TDRMC = Teacher Designed Rubric Measuring Cogency.

** $p < .01$ (2-tailed).

To conduct the test for the difference between the dependent correlations, I converted Kendall's tau into Pearson's correlation coefficient (Walker, 2003). The results for the difference between the correlation between sixth-grade ELA and SBAC and sixth-grade ELA with TDRMC approached significance, $p = .119$. Furthermore, the difference between the correlation between TDRMC and seventh-grade ELA grades and seventh-grade ELA grades and SBAC was statistically significant, $p = .017$. The correlation between TDRMC and seventh-grade ELA grades was higher than the correlation between SBAC and seventh-grade ELA grades.

CHAPTER 5: DISCUSSION

The purpose of this study was to create an instrument (TDRMC) that validly and reliably measured the construct of cogency, a dimension of the multidimensional construct of argument. A moderation session helped me to establish reliability of the instrument. Additionally, the moderation session resulted in modifications to the instrument. The reliability of the TDRMC was then compared to the common standardized assessment tool (SBAC) used to measure argument. The two instruments were compared to determine if the TDRMC scoring guide yielded more variable scores. Finally, the study examined the validity of the TDRMC by correlating its scores with established measures of academic achievement.

Overview of Argument

Starting at the middle school level, constructing argument in an English classroom requires students to move beyond analysis and interpretation of a literary text. Teachers are encouraged to pair literary texts with nonliterary texts that are thematically related to the identified issue in a novel (NGA, 2010). As a result, students are asked to provide evidence from both literary and nonliterary texts in support of their arguments.

The rationale for moving students beyond the primary literary text is to encourage them to connect identified issues to real-world problems. In this way, students will not only read widely and deeply but will also engage in more nuanced thinking by intellectually wrestling with the complexity of the issues. This approach also shifts the construction of argument away from persuasion and prioritizes the importance of appeals to logic. To aid students in argument construction using appeals to logic, Toulmin's

model of argument has been primarily used in Grades 6–12 (Hillocks, 2011; NGA, 2010). Therefore Toulmin’s model should be used to create rubrics.

Summary of the Study

Rubrics are widely used in the classroom, both in K–12 and at the university level. Sometimes teachers create rubrics to assess student writing, yet often there is a mismatch between a teacher’s expectations and students’ understanding of the performance task required of them (Anson, Dannels, Flash, & Gaffney, 2012; Jonsson & Svingby, 2007). Most of the research on rubrics has focused on making transparent the expectations of a writing assignment or its use for student peer assessment (Allen & Tanner, 2006; Hafner & Hafner, 2003; Jonsson & Svingby, 2007). At the same time, research has shown that generic rubrics do not provide assessment of task-specific expectations (Anson et al., 2012; Hafner & Hafner, 2003; Jonsson & Svingby, 2007; Rezaei & Lovorn, 2010).

Previous studies on rubrics have examined the protocol for design, uses in the classroom, and large-scale assessments, yet few studies have examined students’ use of warrants in the construction of argument, and the studies that do exist have primarily focused on the subject domain of science or mathematics (Alcock & Weber, 2005; Bottcher & Meisert, 2011; Sandoval & Millwood, 2005; Simon, 2008; von Aufschnaiter et al., 2008). This current study is an attempt to address this gap in the literature of assessing reasoning in support of evidence in an English classroom.

This study involved the creation of a rubric that was developed from principles of the BEAR assessment system to operationalize the construct of warrants, a dimension of the multidimensional construct of argument (Wilson, 2005). Operationalizing the

construct was an important aspect in construction of an instrument to ensure reliability and validity so that meaningful feedback was provided to both student and teacher.

The new rubric was assessed using a sample ($N = 73$) of seventh-grade students and was evaluated for reliability and validity. The design procedures articulated in Wilson's (2005) framework should ensure the focus is on the item and the instrument; therefore the procedures and results should be generalizable beyond the students in this study.

Limitations

Several issues limited the findings of my study. I obtained a convenience sample from the five sections of English that I taught. My sample size was small ($N = 73$) in comparison to the SBAC instrument, which is used nationally on the standardized tests that accompany the CCSS. I taught the students how to construct argument using Toulmin's model and provided feedback on rough drafts of their essays. Thus findings from this study may be biased.

A lack of transparency in scoring the SBAC—the comparison rubric used in this study—is also a limitation. No document explained how the SBAC instrument was generated and edited as it went through iterations:

Testing experts have raised significant concerns about all (SBAC, PARCC, Pearson) assessments, including the lack of basic principles of sound science, such as construct validity, research-based cut scores, computer adaptability, inter-rater reliability, and most basic of all, independent verification of validity. Here in California, the SBAC assessments have been carefully examined by independent examiners of the test content who concluded that they lack validity,

reliability, and fairness, and should not be administered, much less be considered a basis for high-stakes decision making. When asked for documentation of the validity of the CA tests, the CA Department of Education failed to make such documentation public. Even SBAC's own contractor, Measured Progress, in 2012 gave several warnings, including against administering these tests on computers. (California Alliance of Researchers for Equity in Education, 2016)

This lack of transparency limited my interpretation of the correlation between the TDRMC and SBAC.

An unexpected limitation that arose from this study was that teachers have a very uneven understanding of Toulmin's model of argument as evidenced in feedback that I received from all the raters. Raters 1 and 3 felt restricted by the criteria provided for each descriptor level. As Rater 3 stated in feedback after the moderation session,

I found myself giving a lot of 3s, and consequently Level 3 became a fairly broad category for me. The reason for this had to do with the use of the word "some" in the following: "Some but not all selection of the evidence is relevant, and is thereby not appropriate to argument field." I thought a number of the papers had chosen *some* evidence that worked well and *some* evidence that could be improved upon. The 4 criterion ("selection of evidence is relevant, and appropriate") felt a little too restrictive on this point.

This rater's comment points to the necessity of using Toulmin's model and its use of warrants. In his critique of syllogisms, Toulmin (1958/2003) noted that most arguments start with the claim, which is the same as the end statement of a syllogism. If the claim is accepted, there is no need to prove the validity of the statement, but if the claim requires

evidence, then the evidentiary reasoning must show how the evidence supports the claim. This rater looked at the logic of a student's reasoning without looking at warrants. This again points to the importance of professional development, in particular the importance of a deep understanding of argument theory.

Similarly, Rater 2 expressed concerns about familiarity with Toulmin's model, in particular how to identify a warrant and a backing:

The challenge I'm running into is my unfamiliarity with Toulmin's model. I have read *Teaching Argument Writing* by Hillock, which incorporates the concepts of warrants and backing, but I have not read Toulmin. Essentially, even though I have a sense of the levels of reasoning in the student papers, I am not entirely sure how to identify a warrant or backing. I was wondering if this instrument will be used specifically by teachers trained in the Toulmin model and/or if it could include some definitions and examples of the terms for those not trained.

Some version of Toulmin's model of argument is generally used in Grades 6–12, and the instruction of warrants is very uneven (Warren, 2010) because teachers believe it is difficult to teach. Teachers do need to know the model to effectively use the TDRMC, because the instrument uses Toulmin's model of argument. When teachers struggle with the model, assessment, inferences, and interpretation of the test results are limited. The creation of the TDRMC is a good starting point but also points to the importance of educating teachers about Toulmin's model of argument. In my later discussion of educational and research implications, some of these concerns are addressed.

Discussion of Research Questions

Research Question 1

The first research question set out to establish the reliability of the TDRMC and how it compared with the SBAC. To establish reliability, I trained teachers on the TDRMC and SBAC instruments and facilitated a discussion during the moderation session to address questions and promote clarity. Discussion of the instrument resulted in revisions of the TDRMC. To assess the reliability of the TDRMC, the percentage agreement of all four raters and the reliability of the measure were calculated.

The TDRMC was shown to be more reliable, $\alpha = .816$, than the SBAC, $\alpha = .741$. The percentage agreement showed that the TDRMC is higher, which contradicts the research from Jonsson and Svingby's (2007) meta-analysis of the reliability and validity of scoring guides. They pointed out that fewer levels of descriptors usually results in more agreement among raters; however, the opposite was demonstrated in this study, as evidenced by a comment from Rater 3:

Because the rubric had a number of criteria at each level, it was difficult at times to determine which score to give. In many cases, papers got a 3 for me because of the evidence selected, but had I been scoring them on their evidentiary reasoning alone, they might have gotten a 4.

Rater 1 shared a similar sentiment in her feedback to me regarding the constraints of the evaluative criteria: "I also found myself giving a lot of 3s for the same reasons. Many students had some relevant and/or appropriate evidence."

The TDRMC's higher reliability is consistent with research stating that articulation of the evaluative criteria for the construct being measured is important for

reliability. Thus the results of this test suggest that the TDRMC is more reliable because it is not a generic rubric.

Research Question 2

The second research question investigated whether the TDRMC yielded more variable scores than the SBAC and was thus better at capturing the true variability in the population. A comparison of the variance shows that the TDRMC and SBAC instruments generate equally variable distributions of scores. However, these results suggest that the TDRMC was at least as good at capturing the variability in cogency as the SBAC.

Variability of scores is important because it would have shown that the TDRMC was representative of the academic performance range. If the TDRMC had yielded more variable scores, then it would have been more representative of the academic performance range. In the future, the descriptor levels of the TDRMC could be revised to ensure that they represent the construct. However, teachers do need an understanding of Toulmin's model of argument. Adding more descriptor levels to the TDRMC would ensure that the delineation of the construct of cogency is more representative of the academic performance range. Furthermore, as discussed, incorporating some of the raters' feedback in the revision of the instrument would provide more clarity. Given that this was only the first version of the TDRMC, a few changes to add clarity would most likely capture more variability in the distribution of scores. In sum, although two of the raters struggled with the TDRMC's dependency on Toulmin's model and lacked a clear understanding of the role warrants play in the construction of a cogent argument, the

results suggest that the TDRMC produces scores that are as variable as those produced with the SBAC.

Research Question 3

The third research question examined the validity of the TDRMC. There were two parts to this question. The first part of this question examined whether the TDRMC had validity, and the answer is yes. The TDRMC correlated with the following external measures of academic achievement and writing ability—sixth-grade GPA, first-semester seventh-grade GPA, raw scores for sixth-grade SBAC ELA, and the scores for sixth-grade SBAC Writing. The TDRMC was significantly correlated with all the measures used, which are all highly valued and established indicators of student achievement.

The second part of the question addressed whether the TDRMC was more valid than the SBAC. In all cases but one, the correlations did not differ, suggesting that the TDRMC is as valid as the SBAC, but not more so. The one comparison with a significant difference showed that the TDRMC had a higher correlation with seventh-grade ELA grades than the SBAC. Although findings suggest that there are no statistically significant differences across the two instruments, there was a trend suggesting that the TDRMC is more highly correlated with various performance measures (sixth- and seventh-grade GPA and SBAC ELA) than SBAC. These results hint at the possibility of a practical difference between the two instruments, such that the TDRMC was more highly correlated with other measures of academic performance than the SBAC. The TDRMC is more reflective of the construct of cogency. Having a good GPA and a good grade in an English class is probably reflective of the fact that the student knows how to construct a cogent argument.

Instructional Implications

Implications for Instruction of Toulmin's Model of Argument

The moderation session revealed a misunderstanding of the role warrants play in Toulmin's model of argument. In the moderation session on the TDRMC, a discussion of family as it pertained to the novel *The Outsiders* involved a clarification of the role of warrants in establishing the criterion for what constituted a family. We were reviewing why I gave one of the exemplars a 2. My reasoning was that the evidence was not connected to the argument field, whereas Rater 2 thought that a warrant was a universal rule. The idea of field theory was new to her.

The classical argument structure is similar to Toulmin in that both contain claim, evidence, and reasoning. However, the component of the warrant in Toulmin's model is what differentiates it from the classical argument structure. The warrants are context specific to the argument and do not follow a universal rule (Hillocks, 2011; Lunsford, 2002; Stygall, 1986; Toulmin, 1958/2003).

Future training for the TDRMC should include a half-day of instruction on the model, with the afternoon devoted to practicing on exemplars.

Revision of the TDRMC

The feedback that I received from the raters asked for a list of definitions and examples for the components of Toulmin's model, which is something I would incorporate into training on Toulmin's model. However, only providing this glossary of terms without training would not aid teachers in understanding the model, because the warrant component is what gets modified or ignored. Defining the warrant during the moderation session was not sufficient.

Conclusion

The purpose of this study was to create a reliable and valid instrument to measure cogency in argument. The reliability of the TDRMC was higher than the reliability of the SBAC, and the TDRMC was shown to be more valid in its correlation with real-world measures of academic achievement, such as GPA and writing ability. The moderation session helped the raters to gain some clarity, which resulted in the reliability of the TDRMC. Wilson's (2005) framework made transparent the robust nature of attaining reliability.

The moderation session for establishing reliability of the TDRMC also revealed that teachers, who were chosen because of their leadership positions in the district, were not familiar with warrants, backing, and evidentiary reasoning. This points to the concern that if teachers do not know how to teach warrants, then how will students know? For example, some teachers felt constrained by not being able to give a higher mark for cogency when a student demonstrated the ability to write reasoning, even though the selection of evidence was not entirely relevant in supporting the claim. This restriction, however, contributes to the strength of the TDRMC. Many times, reasoning is scored separately from the selection of evidence, and as a result, what does not get accurately assessed is whether, combined, the selection of evidence is relevant and reasoning is appropriate to the argument field (Hillocks, 2011; Toulmin, 1958/2003).

The most important component of a rubric is the articulation of the evaluative criteria for the construct being measured (Arter, 2000; Hafner & Hafner, 2003; Haladyna & Rodriguez, 2013; Nodoushan, 2014; Popham, 1997b; Rezaei & Lovorn, 2010; Wilson, 2005). Adding more descriptor levels provides clarification of the evaluative criteria and

would be more representative of the construct being measured, because generic rubrics fail to assess the context-specific expectations of a particular writing assessment within a specific discipline (Anson et al., 2012; Hafner & Hafner, 2003; Haladyna & Rodriguez, 2013; Nodoushan, 2014; Popham, 1997b; Rezaei & Lovorn, 2010). The TDRMC contained more descriptors and evaluative criteria; however, it did not generate greater variability in scores than the SBAC. Neither instrument is satisfactory in generating a representation of the academic performance range, so needed are more descriptor levels. For the SBAC, variability may have been limited, because most students scored within the top two levels. However, this does not appear to be a true ceiling effect, because less than 30% of scores are at a Level 4, which is the highest descriptor level for the SBAC. For the TDRMC, there appears to be no evidence of a ceiling or floor effect. The total proportions of scores at the lowest and highest levels were very small. The most common score was the middle level (3) on the TDRMC. This may be a reflection of the fact that raters felt constrained by the criteria articulated in the rubric.

