

STUDENT WORK IN CURRICULUM-BASED ASSESSMENT TASKS: AN
EXPLORATORY VALIDITY STUDY OF A MECHANISM FOR ASSESSING SMP3B

By

Amy Elizabeth Ray

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics Education – Doctor of Philosophy

2018

ABSTRACT

STUDENT WORK IN CURRICULUM-BASED ASSESSMENT TASKS: AN EXPLORATORY VALIDITY STUDY OF A MECHANISM FOR ASSESSING SMP3B

By

Amy Elizabeth Ray

Research in mathematics education has shown that mathematics assessments have become highly consequential for students, teachers, and educational institutions (Hamilton, Stecher, & Yuan, 2012; Swan & Burkhardt, 2012). Even so, traditional assessment methods are often lagging behind reform curriculum design and advancements in instructional methods (Shepard, 2000). One advancement in instructional practices is the promotion of mathematical habits of mind in mathematics classrooms as evidenced by various standards documents and seminal works (CCSSI, 2010; NCTM, 2000; 2014; NRC, 2001; Seeley, 2014). My study focused specifically on one aspect of the Standards for Mathematical Practice, SMP3b, “*critique the reasoning of others*” (CCSSI, 2010, p. 6).

Extending the work of curriculum researchers studying the use of student work embedded in textbook tasks (Gilbertson et al., 2016; Going, Ray, & Edson, in preparation), tasks provided in the student textbook for classwork or homework, I explored the use of student work embedded in curriculum-based assessment tasks, tasks included in curriculum-designated assessment materials. Because curriculum-based assessments can be highly influential on the mathematics promoted in classrooms (Hunsader et al., 2013; 2014), it is important to consider what kinds of opportunities students have to make sense of, or critique, someone else’s mathematical thinking on assessment tasks and how these opportunities compare to tasks in student textbooks.

The purpose of this study was to explore whether or not student work embedded in assessment tasks can be used as a mechanism for assessing SMP3b. I gathered data from text

analyses of curriculum-based assessment materials from five seventh-grade, CCSSM-aligned curriculum series, interviews with students (n=6), and interviews with teachers (n=6). Using these data sources, I investigated (1) the prevalence and nature of student work assessment tasks (SWAT) as compared to student work textbook tasks (SWTT), (2) students' experiences with SWAT, and (3) teachers' perspectives on SWAT and students' written work on SWAT.

My text analyses findings revealed substantial differences between students' opportunities to make sense of someone else's mathematical thinking on curriculum-based assessments as compared to the student textbooks, reinforcing Shepard's (2000) observations about assessment methods often lag behind instructional methods. Also, not surprisingly, my analyses of students' written and verbal work on assessment tasks showed that students often verbally communicated more of their mathematical thinking than they included in their written responses. Even so, there were interesting differences between the types of evidence of thinking students tended to omit from their written work for SWAT and non-SWAT. Students also provided thoughtful perspectives on characteristics of SWAT as compared to non-SWAT.

My analyses of teacher interviews revealed that, while student work embedded in curriculum-based assessment tasks may provide students the potential to engage in SMP3b, more assessment task design work is needed to (1) determine the types and extent of thinking students should provide on written assessment tasks to demonstrate SMP3b and (2) decide how these requirements can be explicitly communicated to students on written assessment tasks in order for students to successfully demonstrate what they know. These results suggest a need for additional attention on the design of tasks for assessing mathematical practices and other habits of mind as well as critical thinking about whether or not written assessments are the most conducive assessment formats for assessing practices which are described as dynamic processes.

Copyright by
AMY ELIZABETH RAY
2018

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER 1: INTRODUCTION	1
Research Questions	6
CHAPTER 2: LITERATURE REVIEW	9
Overview	9
Curriculum-Based Assessments.....	10
Background	10
Assessment Reform Efforts	12
SMP3b: Critique the Reasoning of Others	14
Background	15
Assessing Mathematical Processes and Practices	20
Assessing NCTM Process Standards.....	20
CCSSM-Aligned Assessments.....	23
Student Work	26
Teachers’ Examination of Student Work	26
Students’ Examination of Student Work.....	28
A Study of Student Work Embedded in Student Textbook Tasks.....	32
Summary.....	34
CHAPTER 3: METHODOLOGY	36
Overview	36
Evidence-Centered Assessment Design (ECD).....	36
RQ #1: Exploring the Existence of Student Work in Curriculum-Based Assessments.....	39
Data Collection	40
Rationale for Curriculum Selection.....	40
Collecting Curriculum-Based Assessments.....	43
Refining Criteria for Student Work.....	44
Data Analysis	47
Inter-Rater Reliability (IRR).....	53
RQ #2: Exploring Students’ Experiences with and Perspectives on SWAT.....	58
Data Collection	59
Assessment task selection	60
Clinical interview component	62
Semi-structured interview component.....	63
Exclusion of the final two student interviews from the data analyses	64
Six student interviews.....	65
Data Analysis	68

RQ #3: Exploring Teachers’ Experiences with and Perspectives on SWAT and Students’ Work on SWAT	76
Data Collection	77
Data Analysis	82
Connecting Study Methods to the Stages of ECD	88
Domain Analysis – SMP3b	88
Domain Modeling –Unpacking and Making Sense of SMP3b.....	89
Conceptual Assessment Framework – Criteria for Student Work.....	91
Assessment Implementation – Selecting Assessment Tasks with and without Student Work	92
Assessment Delivery – Interviews with Students and Teachers.....	92
Summary.....	93
 CHAPTER 4: RQ #1: SWAT IN CURRICULUM-BASED ASSESSMENTS.....	95
Overview	95
RQ #1a: Prevalence of SWAT in Curriculum-based Assessments	95
Student Work Criteria Analyses	96
SWAT General Analyses.....	100
Assessment Types	100
CCSSM Content Strands	100
Evidence Types	101
Critique Types.....	101
RQ #1b: SWAT by Curriculum Series.....	102
Big Ideas	108
Assessment Types	108
CCSSM Content Strands	108
Evidence Types	109
Critique Types.....	109
CMP.....	109
Assessment Types	109
CCSSM Content Strands	109
Evidence Types	109
Critique Types.....	110
CPM.....	110
Assessment Types	110
CCSSM Content Strands	110
Evidence Types	110
Critique Types.....	110
Eureka.....	111
Assessment Types	111
CCSSM Content Strands	111
Evidence Types	111
Critique Types.....	111
Go Math.....	111
Assessment Types	111
CCSSM Content Strands	112

Evidence Types	112
Critique Types	112
Comparisons Across Curriculum Series SWAT	112
Assessment Types	112
CCSSM Content Strands	113
Evidence Types	113
Critique Types	113
RQ #1c: Comparison of SWAT and SWTT	114
SWTT General Analyses and Comparison to SWAT	120
CCSSM Content Strands	120
Evidence Types	121
Critique Types	121
Big Ideas	122
CCSSM Content Strands	122
Evidence Types	122
Critique Types	122
CMP	122
CCSSM Content Strands	122
Evidence Types	123
Critique Types	123
CPM	123
CCSSM Content Strands	123
Evidence Types	124
Critique Types	124
Eureka	124
CCSSM Content Strands	124
Evidence Types	125
Critique Types	125
Go Math	125
CCSSM Content Strands	125
Evidence Types	126
Critique Types	126
Comparisons Across Curriculum Series SWAT and SWTT	126
CCSSM Content Strands	126
Evidence Types	127
Critique Types	128
Summary of Chapter Findings	129

CHAPTER 5: RQ #2: STUDENTS' EXPERIENCES WITH AND PERSPECTIVES ON SWAT	133
Overview	133
RQ #2a: Nature of Students' Written and Verbal Work on SWAT and Non-SWAT	134
Student Thinking in Students' Written Work	135
Additional Student Thinking in Students' Verbal Responses	137
RQ #2b: Students' Descriptions of Experiences with SWAT	142
SWAT Required Different Thinking Processes and Final Products	142

Attributes of SWAT	144
Complexity.....	145
Connection to Classroom.....	146
Imagination	147
Comparison	148
Motivation.....	148
Summary of Chapter Findings.....	149
 CHAPTER 6: RQ #3: TEACHERS' EXPERIENCES WITH AND PERSPECTIVES ON SWAT AND STUDENTS' WORK ON SWAT	151
Overview	151
RQ #3a: What SWAT and Non-SWAT Assessed	152
RQ #3b: Teachers' Descriptions of SWAT.....	154
Advantages and Disadvantages of SWAT.....	154
Advantages of SWAT.....	155
Disadvantages of SWAT	157
Key Themes about SWAT and Using Student Work in Assessment Tasks.....	160
Interdisciplinary Practice	160
Connections to Testing	161
Connections to Curriculum/Classroom	164
CMP Curriculum Materials.....	164
Collaboration and the CMP Classroom	165
Emphasis on Explanation.....	168
Teachers as Assessment Writers	169
RQ #3c: Evidence of SMP3b in Teachers' Discussions of SWAT	175
and Students' Work on SWAT	175
Teachers' Initial Reactions to SWAT	177
Initial Thoughts	178
What SWAT Were Assessing.....	178
Revisiting Advantages of SWAT.....	179
Do SWAT Assess SMP3b?	179
Do Teachers See Evidence of SMP3b in Students' Work?	180
Summary of Chapter Findings.....	183
 CHAPTER 7: DISCUSSION	185
Overview	185
Findings Big Picture.....	185
RQ #1: SWAT in Curriculum-Based Assessments.....	186
RQ #2: Students' Experiences and Perspectives on SWAT.....	188
RQ #3: Teachers' Experiences and Perspectives on SWAT and Students' Work on SWAT	189
Potential of SWAT.....	191
Characterize SMP3b in the Context of Curriculum Tasks	192
Align Assessment and Instruction.....	196
Elicit Student Thinking.....	198
Motivate Students to Solve Tasks	199