The objective of the TDRMC is to provide evaluative criteria of the construct being measured, which was cogency. The results of this study show that the TDRMC is correlated with real-world student outcomes and so appears to be a valid and reliable instrument to be used in assessment of how students construct argument. In doing so, the TDRMC has the potential to equip teachers to provide feedback on writing to promote student learning. However, while the results presented in this study suggest that the TDRMC is a promising measure of cogency, future iterations that account for a better representation of the construct will only improve its capability to capture students' ability to construct cogent arguments.

Research Implications

This study addressed the importance of using valid and reliable instruments to assess writing, specifically the construction of a cogent argument, which focuses on reasoning appropriate to the argument field. Previous studies have been conducted on reasoning in the subject domains of math and science, but few studies have been conducted in an English classroom. A valid and reliable instrument is critical, because with the adoption of the CCSS, students are asked to demonstrate their ability to write argument, which is a key indicator of college readiness (Conley, 2007; Conley et al., 2011); thus teachers need to be able to assess writing to ensure that all students are college ready.

Rubrics are widely used for assessment of writing. Many studies have been conducted on the importance of reliability of a rubric, yet needed are more studies that investigate the efficacy of rubrics that delineate evaluative criteria of the construct being measured (Reddy & Andrade, 2010). In particular, more research is needed on the assessment of argument in an English classroom at the middle school level. In this way, alignment of the evaluative criteria on the rubric to the construct (Reddy & Andrade, 2010; Wilson, 2005) will inform instruction of argument for the teacher and will also provide feedback to the student regarding what he or she needs to accomplish to move to the next level of proficiency on the rubric.

Educational Implications

The CCSS (NGA, 2010, Appendix A-1) indirectly refer to Toulmin's model of argument, which is generally used in Grades 6–12. Yet, as this study has highlighted, teachers are not entirely clear what the model is. Professional development on Toulmin's

model of argument is necessary for teachers of writing instruction, especially in Grades 6–8, because academic writing is introduced in middle school. Teachers' feedback after the moderation session pointed to the necessity of ensuring that professional development boosts their knowledge of argument, in particular Toulmin's model, because this model is expected of students.

Given the importance of aligning assessment to what is taught in the classroom, it follows that assessment of argument must emphasize the concept of warrants and the idea that warrants must be specific to the argument field. The SBAC uses more generic descriptors and therefore is not reflective of writing instruction. The TDRMC may be more valid because it provides evaluative criteria for the construct of cogency. However, the efficacy of the TDRMC appears to be better when teachers are knowledgeable of Toulmin's model, and even the raters who participated in this study, who were chosen for their excellence in teaching, demonstrated unfamiliarity with the model.

Thus, as mentioned, teachers in general could be better and more specifically trained in Toulmin's model. At the same time, future iterations of the TDRMC could offer more scaffolding to teachers in the model so that the design of the TDRMC accounts for the reality that teachers will vary in their knowledge of Toulmin's model. One way of addressing this gap in knowledge is to include a short synopsis of Toulmin's model with the TDRMC. In addition, a half-day to full-day workshop on Toulmin is needed because his model differs from the classical argument structure owing to the importance context-specific warrants play in the construction of argument.

As this study has shown, if teachers are not familiar with the construct and the evaluative criteria that are necessary to promote student learning, then the usefulness of

the assessment becomes limited. We rely on these assessments in multiple high-stakes scenarios to evaluate students, teachers, and schools, even though questions have been raised about the validity of nationally standardized assessments like the SBAC (California Alliance of Researchers for Equity in Education, 2016). To be fair, the SBAC is a summative assessment used for accountability purposes, while the TDRMC is a tool for formative assessment. However, the SBAC is not measuring what is articulated in the standards. The TDRMC is more valid and reliable, and therefore it should be used in assessing student proficiency of argument.

The challenge for educators is to invest their time in creating instruments that (a) clearly operationalize the construct of what is being measured and (b) are aligned with what is being taught. If teachers focus on using specific rubrics, they will get specific feedback on their students' academic proficiency in constructing argument, and this information will inform their instruction. A tool such as the TDRMC can teach teachers what is important because it is backed by a theory of argument and cognition. Only then can we hope that evaluative criteria will inform instruction and provide an accurate representation of the academic performance range.

Summary

The ability to construct argument is a key expectation in the seventh-grade ELA classroom. Teachers depend on assessment tools to provide feedback to students on their progress and also to inform teachers of the construct that is being measured so that they can better provide instruction to help students achieve this progress.

The purpose of this study was to create a teacher-informed, valid and reliable instrument to measure the construct of cogency, which is a dimension of the

multidimensional construct of argument. I chose Toulmin's model of argument because it is the dominant model used in Grades 6–12 in California and because it lays out argument in a procedural manner so that the components of argument are clearly connected. In addition, I used Wilson's framework for assessment to create the instrument used in the study. The use of explicit evaluative criteria aligned with the construct allows the teacher to provide meaningful feedback to the student and also informs the teacher of the type of instruction needed to help a student attain the next level of proficiency on the rubric, demonstrating a deeper and more sophisticated understanding of the construct.

The results of this study suggest that the TDRMC may be better than the SBAC at assessing student writing in the area of argument. The pattern of results showed that even in its early stages of development, the TDRMC is performing as well as, if not better than, the SBAC in capturing students' ability to write a cogent argument with reliability and validity. It is likely that the TDRMC's strengths stem from a more accurate operationalization of the construct of cogency, which is linked to the theory of argument, as articulated in Toulmin's model.

With the CCSS (CDE, 2015), argument is promoted as a fundamental writing skill rendering students college ready. Specifically, argument involves complex reasoning that is indicative of a student's ability to do well in college. If we are to continue using national standardized tests to assess argument, then we need instruments that are reflective of this essential construct of cogency. Regardless of the subject domain, instruments must be reflective of the construct so that meaningful feedback can be provided to the student to promote learning.

This study, though small, has opened the door to the possibilities of improving the validity and reliability of writing assessments so that they can meaningfully impact teaching and learning for students on the path to becoming members of a democratic society.

REFERENCES

- Alcock, L., & Weber, K. (2005, January 1). Proof validation in real analysis: Inferring and checking warrants. *Journal of Mathematical Behavior*, *24*, 125–134.
<https://doi.org/10.1016/j.jmathb.2005.03.003>
- Ali, R., & Jenkins, G. (2002). *The high school diploma: Making it more than an empty promise*. Oakland, CA: Education Trust West.
- Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education*, *5*, 197–203.
- Anastasi, A. (1986, January 1). Evolving concepts of test validation. *Annual Review of Psychology*, *37*, 1–15.
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, the ugly. *College Teaching*, *53*(1), 27–30.
- Angier, N. (2013, November 25). The changing American family. *New York Times*.
 Retrieved from <https://www.nytimes.com/2013/11/26/health/families.html>
- Anson, C. M., Dannels, D. P., Flash, P., & Housley Gaffney, A. L. (2012). Big rubrics and weird genres: The futility of using generic assessment tools across diverse instructional contexts. *Journal of Writing Assessment*, *5*(1). Retrieved from <http://journalofwritingassessment.org/article.php?article=57>
- Arter, J. (2000, April 24–28). *Rubrics, scoring guides, and performance criteria: Classroom tools for assessing and improving student learning*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

- Bottcher, F., & Meisert, A. (2011). Argumentation in science education: A model-based framework. *Science & Education, 20*, 103–140.
- Braddock, R. R., Lloyd-Jones, R., & Schoer, L. (1963). *Research in written composition*. Champaign, IL: National Council of Teachers of English.
- Brennan, R. L. (2006). *Educational measurement* (4th ed.). Westport, CT: American Council on Education/Praeger.
- Brown, N. J., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review, 23*, 221–234.
<https://doi.org/10.1007/s10648-011-9161-z>
- California Alliance of Researchers for Equity in Education. (2016, February). *Common Core State Standards assessments in California: Concerns and recommendations* (Research Brief No. 1). Retrieved from <http://www.care-ed.org/>
- California Department of Education. (2015). *English language arts/English language development framework for California public schools kindergarten through grade twelve*. Sacramento, CA: Author.
- Cheesebro, J. (2003). Effects of teacher clarity and non-verbal immediacy on student learning, receiver apprehension, and affect. *Communication Education, 52*, 135–147.
- Civikly, J. M. (1992). Teachers and students making sense of instruction. *Communication Education, 41*, 138–152.
- Clark, I. (2011). Formative assessment: Policy, perspectives and practice. *Florida Journal of Educational Administration & Policy, 4*, 158–179.
- Clinchy, E. (1998). Democratizing America's public schools. *Principal, 77*(5), 13–14.

- Conley, D. T. (2007). *Redefining college readiness*. Eugene, OR: Educational Policy Improvement Center.
- Conley, D. T., Drummond, K. V., de Gonzalez, A., Rooseboom, J., & Stout, O. (2011). *Reaching the goal: The applicability and importance of the Common Core State Standards to college and career readiness*. Eugene, OR: Educational Policy Improvement Center.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. <https://doi.org/10.1037/h0040957>
- Deane, P. (2011). *Writing assessment and cognition* (Research Report No. RR-11-14). Princeton, NJ: Educational Testing Service.
- Diederich, P. B., French, J. W., & Carlton, S. T. (1961). *Factors in judgments of writing ability*. Princeton, NJ: Educational Testing Service.
- Draney, K. (2009, June). *Designing learning progressions with the BEAR assessment system*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287–312. [https://doi.org/10.1002/\(sici\)1098-237x\(200005\)84:33.0.co;2-a](https://doi.org/10.1002/(sici)1098-237x(200005)84:33.0.co;2-a)
- Eemeren, F. H., Grootendorst, R., & Snoeck, H. A. F. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Erlbaum.
- Embretson, S. (2007). Construct validity: A universal validity system or just another test evaluation procedure. *Educational Researcher*, 36, 449–455.

- Emmel, B., Resch, P., & Tenney, D. (1996). *Argument revisited; argument redefined: Negotiating meaning in the composition classroom*. Thousand Oaks, CA: Sage.
- Fritzberg, G. J. (2004). No child left behind? Assessing President Bush's assessment law. *Educational Foundations, 18*(3), 7–24.
- Fulkerson, R. (1996). *Teaching the argument in writing*. Urbana, IL: National Council of Teachers of English.
- Gere, A. R. (1980). Written composition: Toward a theory of evaluation. *College English, 42*(1), 44–58. <https://doi.org/10.2307/376032>
- Godshalk, F. I., Swineford, F., Coffman, W. E., Educational Testing Service, & College Entrance Examination Board. (1966). *The measurement of writing ability*. New York, NY: College Entrance Examination Board.
- Graff, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.
- Greenberg, K. L. (1992). Validity and reliability issues in the direct assessment of writing. *WPA: Writing Program Administration, 16*(1–2), 7–22.
- Greve, C., Morris, W., & Huot, B. (2018). Use and misuse of writing rubrics. In J. I. Lontas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching*. Mahwah, NJ: John Wiley.
<http://doi.org/10.1002/9781118784235.eelt0545>
- Guion, R. M. (1977). Content validity—the source of my discontent. *Applied Psychological Measurement, 1*(1), 1–10.

- Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education, 25*, 1509–1528.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.
- Haswell, R. (2014). Paul B. Diederich? Which Paul B. Diederich? *Journal of Writing Assessment, 7*(1).
- Hillocks, G., Jr. (2011). *Teaching argument writing: Grades 6–12*. Portsmouth, NH: Heinemann.
- Howell, R. J. (2014). Grading rubrics: Hoopla or help. *Innovations in Education and Teaching International, 51*, 400–410.
- Huot, B. A. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.
- Huot, B., O'Neill, P., & Moore, C. (2010). A usable past for writing assessment. *College English, 72*, 495–517.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review, 2*, 130–144.
- Keith, W. M., & Beard, D. E. (2008, March 17). Toulmin's rhetorical logic: What's the warrant for warrants? *Philosophy and Rhetoric, 41*(1), 22–50.
<https://doi.org/10.1353/par.2008.0003>
- Kennedy, K. (2005). *The BEAR assessment system: A brief summary for the classroom context* (BEAR Report Series No. 2005-03-01). Berkeley, CA: University of California.

- Kneupper, C. W. (1979, March 8). Teaching argument: An introduction to the Toulmin model. *College Composition and Communication*, 29, 237–241.
<https://doi.org/10.2307/356935>
- Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- Leeman, R. (1987). *Teaching critical thinking in the argumentation course*. Paper presented at the Speech Communication Association, Boston, MA.
- Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice*, 16(2), 14–16.
- Lunsford, K. J. (2002). Contextualizing Toulmin's model in the writing classroom. *Written Composition*, 19, 109–174. <https://doi.org/10.1177/074108830201900105>
- McCann, T. M. (2010). Gateways to writing logical arguments. *English Journal*, 99(6), 33–39.
- Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practice*, 16, 16–18.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955–966.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012–1027.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *The American Psychologist*, 50(9), 1–9. <https://doi.org/10.1037/0003-066X.50.9.741>

- Nardi, E., Biza, I., & Zachariades, T. (2012). "Warrant" revisited: Integrating mathematics teachers' pedagogical and epistemological considerations into Toulmin's model for argumentation. *Educational Studies in Mathematics*, 79, 157–173. <https://doi.org/10.1007/s10649-011-9345-y>
- National Center for Education Statistics. (2004). *The condition of educations 2004 in brief*. Washington, DC: U.S. Department of Education.
- National Governors Association Center for Best Practices Council of Chief State School Officers. (2010). *Common Core State Standards initiative*. Washington, DC: Author.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Nodoushan, M. A. (2014). Assessing writing: A review of the main trends. *Studies in English Language and Education*, 1(2), 119–129. <https://doi.org/10.24815/siele.v1i2.1831>
- Pellegrino, J. W., Chudowsky, N., Glaser, R., & National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Peter D. Hart Research Associates & Achieve Group Inc. (2005). *Rising to the challenge: Are high school graduates prepared for college and work? A study of recent high school graduates, college instructors, and employers*. Washington, DC: Peter D. Hart Research Associates.
- Popham, W. J. (1997a). Consequential validity: Right concern—wrong concept. *Educational Measurement: Issues and Practice*, 16, 9–13.

- Popham, W. J. (1997b). What's wrong—and what's right—with rubrics. *Educational Leadership*, 55, 72–75.
- Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York, NY: Knopf.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, 35, 435–448.
- Rex, L. A., Thomas, E. E., & Engel, T. (2010). Applying Toulmin: Teaching logical reasoning and argumentative writing. *The English Journal*, 99(6), 56–62.
- Rezaei, A. R., & Lovorn M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18–39.
- Roberts, R., & Gott, R. (2010). Questioning the evidence for a claim in a socio-scientific issue: An aspect of scientific literacy. *Research in Science and Technological Education*, 28, 203–226. <https://doi.org/10.1080/02635143.2010.506413>
- Rothman, R. (2012). A common core of readiness. *Educational Leadership*, 69(7), 10–15.
- Rowland, R. C. (2008, May 1). Purpose, argument fields, and theoretical justification. *Argumentation: An International Journal on Reasoning*, 22, 235–250.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sandoval, W. A., & Millwood, K. A. (2005, January 1). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

- Sandoval, W. A., & Reiser, B. J. (2004, May 1). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345–372. <https://doi.org/10.1002/sce.10130>
- Shepard, L. A. (1997). *Measuring achievement: What does it mean to test for robust understanding?* Princeton, NJ: Educational Testing Service.
- Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63(3), 66–70.
- Simon, S. (2008). Using Toulmin’s argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, 31, 277–289.
- Smarter Balanced Assessment Consortium. (2015, July). *Content specifications for the Summative Assessment of the Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Retrieved from https://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA_Content_Specs.pdf
- Stygall, G. (1986). Toulmin and the ethics of argument fields: Teaching writing and argument. *Journal of Teaching Writing*, 6, 93–107.
- Timperley, H. S., & Parr, J. M. (2009). What is this lesson about? Instructional process and student understandings in writing classrooms. *British Curriculum Foundation*, 20, 43–60.
- Toulmin, S. (2003). *The uses of argument*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511840005> (Original work published 1958)

- Trauth-Nare, A., & Buck, G. (2011). Assessment for learning. *Science Teacher*, 78(1), 34–39.
- Turley, E. D., & Gallagher, C. W. (2008). On the “uses” of rubrics: Reframing the great rubric debate. *English Journal*, 97(4), 87–92.
- U.S. National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and the secretary of education*, United States Department of Education. Washington, DC: U.S. Department of Education.
- von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008, January 1). Arguing to learn and learning to argue: Case studies of how student’s argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45, 101–131.
- Walker, D. A. (2003, November 1). JMASM9: Converting Kendall’s tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, 2, 525–530.
- Warren, J. E. (2010, July 1). Taming the warrant in Toulmin’s model of argument. *English Journal*, 99(6), 41–46.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York, NY: Taylor & Francis.
- Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR assessment system. *Higher Education: The International Journal of Higher Education and Educational Planning*, 52, 635–663.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*, 181–208.

https://doi.org/10.1207/S15324818AME1302_4

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*, 193–219.

Yancey, K. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50*, 483–503.

<https://doi.org/10.2307/358862>

APPENDIX A: UNIVERSITY OF SAN FRANCISCO IRB APPROVAL*IRBPHS - Approval Notification*

To: Mildred Gonzalez-Balsam
From: Terence Patterson, IRB Chair
Subject: Protocol #437
Date: 05/22/2015

The Institutional Review Board for the Protection of Human Subjects (IRBPHS) at the University of San Francisco (USF) has reviewed your request for human subjects approval regarding your study.

Your research (IRB Protocol #437) with the project title **Measuring Argument** has been approved by the IRB Chair under the rules for expedited review on **05/22/2015**.

Any modifications, adverse reactions or complications must be reported using a modification application to the IRBPHS within ten (10) working days.

If you have any questions, please contact the IRBPHS via email at IRBPHS@usfca.edu. Please include the Protocol number assigned to your application in your correspondence.

On behalf of the IRBPHS committee, I wish you much success in your research.

Sincerely,

Terence Patterson, EdD, ABPP

Professor & Chair, Institutional Review Board for the Protection of Human Subjects

University of San Francisco

APPENDIX B: INFORMED CONSENT

Dear Parents:

My name is Millie Gonzalez-Balsam, and I am a doctoral student in the Learning and Instruction Department in the School of Education at the University of San Francisco. I am sending this letter to explain why I would like for your child to participate in my research project. I am studying argument and would like to design an instrument that is reliable and valid in measuring cogency, a feature of the multidimensional construct of argument.

With your permission, I will ask your child to construct a written argument in response to a prompt. Your child's participation in this study is completely voluntary and will not affect his or her grades in any way. Your child may quit this study at any time simply by saying "stop" or "I do not wish to participate."

The study will be conducted at [school name removed] on Monday from November 7 to November 17 during their class period with me. During this time students will complete a rough draft of their response, which will receive feedback. They will then submit a final draft of their essay on December 5. A grade will be assigned for their class work, but for the purposes of the study, an additional grade, which will not be recorded on their report card, will be anonymously recorded for the study. This study will not interfere with the curriculum since the unit on *The Outsiders* novel will be used. There are no known risks involved in this study. To protect your child's confidentiality, your child's name will not appear on any record sheets. The information obtained will not be shared with anyone, unless required by law. I and my faculty sponsor, Dr. Nicola

McClung, will maintain the records. If you have any questions, please contact me at [contact information removed].

This letter will serve as a consent form for your child's participation and will be kept in the Learning and Instruction Department at the University of San Francisco. If you have any questions about this study, please contact Dr. Nicola McClung, faculty sponsor of this project, at [contact information removed]. If you have any questions about your child's rights as a participant, you may contact the University of San Francisco IRB at IRBPHS@usfca.edu.

Please have your child return this form to Millie Gonzalez-Balsam.

Sincerely yours,

Millie Gonzalez-Balsam

APPENDIX C: RATER CONSENT

September 7, 2016

Dear Teacher:

You have been asked to participate in a research study conducted by Millie Gonzalez-Balsam, a graduate student in the Department of Learning & Instruction at the University of San Francisco. This faculty supervisor for this study is Dr. Nicola McClung, a professor in the Department of Learning & Instruction at the University of San Francisco.

The purpose of this study is to create an instrument that validly and reliably measures the construct of cogency, which is one dimension of the multidimensional construct of argument. The instrument will include a construct map, a prompt, a scoring guide, and student responses. The overall construct of cogency, a dimension of the multidimensional construct of argument used in the study, will be based on Toulmin's model of argumentation, which is identified by the Common Core (2010) and guides the teaching of argument in high schools and universities in the United States (CCSS, 2015; Graff, 2003; Lunsford, 2002; Kneupper, 1979).

During this study, the following will happen: You will participate in a moderation session. The moderation session establishes interrater reliability. Thirty papers randomly selected will be graded. All raters will grade 10 papers at a time, once using the SBAC scoring guide and once using the TDRMC scoring guide. Discussion and interrater reliability will be recorded for each set of 10 papers and presented in my dissertation to explain the rationale for the scoring and the possible iterations of the TDRMC scoring guide. In addition, revised versions of the TDRMC scoring guide accompanied with

discussion and rationale for revisions will be recorded, since the moderation session makes transparent the revisions of the scoring guide. Once interrater reliability has been achieved, I will grade the remaining 90 papers twice, once with the SBAC scoring guide and once with the TDRMC scoring guide.

Your participation in this study will involve approximately 6 hours during one session. The location of the moderation session will be [location removed].

The research procedures described will not incur any risks or discomforts to you as a participant. If you wish, you may choose to withdraw your consent and discontinue your participation at any time during the study without penalty.

Sincerely,

Millie B. Gonzalez-Balsam

Statement of Consent

I read the above consent form for the project entitled “Measuring Cogency in Argument in a Seventh-Grade English Classroom,” conducted by Millie Gonzalez-Balsam of the University of San Francisco. The nature, demands, risk, and benefits of the project have been explained to me. I am aware that I have the opportunity to ask questions about this research. I understand that I may withdraw my consent and discontinue my participation at any time without penalty.

Participant Name

Signature of Participant

Date

APPENDIX D: SITE CONSENT

Millie Gonzalez-Balsam
University of San Francisco
San Francisco, Ca. 94117
School of Education – Learning & Instruction Department

Intent of this Memorandum of Understanding: This MOU outline the study of Millie Gonzalez-Balsam will be conducting within the Palo Alto Unified School District (PAUSD). It articulates how PAUSD and Millie Gonzalez-Balsam will work together to gather data and information to support her dissertation study. Millie Gonzalez-Balsam’s research will be supervised by Dr. Nicol McClung, her Dissertation Chair from the Learning & Instruction Department in the School of Education at the University of San Francisco.

Research Questions: 1. Do the steps involved in the moderation session, which help to ensure calibration of scores and inter-rater reliability, help with fidelity of the construct of cogency? 2. Does a Teacher-Designed Rubric Measuring Cogency (TDRMC), a dimension of the multi-dimensional construct of argument, capture variability in student performance, and how does that compare to the distribution of scores from SBAC? Distribution is referring to shape, and spread of scores. 3. Specifically how does the TDRMC correlate to other established measures of writing and academic ability such as grades, SBAC scores, and GPA?

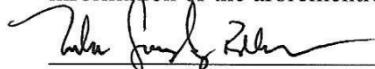
Goals: In the context of the current study, it is thought that a valid and reliable measure of cogency will be informative for both the student and the teacher. Students will obtain a clearer picture of their level of proficiency and additionally have an understanding of what they need to do to move forward. Teachers will also have a better understanding of instruction to aid students in moving them forward.

Date Collection:

- Access to Smarter Balance Assessment Scores for Language Arts
- Permission to interview 6 students
- Permission to use student essays

I will ensure confidentiality of all study participants and student records through the use of pseudonyms.

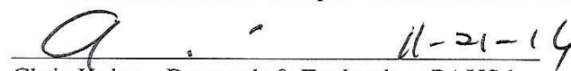
PAUSD agrees to provide data and access to all of the information above for the purposes of information of the aforementioned dissertation study.

 11-21-16

Millie Gonzalez-Balsam, Ed.D Student USF

 11-21-16

Katherine Kinnaman, Principal –Jordan Middle School

 11-21-16

Chris Kolar – Research & Evaluation, PAUSD

APPENDIX E: TRANSCRIPTION OF MODERATION SESSION FOR SBAC

Moderation Session: SBAC (1:05:22)

ME: Let's take a look at the rubric first. This is the category of "elaboration of evidence" for Smarter Balance [SBAC], and for 4 points, it says the response provides thorough and convincing support/evidence for the arguments and claim that includes the effective use of sources, facts, and details. The response clearly and effectively expresses ideas using precise language. So the bullet points are it has to include the following: comprehensive evidence from sources is integrated, references are relevant and specific, effective use of a variety of elaborative techniques which is, um, further specified, elaborative techniques may include the use of personal experiences that supports the argument. And, vocabulary is clearly effective for the audience, and purpose, effective appropriate style enhances content. Umm, any questions about that? Then for 3, the response provides adequate support, evidence for the argument and claim that includes the use of sources, such as facts and details. Umm, the response adequately expresses ideas employing a mix of precise with more general language, so what I'm noticing is that instead of clearly and effectively expresses ideas, it employs a mixture of precise with more general language. So it's parsing it out and it's qualifying it. Adequate evidence as opposed to comprehensive, umm, evidence from sources, some sources may be general as opposed to a 4, which is references are relevant and specific. Adequate use of elaborative techniques, umm, as opposed to effective use of elaborative techniques, and vocabulary is generally appropriate for the audience. Generally as opposed to clearly, and again, generally appropriate style is evident. So this is more general with a mix of precise language and some of the elements of a 4, but it's more now trending towards general. Then for number 2, to get 2 points, the response provides uneven cursory support/evidence for the argument and claim that includes partial or uneven use of sources. It expresses ideas unevenly using simplistic language. So for that the response would provide minimal support/evidence, umm, for the argument and claim that includes little or no use of source, facts, and detail. The responses' expression of ideas is vague, lacks clarity, or is confusing. Some evidence from sources may be weakly integrated, imprecise, or repetitive. So, weak integration, imprecise, repetitive, and vague references. Weak or uneven use of elaborative techniques, so the development may consist primarily of source summary, umm, as opposed to what you got in a 3 and a 4, and it may rely on emotional appeal. Vocabulary use is uneven or somewhat ineffective for the audience and purpose, umm, and it's inconsistent or weak attempt to create appropriate style. For a 1, the response provides minimal support/evidence for the argument and claim that includes little or no use of sources, facts, and details. The responses' expression of ideas is vague, lacks clarity, or is confusing. I'll have to say I don't really have 1s, umm, from the pool that I have, so that's a hint. Are there any questions about these?

ALL RATERS: No.

ME: So, umm, what I going to hand out to you are, these papers that I'm handing out, they do not correspond to the level at all. I'm numbering them so we know what

paper we are looking at. So, I'm going to give you a set of papers, so that we can practice looking at the rubric. And I intentionally chose papers where it's kind of between, so we really have to determine what the score would be. OK, so we're just kind of practicing with these. So that's the first set and I'm going to give you a whole bunch. That's one, one I have, my own. Umm, I didn't staple them, sorry.

I pass out four sets of papers to each rater to practice. We read the first paper silently together.

R3: Is there any chance we could have the prompt? Is the prompt the same for all of these?

R1: Do you want us to focus on elaboration of evidence now?

ME: The only thing you're looking at is that category. Umm, yeah, the prompt is the same. The prompt is "Argue for what constitutes a family." What constitutes a family is the question, and this came right on the heels of reading *The Outsiders*. We read *The Outsiders*, and then we read the *New York Times*, a 22-page article from the *New York Times* which we read in class and marked up and annotated. And it provided several portrayals of different types of families. So you'll see in, umm, we looked at, there was a gay family, a couple that were not married, but living together. There was examples of incarcerated, you know, family members, and you know that, there was an example and then there was another example of extended kin. So the extended kin family where friends become family due to circumstances. And so we looked at all of those and we defined what a family was, and we wrote down what are the values and assumptions that are attached to your definition of family in order to establish a criterion for what constitutes a family. So they, so then, they, so, then I said so now, after having read *The Outsiders* and the *New York Times* article, what I want you to do is to argue for what constitutes a family. And just left it at that.

R3: So, OK. Thank you.

We all silently read the same essay.

ME: Let's look at the exemplar first. Put the paper that we were all reading away. We will get back to that paper after we look at the exemplars.

I pass out papers that are between a 2 and a 3.

R2: What is the . . .

ME: So I'm just saying these are between a 2 and 3, and let's look at why.

R2: How is that helpful if we have to just pick one score? Like what's the objective of knowing it between . . .

ME: Because I think it really, if we look at how it's between a 2 and 3, it can force us to really try and understand what it would be. To try, and gain clarity.

R2: As a 2 to 3 based on your opinion.

ME: Based on my opinion.

R2: These are not SBAC scored papers.

ME: Well, no they are SBAC scored papers.

R2: They're not from the SBAC bank of scored papers.

ME: No, these are my students, and so I'm saying they're between a 2 and a 3, and I want to hear from you, and to tell me if you think it's a 2 or a 3 so we can really identify what level they're at. And then we're going to score the other papers.

R3: So the one we just scored was . . .

ME: That's going to be the practice one.

R3: Oh, that's the practice one. OK, and this is the calibration.

ME: So this is going to be between a 3 and a 4.

I pass out to all raters the set of papers that are between a 3 and a 4 and a set of papers that are between a 4 and a 5. I then notice a mistake and ask for the papers labeled between a 4 and a 5 to be passed back to me because they will be used in the calibration session for my instrument, TDRMC.

ME: Read them on your own, and look at the rubric, and I'll tell you why this was a 2.

(29:37) ME: Are we ready?