Provide Benefits for Students	199
Limitations of SWAT.....	200
Expectations for Evidence	201
Evidence of Thinking Not Captured and Components of SMP3b Not Assessed on SWAT	207
Revisiting Validity	212
Summary of Implications, Next Steps, and Final Thoughts.....	214
APPENDICES	217
Appendix A: Criteria for Student Work Revision Iterations.....	217
Appendix B: Student Work in Curriculum Materials Analysis Codebook.....	218
Appendix C: Research Participant Information and Assent Form.....	223
Appendix D: Research Participant Information and Consent Form	225
Appendix E: Clinical and Semi-Structured Interview Questions for Student Interview	227
Appendix F: Clinical Interview Assessment Tasks	229
Appendix G: Research Participant Information and Consent Form	232
Appendix H: Semi-Structured Interview Questions for Teacher Interview.....	234
Appendix I: Students’ Written Work from Clinical Interview Assessment Tasks.....	236
Appendix J: SWAT Analyses by Assessment Type Additional Findings	240
REFERENCES	246

LIST OF TABLES

Table 3.1	Features of the Eight Assessment Tasks Used in Clinical Interviews with Students	61
Table 3.2	Background Information on Student Participants	66
Table 3.3	Evidence of Student Thinking in Students' Written Responses	70
Table 3.4	Evidence of Student Thinking in Students' Verbal Responses Not Found in Written Responses	73
Table 3.5	Emerging Themes from Students' Experiences with SWAT	75
Table 3.6	Background Information on Teacher Participants	78
Table 3.7	Categories of Teachers' Descriptions of What Assessment Tasks Assessed	84
Table 3.8	Emerging Themes from Teachers' Experiences with and Perspectives on SWAT	86
Table 4.1	Number of Assessment Tasks by Assessment Type and Curriculum Series	97
Table 4.2	Percentage of Assessment Tasks by Student Work Criteria and Curriculum Series	97
Table 4.3	Number of SWAT by Curriculum Series and Assessment Type	103
Table 4.4	Number of Swat by Curriculum Series and CCSSM Content Strand for Grade 7	104
Table 4.5	Number of Instances of Student Thinking in SWAT by Curriculum Series and Evidence Type	106
Table 4.6	Number of Instances of Critiques in SWAT by Curriculum Series and Critique Types	107
Table 4.7	Number of SWTT and SWAT by Curriculum Series	115
Table 4.8	Number of SWTT by Curriculum Series and CCSSM Content Strand for Grade 7	117
Table 4.9	Number of Instances of Student Thinking in SWTT by Curriculum Series and Evidence Type	118

Table 4.10	Number of Instances of Critiques in SWTT by Curriculum Series and Critique Type	119
Table 5.1	Instances of Student Thinking Codes Evident in Students Written Responses	135
Table 5.2	Instances of Student Thinking Codes Evident in Students' Verbal but not Written Responses	138
Table A.1	Number of SWAT by Assessment Type and CCSSM Content Strand for Grade 7	240
Table A.2	Number of Instances of Student Thinking in SWAT by Assessment Type and Evidence Type	241
Table A.3	Number of Instances of Critiques in SWAT by Assessment Type and Critique Type	242

LIST OF FIGURES

Figure 1.1	Changing Conceptions of Curriculum, Learning, and Assessment	3
Figure 2.1	New Theories of Curriculum, Learning, and Assessment	13
Figure 2.2	CCSSM Description of SMP3	16
Figure 2.3	SMP3b Proficiency Matrix	19
Figure 2.4	Mathematical Processes Assessment Coding Framework (MPAC)	22
Figure 2.5	ECD General Approach	24
Figure 2.6	The Reflective Cycle of Error Analysis	30
Figure 2.7	Criteria for Curriculum-Generated Student Work	33
Figure 3.1	ECD Layers	37
Figure 3.2	SMP3 Example Task from Big Ideas	41
Figure 3.3	SMP3 Example Task from Go Math	41
Figure 3.4	Connections Between Criteria for Student Work for SWAT and SMP3b	45
Figure 3.5	Assessment Types Descriptions	49
Figure 3.6	CCSSM Content Strand Descriptions	50
Figure 3.7	Evidence of Student Thinking Categories and Corresponding Codes	50
Figure 3.8	Critique Categories and Corresponding Codes	52
Figure 4.1	Assessment Types Across Curriculum Series	96
Figure 4.2	SWAT by Curriculum Series	100
Figure 4.3	SWAT by Assessment Type	10
Figure 4.4	SWAT by CCSSM Content Strand	101
Figure 4.5	SWAT by Evidence Type	101

Figure 4.6	SWAT by Critique Type	102
Figure 4.7	Comparison of SWAT by Curriculum Series and Assessment Types	103
Figure 4.8	Comparison of SWAT by Curriculum Series and CCSSM Content Strands	105
Figure 4.9	Comparison of SWAT by Curriculum Series and Evidence of Student Thinking Types	106
Figure 4.10	Comparison of SWAT by Curriculum Series and Critique Types	107
Figure 4.11	SWTT by Curriculum Series	115
Figure 4.12	Comparison of SWTT by Curriculum Series and CCSSM Content Strand	117
Figure 4.13	Comparison of SWTT by Curriculum Series and Evidence of Student Thinking Types	118
Figure 4.14	Comparison of SWTT by Curriculum Series and Critique Types	120
Figure 5.1	Student Thinking Codes Evident in Students' Written Work Sorted by Task and Student	136
Figure 5.2	Student Thinking Codes Evident in Students' Written Work Sorted by Task and by SWAT and Non-SWAT	137
Figure 5.3	Student Thinking Codes Evident in Students' Verbal but not Written Responses Sorted by Task and Student	139
Figure 5.4	Student Thinking Codes Evident in Students' Verbal but not Written Responses Sorted by Task and by SWAT and Non-SWAT	140
Figure 5.5	Students' Designated Easiest and Hardest Tasks	145
Figure 6.1	Summary of Teachers' Descriptions of What Interview Assessment Tasks Assessed	153
Figure 6.2	Advantages and Disadvantages of SWAT	155
Figure 6.3	Snapshots of Evidence of SMP3b in Teachers' Thoughts about SWAT and Students' Work Across Interview Segments	176
Figure 7.1	Student Thinking Codes for Written Work and Transcripts Pairings	189

Figure 7.2	Assessment Task That Fails Criterion #3	193
Figure 7.3	Preference Critique Type Example	197
Figure 7.4	Student Response Expectations Provided in Assessment Tasks	202
Figure 7.5	Task B and Anna’s Written Work on Task B	203
Figure 7.6	Task H and Jane’s Written Work on Task H	204
Figure 7.7	Ms. Shirley’s Revised Expectations for Task F	206
Figure A.1	Criteria for Student Work Iterations	217
Figure A.2	Criteria for Student Work	219
Figure A.3	Assessment Type Descriptions	220
Figure A.4	CCSSM Content Strand Descriptions	220
Figure A.5	Evidence of Student Thinking Categories and Corresponding Codes	221
Figure A.6	Critique Categories and Corresponding Codes	222
Figure A.7	Task A	229
Figure A.8	Task B	229
Figure A.9	Task C	229
Figure A.10	Task D	230
Figure A.11	Task E	230
Figure A.12	Task F	230
Figure A.13	Task G	230
Figure A.14	Task H	231
Figure A.15	Student Work on Task A	236
Figure A.16	Student Work on Task B	236
Figure A.17	Student Work on Task C	237

Figure A.18	Student Work on Task D	237
Figure A.19	Student Work on Task E	238
Figure A.20	Student Work on Task F	238
Figure A.21	Student Work on Task G	239
Figure A.22	Student Work on Task H	239
Figure A.23	Comparison of SWAT Assessment Types by CCSSM Content Strands for Grade 7	240
Figure A.24	Comparison of SWAT Assessment Types by Evidence of Student Thinking Types	241
Figure A.25	Comparison of Assessment Types in SWAT by Critique Types	242

CHAPTER 1: INTRODUCTION

As a graduate student researcher on the Connected Mathematics Project (CMP), I have had the opportunity to learn about, study with colleagues, and work with teachers around curricular components that are integral to the CMP curriculum. One of these curricular components is student work embedded in student textbooks in both classwork and homework tasks (CMP, 2018b; Gilbertson et al., 2016; Going, Ray, & Edson, in preparation). My (and my colleagues') investigations of student work in the research literature have shown that student work can be used as a learning tool for teachers (e.g. Silver & Suh, 2014) and for students (e.g. Rittle-Johnson, Star, Durkin, 2009).

From numerous conversations with teachers at CMP professional development sessions, I learned that CMP teachers see many benefits of student work embedded in the student textbook. Teachers have shared with me that these types of mathematical tasks promote students owning mathematical ideas, valuing mistakes, taking mathematical risks, engaging in discourse, self-correcting errors or misconceptions, and providing peer-to-peer feedback. Teachers also have shared that student work embedded in student textbooks provide teachers insight into the types of strategies their own students might use to solve mathematical problems. The research literature shows, and CMP teachers have expressed, that student work can be valuable for student and teacher learning of mathematics. However, because assessments often heavily influence instruction (Swan & Burkhardt, 2012), I was interested in studying the role of student work in assessment materials. In this dissertation, I explored student work embedded in curriculum-based assessment materials and investigated whether or not student work assessment tasks (SWAT) could be used to assess students' abilities to analyze someone else's mathematical thinking (SMP3b, CCSSI, 2010).