ALL RATERS: Yeah.

ME: Let's look at a 2. So if we look at a number 2, on the rubric it says that the responses provide uneven, cursory support for the argument and claim, umm, and little or no use of sources, but there are sources in here, wouldn't you agree? And, there are facts and details. Umm, it's kind of the integration, umm, it's kind of vague, it's a little bit confusing, umm, so if we look at. In the introduction for example, she has a quote in there when there is not supposed to be a quote. The introduction is really just to introduce the information. But let's look at body paragraph 1. She, umm, talks about, her topic sentence is that "Family is a group of people who may not always agree with you, but who will always love you." And yet the criterion is family is a group of people who will love you unconditionally, they will love you no matter what. So there is kind of a clear criterion, would you guys agree?

R2: What do you mean by a clear criterion?

ME: Well, umm, for what constitutes a family, she's defining what constitutes a family, that they'll be there unconditionally for you. And then in body paragraph 1, we look at how, this is the scene where. Well, first of all, it's not really clear where the scene is—where the scene takes place. I know where the scene takes place, but are you clear on where the scene takes place?

R3: Not offhand.

ME: Do you know the book?

R3, laughing: Yes, I know the book, but I haven't read it in like 5 years.

ME: But I'm gonna tell you, but do you have a clear idea where the scene takes place?

R3: No.

ME: And then so when we look at elaboration of evidence. Well. I gave it a 2 because I'm not really clear on how this is supporting the claim.

R1: Yeah.

R3: I'm in agreement.

ME: Any other comments?

R3: There's a lot of summary, that's one thing that gets called out here. Weak or uneven use of elaborative techniques, development may consist primarily of source summary, and I see that in a couple places here including, uhh, the beginning of body paragraph 1, and then again in, yeah . . . maybe a little bit again in body paragraph 2 in the middle.

ME: And kind of, umm, not really emotional appeal, so I wouldn't say that it was emotional appeal.

R1, *inaudibly expresses agreement.*

ME: But it is repetitive; it's kind of vague.

R3: Yeah.

ME: Umm.

R2: It's weakly integrated.

R3: Yeah, for sure.

ME: And weakly integrated, I'm taking because this rubric is pretty general. Umm, so OK, so it's weak integration of sources as well.

R1: Umm, hmm, yeah.

ME: So, summary of material and weak integration.

R2: So like you said, I don't see how the evidence supports the argument . . .

ME: Right.

R2: . . . which I'm not sure on the rubric [*inaudible*].

ME: Could you speak up?

R2: Sorry, the response provided minimal support evidence for the argument.

ME: Right.

R3: Umhum.

ME: OK, so we're in agreement that this would be a 2, on this paper.

R3: Yes.

ME: OK, let's put this to the side, and let's look at a 3. And a 3 is a little bit, umm, better, umm, and we're looking at the right-hand side, remember the SBAC one. So it's providing support, and yet what we were, what I was explaining before is that it's a mixture of, umm, precise with general and the sources are adequately integrated, but, umm, some references may be general. So, maybe not specific to, I'm taking

it to, I'm interpreting it as maybe not specific in support of, umm, specifically supporting the claim, but it's there, there's evidence there. And the vocabulary is generally appropriate. Generally appropriate style is evident. Although, that's not really clear what that means, generally appropriate style is evident, so, but I am looking at this rubric and trying to match it to this and saying my understanding of generally appropriate style would be that for the argument task. For the task so . . .

R3: Yeah, yeah . . .

(38:08) ME: OK, so why don't we read this. So to read this, probably the best is to look at probably, umm, the last sentence in the introduction to see what their thesis is and then just to read the body paragraphs cause you don't really need to read the conclusion, since we're just looking at the elaboration of evidence. If only we had to grade like that, right? Would be a little bit quicker. . . . Just let me know when you're ready. You're ready? OK. So, I say based on the rubric again that there is a mix of precise with more general language and, umm, and I'm looking at in body paragraph 1 where he says, he sets his quote up as "as said in the novel," again, I think it's weakly integrated. Umm, you know it's integrated, but it's still kind of weak; however, what puts it at a 3 for me is more of the, umm, it expresses ideas in a little bit more of a precise manner. Right? Umm, so that it does lends more support for the claim. So, I would say for example, after the first quote in body paragraph 1, "Unconditional love is what made the gang, Johnny's and everyone else's family. Johnny and Ponyboy see that unconditional love is what makes the gang, a pack of kids who will stand up for each other no matter what, such as what a family would do for each other." So there's an attempt to make a connection there that you know a family, since it's defined that, umm, unconditional love between people is what constitutes a family is what's said in the sentence in the introduction. So, he's kind of making support for that specific claim that he made in the introduction. You see that there, and then in body paragraph 2, umm, after, no I'm sorry in body paragraph 1, where he refers to Mr. Hill and Ms. Perez, the unmarried couple, umm, talking about "Mr. Hill wants the best for his children, he knows what he has to do to get the best for his children and is showing unconditional love even though he is not related to the children by blood. He has high expectations for the children. He doesn't push them too hard that they go crazy, and he's showing unconditional love towards them that way."

So I would say, again that's kind of going back to the support where he says, "it's unconditional love is between people," you know he doesn't specify, and so this example of an unmarried couple with children kind of supports that. What do you, would you guys agree?

R3: I think the analysis in the first body paragraph is stronger than the last example we saw.

ME: Yeah, and then in body paragraph 2, umm, he talks about incarcerated families and he talks about how the incarcerated parent is showing unconditional love by "keeping connected with his family at all costs." So, he really does make an effort and he talks about how he talks with his daughters very seriously just as a parent should. Umm . . .

R2: I thought that this paragraph to really reflect unconditional love.

ME: The second body paragraph.

R2: I mean I thought that I saw that this person was trying to definitely analyze the quotes and link them back to unconditional love, but maybe there was a misunderstanding of unconditional love or . . .

ME: Right . . .

R2: Umm, so I wasn't sure what you wanted to do with that because it's stronger than the previous paper but I didn't actually think the actual evidence related to unconditional love.

ME: And, in the rubric, it says that it's employing a mix of with precise and more general, so maybe body paragraph 1 would be where it's more precise and specific in connecting the evidence to the claim, but in body paragraph 2, kind of falls short, but still stronger than the paper we saw as a 2. So that's where it's really uneven. And, again, we're talking about the evidence. Right, it's uneven in the analysis. But, I'm also just going off of what the rubric says, right? And so we have to limit ourselves to the language in the rubric and what it is asking us to do instead of just interpreting the rubric. Does that make sense?

R2: Yeah, that's where I'm looking at where it says "adequate support or evidence for the argument." Would it be adequate support if it isn't showing, if it isn't really about unconditional love?

If one paragraph is and one isn't as strong, I think that would be adequate. It wouldn't be a 2, would it?

R2: No, I don't think it's a 2, I'm just bringing that up.

ME: No, no, I agree with you.

R1: It's not a true 3, because I was thinking it's not a 2, it's in between a 2 or a 3, because it doesn't relate. It's not proving its topic sentence or its claim which elaboration . . . if you're elaborating it should. Your evidence should connect, but it's not quite connecting.

R3: I think part of the problem is the topic of this paragraph is a little unclear because it seems like in the last sentence of the first paragraph, "However in order to decide if my argument is valid, we need to examine the other point of view." That's setting us up for a counterargument.

ME: Right.

R3: But then the topic sentence appears to be taking us right back to the argument.

ME: But, it's not exactly, because the way I actually taught this was when you are doing counterargument, you first have to say when it comes that you acknowledge that, yeah, "most of us will agree, however, where this agreement usually ends, however, is on whether unconditional love is actually the meaning of a family. Whereas some are convinced that relatives are the only thing that contributes to a family, others maintain that unconditional love is the key to a family." So, I would argue that he, the counterargument when you are establishing a

counterargument, you don't just jump into the counterargument, you have to acknowledge what is it that most people agree, and then jump into it by saying, however, and that is where this argument ends is pivoting now towards the counterargument.

R3: I think that what's unclear to me is whether the quotes are in support of the rebuttal or of the counterargument. Umm, so I wasn't clear on whether the quotes are intended to show us how some people might argue that relatives are . . .

ME: You know what this is unclear, what he, what this student did with the counterargument was provided examples to support their position as opposed to critique of whether or not, they did not provide critique. What they should have said probably is that where you know some people look at children of incarcerated families as those who are going to be at high risk, blah, blah, blah, blah, blah. And therefore those kinds of families are not very strong, blah, blah, blah, and then provided evidence to support the counterargument, and then the rebuttal. I think what's missing here is the counterargument and then the rebuttal.

R3: Yeah.

R2: I had another question about the language, umm, with the, "you," "my," not being the third person. Are we . . . and generally the way I've taught formal argument writing is no second or first persons, but are you interpreting this is that still, I'm looking at the point that vocabulary is clearly appropriate. Vocabulary is generally appropriate.

ME: The vocabulary is more I think for the SBAC rubric is more looking at the vocabulary that is related to the topic, not so much the grammar, because they have a separate category for spelling and conventions.

R2: So then that's style, so we don't need to take it into consideration, maybe the style, the more formal "you."

ME: You can use "I" in academic writing. That changed. When I started my doctoral program 6 years ago, I was telling my students this that we were in class, we had to write papers every class session for this one particular class on research methods, and it was explaining what was going on in a study. Every class we had to come in with a five-pager that really talked about why the study, the purpose, the methodology, whether it was valid, reliable, you know, looking at whatever it was that we were studying that week in that particular statistics class, and I wrote, "The writer agrees with the conclusion of the researchers," and my teacher said, "What the hell is that? You can use 'I.'" All of us in the class, had, our mouths just fell open. You can use "I," that rule has changed, but you can use "I" in academic papers, and that was 6 years ago. But, still, I still would argue for that this is "adequate," considering that it is a mixture, and in SBAC, they are just starting, well, there is no counterargument really. There is an attempt at a counterargument, but it's not successful. So, would that make it, I mean it wouldn't make it a 4. Therefore, their categories are 1, 2, 3, and 4. Right? And if we had to just do a 1, 2, 3, and 4, but not in between, and you had to choose between a 3 and a 2, this is why I gave it a 3. Do you see what I am saying?

R3: Umhum, yeah.

ME: Yeah, that's why I gave it a 3. OK, so now, let's read the 4. So these are tricky, aren't they? But let's look at again, the rubric, and please jump in. So it clearly and effectively expresses ideas using precise language, the evidence from sources is integrated, references are relevant and specific, elaborative techniques are effective, vocabulary is appropriate, and effective, appropriate style enhances content. So, I gave it a 4 because, umm, it says, in the criteria, in their introduction, at the end, "If we define a family by only the nuclear family, we would be excluding other families that do not fit the rules of the nuclear family." And the sentence before that is, "I think that unconditional love holds a family together and is what a family should be defined by." So, we look at the body paragraph 1, and we are looking at, "This is a close-knit group of friends that make up a gang known as the Greasers, they show each other unconditional love." Umm, and in here, the scene is showing when they both, Ponyboy and Johnny, meet Cherry for the first time in the movie theater, and Dally's rude to Cherry, umm, and Ponyboy is not like Dally. He doesn't have that kind, he's a really polite boy who would never talk like that to a girl, but he sticks up for Dally, because it's like, hey, we're in a gang, and you stick up for each other, we're like brothers. Umm, they care and love for each other. Umm, maybe not the strongest support, but again, the definition, the criteria is that, it's not just the nuclear family that's a family, it's people that you know, unconditionally love and support each other, and he's unconditionally loving and supporting Dally, even though he knows that what he did was wrong. Right? So, in the second example, putting Ms. Hill and Mr. Perez—they are putting off, putting your kids before your marriage, and it talks about how she really does fret about the money and, umm, looks at the pennies, you know, pinching pennies. They want to get married, but they're putting that off, because what comes first is their kids. And this couple's not married, it's her kids, and not Mr. Hill's kids. Then in the second one, it gets a little tricky, right, because in body paragraph 2, umm, there is a bit of a critique about incarcerated parents saying, "Obviously incarcerated parents do not fit the strict expectations of the nuclear parents, of the nuclear family." So there's a little bit of a criticism there, acknowledging that, yeah, and these parents try to keep their family together, umm, but then, the student acknowledges that "Studies have shown that even accounting for factors like poverty, the children of incarcerated parents are at heightened risk of serious behavioral problems, of doing poorly in school or dropping out, of getting into trouble with the law and starting the cycle anew." And that's evidence to support why incarcerated families, you know, children of incarcerated families have a really hard time. Umm, but, it says where the argument comes to an end is when we look at specific people and how the loss of their family makes incarcerated parents want to stay even closer to their kids' education and welfare. And then the rebuttal is where Mr. Singh tries his hardest to contact his kids through phone, letters, and emails despite the challenges of being in jail, and it would be hard to say that he wasn't part of the family because he was in jail. And umm, I believe the child is trying to attempt to say, the efforts really show that he's really trying to keep the family together and there's that

unconditional love binding them together. Although the second is not as strong, the second piece of evidence. Does anyone want to chime in?

R3: I think it's a successful counterargument paragraph. Umm . . . I was struck by the relevance of this quote here, uhh . . . "Even accounting for factors like poverty, etc. kids of incarcerated parents are at heightened risk." As you pointed out, it's a nice acknowledgment of the nuclear side, and a rebuttal of that. I think the second quote is fine as well. I mean, it, umm, I'm not familiar with that article, so maybe there's something even better that he or she could have picked but some might draw the conclusion that cohabitating couples do not recognize the importance of marriage, but on the contrary, the two do know the importance of marriage, "Ms. Perez must first . . ." So, a nice pivot in that sentence, and then, you know, whoever this is ends with a strong conclusion. I think if I had read this paragraph first, you know, before the other ones, I would have had a clearer sense of where the other ones were trying to go, you know, in a way, cause this feels successful to me.

R1: I think that last part is a bit of a departure from this nuclear family, you have to have a nuclear family versus not a nuclear family, he's going back to unconditional love as like the reason, like the emphasis. So in that way, it's a little bit veering off.

ME: I'm not sure I understand.

R1: Well, so this paragraph is all about . . .

ME: Which one, body paragraph . . .

R2: This body paragraph 2, it's about maintaining a family, meeting the rules of a nuclear family versus a family that's not a nuclear family but still is a family. But this last evidence is going back to unconditional love. It's not really proving the point of, you know, the example I can see, but his explanation of the example or her explanation is not quite, it's going back to unconditional love, which is still related but it's not extra evidence that proves his point or her point that you don't have to be a nuclear family.

R3: That's a good point.

ME: But don't you think that in, I'm going to push back a little, in the introduction, umm . . . "when it comes to the topic of family, there is no one written rule and finding the happy medium can be hard. In the end, however, most of us do. There are always two sides to defining family. One side holds strongly to the idea that a nuclear family is the best, and families that fail to meet these standards are not a real family. The other side believes that no matter what the family looks like, it is defined as a group of people that express an unconditional amount of love or love that has no limits or boundaries for each other. I think that this strong unconditional love that holds a family together is what a family should be defined by. If we define a family only by the nuclear family, we would be excluding other families that do not fit the rules of the nuclear family."

R2: I agree with [R1], because I understand that and I think where this, I think the rest of this argument, counterargument paragraph is about, that is about how it doesn't fit

the image, but then in the analysis of this piece of evidence, they didn't focus on the fact that the family doesn't look like a nuclear family; they went right back to focusing on unconditional love, whereas the rest of the paragraph was about the family not looking like a nuclear family. So then this particular analysis seems to be more about unconditional love, whereas the rest of the paragraph was about not being, nuclear families not being the only way. So they didn't talk about, they could have referenced the fact that this is an unmarried couple, and that the kids belong to one of the people, and therefore it's not a traditional nuclear family, and they missed that opportunity to describe this untraditional family.

R3: But that's a minor point in an otherwise thorough and convincing paper.

R2: Yes.

R1: Yeah, I didn't bring that up to say it's not a 4; I think I was addressing your question of body paragraph 2 [*laughs*], I forgot the question.

ME: I was saying in body, I agree with you, I agree with you. I was saying in body paragraph 2, that last piece of evidence wasn't as strong as the first piece, but it still, I thought, would make it a 4.

R3: Yeah, yep yep.

ME: It's definitely not a 3. Right?

R2: So is your phil——, I've done other scoring projects where if it's not the higher score, you give it the lower score, and that's not our philosophy that we're using today. Like I've done scoring where if it's not a 3 all the way, you give it the 2.

R1: Like if it's a 2 and a half.

R2: You go 2.

ME: Right.

R2: But we're not doing that.

ME: No, because SBAC doesn't, you know why, the reason is because on a SBAC rubric, you have 1, 2, 3, 4.

R2: Right.

ME: So, I'm sticking to it. So, if I had to choose, I would give this a 4, because there's nothing in between.

R2: Right, and I've done scoring things where you do [*inaudible*].

ME: No, I know, I understand, no, we're not doing that.

R2: OK, we're not doing that.

R1: I was thinking about that too, and then I reminded myself that this SBAC is for Grade 6–11. For seventh graders, you know, this could be appropriate for a 4.

R3: Yeah.

R2: No, it's strong.

R1: Versus for an 11th grader.

ME: OK, so why don't I grab those papers from you, or actually hold on to them, cause now you know what, you can refer to those, right? Now, we're going to practice, and I think that I know that we're a little bit behind, but, umm, so this is what I'm envisioning, that we're going to practice on, there's about, I think six papers each person will get, and they're at the various levels we just looked at, a 2, a 3, or a 4. Umm, and there's two papers that are going to be at the same level. So there's two that are going to be at a 2, two at a 3, and two at a 4. So we need to look at those papers and see what we think they are, talk about it a little more, and then I'm going to give you the first set of 10 papers where we're just going to look at SBAC, and the only time, and . . . and this I think we should probably just score like you do. You score it and it's pretty quick to do that, and then you'll give me your scores and I'll compute them.

Group takes a 10 minute break.

Moderation Session: Practicing on Exemplars (30:46)

ME: OK, let's start with "A," what did you guys give this one?

R1: I gave it a 3. I felt like it was a mixture.

R2: I gave it a 3 also.

R3: I gave it a 3 to 4.

ME: I gave it a 3 to 4 too. I said this paper was between a 3 and a 4. Umm, so now . . .

R1: Wait, I thought we're not doing in-between papers.

ME: No, no, we're not, but when I chose these, because it was really hard to find, anyway so. Why did you give this a 3?

R3: One reason that I didn't come flat out on either the 3 side, so one reason I had the 3 to 4, and not the 4, umm, is because for instance in the first body paragraph, the lead-in to the quote, so the integration of quote, umm, I thought was kind of weak. It devolves into summary, here there's way too much setup for that quote from Ponyboy. And then, umm, in the second body paragraph, I felt that the thinking wasn't very, umm, it wasn't directed enough, so, "These people think that incarcerated families do not count as a traditional nuclear family, because a parent is in jail." Then we have this quote. And then the author says, "The thinking of this point of view is that children who have these kinds of problems usually have bad parents to cause the child's troubles." Umm . . . and I don't know if that really is the thinking, that that quote establishes, uhh, yeah, so it felt like the thinking here in this counterargument paragraph is a little bit questionable, and then there's no strong conclusion.

R2: I gave it a 3. I thought that the first paragraph was at the 4 level, but the second body paragraph was about the behavioral problems in children more than about family and unconditional love. It seemed like it was more about the impact of parental choices on children, and so I looked at the rubric when we had talked about the

earlier essay, that it was that, umm, what was the mix, so adequate. So one strong paragraph, one weaker paragraph, put it at a 3 level.

ME: Right, OK.

R1: I agree with what the previous two said. I went with a 3, too, because I was just thinking about the calibration, and it's more than a 2, but it's not a 4, because the evidence, while they had cited evidence from both sources, I didn't think the integration was very effective, and it was a little bit unclear some of their reasoning, and it didn't quite support their claim or their thesis. So it was a 3 for me, and I just kind of estimated, because I think it was adequate, right, it wasn't uneven, nor was it thorough, but it was adequate, it was kind of good enough.

ME: And I said it was 3 to 4 for the same reasons as [R3], but I would say it's more a 3, because of the mix of precise with general, and the sources were not really integrated. They were integrated in a very general way. I thought, umm, and also the elaboration was weak, it was weak. OK, do we want to move on to "B?"

R3: I said 3 to 4, but on this one I was leaning towards a 4. So, I'm going to come out and say 4.

R2: 4.

R1: 4.

ME: 4. OK, and why? That's good.

R3: I thought the analysis was what pushed me in, so like the first body paragraph in particular, the second quote, this quote shows, "Ms. Perez puts her kids before her own marriage, which is showing a great amount of unconditional love. Spending money on the well-being of your kids, instead of spending money on something you really want, is very noble as well. This quote also shows that she cares about her kids even though they might not all be related." It's a nice explanation of the quote; it connects clearly to the unconditional love theme. And the second body paragraph, I thought there was a nice . . . there was a thorough analysis of the first quote. "Some people would claim that with only one parent taking care of kids, it cannot be considered a family." And acknowledgment of the other side, "While they might be right that the family might have an empty spot, it can still be considered one, since they have gone above and beyond to communicate with each other. Trying to really stay connected with your family, even if you might be sitting hundreds of miles behind bars," etc. I thought it was a nice, you know, nodding at the other side, and then refuting it. Umm, so I liked the analysis. They're real convincing.

R2: Yeah, the quotes were well integrated, and I was very solidly a 4, because I felt that this writer really stayed focused on the claim, and did not stray from what they were trying to prove.

R1: I rated it a 4, so I kind of kept in mind our calibration of estimating upwards so if they were an in-between I went with the higher. I did a 4 for the same reasons really because they were focused on their point. So, each paragraph had a point,

and the evidence followed that point or tried to, and then their commentary tried to explain it and stuck with it. So I gave it a 4.

ME: I said a 4 for the same reasons, umm, although I will say, as [R3] and [R1] pointed, that it's not exactly a 4, but I feel constrained by the language of the rubric because it said they are expressing their ideas using precise language, they are and they are using it in support of the claim. I don't see anything in here about analysis, so I can't really say that, but I'm just going by the language here and saying that they are using precise language, they are supporting their claim. The evidence is supporting their claim.

R3: I think where the analysis piece comes in for me is, it says the response provides thorough and convincing support, evidence for the argument, and we could also say that only about the evidence, but for evidence to be thorough and convincing without an analysis is . . . you know then we could also . . . I don't know, that's where I think the analysis piece has to come in from this rubric.

ME: I agree with you, but I'm looking at it on face value. And if I were to look at the rubric just on face value, I wouldn't see anything, I would just, for me what I'm seeing is convincing support and evidence for the argument. Umm, I don't the analysis piece. I'm sure maybe we should intuit that, and maybe I'm being too rigid, but if we are looking at rubrics that tell teachers what to grade—this is what they're telling teachers what to grade.

R1: I was reading it as part of the second bullet, elaborative technique, cause you don't just want to have a quote, right? You kind of have to explain it. And that's part of elaborating in your paragraph.

R3: And the title might back that up—elaboration of evidence.

ME: But I feel that the fault for me is that it could be clearer. And so we're pretty seasoned teachers. We know our stuff, we've done a lot of PD [professional development], but if you are, but I'm saying in general, you put that in front of a teacher, and they're looking at that, that's an interpretation we're all making, but the fact is that it is still an interpretation, right? OK, so let's move on to the next one.

We look at "C."

R3: That me again?

ME: Yeah.

R3: I picked the wrong seat.

Laughter.

R3: I gave this one a 3.

R2: I gave it a 3.

R1: Yeah.

ME: I did, too. Anyone want to explain why or just move on?

R2: I mean I was a 2 to 3, but I thought again it was that “adequate,” where it was a solid paragraph, and a weaker paragraph for the same reasons we discussed in our calibration.

All raters agree.

ME: And so, which did you think was the adequate paragraph and which was the weaker one?

R2: I thought the first paragraph was the adequate one . . .

ME: The second one has only one example.

R1: I put, so at first, I thought it was a 2, but then as I looked at this, I was reminded that, oh, I should estimate up. So, it would be a 3. But I think at first I picked 2, because I felt body paragraph 2 was very weak; it wasn’t as weak as the others we looked at because I thought there were a lot of generalities sweeping, and not enough evidence.

ME: Right.

R1: And they tried to incorporate evidence, but because what they’re saying is . . .

ME: A huge generalization . . .

R1: . . . yeah, so it needs, it requires more backup, but I would put 3.

ME: I’m closer to a 2 on this one, but because of the fact that body paragraph 2 doesn’t have a lot of, umm. It says basically [*referring to rubric*] “uneven use of sources,” and for me, and, “the response provides minimal support” argument. I feel that body paragraph provides minimal and it’s the counterargument.

R1: Umhm.

ME: And the counterargument is really important. Right? But, we’re only looking at “elaboration of evidence,” so we need to focus on that. I still, I feel like I would give this a 2.

R2: What saved this paragraph for me was that the coherency within the paragraph, that the counterargument was a question of what family is best for kids, and even though it was general it stuck, the evidence did support what they were, it was weak evidence, but they were trying to talk about what was best for kids, and the evidence they cited addressed what they were trying to the point they were trying to make. And they wrapped it up with “no reason why nontraditional families are not less suited.” So, I thought there was a coherency in the paragraph even though overall it wasn’t great. That’s what pushed me to that 3 level.

R3: Yeah, and similarly to that [*inaudible*].

ME: OK, you’ve won me over for the 3, because of what you said about the, there’s cohesion, and when and so that would differentiate it from a 2, in which we’re saying “it’s uneven and cursory, partial and uneven,” which would not create a cohesive paragraph. Right, so, OK.

R1: And, overall, there’s cohesion too.

All raters agree.

ME: OK, "D."

R3: So, I gave this one a 3 as well.

R2: I gave this a 3 to 4, but I ended up giving it a 4. I gave it a 2, because I thought the student had a lot of evidence, but I wanted them to explain it a lot more, and I felt like it was vague.

ME: I was a 2 to 3 on this one, and so I purposefully chose ones that were tricky, right? Because we really then have to go to the rubric. I'm in agreement with both [R3] and [R1], I mean and you [R2]. But, [R2], I wouldn't, umm . . . OK, let's see.

R2: Well I was very much on the fence of a 3 to 4, and I started looking at our calibration and it said up to a 4, I could easily be reasoned to a 3, but I don't see it as a 2.

R1: Maybe now I need to reread it, maybe I could see a 3 because it's like a mixture of accurate and general.

R3: One of the consistent flaws was the fact that the evidence didn't seem like the most relevant and thorough, and so I thought it was adequate. In the first body paragraph, "In this next example from *The Outsiders*," etc., "I had grown up with them and they had accepted me even though I was younger, because I was Darry and Soda's kid brother and I kept my mouth shut good." Umm, I didn't feel like that was the most convincing piece of evidence that the writer could have chosen. Umm, although now I'm wondering what wa . . .

R2: That's interesting because . . .

R3: Yeah, why did I think that . . .

ME: Yeah, because they're saying they love him and saved him from being beat up by the Socs even though he is different from them.

R3: Yeah . . .

ME: And they acknowledge that. They show him, they accept him for who he is, and love him, and they put themselves on the line for Ponyboy by saving him from the Socs. I thought that was strong.

R2: The integration of quotes bothered me, but I didn't want to base my whole score just on the integration of quotes.

ME: And then the second example, you know he's pushing, he's talking about himself, and his three kids, and he has unconditional love, even though he gets home late, he makes an effort, can't see them, he works hard to keep them together. So, I felt like that was strong. And then in body paragraph 2, umm, let's see, "others maintain the traditional. . ." OK, this is about an unmarried couple, he says that the main, that people who object are saying that the traditional nuclear family is the best kind. And he goes into an example showing a family that's unmarried with kids and a house, and they still love each other. And it shows that marriage doesn't need to be a big deal, and that they love each other. So, I kind of felt like that was weak even though there was evidence there. They planned to go, the

evidence really didn't support what he's saying. They're just talking about that they planned to go for one more year before getting married. OK, so that was kind of weak for me. Then in the next example, about the incarcerated man, Sing, who is in, you know, umm, even though they are, "Many people believe, blah, blah, blah, blah, blah," and this next example it's talking about a man who has been incarcerated but is insistent on keeping in touch with his family. But this is not really a counterargument, it's more support for the claim. So it is kind of uneven, umm, and that's why I'm thinking it could be a 2, because it's more uneven in body paragraph 2.

R2: Well, I thought that the first body paragraph was strong, the evidence was well selected, and even though, and the umm, explained.

ME: But a 3 is a mixture of "precise with general language," whereas a 2 is uneven, cursory. So, to me, I'm thinking that it is uneven?

R2: I'm not sure so what are you seeing as uneven.

ME: In body paragraph 2, it's uneven, it's not matching the strength of body paragraph 1.

R2: But the other 3s we gave were the same.

R1: Where one body paragraph was strong

ME: And the one was uneven? Hmm.

R3: I'm retracting what I said about the evidence; it seems fine to me.

ME: So maybe it's a 3 then—it is a mixture of precise and general.

R2: The thing that I . . . my weakness that I thought of in the second body paragraph was the analysis after the second quote is where it really fizzled off.

ME: Yeah, yeah.

R2: That was weak.

R3: I could see it could be elaborated . . . definitely. It's just one sentence, right?

R2: Right.

R3: And then the paragraph is over.

R2: And . . . informal "marriage doesn't have to be a big deal."

R1: I think why I gave it a 2 because I felt like all their explanations for evidence were vague. It could have been more specific. [*R3 and R2 signify agreement*] But I think I could also give it a 3, because it's like halfway, so if I were to, it's like an in-between.