As evidenced by recent state and national education policies, increased accountability measures related to standardized testing show “measurement of outcomes using achievement tests drives education policy and practice today” (Hamilton, Stecher, & Yuan, 2012, p. 157). Swan and Burkhardt (2012) stated, “In a target driven system where examination results have serious consequences: What You Test Is What You Get (WYTIWYG)” (p. 4). Thus, mathematics assessment can be viewed as a major influence on the mathematics that is taught in classrooms.

Even while mathematics assessments have become highly consequential for students and teachers and influential on mathematics teaching, Shepard’s (2000) research in mathematics education assessment has challenged dominant views of testing and assessment in mathematics and championed for a disruption of traditional assessment that is informed by reformed curriculum work and contemporary cognitive and constructivist learning theories. She introduced an emerging paradigm of curriculum theories, learning theories, and classroom assessment that confronts the historical traditions of mathematics assessment (see Figure 1.1). In her view, “the present dissonance between instruction and assessment arises because of the misfit between old views of testing and a transformed vision of teaching” (Shepard, 2000, p. 15). In other words, mathematics assessment is currently lagging behind advancements in other mathematics education arenas of curriculum and instruction.

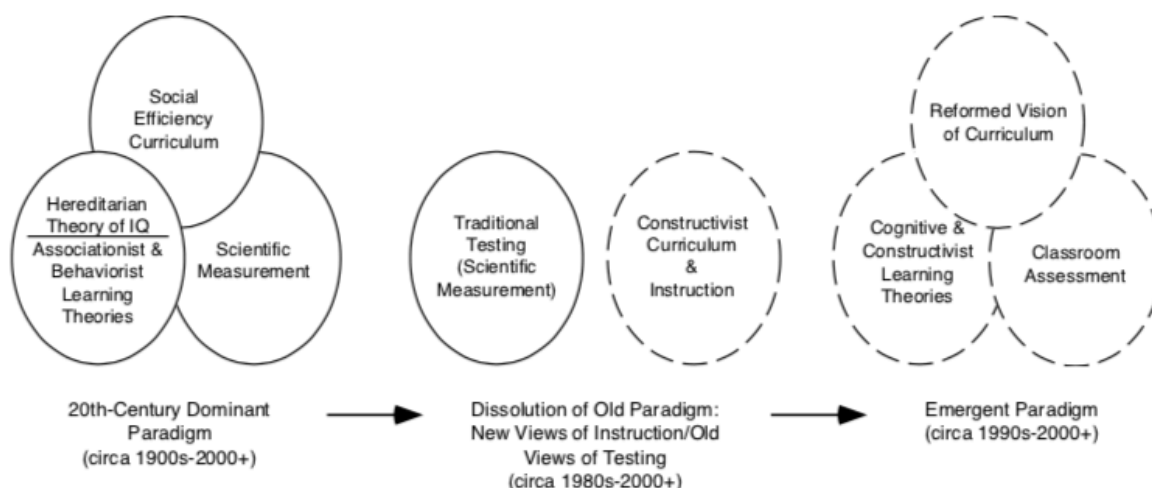


Figure 1.1 Changing Conceptions of Curriculum, Learning, and Assessment (Shepard, 2000, p. 5)

One of the recent advancements in mathematics teaching is a focus on broadening the types of mathematical proficiencies or habits of mind that educators should seek to promote in classrooms and use as indicators of a student’s ability to do and understand mathematics (National Research Council, 2001; Seeley, 2014). In the policy arena, the National Council of Teachers of Mathematics (NCTM), the National Governors Association Center for Best Practices, and the Council of Chief State School Officers have all drawn attention to expanding mathematical proficiency beyond the ability to perform mathematical procedures quickly and efficiently (CCSSI, 2010; NCTM, 2010; 2014).

More specifically, the Common Core State Standards for Mathematics (CCSSM), a standards document widely adopted across the United States, contains both content and practice standards, supporting the argument that how students do mathematics is equally important to what mathematics students learn. Content standards in the CCSSM identify key mathematics content students should learn at each grade or subject level. The practice standards describe how students can and should learn mathematics in meaningful ways and demonstrate their mathematical understandings (CCSSI, 2010). The CCSSM practice standard of particular interest

in this study is Standard for Mathematical Practice (SMP) #3: “Construct viable arguments and critique the reasoning of others” (CCSSO, 2010, p. 6). The second half of this practice, “critique the reasoning of others” (SMP3b), suggests that students should be encouraged to analyze and evaluate someone else’s mathematical thinking as a part of learning mathematics and demonstrating their mathematical knowledge (CCSSO, 2010, p. 6).

Since the adoption of the CCSSM by most states, authors of mathematics textbooks have attempted to align their textbooks with the CCSSM. Beyond attending to new content standards in the CCSSM by adding to, omitting from, or adapting previous editions, curriculum writers have also attempted to address (in varying degrees) the practice standards in new textbook editions by listing practice standards addressed in book chapters or by integrating practice standards language into tasks for students. One way curriculum writers have created opportunities for students to engage in the practice of making sense of another student’s reasoning, or SMP3b, is by embedding student work into both classroom and homework tasks in student textbooks. These types of student work tasks often involve error analysis, comparing students’ strategies, or determining the validity of fictional students’ conjectures or strategies.

Gilbertson and colleagues (2016) investigated these types of problems in their development of a framework for curriculum analysis which they called *student work tasks*. The researchers studied the existence and character of student work tasks in three different seventh-grade curriculum series intended to be aligned with the CCSSM. The authors positioned student work as one avenue for promoting students’ ownership of engaging in mathematical thinking and generating mathematical ideas (in contrast to the textbook or the teacher as the mathematical authority in a classroom). In looking for instances of student work within the student textbooks,

the researchers looked for tasks where “the author of the work [in the task] is uniquely positioned as external to the classroom interaction” (Gilbertson et al., 2016, p. 4).

In this study, I extended and broadened Gilbertson and colleagues’ curriculum analysis work by using their criteria for student work to determine what opportunities students have to engage with student work in curriculum-based assessments, the assessments designated by each curriculum series, from middle school mathematics curricula intended to be aligned with CCSSM. My work acknowledged prior research in mathematics education that has shown assessment as extremely influential on mathematics instruction while also examining whether or not Shepard’s (2000) described dissonance between assessment and instruction was evident when comparing CCSSM-aligned curriculum-based assessments and student textbook curriculum materials.

The purpose of this study was to explore whether or not student work embedded in assessment tasks can be used as a mechanism for assessing SMP3b. I investigated student work in curriculum-based assessments using three different approaches. First, I explored the existence and nature of SWAT found in curriculum-based assessments. I also compared SWAT to the student work textbook tasks (SWTT) based on the criteria from Gilbertson and colleagues (2016) in order to understand whether or not analyzing and critiquing the reasoning of others was promoted for students differently across the student textbooks and the curriculum-based assessments.

Second, extending beyond text analyses of curriculum-based assessments and student textbooks, I also explored students’ experiences solving assessment tasks, both SWAT and non-SWAT, and their perspectives on SWAT. Third, I investigated teachers’ experiences with and perspectives on SWAT and students’ work on SWAT. Positioning teachers as professionals

within their field, I examined whether or not they viewed student work embedded in assessment tasks as a mechanism for assessing students' abilities to analyze and critique the mathematical reasoning of others, or SMP3b. Broadly, this exploratory validity study was meant to build on existing work in both curriculum design and assessment research by utilizing an advancement in curriculum design, student work embedded in textbooks, to inform assessment design and explore whether or not student work, a component of curriculum materials that provides insight into someone's thinking and requires the reader to make sense of it in some way, can be used as a mechanism for assessing SMP3b.

Investigating student work embedded in curriculum-based assessments through exploring curriculum materials, students' perspectives on SWAT, and teachers' perspectives on SWAT has the potential to challenge traditional forms of assessment that focus predominantly on measuring students' mastery of mathematical content as opposed to assessing students' proficiencies in analyzing and critiquing the reasoning of others. Thus, in this study, I acknowledged the powerful influence assessments have on teaching and learning and conjectured that utilizing student work in curriculum-based assessments will further promote the practices of analyzing and critiquing the reasoning of others described in mathematics standards and principles documents and allow teachers to assess students' abilities to engage in these practices.

Research Questions

This study was guided by three research questions:

1. What is the frequency and nature of SWAT in curriculum-based assessments?
 - a. How prevalent are SWAT in curriculum-based assessments?
 - b. How do SWAT vary across curriculum series?
 - c. How do SWAT compare to SWTT in corresponding student textbooks?

2. How do students talk about and make sense of SWAT?
 - a. What is the nature of students' written work and verbal responses on SWAT and non-SWAT?
 - b. How do students describe their experiences with SWAT?
3. How do teachers talk about and understand SWAT and non-SWAT based on the tasks and students' written work on the tasks?
 - a. What do teachers think that SWAT and non-SWAT assess?
 - b. How do teachers describe SWAT?
 - c. What evidence of SMP3b do teachers notice in SWAT based on both the written assessment tasks as well as students' written work on the tasks?

The first research question focused on the existence of student work in assessment materials based on the criteria for student work introduced by Gilbertson and colleagues (2016) as well as comparisons between SWAT and SWTT. The second research question was motivated by my desire to explore and understand students' experiences working on SWAT and non-SWAT as well as their perspectives on assessment tasks that provide information about a person's thinking and ask them to make sense of it. The third research question focused on teachers' experiences with SWAT and non-SWAT and their perspectives on SWAT and students' work on SWAT. In answering the third research question, I explicitly explored whether or not student work embedded in assessment tasks can serve as a mechanism for assessing SMP3b from the viewpoint of teachers.

In Chapter 2, I provide a review of literature detailing (1) research conducted on curriculum-based assessments and why investigating these assessments is important, (2) research related to the Standards for Mathematical Practice, specifically SMP3b, and (3) research on the

use of student work for both student and teacher learning. In Chapter 3, I describe my guiding perspective of assessment as evidentiary argumentation and I detail my methods for answering the research questions focused on a text analysis, clinical interviews with students, and semi-structured interviews with students and teachers. I connect my methodologies to evidence-centered assessment design (ECD) used more formally in educational assessment design research (Gotwals & Songer, 2013; Mislevy & Haertel, 2006; Mislevy, Steinberg, & Almond, 2003). In Chapter 4, I detail findings from text analyses and an exploration of student work in curriculum-based assessment tasks. In Chapter 5, I present findings from clinical and semi-structured interviews with students and my exploration of students' experiences with and perspectives on SWAT. In Chapter 6, I present findings from semi-structured interviews with teachers and my exploration of teachers' experience with and perspectives on SWAT and students' work on SWAT. In Chapter 7, I discuss all of the findings within and across the research questions as well as implications for these results.

CHAPTER 2: LITERATURE REVIEW

Overview

In this chapter, I present a review of research literature in mathematics education on curriculum-based assessment, SMP3b, and student work. The research literature on assessment in mathematics education is generally quite vast and highly varied. For the purposes of this study, I focus on assessment literature related to curriculum-based assessments because these assessments often dictate what mathematical content and practices teachers promote in classrooms. I also highlight current reform efforts aimed at framing assessment, broadly, as integral to the teaching and learning process, not just as an evaluative tool for use after teaching and learning.

Then, I focus on standards writers' characterization of SMP3b, "critique the reasoning of others" (CCSSI, 2010, p.6), as well as other standards writers' and researchers' characterizations of this practice. I also detail research related to assessing the SMPs and other mathematical habits of mind. Lastly, I present a summary of the research on the use of student work in mathematics education. This research has addressed how student work has been used to support professional learning for teachers either at the individual level or in professional learning communities. This research has also focused on the benefits for students of learning from worked examples. Research on the benefits for students of learning from worked examples provided a first step in understanding how the use of student work can be beneficial for student learning. I conclude my overview of research on student work by highlighting the work of one research group focused on student work embedded in curriculum materials. For these researchers, student work as a context for student learning extends beyond a student viewing worked examples by

attributing the work to a person, providing insight into a person's thinking, and requiring the reader to make sense of the person's thinking.

My study lies at the intersection of the three topics of curriculum-based assessment, SMP3b, and student work with a specific focus on exploring whether or not student work embedded in curriculum-based assessments has the potential to serve as a mechanism for assessing students' abilities to critique the reasoning of others.

Curriculum-Based Assessments

In the following paragraphs, I present research literature on curriculum-based assessments as well as reform efforts addressing the assessment lag discussed by Shepard (2000). I argue it is important to acknowledge how curriculum-based assessments can be extremely influential on the types of mathematical content and practices promoted in classrooms.

Background

Increased emphasis on the importance of accountability through assessment has positioned assessments as powerful indicators of the mathematics that should be taught and learned in classrooms (Hamilton, Stecher, & Yuan, 2012; Swan & Burkhardt, 2012). Swan and Burkhardt (2012) stated, "In a target driven system where examination results have serious consequences: What You Test Is What You Get (WYTIWYG)" (p. 4). Within the classroom, teachers do not always have access to the high-stakes assessments that are required for students due to concerns of testing security. Teachers often are only provided sample problems or information about the mathematical content that will be covered on these assessments. In contrast, teachers have direct access to assessments provided in curriculum materials. Because the assessments provided in curriculum materials are the assessments teachers have access to,

curriculum-based assessments serve as an influence on the mathematics that is taught in classrooms.

In their study of assessments accompanying elementary, middle, and high school curriculum materials, Hunsader and colleagues (2013; 2014) created a framework for analyzing curriculum materials for opportunities students have to engage with mathematical processes (NCTM, 2000). In describing why it is important to consider the assessments provided with written curriculum materials as influential, the authors stated:

We believe that the assessments accompanying the textbook also exert a powerful influence on student learning. Just as the textbook helps translate the intended curriculum of standards into the written curriculum that guides instruction, the chapter tests included in teacher resources accompanying a textbook help translate for the teacher what the textbook authors or curriculum developers believe is important for students to master, both in terms of *what* and *how*. (Hunsader, Thompson, & Zorin, 2013, p. 2).

The authors acknowledged that the assessment materials provided with the curriculum series do not account for all the ways in which teachers assess their students. Even so, the researchers detailed a number of studies highlighting that “many teachers rely on the assessments that accompany their published textbooks as a primary means of assessment” (Hunsader et al., 2014, p. 798). Their review of the research literature on assessments accompanying curriculum materials led Hunsader and colleagues to conclude that curriculum-based assessments were worthy of attention to determine “whether they emphasize the goals desired for mathematics teaching and learning, both in terms of content and processes” (p. 798). The authors also described how most research previously conducted on assessments has focused on alignment of

content standards and assessments (Hunsader et al., 2013; 2014) and does not address the NCTM mathematical processes the authors wanted to study in the assessment materials.

The work of Hunsader and colleagues (2013; 2014) and their focus on curriculum-based assessments and NCTM process standards relates closely to my study of student work as a possible mechanism for assessing SMP3b in curriculum-based assessment tasks as SMP3b is a component of the CCSSM Standards for Mathematical Practice (CCSSI, 2010). Their study emphasized the influential nature of curriculum-based assessments on the mathematics that is promoted in classrooms and the importance of analyzing these assessments for the opportunities students have to engage in mathematical habits of mind, processes, and practices (CCSSI, 2010; NCTM, 2000; Seeley, 2014).

Assessment Reform Efforts

Even as an emphasis on accountability as tied to assessment has grown, other work in mathematics education has focused on closing the gap between teaching and assessment as opposed to widening it. In their seminal book *Understanding by Design*, Wiggins and McTighe (2005) detailed the process of backwards design for teaching and learning. The authors described a planning and teaching process that focuses on learning outcomes for students as key indicators of what should be taught in the classroom. In other words, the authors started from the content knowledge students would be assessed on to design teaching and learning experiences. Their work focused on aligning assessment and teaching practices so that student outcomes are clear and so that learning experiences provide opportunities for students to learn the content knowledge that will be assessed.

Shepard (2000) also focused on alignment between assessment methods and teaching and learning experiences, describing the incompatibility of the social-constructivist conceptualization

of instruction and traditional forms of assessment. She posited, “If instructional goals include developing students’ metacognitive abilities, fostering important dispositions, and socializing students into the discourse and practices of academic disciplines, then it is essential that classroom routines and corresponding assessments reflect these goals” (Shepard, 2000, p. 8). Shepard characterized an emergent, constructivist paradigm that highlighted the intersection of curriculum theories, learning theories, and classroom assessment as a conceptual framework for challenging traditional modes of assessment (see Figure 2.1).

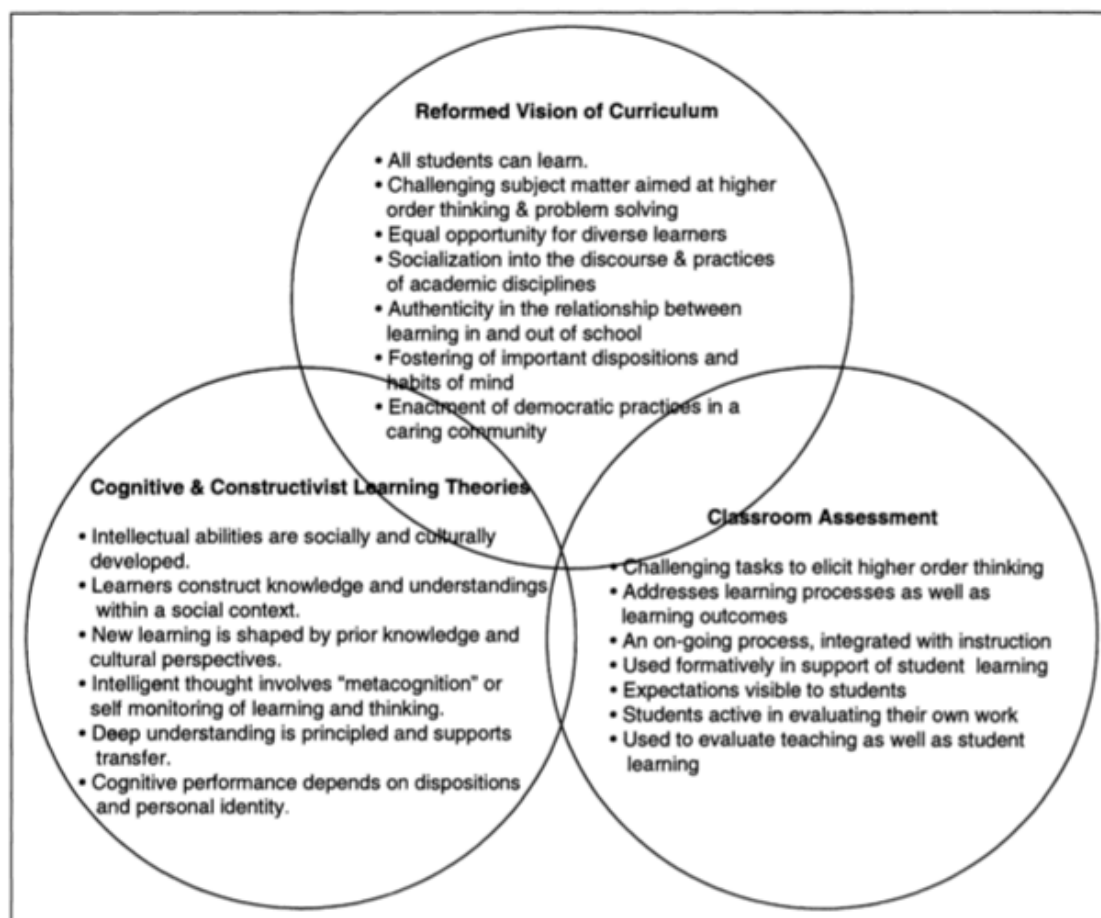


Figure 2.1 New Theories of Curriculum, Learning, and Assessment (Shepard, 2000, p. 6)