ME: Yeah, and I did have it as an in-between, a 2 and a 3. Umm, but I could see it going into a 3. He doesn't really have, he's not really repetitive. Umm, but it is weakly integrated, and it's vague, and that's a 2.

R1: But then if it's a mixture . . .

ME: I mean, if we look at the ones that are a 3, OK, the first one is a 3, and the elaboration of evidence is a bit more developed. So if we look at "A" and you

look at body paragraph 2, and he, after the quote about incarcerated parents, “The thinking of this point of view is that children who have these kind of problems, usually have bad parents who cause the child’s troubles, however many parents like in jail like Sing who’s profiled . . .” like you know. There’s a, the elaboration is more developed. That’s what I would say, whereas I don’t feel that it’s developed in this one, so it’s going to be more vague. And I feel like that even in both body paragraphs, not just the second one. You have one sentence for that quote, “This quote shows how the gang accepts Ponyboy for who he is, and even though he is not like them he is still loved by them.” It’s a good sentence, but in this paragraph, the elaboration here of that is, well he also has, he has two sentences: “Despite tough relationships, Darry still loved and cared about Ponyboy. People show affections towards their friends and siblings unconditionally. This loving and caring is further illustrated in an article from the *NY Times* titled ‘The Changing American Family.’ One of the nontraditional families profiled is the family of Sing, an incarcerated parent who has been in jail for drug charges.” I felt there is more of an attempt to more clearly integrate it and there’s more of an attempt to develop, umm, you know, elaboration of evidence in “A.” And in “D,” I don’t see that as much.

R2: Hmmm . . .

R3: So you’re coming down on a 2 for “D,” right?

ME: Yeah, because I don’t see where one paragraph is better than the other, I think that they’re both uneven, and vague, and the sources are weakly integrated.

R3: It does say that in the 2 rubric, “The response provides minimal support or, evidence for the argument, and the claim that includes little or no use of sources.” And I don’t know that “little or no use” fits this, umm . . . because unlike “C,” this person does have two quotes per paragraph.

ME: True, right.

R3: Umm . . . so that’s where some of the vagueness of this rubric comes in.

ME: OK, you won me over with that. Because, there are sources in here—right! So, we’ll make this a 3.

R2, directed to R1: You all right with this?

R1: Yeah, I just have to remind myself to grade up or down.

ME: But, it’s not even, you know, the rubric is constraining it. The rubric is telling you that you have to do this, right?

R1: Right.

ME: OK, so let’s put these away. No, hold on to the exemplars, that’s what you want to do.

R2: Oh, good.

R1: The last set.

ME: Yeah, the ones I gave before . . .

R2: Keep all of them to look at for reference, or not?

ME: Yes, why don't you keep them all.

R2: Perfect, good.

ME: So, OK. So now, you have a stack in front of you and you're going to be rating the first 10, so it would be numbers 1 to 29, and they're not in numerical order, remember, because these have been randomly selected. So, we're only doing those, and we're using the Smarter Balance rubric. So just do the first 10, and then please put your scores here next to the paper that you graded. And remember we're looking at the thesis sentence pretty much, not the conclusion, and then body paragraphs 1 and 2. And you can write on these if you want [*indicating writing on the essays*].

APPENDIX F: TRANSCRIPTION OF MODERATION SESSION FOR TDRMC

Moderation Session: TDRMC (1:37)

ME: The warrants are what, are the hypothetical bridge between the evidence and the claim. Right? But the reasoning has to be evidentiary. So that you can deduce the reasoning based on the evidence. Does that make sense? OK, so your reasoning really needs to show how the evidence is in support of the claim—that is what is meant by “evidentiary reasoning.” Awareness of audience is that, in addition to the reasoning and sometimes it’s combined with the reasoning, the backing includes underlying values and assumptions for the criteria of the warrants, which in this case is, if they say that “unconditional love is what constitutes a family,” that has to be, you know, implicit in the backing in how, which is part of their reasoning, when they talk about how the evidence is supporting the claim. OK, if that doesn’t make any sense, ask me questions. This is . . . and I’m using this, I’m basing this on Toulmin’s model. OK, so let’s say for example, I can show you an example.

R1: Well you’re going to give us . . .

ME: Yeah, we’re going to look at exemplars.

Deductive and inductive reasoning—the writer again . . . so at a level 4, they’re able to use inductive and deductive reasoning again, but now, there’s qualifiers attached to what they can do. So, for the selection of evidence and use of warrants—it’s relevant but, and appropriate to the argument field is a level 4, but the warrant does not clearly, so it means that the evidentiary reasoning in support of the claim is not as clear and strong as it would be for a level 5. And, the backing includes some underlying values. The key words are in bold font, “**it doesn’t clearly,**” and for the backing, “**some evidence of underlying values and assumptions.**” In a level 3, again, the writer is able to do inductive and deductive reasoning, but here now for the selection of evidence and use of warrants, “**some but not all selection of the evidence is relevant,**” and is thereby because of that, because there is a mix, because there is “**some, but not all,**” then it results in the evidence not being appropriate to the argument field. And, the warrants therefore are incomplete, “**they provide incomplete evidentiary reasoning,**” because some of the evidence is relevant and some of it is not. As opposed to a level 4, where the selection of evidence is relevant and appropriate. OK, but when you have that mix in level 3, then the warrants are going to provide incomplete evidentiary reasoning in support of the claim. Does that make sense so far?

R3: I see how it’s scaled.

ME: And, awareness of audience, backing does not include underlying values and assumptions for the criteria of the warrants. OK, so here it doesn’t. In a level 2, the writer does again demonstrate the ability to use inductive and deductive reasoning, but in this case the selection of the evidence is not connected to the argument field, so it’s not connected at all. Even though there’s evidence it’s connected to the argument field, which is, you know, what constitutes a family. And the warrants do not provide evidentiary reasoning, and again, the backing

does not include underlying values. Do you see a difference between this and Smarter Balance?

We look at exemplars.

ME: So, let's look at ones that are at a 2 to 3.

ME: Let's start with number 1.

R1: So, you're going to tell us why you scored it a 2.

ME: After you read it.

We read the text.

ME: I would say that this is a 2. The evidence is not connected to the argument field. It is connected to the argument field about what constitutes a family but not really to the claim. The warrants do not provide evidentiary reasoning in support of the claim. He talks about, "Some people say you can only be family if you are blood related. The Greasers show that with unconditional love you can be family no matter what. They're always standing up for each other, that's what makes them family." And then he refers to the scene in the hospital. "The nurses and the doctors told us that we could only talk to family, but Darry finally got it through the guy's head that we're about as much as family as Dally and Johnny." This quote shows that even if you're blood related, being a close friend you can still be known as part of their family." And, while there's evidence there, it's not really connected to the claim. I mean the claim is even unclear. If you look at the thesis—"On the other hand, nuclear families are supposedly the best kind of family. They're the kind of family that can go over any obstacle facing them and their family." So, "On the other hand," he says, "some people argue you need more than unconditional love to have a good family." It's actually not clear what his claim is. And, so then it's not really clear what he's really arguing. And then he talks about an incarcerated person.

R2: Just to be clear, what would you say are the warrants in this paragraph?

ME: I'm not sure if I see—so, the warrants connect the evidence to the claim. And, that would be, warrants is another way of saying reasoning.

R2: Right—I'm just wondering if the way I learned it and taught it is the way . . . how would you define a warrant, give me an example of a warrant. I've taught it as a general rule about life like "even if you are not blood related, you can be known as family."

ME: But the warrants have to be very specific to the task at hand, they can't be really general, so warrants are always going to be specific to the argument that you're making and that's why I say argument field. And so the argument field here is looking at the issue of what *constitutes* a family. And so the warrants have to be specific to that. Does that make sense?

R2: Mmmhmm . . .

ME: And the reasoning has to be deduced from how the evidence is supporting the claim. But his claim is unclear, so it's not really clear what he's trying to say. He may

have evidence, and this is where I think a lot of times teachers get stuck sometimes when they're reading essays, is they'll see there's clearly evidence here and if they can write it well, then they get a high score, but the problem is really parsing out what is the kid constructing. Is there a claim, and even if there is evidence, what is it in support of? It kind of falls apart, so he's able to demonstrate the ability to define theme. You know, maybe he doesn't really redefine theme in this case when it's about . . . and, so the whole point of doing this moderation session is maybe I need to tweak my instrument a little bit, and say that at a level 2, you may be able to define what a family is, but you may not be able to redefine it, meaning as a result of looking at examples, and having an analysis, is there a creation of new knowledge going on in the paper? And I'm not sure that's happening as a result of his analysis, and selection of evidence, etc. Because, it's not really clear what he's arguing. He's not even really able to define family, because in his, you know, first paragraph, in the introduction, he says, "In discussions of family one controversial issue has been that unconditional love is all you need in a family. On the other hand, some people argue that you need more than unconditional love to have a good family. On the other hand, nuclear families are supposedly the best kind of family. They are the kind of family that can go over any kind of obstacle facing them and their family." So actually maybe a 2, "the writer doesn't demonstrate the ability to define the theme, and cannot redefine the theme."

R2: That was my next question: How are you defining theme in this rubric? What do you mean by "theme"?

ME: In this rubric, it would mean the writer is able to define family, because family is the theme, and we're talking about what constitutes a family. Right? And so he's not really clear on defining what a family is. He's giving you all these options, it's kind of like a multiple-choice . . .

R3: It does have that feeling.

ME: So there's really not a clear definition, and he doesn't really redefine it as a result. So, when we do this, what I'm saying is, why this is good, is to help me to tweak the instrument. So I'm going to ask, so would you guys agree that maybe a level 2, is that "the writer does not define the theme as it is defined traditionally, and does not demonstrate the ability to redefine the theme that utilizes an analysis"? Because there are clearly examples, but they're not being utilized to help. Does that make sense? So. I'm going to ask you then, to change where it says level 2, "the writer does not demonstrate the ability to define the theme, and does not demonstrate the ability to redefine the theme." So, part of this moderation session is it helps me when I hear back from you, and I'm talking through this, this is what it's supposed to do, it helps me to tweak the instrument, to really reflect the construct, what a cogent argument is, and that you can't really have selection of evidence not connected to the argument field and not providing evidentiary reasoning if you do not also have the ability to define a theme and redefine it. Does that make sense? Thank you, that helps with a level 2, and so, do you think that he's a level 2, then?

R2: What do you, just to get all the terms in the rubric—when you’re talking about “backing,” what are you referring to?

ME: Backing provides further support for the warrants. So the warrants are really more, how, it’s the hypothetical bridge that links the evidence to the claim, and then the warrants provide the underlying values and assumptions for the criteria. So, if we say, if you have defined family basically as any group of people that provides unconditional love and support for each other, and that’s the criterion that you are using to set up, you know, and if we’re also looking at the underlying values and assumptions for what constitutes a family, that would be, that’s the criterion and that would be the backing. The warrant shows how the evidence supports the claim, and it is looking at the reasoning that is deduced from the evidence in support of the claim: Is it logical? Does that make sense?

R2: Yes, I understand the definitions—I’m just wondering in practice, and so in this paper, would you say there are new warrants and backing?

ME: I would say—there are warrants, but they don’t really provide evidentiary reasoning, because the claim is confusing. Right? And so if we look at the first body paragraph 1, this quote shows that, OK, so he’s saying, well first of all in the introduction, it’s really not clear what the claim is even though in his topic sentence for body paragraph 1, he’s saying that “most people say that you can only be family if you are blood-related, but in *The Outsiders*, the Greasers show that with unconditional love you can be family no matter what.” But, that’s really specific to the example of the Greasers, it’s not really a topic sentence that’s also going to encompass the other example, which is the incarcerated parent. So, it’s not, you know he has, what I see in this paper are examples, reasoning, but there’s nothing connected to a claim, because a claim is nonexistent. So, even though there are warrants, they don’t provide evidentiary reasoning, umm . . . it’s not cohesive, it’s not a cogent argument, and does it include underlying values and assumptions for the criteria of the warrants? Again, this is where I feel it becomes a 2, because if there’s not a claim, and you have just evidence, but there’s not a connection, what are you making? What is happening?

Which one do you have for a 3? There’s a couple that’s a 2 to 3, can we look at a 3 to 4?

We read, to ourselves, an example of a level 3 to 4 paper.

ME: OK, this one is pretty clear. If you look at body paragraph 1, after the first quote, the student says, “It doesn’t matter what a family looks like, it can be any group of people that truly loves each other.” Then we have this example of talking about how Johnny is described as looking like a puppy that’s been kicked too many times. The quote is, “He would have run away a million times if we hadn’t been there. If it hadn’t been for the gang, Johnny would never have known what love and affection are. The people who love and take care of you are your family, and not just the people you are related to.” I feel like that is a statement, and I’m not clear how that reasoning is deduced from the evidence and how it supports the claim. That warrant to me is incomplete, and there is no backing that includes underlying values and assumptions.

R2: So, could you give me an example of what a solid warrant would be—what would you be looking for right there that would show the evidentiary reasoning?

ME: Let's look at another paper that I think would be a 5. Can you look at the paper that says "Many think of the nuclear family, as the right family." OK. So, umm, I'm going to read body paragraph 2, number 1, "When people think of family, most think of two parents and children all living in a nice home, well it's not like that. Any group of people can be a family as long as it has the one ingredient, unconditional love. Another type of nontraditional family is friends that have become very close and that have unconditional love for one another. In the novel *The Outsiders* by S. E. Hinton, the Greasers are described as the poorer ones, yet they act like a close-knit family. The Greasers either do not have a family or one that does not really care for them, so they are each other's brothers. 'We're almost as close as brothers, when you grow up in a neighborhood like ours, you get to know each other real well. If I had thought about it, I could have called Darry and he would have by on his way home or Two-bit Matthews. One of them would have got me and this car if I had just asked them but sometimes I don't just use my head.' So, if someone in the gang would be hurt, anyone would come by right away. They really have to stick together in a tough neighborhood like theirs. The way that the Greasers show love and care for one another demonstrates that the Greasers are truly like a family containing lots of unconditional love." So, it's talking about how they're supporting each other, and they have to watch each other's backs, and it is nontraditional. "Another example of unconditional love is illustrated in the *New York Times* article named 'The Changing American Family' by Natalie Angier. Ms. Perez is a divorced mother with three kids. Ms. Perez now lives with Mr. Hill, the father of her third son. Mr. Hill and she are worried about money issues. 'Nevertheless, she frets incessantly about the future. She'd like to go back to school, and set aside something for her children's college education. She won't buy cereal that is not on sale, and the last thing she wants to spend money on right now is a wedding.' Ms. Perez is willing to sacrifice anything for her children, like her wedding and even cereal. She's trying her best to make sure that her children have a good school education. She always puts her children first, even if it means giving up many things. This shows lots of unconditional love, which is the only thing a family needs to be a family. To conclude, a family is not only based on the look, but what lies beneath the surface. The only ingredient needed to make a family is unconditional love; however, not all people feel this way about nontraditional family styles, so we have to examine the other point of view as well."