She argued for both improving the content of existing assessments as well as utilizing assessment as part of the learning process. Shepard (2000) stated, “Our aim should be to change our cultural practices so that students and teachers look to assessment as a source of insight and

help instead of an occasion for meting out rewards and punishments” (p. 10). While Wiggins and McTighe (2005) and Shepard (2000) called for change in the ways teachers plan for, create, and utilize assessments, many of their ideas have implications for curriculum writers that design curriculum materials. More specifically, the call to connect teaching, learning, and assessment as opposed to viewing assessment as a separate component has implications for the design of curriculum materials including both the student textbooks and the accompanying curriculum-based assessment materials.

Connecting the research literature on curriculum-based assessments and reform efforts in mathematics assessment to my study, the literature reinforces the importance of considering assessments, generally, and curriculum-based assessments, specifically, as highly influential on the mathematics that students learn. Curriculum-based assessments communicate to teachers what curriculum writers view as important mathematics for students to know and do. My goal of comparing the opportunities students have to engage with student work in curriculum-based assessments to the opportunities found in the student textbook aligns with reform efforts in mathematics focused on promoting similar practices in class and on assessments, thereby, in a small way, confronting the gap between a reformed vision of curriculum and traditional assessment methods (Shepard, 2000).

SMP3b: Critique the Reasoning of Others

In the following paragraphs, I present an overview of reform efforts focused on promoting mathematical habits of mind, proficiencies, and processes (NCTM, 2000; 2014; NRC, 2001; Seeley, 2014) that have led to the CCSSM SMPs (CCSSI, 2010). I also revisit Hunsader and colleagues’ (2013; 2014) studies that explored the opportunities students had to engage with key mathematical processes on curriculum-based assessments as well as literature on CCSSM-

aligned assessments to highlight efforts in designing tasks that assess mathematical processes and practices.

Background

In her article, *Developing Mathematical Habits of Mind*, Cathy Seeley (2014) detailed the concept of mathematical habits of mind and summarized researchers' and standards writers' attempts to capture and describe these habits in the National Research Council's *Adding It Up* (2001), NCTM's process standards (2000), and CCCSM SMPs (2010). She stated, "For years, mathematicians, educators, and other experts have tried to describe the heart of what it means to do mathematics and think mathematically" (Seeley, 2014, p. 247). Mathematical habits of mind often involve "a person's ability to solve mathematical problems, especially those that go beyond simple word problems related to a recently learned procedure" (Seeley, 2014, p. 248). She also described how problem solving is closely connected to explanation and communication of mathematical thinking. However, the bounds of what counts as a mathematical habit of mind are unclear. Seeley stated, "There is no one correct or complete list of mathematical habits of mind. Many descriptions overlap or address similar aspects of the nature of mathematics" (Seeley, 2014, p. 248). For the purposes of this study, I focused on one specific habit of mind detailed in the CCSSM SMP, SMP3b, which states that students should be able to "critique the reasoning of others" (CCSSI, 2010, p. 6).

The writers of the CCSSM provided a paragraph description of each SMP to show what the practices would look like for students and to "describe ways in which developing student practitioners of the discipline of mathematics increasingly ought to engage with the subject matter as they grow in mathematical maturity and expertise" (CCSSI, 2010, p. 8). The full SMP3 is: "Construct viable arguments and critique the reasoning of others" (CCSSI, 2010, p. 6).

Below, I have provided the description of SMP3 written by the CCSSM authors and highlighted the portions that explicitly relate to the components of SMP3b: “others” and the practice “critique the reasoning of others” (see Figure 2.2).

Mathematically proficient students understand and use stated assumptions, definitions, and previously established results in constructing arguments. They make conjectures and build a logical progression of statements to explore the truth of their conjectures. They are able to analyze situations by breaking them into cases, and can recognize and use counterexamples. They justify their conclusions, **communicate them to others**, and **respond to the arguments of others**. They reason inductively about data, making plausible arguments that take into account the context from which the data arose. Mathematically proficient students are also able to compare the effectiveness of two plausible arguments, distinguish correct logic or reasoning from that which is flawed, and—if there is a flaw in an argument—explain what it is. Elementary students can construct arguments using concrete referents such as objects, drawings, diagrams, and actions. Such arguments can make sense and be correct, even though they are not generalized or made formal until later grades. Later, students learn to determine domains to which an argument applies. **Students at all grades can listen or read the arguments of others, decide whether they make sense, and ask useful questions to clarify or improve the arguments.**

Figure 2.2 CCSSM Description of SMP3 (CCSSI, 2010, p. 6)

These few highlighted phrases do not provide a very robust description of the practice. From this description, critiquing seems to involve responding to others, asking questions, and/or making sense of an argument. In order to gain a more robust description of SMP3b, I turned to

other habits of mind resources described by Seeley (2014) and additional resources related to mathematical processes and practices.

In her discussion of habits of mind, Seeley (2014) connected SMP3 to general mathematical practices of reasoning and explanation. She detailed how these practices relate to habits of mind discussed in previous standards documents. Specifically, Seeley related SMP3 to the Reasoning and Proof NCTM Process Standard drawing on the work of Koestler, Felton, Bieda, and Otten (2013). In their book, *Connecting the NCTM Process Standards and the CCSSM Practices*, Koestler and colleagues (2013) connected SMP3 most closely to the Reasoning and Proof NCTM Process Standard which focuses on developing, investigating, and vetting conjectures, arguments, and proof (NCTM, 2000). In my review of the NCTM *Principles and Standards for School Mathematics* (2000), SMP3b, the second half of SMP3, most closely connects with the Communication Standard. The description of this process standard for Grades 6 – 8 states that students should be able “analyze and evaluate the mathematical thinking and strategies of others” (p. 268). So, a *critique* involves analysis and evaluation and *reasoning of others* can include mathematical thinking and strategies of others. However, I agree that elements of reasoning and proof, as described in the Reasoning and Proof NCTM Process Standard (2000), are essential for students to be able to make sense of someone else’s mathematical thinking and describe their own thinking about someone else’s thinking.

Koestler, Felton, Bieda, and Otten (2013) also made explicit connections between SMP3 and the adaptive reasoning proficiency strand detailed in *Adding it Up* (NRC, 2001). The National Research Council (2001) defined the adaptive reasoning proficiency strand as the “capacity for logical thought, reflection, explanation, and justification” (p. 5). A key component of adaptive reasoning is “the ability to justify one’s work” which can involve formal proofs or

informal explanations. Either way, “students need to be able to justify and explain ideas in order to make their reasoning clear, hone their reasoning skills, and improve their conceptual understanding” (NRC, 2001, p. 130). Thus, the practice of critiquing for students must extend beyond viewing someone else’s mathematical thinking, to reflecting on, explaining, or justifying their critique of the other’s thinking.

In order to build a more robust understanding of SMP3b as described in the research literature, I ventured beyond the resources described by Seeley (2014). In their resource for teacher leaders, *The Common Core Mathematics Standards: Transforming Practices Through Team Leadership*, Hull, Miles, and Balka (2012) described student proficiencies for each component of the SMPs at the levels of initial, intermediate, and advanced. The authors stated:

The mathematical practices are not skill-based content that students can learn only through direct teaching methods but rather ones that emerge over time from opportunities and experiences provided in mathematics classrooms. These opportunities and experiences must include challenging problems, student collaborative groups, interactive discourse, and adequate time – clearly not an easy task (Hull, Miles, & Balka, 2012, p. 51).

Because the authors viewed the mathematical practices as emerging in students’ mathematical development, they created a proficiency matrix that highlighted proficiency scales for each component of the eight SMPs. The purpose of this proficiency matrix is for teachers and leaders “to consider and gauge students’ progress for each of the practices as the students demonstrate proficiency for each indicator” (Hull, Miles, & Balka, 2012, p. 59).

Figure 2.3 below details the how the authors made sense of different degrees of proficiencies for SMP3b from initial to advanced. An initial proficiency level requires the

student to be able to talk about and understand mathematical ideas that may be different from their own. Extending beyond understanding, the intermediate level of proficiency requires the student to be able to explain someone else’s mathematical work as well as be able to evaluate the viability of someone’s solution. Lastly, the advanced degree of proficiency for SMP3b requires the student to be able to analyze, evaluate, and compare multiple solution strategies as well as explain the reasoning behind the solutions.

Initial	Intermediate	Advanced
Understand and discuss other ideas and approaches.	Explain other students’ solutions and identify strengths and weaknesses of the solutions.	Compare and contrast various solution strategies, and explain the reasoning of others.

Figure 2.3 SMP3b Proficiency Matrix (Hull, Miles, & Balka, 2012, p. 52)

In addition to different degrees of proficiency, the authors provided instructional strategies teachers could use to engage students in various degrees of the SMPs and promote each SMP in their classrooms. For example, Pair-Share was described as a strategy that could engage students in SMP3b at the initial level. Pair-Share requires “teachers merely ask a question or assign a problem and allow students to think and work with a partner for one to three minutes before requesting an answer to the question or problem” (Hull, Miles, & Balka, 2012, p. 54). This strategy would require students to be able to discuss approaches to questions or problems with other students as described in the proficiency matrix (see Figure 2.3).

In connection to my dissertation study, Hull, Miles, and Balka’s (2012) proficiency matrix offers an elaborated view of SMP3b involving degrees of proficiency not explicitly described in the CCSSM standards document. While the authors focused on the students’ proficiency levels for the SMPs as evidenced by their experiences and demonstrated abilities during in-class activities, teachers could also make use of the matrix for assessing students’ proficiency levels on written assessments. This would require that teachers determine the written

evidence that students would need to be provide in their solutions at the various proficiency levels. Thus, in determining whether or not student work can serve as a mechanism for assessing SMP3b, it is important to consider that assessment of practices could also account for degrees of proficiencies as opposed to the extremes of either fully proficient or not proficient for SMPs.

Assessing Mathematical Processes and Practices

While the work of Hull, Miles, and Balka (2012) presented possibilities for assessing students' proficiency levels for SMPs, other researchers and testing consortia have explored how to assess mathematical processes and practices on written assessments. In the following paragraphs, I return to the work of Hunsader and colleagues (2013; 2014) and their creation of a framework to analyze students' opportunities to engage in NCTM process standards on curriculum-based assessments. I also discuss the advent of two testing consortia that are developing CCSSM-aligned assessments and have explicitly addressed assessing SMPs in their assessment design.

Assessing NCTM Process Standards. Seeley (2014) noted that “in contrast to widespread lack of attention on assessments in the past to mathematical habits of mind and mathematical processes,” curriculum and assessment developers have begun to attend to standards and considered implications for curriculum and assessment design in order “to support the mathematical thinking and habits of mind we value” (p. 253). Hunsader, Thompson, and Zorin (2013) created a framework to analyze assessments at the level of test items for whether or not they provided opportunities for students to engage in the NCTM process standards, except for problem solving (NCTM, 2000). The authors felt that problem solving was not feasible to include in their framework because problem solving “depends largely on their prior experiences with similar items, which would imply relating the item to classroom instruction (the

implemented curriculum) or to what is present in the textbook (the written curriculum)” (Hunsader et al., 2013, p. 5). The authors developed the Mathematical Processes Assessment Coding Framework (MPAC) (see Figure 2.4) so that the criteria for meeting process standards could be determined for each item in a test “by a simple reading of the item, without knowledge of the elementary curriculum, instruction, or students’ prerequisite knowledge” (p. 6).

The authors analyzed three elementary grades curriculum materials and found that “Mathematical processes other than connections [were] not heavily emphasized in the tests accompanying the published curricula” (p. 20). Additionally, when the tests did provide opportunities for students to engage in processes, there was “variability across tests, grades, publishers, and content domains” because for some tests, “students had many opportunities to engage with the processes; on others, they had few opportunities” (p. 20). A larger subsequent study of elementary, middle, and high school curriculum materials using the developed framework reinforced these preliminary findings as the authors concluded, “inconsistent emphasis is placed on the mathematical processes within assessments accompanying commercial textbooks in the USA” (Hunsader et al., 2014, p. 797). The authors felt that the results of their study would be informative for teachers as they promote and assess the NCTM process standards because “if teachers believe that processes are important, and analysis with the framework indicates that they are not present in the assessed curricula of classroom tests, then teachers can use that information to find other avenues for assessing students’ proficiency with these processes” (p. 20).

<i>Reasoning and Proof</i>	
<i>N</i>	The item does not direct students to provide or show a justification or argument for <u>why they gave that response</u> .
<i>Y</i>	The item directs students to provide or show a justification or argument for <u>why they gave that response</u> . ('Check your work' by itself is not justification.)
<i>Opportunity for Mathematical Communication</i>	
<i>N</i>	The item does not direct students to communicate <u>what they are thinking</u> through symbols (beyond a numeral answer), graphics/pictures, or words.
<i>Y</i>	The item directs students to communicate <u>what they are thinking</u> through symbols, graphics/pictures, or words.
<i>V</i>	The item <u>only</u> directs students to record a vocabulary term or interpret/create a representation of vocabulary.
<i>Connections</i>	
<i>N</i>	The item is not set in a real-world context and does not explicitly interconnect two or more mathematical concepts (e.g., multiplication and repeated addition, perimeter and area).
<i>R</i>	The item is set in a real-world context outside of mathematics.
<i>I</i>	The item is <u>not</u> set in a real-world context, but explicitly interconnects two or more mathematical concepts (e.g., multiplication and repeated addition, perimeter and area).
<i>Representation: Role of Graphics</i>	
<i>N</i>	No graphic (graph, picture, or table) is given or needed.
<i>S</i>	A graphic is given but no interpretation is needed for the response, and the graphic does not explicitly illustrate the mathematics inherent in the problem. (superfluous)
<i>R</i>	A graphic is given and no interpretation is needed for the response, but the graphic explicitly illustrates the mathematics inherent in the problem.
<i>I</i>	A graphic is given and must be interpreted to answer the question.
<i>M</i>	The item directs students to make a graphic or add to an existing graphic.
<i>Representation: Translation of Representational Forms (codes are bi-directional as appropriate)</i>	
<i>N</i>	Students are not expected to record a translation between different representational forms of the problem.
<i>SW</i>	Students are expected to record a translation from a verbal representation to a symbolic representation of the problem or vice versa.
<i>GS</i>	Students are expected to record a translation from a symbolic representation to a graphical (graphs, tables, or pictures) representation of the problem or vice versa.
<i>WG</i>	Students are expected to record a translation from a verbal representation to a graphical representation of the problem or vice versa.
<i>TG</i>	Students are expected to record a translation from one graphical representation of the problem to another graphical representation.
<i>A</i>	Students are expected to record two or more translations among symbolic, verbal, and graphical representations of the problem.

Figure 2.4 MPAC Framework (Hunsader et al., 2013; 2014)

Also, because assessments are highly consequential and are often tied to grades, students may not value mathematical processes promoted in other parts of the curriculum materials and “develop a narrow view of what it means to be mathematically competent” (p. 21). The authors acknowledged that perhaps the written assessments accompanying the curriculum materials may not always be the best format for assessing mathematical processes. Even so, their analyses of these written assessments for students’ opportunities to engage with NCTM processes could provide “educators a means to consider where assessment changes might be needed” (p. 21).

My study of student work embedded in curriculum-based assessment tasks extends the work of Hunsader and colleagues (2013; 2014) as I not only identify the potential for students to engage with SMP3b in curriculum-based assessment tasks by reviewing curriculum materials using specific criteria, but I also explore students’ and teachers’ experiences with assessment tasks with embedded student work. Thus, while Hunsader and colleagues focused on the potential for engagement with mathematical processes as determined by features of the curriculum-based assessments, I explore both the potential of student work as a mechanism for assessing SMP3b as well as the implementation based on students’ and teachers’ actual experiences with tasks.

CCSSM-Aligned Assessments. The writers of the CCSSM stated, “Mathematical understanding and procedural skill are equally important, and both are assessable using mathematical tasks of sufficient richness” (CCSSI, 2010, p. 4). Therefore, the authors expressed that both CCSSM content and practice standards are assessable and should be assessed. Due to the widespread adoption of CCSSM, two testing consortia have been tasked with producing assessments aligned with CCSSM – the Partnership for Assessment of Readiness for College Careers (PARCC), and Smarter Balanced Assessment Consortium (SBAC) (Chandler, Fortune,

Lovett, & Scherrer, 2016; Herman & Linn, 2013; Hull, Balka, Miles, 2013; Schoenfeld, 2013). According to Herman and Linn (2013), “both consortia have adopted Evidence-Centered Design (ECD) as their approach to summative assessment development and design” (p.6). This approach (see Figure 2.5) involves creating claims and assessment targets from the CCSSM content and practice standards which are then used to create item specifications. Then, using these specifications, items are created, tested, and refined before being used on actual assessments.