R1: So that's a 5.

ME: So, let's look at the counterargument. It starts off again, and you can tell I use a lot of Gerald Graff templates to get them started. "When it comes to the topic of the modern family, most of us readily agree that the main ingredient in a family really is unconditional love. Where this agreement usually ends however is on the question of what the family looks like. Whereas some are convinced that unconditional love is the thing that matters most, others maintain that a true family must look like the ultimate nuclear family. Nontraditional families like

families with an incarcerated parent struggle in their everyday lives especially in regard to raising their children. It is hard to keep in contact and focus with your mom or dad in jail. Studies have shown that even accounting for factors like poverty, the children of incarcerated parents are at heightened risk of serious behavioral problems or doing poorly in school or dropping out, substance abuse, getting in trouble with the law and starting the cycle anew. This shows that a family with an incarcerated family member is not the best model in raising children as one parent is never there to support their children, nor is setting a good example. This may cause their kids to do poorly in school or get in trouble with the law themselves. Many incarcerated parents try super hard to stay in touch with their family, especially their children through e-mails, phone calls and even letters to not lose contact. Parents put so much effort in trying to talk with their children, which shows they care about their children a lot, but it also shows unconditional love. Another idea that similarly illustrates examples of a nontraditional family as a bad lifestyle is the sadness and loneliness that comes along with having an incarcerated parent.” So, she is kind of repeating. “Families are left to cope as best as they can not only with the deafening absence, the economic hardship, the grief and loneliness that separation from a love one can bring, but also with the stigma that accompanies a criminal conviction, the feeling of humiliation, debasement, and failure.” This makes it obvious that not only are the children having trouble but the whole family has this loneliness and humiliation laid on them. This is why this type of family absolutely does not work out. Although it may be a struggle for some families, others try to keep each other happy, and always positive with what they have even though sometimes it’s the hardest thing to do. Parents in jail are even preparing conversation cards before talking to their children and sending them homemade gifts to show they are putting an effort into connecting with them. As a conclusion, “nontraditional families are putting a lot of effort into making the family work. However, this sometimes fails to succeed not providing good support for the children and could also create an unfortunate lifestyle overall.” So, they’re acknowledging both points, you know, umm . . . what do you think?

R2: Tell me what is a warrant in this paper. I guess I’m really not quite following what the warrant is.

ME: OK, let’s do it with body paragraph 1. So, here the evidence, after the first example, “If someone in the gang would be hurt anyone would come by right away.” You know, and they’re saying, look this is how they feel about each, they’re going to be there for each other. And, they really have to stick together in a tough neighborhood like theirs. This way, the way that the Greasers show love and care for one another demonstrates that they are a family containing unconditional love. I don’t know, maybe that’s not so strong. Do you think it’s repetitive, or do you feel like it’s . . . I mean for me, the reasoning is deduced from the evidence, and it’s also talking about, you know in this way, they are demonstrating how they are a family, even though, as the topic sentence says, “It’s a nontraditional type.” Meaning that they’re, their parents, you know they’re talking about the Greasers as a whole, either do not have a family or one that does not really care for them. So, she’s including the Curtis brothers who do not have a mother and father or the

kinds of configurations of families where they don't really have anyone who doesn't care for them.

R3: The analysis here is on point. I think I was thinking of a warrant like you [R2] with it sort of almost like being a truism, or just a general. Maybe we have kind of a limited view of warrant but, like more of a general statement about life for instance.

R2: I'm not really sure what I'm looking for.

ME: Well, in the warrants, what you're looking for, in the literature on warrants and argument, warrants are argument and task specific, so they just cannot be a general truism. They have to be appropriate to the argument field, and they have to be, the reasoning has to be deduced from the evidence, and so when you're reasoning, this is the logical conclusion that you're coming from with the evidence. Right? And so when you're saying nontraditional families, and you're giving an example which is here, then your reasoning has to stick to the definition of what constitutes a family in the first place, which is articulated in the topic sentence. So it's almost like when you're doing the model for Toulmin, you go down with the claim, evidence, the warrant, and the backing, and then you kind of have to go back up again. Is the backing really including the underlying values and assumptions that you can get from the reasoning, and is the reasoning really showing how the evidence supports the claim?

R2: Right, and I read that book. So, for example, in this paragraph where they really have to stick together in a tough neighborhood like theirs—would that be a warrant, that people have to stick together in a tough neighborhood?

ME: No, the warrant would be, "If someone in the gang would be hurt, anyone would come by right away." Maybe this isn't such a clear example.

R2: Because that would be a restating of the plot—that that's the way they operate.

ME: There is another paper, so that's why I said it was between a 4 and a 5. So, can we look at another? This one starts with "Sometimes we judge things by the way they look on the outside." Do you see that one? OK. "Many people do not consider extended families or groups of families as a valid family unit. Although that kind of family may not be the typical family, it might be the one where we feel the most love, because of the people in the family, not the look of the family. For example, in the novel *The Outsiders*, by S. E. Hinton, a group of boys only have each other as family. The house where the Curtis boys live is a refuge for the other boys to feel safe and loved because they do not have to worry about being judged for who they are by their family. Because of the Curtis brothers, the boys feel safe to make themselves at home whenever they walk into the Curtis boys' home. 'We just always stick our heads into each other's house and holler, "hey," and walk in. Our front door is always unlocked in case one of the boys is hacked off at his parents and needs a place to lay over and cool off.' This shows that even though the Curtis boys are not related to their friends by blood, their home is still a place where the other boys can drop in and feel loved." Another example . . . so I would say that that is the warrant, and implied in that is also, umm . . . the

criterion for the warrant, which is that even though they're not related by blood, they're still there for each other, and they're providing unconditional love for each other. Does that make sense?

R2: OK, ummhummm, yeah.

ME: OK, so maybe this would be a better 5, and then the next example is about an extended family, "This article talks about all different types of family . . ." OK, "In one family, Caleb, the son of Becky Reese became very ill and died of lung cancer. Caleb's best friend, Matt became a son to Becky. 'Through that ordeal, the nine-month period, I became like a full-fledged member of the family.' The article then goes and says that Matt and Becky were having family dinners and talking on the phone every day. This shows that even though Matt is not Becky's biological son, the bond between Matt and Caleb brought Becky and Matt together so close they are like family." Is that clearer?

R3: I think that's helping us understand what a warrant is.

R1: So, this would be a 5?

ME: Yeah.

R3: And the warrants provide the evidentiary reasoning in support of the claim. OK, so it's the reasoning in support of the claim.

ME: And the backing does include the underlying values and assumptions for the criterion of the warrants, which is that unconditional love is what binds a family together, and in both examples, you see that unconditional love is implied there.

R2: So, lots of kids never define what they mean by unconditional love—does that play into your scoring? Lots of kids never . . . without explaining it.

ME: It's explained in the introduction, but we're not looking at that, we're just looking at elaboration of evidence. Right? And so . . .

R1: Sometimes in order to see if they fully elaborated, like I've had that question, well what do you mean by unconditional love, because I'm not sure what you mean. Or are you confusing, or is my definition different from your definition? And that's when I have had like, well what's your definition, and some kids just don't define it. So, for me, it becomes harder to judge their elaboration of evidence.

ME: Well, in this assignment, and it's seventh grade, remember . . .

R2: Umhmm.

R3: Umhmm.

ME: This was their first assignment for writing argument, so what we were looking for was can you define, can you establish a criterion, right for, umm . . . for what constitutes a family? What is the criterion, and unconditional love, so we didn't get as far as what you guys are talking about because it is their first . . .

R2: Well, this . . . when you say underlying values so we're going to infer the underlying values if they mention it. Doesn't really matter if they understood it or not.

R3: OK.

ME: Well, what we discussed was unconditional love is, it doesn't really matter because what they were looking at is, here's the traditional family is the nuclear family, but does the nuclear family always contain unconditional love? Can other family units or can different configurations of family contain unconditional love? Does that make sense?

ALL RATERS: Yes.

We score the papers. Then we discuss our scores.

ME: No clear claim. So basically the student was not able to use the language in the rubric. The student did not demonstrate the ability to define the theme and did not demonstrate the ability to redefine the theme, which in this case is family. So, no deductive and inductive reasoning, and as a result of reading this paper, you guys helped me to tweak my instrument so that the number 2 now does not demonstrate the ability to define the theme, which in this case is family, and the writer does not demonstrate the ability to redefine family. Even though there's evidence, selection of evidence is not connected to the argument field of family because the warrants do not provide evidentiary reasoning in support of the claim, and their underlying values as a result, well, you know what—do you think there are some underlying values and assumptions? I think that there are.

R2: I think that there are—right, that you don't have to be related.

ME: Yeah. I guess a 2 would be no inductive or deductive reasoning, and that there is no . . . the warrants do not provide evidentiary reasoning in support of the claim, but you can have underlying values and assumptions for the criteria of the warrants. Right? Does that make sense?

R3: Say it, say it one more time [to me].

ME: The writer does not demonstrate the ability to define the theme or to redefine it. Selection of evidence is not connected to the argument field, and the warrants do not provide evidentiary reasoning to support the claim.

R3: Yeah.

ME: OK, there's no connection, but the backing does include underlying values and assumptions for the criteria of the warrants. So, the warrants are not connected to, are not really linking the evidence to the claim, because the claim is unclear. Right?

R3: Yes.

ME: And as a result. So that's good, but there are still, but he still knows that unconditional love is what constitutes a family. Right?

R3: So the backing may not include, may not include values and assumptions? Or the backing, may or may not.

ME: Well, actually, the backing, a lot of times they are able to say that unconditional love is what constitutes a family. Right? And that would be the underlying values and assumptions for the warrant, but they are not providing evidentiary reasoning, and

they're not really creating new knowledge by, you know, by basically redefining the theme. Does that make sense? [R1], you look puzzled.

R1: I'm just not, are we talking about changing the wording then? Or no?

ME: Yeah, this is what this whole session is about—it helps me to tweak the instrument, so that it's more reflective of the construct.

R1: So, [R3] suggested “may or may not,” are we . . . ?

R3: I think it sounded like, maybe “backing does include underlying values and assumptions”?

ME: I would say it does include, yeah, or should we say “may or may not” to make it even, because there may be times when it's not really . . . ?

R2: I was trying to imagine the case where a student does not include the underlying values but could potentially have the rest—I don't even know if that's a possible combination. Could you evidentiary reasoning if you didn't have the underlying assumptions? Is that possible?

ME: I think we'll find out going through the papers, but maybe for now say that the “backing may or may not include.” “Backing may or may not, because there is a possibility that they may or may not” [*at this point, we are all rewording my instrument*].

R3: That gives us a little more room, because the 1 is off topic.

R1: Then level 3, should that be, does not still?

ME: Level 3, then should be “may or may not” as well or “backing mostly.”

R3: OK.

R2: Some underlying.

ME: Backing mainly or is that or do we see that in writing?

R3: It's hard to . . .

ME: Why don't we determine that once we get to a 3?

R3: OK.

ME: OK, let's put the 2 aside and then look at a 3 to 4. Umm . . . so this one starts with “Since the 20th century the American family has . . .” What number is that? OK. “Since the 20th century, the American family has been rapidly evolving . . .” and I'm going to go to [*starting to quote the first piece of evidence for example 1 in body paragraph 1*] “When you are part of a family, you look out for each other and make sacrifices out of unconditional love as seen in the quote from *The Outsiders* when Dally is giving directions and advice to Johnny and Ponyboy following the murder of Bob. ‘He handed me his worn brown leather jacket with the yellow sheep-wool's lining. “It'll get cold where you're going and you can't risk being loaded down with blankets.”’ This shows how Dally was putting himself in danger by helping Johnny after he murdered Bob. His sacrifice shows unconditional love for Johnny and Ponyboy because the police was after them,

and Dally was risking being an accomplice in the murder case. Family members will make sacrifices either small or large to show their unconditional love for each other.” So, I see in here, so far, a warrant and, a backing. And, the backing is “Family members will make sacrifices either small or large, to show their unconditional love for each other,” which is a value and an assumption about family. Right? And the warrant, “His sacrifice shows unconditional love for Johnny and Ponyboy because the police was after them and Dally was risking being considered . . .” That’s actually the evidentiary reasoning that comes out of the evidence.

R3: Yeah.

ME: And then the backing is that next part? Would we all agree with that? So then, it goes on to the next example about Ana Perez, and not wanting to spend money on a bunch of stuff because she wants to set aside something for her children’s college education. This quote shows Ana’s unconditional love for her children and their future as she is putting both her schooling and her wedding second for her children’s college education, and unconditional love is the key aspect of a family. So, I would say for a level 3, as we’re looking at this, it says “backing does not include,” umm, “backing does,” I would say, “backing mainly includes” . . . right? Because I don’t really see a backing in that one—I see more evidentiary reasoning after the quote.

R2: I thought you said unconditional love was the backing—was that assumption?

ME: Well, it is the . . . underlying value and assumption for the criteria of the warrants. Yeah, but I felt that in the first example, it was a little clearer.

R2: Umhmm.

ME: Family members will make sacrifices as opposed to, umm, there’s more of a development to that, as opposed to this one where it says “Unconditional love is the key aspect . . .” She’s just stating it—showing and not telling. No, telling not showing.

R2: Plus I was wondering if it was warrantable because it was some underlying, it’s not an absence of underlying values.

ME: So, you’re saying for a 4?

R2: Well, I was wondering if it was the 4 level because it’s “some underlying values,” right, because the first one was stronger, at least so far in this paper. There isn’t an absence of underlying values. Because the 3, seems, at least the way it is right now . . .

ME: But if we have a 2, “may or may not,” then the 3 has to change, because I don’t know if I would call this a 4. This paper—would you?

R2: That’s kind of where I was in my head, that maybe we should keep going.

ME: So, you think this is a 4?

R2: Well, the evidence is not. Like I don’t think the reasoning follows necessarily from the evidence.

ME: That's what I was thinking with this second example in body paragraph 1—I thought it was stronger in the first example, and if we look at the second paragraph, “When it comes to the topic of . . . , whereas some are convinced that unconditional love is the most important aspect of being considered a real family, others maintain that a family must be traditional in functions to be really considered a family.” I think what she's trying to say is to look like the nuclear family. “A *New York Times* article profiles a homosexual couple and talks about the controversy over same-sex marriages.” And, she has the quote, “Angier later goes on to say children of same-sex parents are academically and emotionally indistinguishable from those of heterosexual parents, therefore contradicting the previous statement. This shows that same-sex parents are just as capable of raising a family as there is no concrete evidence of their children being academically, emotionally, or socially challenged.” And, that steers off the topic, now we're talking about whether or not children are, you know, the well-being of children as opposed to unconditional love. I mean maybe you could say what's implied is if a child is stable, then there's unconditional love, but it's not clearly stated.