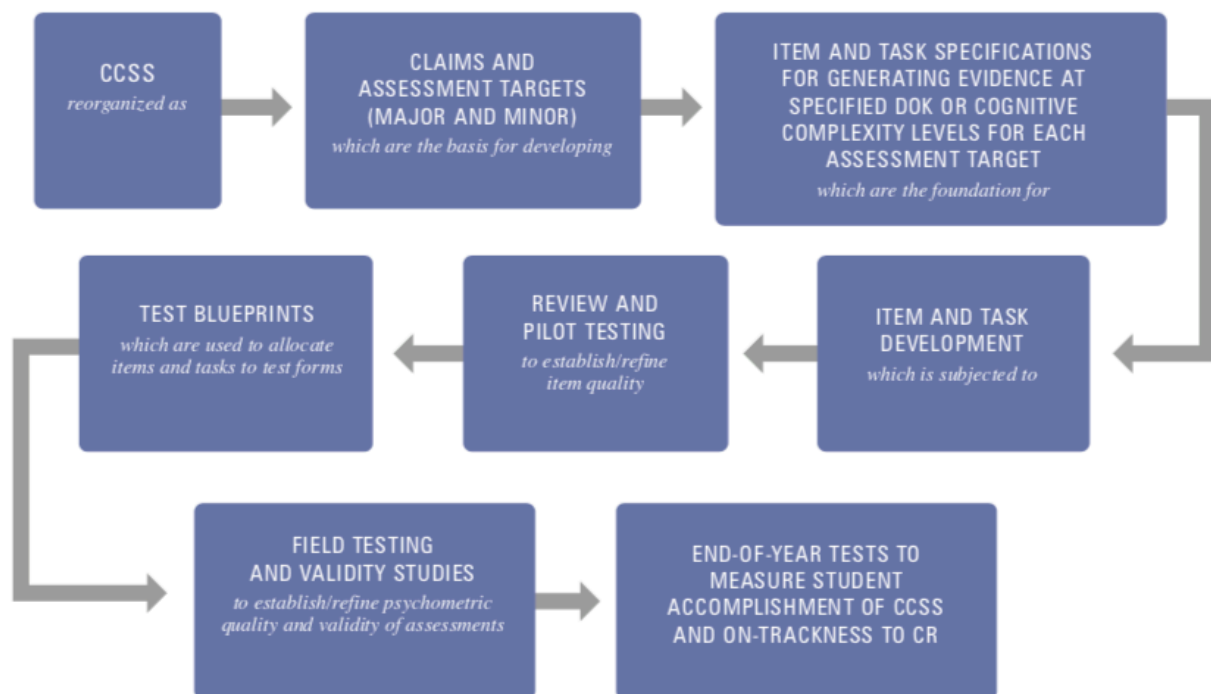


Figure 2.5 ECD General Approach (Herman & Linn, 2013, p. 7)

Specific to SMP3b and the practice of students critiquing, the PARCC consortium has a claim derived from the CCSSM that states, “Students express mathematical reasoning by constructing mathematical arguments and critiques” (Hermann & Linn, 2013, p. 8). Similarly, the SBAC consortium has a claim that states, “Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others,” drawing

language directly from SMP3b (p. 8). Therefore, both consortia assert that their assessments will assess SMP3b. Due to the consequential nature of assessments:

If the tests produced by the consortia provide students with opportunities to demonstrate such mathematical habits of mind [in reference to SMPs], the tests will serve as a level for moving the K-12 system in productive directions. But, if they consist largely of short answer questions aimed at determining students' mastery of facts and procedures, they will impede the kind of progress we have been making over the past 25 years (Schoenfeld, 2013, p. 24).

Although the use of ECD as a design framework indicates the consortia's intentions at the general levels of what will be assessed, "key questions remain about how well these intentions will be realized" (Herman & Linn, 2013, p. 13). Much of the development of the assessments along with the technological advancements informing the assessment writing is hidden from view of the public in favor of test security. Therefore, even though these testing consortia have indicated that their developed assessments will assess the SMPs, is it unclear exactly how these practices will be assessed on assessment tasks.

In connection to my study, in order to promote SMPs as important mathematical habits of mind, it is important to consider how these practices are assessed. The writers of the CCSSM intended for both content and practices to be assessable and two testing consortia are using tenets of ECD to design assessment tasks that assess these practices. However, teachers do not have access to these assessment tasks due to test security concerns and the business of high-stakes testing. Even so, the work of these consortia reveals that assessment of mathematical habits of mind is a key component of these CCSSM-aligned assessments and a broader goal for overall mathematics assessment design. Therefore, it is important to consider how the assessment

materials more readily available to teachers, including curriculum-based assessments, assess these practices. For the focus of my study, I am exploring student work in curriculum-based assessment tasks as a possible mechanism for assessing SMP3b.

Student Work

For the purposes of this review of literature related to research on student work in mathematics, I focus on studies of teachers' examination of student work and students' examination of student work or worked examples. I also discuss one research group's conceptualization of student work and their study of student work embedded in student textbook tasks in three seventh-grade mathematics curriculum series.

Teachers' Examination of Student Work

A large number of studies of teachers' examination of student work often situate this teacher practice as a professional development component in a teacher learning community (Cameron, Loesing, Rorvig, and Chval, 2014; Driscoll and Moyer, 2001; Flowers, Mertens, & Mulhall, 2005; Kazemi & Franke, 2004; Silver & Suh, 2014; Slavit & Nelson, 2010). In either individual or community professional development situations, studies have shown that teachers' examination of student work has been used for the purposes of (1) advancing teachers' understandings of students' thinking (An and Wu, 2012; Driscoll and Moyer, 2001; Ghouseini & Sleep, 2011; Herbel-Eisenmann & Phillips, 2005; Kazemi & Franke, 2004; Kersaint & Chappell, 2004; McDonald, 2002; Silver & Suh, 2014) and (2) enhancing teachers' abilities to assess and improve teacher quality (Boston, 2014; McDonald, 2002; Ryken, 2009; Sandholtz, 2005). In the following paragraphs, I describe two studies that most clearly illustrate these two purposes. I also connect the purposes of teachers' examination of student work to my current study of the use of student work in assessment tasks as a mechanism for assessing SMP3b.

Teachers often examine students' written work in order to make sense of and gain deeper understanding of students' thinking. Silver and Suh (2014) studied the use of student work for teacher learning with teachers of grades 7-11 during a 3-year professional development initiative. Their analysis focused on teachers' engagement with one mathematical task. First, teachers were asked to complete the mathematical task, predict students' solutions, and make comparisons with a partner in the professional development setting. Then, teachers collected their own students' work on the task. This student work was compiled for use and the professional development leaders led an exploration of the student work. Teachers were encouraged to analyze the student work by grouping the student work based on characteristics of students' strategies or solutions.

The authors indicated that student work was not an automatic source of teacher learning, but instead teachers' examination of student work in order to make sense of student thinking needed to be supported by explicit criteria for examining student work. The authors described the pervasive orientation of viewing student work for evaluative purposes that teachers held and the purposeful reorientation needed to make examination of student work a learning experience for teachers. Silver and Suh (2014) found that teachers' examination of student work, when framed purposefully, allowed teachers to gain insight into student thinking once they stopped viewing student work in only an evaluative way.

According to the research literature, teachers also examine student work as a teacher practice in order to assess and improve their own teaching. For example, Sandholtz (2005) described her own self-study of her teaching practice by examining her students' work in which her students, who are practicing teachers, analyzed the work of their own K-12 students. Thus, analysis of student work as a component of teacher learning in Sandholtz's study was focused on assessing her teaching and determining improvements for the future. A key component of the

author's findings is as followed: "Teaching strategies that engage students in analyzing their own practice ... present opportunities for authentic assessment of student understanding of the process" (Sandholtz, 2005, p. 120). In other words, driven to improve her own teaching, Sandholtz reviewed students' work from her course in order to determine teaching strategies that could help future teachers (her students) be more successful in making sense of their own teaching and learning practices.

Overall, the research literature on teachers' examination of student work for teacher learning provides insight into student work as a context for teachers to gain insight into their students' thinking and assess their teaching practices in order to improve their instruction. Connecting this research to my own study with a focus on student work embedded in assessment tasks as a mechanism for assessing SMP3b, I explore both student work embedded in assessment tasks as well as students' generated work on these tasks. It was important for me to gain teachers' perspectives on these tasks as well as students' written work on these tasks because, as the research shows, teachers have used student work both to gain insight into students' thinking as well as to assess their own teaching. I posit that SWAT could be educative for teachers in two ways. First, SWAT could provide teachers with insight into possible student strategies or habits of mind. Second, SWAT could provide teachers with a tool for assessing their teaching of SMP3b in order to inform future teaching. Thus, SWAT could serve as both a professional learning tool for teachers as well as an assessment tool.

Students' Examination of Student Work

Few studies have explored how students could benefit from examining student work in ways similar to what is described in the studies of teachers' examination of student work. Even so, numerous studies of students' examination of worked examples have focused on the benefits

for students of learning from examples. Specifically, these studies have focused on (1) students' examination of problems with error analysis (Lannin, Townsend, & Barker, 2006), (2) the role of student work in classroom discussions (Ely & Cohen, 2010), (3) and how students can learn from worked examples (Atkinson, Derry, Renkl, and Wortham, 2000; Booth, Lange, Koedinger, & Newton, 2013; Rittle-Johnson, Star, Durkin, 2009; Rittle-Johnson & Star, 2011; Rittle-Johnson, Star, Durkin, 2012; Rittle-Johnson & Star, 2007; Rowland, 2008; Star, Pollack, Durkin, Rittle-Johnson, Lynch, Newton, & Gogolen, 2014; Star & Riddle-Johnson, 2009). In the following paragraphs, I describe studies that most clearly addressed these three themes. I also connect these themes to my current study of the use of student work in assessment tasks as a mechanism for assessing SMP3b.

According to a number of studies, students often examine student work in order to identify and correct errors. For example, Lannin, Townsend, and Barker (2006) developed a reflective cycle of error analysis (see Figure 2.6) framework in their work analyzing the reasoning of two twelve-year-old students to understand how students both recognize and reconcile their own errors. Findings from their study showed that if teachers want to promote students as problem solvers, recognizing and using errors to solve problems needs to be more commonplace practices in mathematics classrooms. The authors indicated, "we must further understand how we can encourage students to make use of their errors rather than simply ignore them" (Lannin, Townsend, & Barker, 2006, p. 38).

Identifying and correcting errors is a key component of real problem solving and is a complex practice used in many professions. Thus, normalizing errors as learning opportunities as opposed to experiences to avoid not only provides productive problem solving situations but also engages students in practices used in the real world. Error analysis as described in this study

varies slightly from the student work of interest in the current study because Lannin, Townsend, and Barker (2006) focused on students analyzing their own errors as opposed to analyzing another student's work. Even so, the authors provided insight into how error analysis allows students to engage in authentic mathematical experiences that mirror how mathematics is used in numerous occupations.