R2: Well, what I've found, and I don't know how this plays out, because the rebuttal was often in these papers about the environment it created for children, then they picked the evidence related to that, so, what I found is the rebuttal, the rebuttal wasn't really in rebuttal to the fact of unconditional love, like the rebuttal was already off. Does that make sense? Like they're saying, that somebody may argue that unconditional love is the most important, or, sorry, somebody might disagree that unconditional love is the most important ingredient because they might argue that you need a healthy environment. Right? So, then she's trying to prove or that he's trying to prove that this is not an unhealthy environment. So, in that case, it is in topic, it is supporting the claim. Right? Because this person is trying to prove that alternative families are not unhealthy.

R3: Yes.

ME: Right, OK.

R2: Is that right? I said it in a weird way, but I felt that like it's not explained very well, but the point they're trying to make was that this is not an unhealthy environment. But, then it switches to unconditional love.

R1: It's kind of all over the place.

R2: But the main claim in this paragraph was nontraditional families provide a healthy environment for emotionally stable children.

R3: Yes. And the second quote's reasoning doesn't really do that.

ME: So, then it would be a 3, because “some, but not all selection of the evidence is relevant.” And, but the backing, does the backing not include underlying values and assumptions?

R2: I think it does. It can be bound together with unconditional love.

ME: Yeah, so we would take the “not,” out for 3, the backing. I would say the backing “mostly,” not, the backing somewhat includes underlying, I would put a qualifier there. It’s “somewhat,” because I’m still kind of struggling with whether or not it does. So the backing “somewhat includes underlying values and assumptions.” And, how is that different from “may or may not”? Well, “somewhat” implies “some of the time.”

R3: And there’s an attempt throughout, right. A consistent attempt, where maybe in the 2 . . .

ME: So it’s about intention? There is an intention, I’m writing this down on the side, you might want to do that on your instrument.

R1: How does “somewhat” differ from . . . ?

ME: “May or may not”?

R1: No, “some” in level 4.

ME: Then, maybe we need to look at a level 4 to determine, umm, backing, umm, “unevenly includes” or “inconsistently” . . .

R3: That speaks to what we just saw in this example. Right? Where it did and then it didn’t.

ME: So then would it be “inconsistently” or “unevenly”?

R3: I like “uneven,” but I don’t know if I have a strong preference.

ME: What do you guys think?

R2: Well, then it gets at the 3, that sometimes has a difference . . .

ME: I’m sorry . . . ?

R2: If 3 was going to be “sometimes,” then how is “unevenly” different from “sometimes”? Like working our way down to the 3.

ME: Umm, because . . . a 4 includes “some,” a 3, “unevenly.”

R2: Oh, the 3 is “unevenly.”

ME: Yeah, so how is “some,” I still think “some” and “unevenly” are . . .

R1: Yeah, that’s still unclear to me.

R2: I like the idea of “attempts,” because I feel like the 3 attempts it, but it may not be clear.

ME: Well, maybe, say that—“backing is attempted, but not clearly expresses,” how does that sound? “The backing attempts to provide underlying values and assumptions for the criteria of the warrants, but it is not clearly expressed.”

R1: Are you saying that for a 3, or a 4?

ME: For a 3.

R1: “Attempts but not clearly . . .”

ME: “Backing attempts to include underlying values and assumptions for the criteria of the warrants, but it is not clearly expressed.”

R1: Yes, but then if we say that, is a 4, “clearly”? And then a 5 would be . . .

ME: So then a 3, let me just get the wording on number 3, because I want all of you, so let’s hear this first—“The backing **attempts** to include underlying values and assumptions for the criteria of the warrant, but it is not clearly expressed.” And a number 4 would be “backing includes **some** underlying values and assumptions for the criteria of the warrant, and is **clearly** expressed.” So the “**some**” that it does include is at least “**clearly expressed.**”

R2: Do you want to say “**effectively**” at all instead of “**clearly expressed**”?

ME: But it is not effectively expressed?

R2: No, sorry, I was trying to weave in “**effectively**” into the backing, “**effectively includes,**” but that doesn’t sound . . .

ME: Yeah, because that would be, kind of, how is that different from a 4?

R2: Well, I was thinking that was a 4.

ME: I mean how is that different from a 5? Because a 5 is, includes “**underlying values.**”

R2: Because it has “**some.**”

ME: Oh, OK.

R2: Well, I think we have two different things going on here, like quantity, and quality. Right? For your use, do you have a preference? Are you going for quality or quantity?

ME: Both. You need quality and quantity because the quality is talking about how effective, and how clear it’s being expressed, the quantity is how consistent. Right?

R2: Right.

ME: So, “**backing includes some underlying values and assumptions for the criteria of the warrants that clearly expresses, that is clearly expressed,**” or “**backing is clearly expressed and includes some,**” how about changing the order: “**backing is clearly expressed and includes some underlying values and assumptions for the criteria of the warrants,**” and for 3, “**backing is not clearly expressed.**”

R2: OK, here’s my question: Is the backing the underlying value, or is the backing something else that also includes underlying values? The way it sounds right now is the backing is something, and inside that backing you may have underlying values.

ME: The backing only includes the underlying values and assumptions for the criteria of the warrants. OK, so maybe some backing is provided.

R2: OK, so would you want to say, clearly expressed backing, because on the 4, right now, the way it sounds right now is that there's a backing and the backing is the underlying values.

ME: Clearly expressed backing includes . . .

R3: That's nice.

ME: . . . some of the underlying values and assumptions for the criteria of the warrants.

R2: OK, here is what I am realizing, could you then have . . .

R3: Whew! Complicated.

ME: OK, how about just saying "**the quantity and the quality of the backing,**" "**the quality of the backing clearly expresses some of the underlying values and assumptions for the criteria of the warrants.**"

R1: Wait, say that again.

ME: "**The quality of the backing includes some underlying values and assumptions for the criteria of the warrants.**"

ME: No, we're talking about effective being expressed clearly, and whether or not there's enough of it. So, umm . . .

R3: I think, umm . . . so the 5 is "**backing includes underlying values and assumptions for the criteria of the warrants.**" That's just kind of a statement. You don't have any quantifiers.

R1: Maybe a 5 is "**clear and consistent.**"

ME: "Includes clear and consistent underlying values."

R1: Right, because consistent is like it's always there, it's not absent in one paragraph.

All the raters agree.

ME: OK, so, "**backing includes clear and consistent values and assumptions for the criteria of the warrants,**" and a 4 would be "backing includes . . ."

R1: Maybe it's just "**clear.**"

ME: "**Clear, but not consistent.**"

R1: I think "consistent" is like if it's all running throughout and it's all clear, that's the best. Isn't that what we want? So, that would be a 5. So, then, a 4 is missing something, so what's the qualifier?

ME: So, the backing includes "**clear, but inconsistent.**"

R3: "**Clear, but possibly inconsistent.**"

R2: So they could have an incorrect backing, but if it were clear, it would be OK?

ME: No.

R3: No.

R1: No.

ME: Backing includes clear, how about . . .

R1: **“Mostly clear and consistent, but not always.”**

ME: Yeah, we could say that, **“mostly clear and consistent.”** So a number 5, **“backing includes clear and consistent underlying values and assumptions for the criteria of the warrant,”** for a number 4, **“backing includes mostly clear and consistent underlying values and assumptions for the criteria of the warrant,”** a number 3 . . .

R2: Is it too vague to say “some clear and consistent,” I mean for a 3, would you still want them to be clear and consistent?

R3: Well, we just saw our example here, which was not consistent.

ME: I would say here that the backing would include uneven, well, it would have to be uneven, but we could say inconsistent underlying values, unclear and inconsistent underlying values. OK?

R2: So, “backing includes unclear . . .”

All raters ask, would you do or . . . and/or?

R2: No. Because of what you just said, if it’s clear, but inconsistent or inconsistent, but clear. So, yeah, so you need both.

ME: Yeah, so I would say it’s unclear and inconsistent.

All the raters agree.

R3: And then the 2.

ME: **“Unclear and inconsistent assumptions for the criteria of the warrant,”** and then the 2, we would have to take the **“may or may not,”** or do we leave that? **“Backing may or may not,”** no, we could leave that 1 is **“inconsistent and unclear.”** So, they still have backing in there, but it’s the quality is clear, and the quantity is inconsistent. And a 2 is that it **“may or may not contain, backing”** **“may or may not include,”** keep the language consistent, **“underlying values and assumptions for the criteria of the warrants,”** so it’s the possibilities that it may or may not even be there. Right?

R1: Is it possible for you to revise and . . . ?

ME: Resend it to all of you—sure.

R1: That way we have, you know, all . . .

ME: I will do that when I get home, because I want to push you to try to get this to me.

We check in with each other after scoring 10 papers and have questions regarding what constitutes a 4.

ME: “There has to be unconditional love in it. For there to be unconditional love you have to go through the thick and thin with them. In the novel, Cherry thanked Ponyboy and Johnny for not jumping into Dally’s dirty talk, but then Johnny stood up for Dally by saying that he’s not a bad person.” She has her quote. “No matter what your friends do, you always have to stick up for them because you are their

friend.” OK, umm, another example of unconditional love is described in Ms. Perez and Mr. Hill’s family: “In the *New York Times* article, “The Changing American Family,” by Natalie Angier, an unmarried couple who are very committed to each other and their children is described. Ms. Perez and Mr. Hill are in a relationship, but they are not married. They have one kid from their relationship and she has two other children from a previous relationship. ‘I come home every night,’ he said, ‘they might be asleep when I get home, but I’m here every night. I’m always pushing them hard to do their very best because, maybe sometimes a little too hard.’ When parents push their children hard, they are doing that because they care about them and they want their kids to succeed, even if the parent is not the biological parent. In spite of what I think, we have to look at the other side of the story in order to fully understand the topic.” The claim here is that “For there to be unconditional love you have to be able to go through the thick and thin with them,” you know, does that? Then, in the second paragraph, “When it come to the topic of family, most of us will agree that to be considered a family, you have to show unconditional love. Where this agreement usually ends is whether you have unconditional love in a nontraditional family. Whereas some are convinced that nontraditional families can provide unconditional love and support, others believe that in only a traditional family there can be unconditional love. Angier describes Mr. Wayser and Mr. Schulte, a pair of gay parents who have adopted six children. Each of them has their role of taking on a parent. Mr. Wayser works as a lawyer. Mr. Schulte is the one who stays at home with their kids. ‘Some critics have expressed concern that the children of gay parents may suffer from social stigma, and the lack of conventional adult role models, or that same-sex couples are not suited to the monotonous rigors of a family life.’ Some people might say that all children that have gay parents may suffer from social stigma, but I believe it doesn’t matter what type of family you are in, because in every family there is unconditional love. As well as that, new studies from Stanford University show that the children suffer from social stigma because of the split between the children’s biological family.” So, she’s talking about divorced parents. “Another family that faces . . .” OK, we have incarcerated parents, “seven times more likely to go in jail, that live in low-income neighborhoods, plus if a parent is arrested and thrown in jail, studies have shown that the children have a high risk of behavioral problems.” We’ve seen this quote many times before, so I’ll go to the reasoning. “Families with a loved one in jail will feel bad about themselves for many different reasons. Sing is a man who has been in jail for fifteen years. He has one son and one daughter that he stays in touch with so he can still be in their lives. Similarly, there is a man named Rob who has been in prison for four years. He has three teenage daughters. He makes sure he calls each girl once a week, and sends drawings and crocheted gifts to them so he can keep in contact with them.” So, what do you think of that? It’s kind of mostly clear and inconsistent, I mean mostly clear and consistent, for . . .

R2: Backing?

ME: Or, do you think it’s a 3?

R2: My thought it was only a 3, on this one.

R1: I thought it was, yeah, the whole thing I would rate it a 3.

R2: And the reason I was thinking was that it was incomplete evidentiary reasoning. Which one did we rate as a 5? Was it the one, “Sometimes we judge things by the way they look on the outside”?

R1: Yeah, I think it was that one.

ME: That’s a 5?

R3: Yeah.

ME: OK, that’s a 5. I think the 4, then, is the one, “Many think of the nuclear family as the right family.” Did we give that a 4?

R2: Which one is the 5? “Sometimes we judge things by the way they look on the outside”?

ME: Yeah. “Sometimes we judge things by the way they look on the outside” is a 5. I think this one, “Many think of the nuclear family as the right family.” What number is that?

R3: Number 4.

ME: I think this is the 4.

R3: Was that a 5? No, I don’t think so.

R1: OK, “Sometimes we judge . . .” You did point out that backing as a 5, because it’s clear.

ME: OK, then the “Many think of the nuclear family as the right family . . .” I think this is a 4.

R3: OK.

ME: So, I’m just going to quickly go, OK, I’ll read this really quick. “When people think . . .”

R2: Could we just read it in our heads?

ME: Oh, sure, sure. The “Many think of the nuclear family as the right family,” that would be a 4.

R2: I thought it was on track to be a 5, but then there was . . .

ME: So, just to be clear, because I want to make clear that we have our exemplars. We have identified a 5, which is “Sometimes we . . .” What number paper?

R1: It’s number 3.

ME: I’m writing paper number 3, so it’s not confusing. Then the 4 is the one, “Many think of the nuclear family as the right family . . .” And, what paper is that for you?

R1: It’s number 4.

ME: And then the number 1 paper for you is a 2. And then the 3 paper is which one?

R3: I thought it was the one with the Schulte-Wayser.

ME: Oh, “The Schulte-Waysers are your typical American family . . .” That’s a 3.

At this point, there is a question as to whether we read this paper.

R3: Which one did you have as a 3?

ME: Oh, and that’s paper number 7.

R1: Yeah.

ME: But I think that that one, the Schulte-Wayser, could also be a 3. And I want you to take a look at that, if you go, “The people who love and take care of you are your family, not just the people you are related to. Johnny’s mother and father hate him, and the gang love him. There’s not even a choice in my opinion.” I’m like, what the heck is that? You know. That could be, that could actually be a 2. Did we identify a 2?

R3: Yeah, the 2 is the first one, “Unconditional love is a great thing . . .”

ME: Oh, right, so this one, this could also be a 3. You can put that in your pocket. What number is that paper for you?

R3: The Schulte-Waysers is a 3.

ME: OK, so we have . . . “The 20th century . . .” is a 3. I’m just putting that other one in there too, as a 3. So, now we have identified the papers that are the exemplars, that are going to help you, guide you, and tweaked the instrument. And I don’t know if you need the other papers, so maybe you can give me the ones that you’re not using so you’re not, and you keep the pile.

R1: I’m just going to hang on to all of them. Is that OK?

ME: I’d rather you didn’t. So, just hang on to the exemplars, and then keep the ones that you have to score.

R3: And you’ll send us the updated instrument.

ME: And I’ll send you the updated instrument. Hang on, don’t leave yet. So you all have your exemplars, you all have your paper with the scores.

R1: And if we come up with rubrics, or suggestions or questions, do you want us to keep track of it . . .

ME: If you have some questions, yeah, you could either call me and, email me.