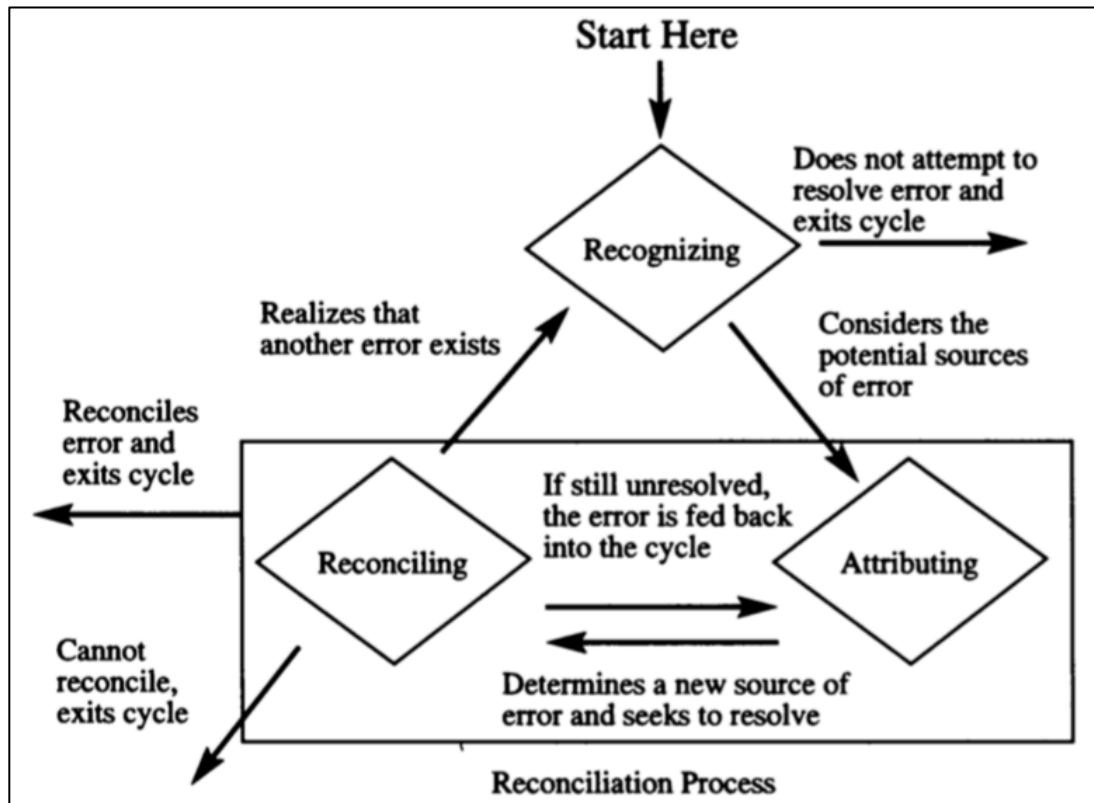


Figure 2.6 The Reflective Cycle of Error Analysis (Lannin, Townsend, & Barker, 2006)

In addition to error analysis, students often also examine student work in mathematics during classroom conversations. Drawing on the five practices of anticipating, monitoring, selecting, sequencing, and connecting from Smith and colleagues (2009), Ely and Cohen (2010) described selecting and sequencing student work on a rich and student-centered mathematical task in order to teach compound probability. Their work was guided by the question, “What is the best way to use student-generated work on [complex] tasks to guide a productive whole-class discussion?” (Ely & Cohen, 2010, p. 210). In implementing Smith and colleagues’ practices, the

authors discussed the importance of selecting and sequencing student work examples to be shared in whole-class discussion that not only support the mathematical goal of the lesson, but also “display important realizations about the task and challenge misconceptions observed during the monitoring phase” (Ely & Cohen, 2010, p. 211). Thus, student work can be purposefully selected for use in whole-class discussions to reveal important mathematical strategies and ideas. Specific to their lesson on compound probability, the authors found that students’ engagement with the mathematics and students’ learning from using student work in whole-class discussions made the complex task worthwhile.

Lastly, beyond the examination of worked examples, research has shown that students’ comparison of worked examples can prompt students to analyze the viability of various methods for their own use in solving future mathematical tasks. Many of the studies focused on students’ examination of student and worked examples introduce the importance of comparison of examples for improving student learning. Rittle-Johnson and Star (2011) described a series of five studies in which the researchers utilized comparison of worked examples. Students were presented with paired worked examples side by side and provided explicit directions to find similarities and differences. In one of the studies of seventh-grade students solving multistep linear equations, the authors made comparisons between students that viewed worked examples as described above and students that studied individual worked examples in a sequential order.

The authors found that the students that were prompted to compare two student work methods side by side gained greater procedural fluency than the students that only viewed worked examples. Their findings indicated, “those who compared methods often compared the similarities and differences in solution steps across examples and evaluated their efficiency and accuracy; these students were also more likely to use alternative methods when solving practice

problems during the intervention” (Rittle-Johnson & Star, 2011, p. 212). Therefore, when students examine multiple instances of student work, they often engage in comparison that supported their future solving of mathematical tasks.

In examining student work, the research literature shows that students often have the opportunity to learn from their own errors, learn from other students’ work in classroom discussions, and improve their mathematical understandings by comparing worked examples. Extending to the current study, perhaps the examination of someone else’s work on an assessment could provide students opportunities to demonstrate their ability to identify and correct errors, talk about someone else’s mathematical thinking, and make comparisons between multiple instances of student work. In other words, student work could serve as a mechanism to assess students’ abilities to critique the reasoning of others in a number of ways already highlighted in the literature around learning from student work. In my study, I am interested in understanding how student work can be used as an assessment tool that can possibly afford many of the same benefits of student work illustrated above.

A Study of Student Work Embedded in Student Textbook Tasks

Gilbertson and colleagues (2016) investigated instances of student work in curriculum series in their development of a framework for curriculum analysis. The researchers studied the existence and character of student work exemplars in three different seventh-grade curriculum series intended to be aligned with the CCSSM. Focusing their analysis on chapters in the student texts that explored the mathematical idea of similarity, the researchers looked specifically at student work embedded in the curriculum materials as one way for curriculum writers to promote the practice of students analyzing, critiquing, and/or reflecting on the mathematical reasoning of another (see Figure 2.7).

Gilbertson et al.'s criteria for tasks embedded with student work were created by the researchers as they examined the curriculum texts and refined their ideas about student work by comparing various examples and non-examples and reaching a consensus about the distinguishing characteristics of student work problems in contrast to tasks that used real-life contexts or only required the student reader to complete a mathematical task for a character in the problem. The authors positioned student work as one avenue for promoting students' ownership of engaging in mathematical thinking and generating mathematical ideas (in contrast to the textbook or the teacher as the mathematical authority in a classroom), so in looking for instances of student work within the curriculum texts, the researchers looked for tasks where "the author of the work [in the task] is uniquely positioned as external to the classroom interaction" (Gilbertson et al., 2016, p. 4). This characteristic is reflected in the first criterion.

The three criteria of Curriculum-generated Student Work are:

1. The mathematical task must mention at least one person (the *character*) to which the work is attributed.
2. The task must include a character's thinking or actions or prompt the reader to determine the character's thinking or actions. Thinking might include a written mathematical claim, a conjecture, a strategy, some form of reasoning, an observation or measurement, an algorithm, or a reflection on a mathematical idea.
3. There must be an expected activity for the reader of the text. These activities might include analyzing, critiquing, or reflecting on the mathematical thinking/actions of the character in the written materials.

Figure 2.7 Criteria for Curriculum-Generated Student Work (Gilbertson et al., 2016)

The second and third criteria for student work require that the tasks the authors looked for not only included evidence of a character's thinking or actions (Criterion 2), but also required the

reader to engage in making sense of the character's thinking or actions (Criterion 3). This third criterion helped eliminate tasks that included a character and some information about the character's thinking or actions, but the task for the reader was to only complete the problem for the character without making sense of, analyzing, or reflecting on the work of the character in the task. The authors suggested that all three criteria taken together helped them identify tasks in the curricula that "closely reflect what might be generated in the classroom as student work during the course of a discussion" (Gilbertson et al., 2016, pp. 5-6).

Gilbertson and colleagues' (2016) curriculum analysis framework provided criteria and language for what counts as student work in mathematics tasks. In my own study, I adapted Gilbertson and colleagues' criteria for student work in order to analyze the instances of student work that occurred in the curriculum-based assessments. I also extended their work by considering not only the opportunities or potential students had to engage with student work tasks on assessments, but also students' actual experiences with these tasks and teachers' perspectives on SWAT and students' work on SWAT.

Summary

My study is informed by mathematics education research literature on curriculum-based assessments, SMP3b, and student work. Assessment research literature indicates that prior studies of curriculum-based assessments (Hunsader, 2013; 2014) reinforce Shepard's (2000) observation that assessments lag behind advancements in instructional practices. Even so, reform efforts focused on improving assessment such as backwards design (Wiggins & McTighe, 2005) show promise for the development of assessments that assess "the mathematical thinking and habits of mind we value" (Seeley, 2014, p. 253). The CCSSM SMPs represent one set of habits of mind informed and influenced by previous attempts to capture "what it means to do

mathematics and think mathematically” (Seeley, 2014, p. 253) including previous mathematics standards and seminal works (NCTM, 2000; 2014; NRC, 2001). One of the habits of mind of particular interest in this study is SMP3b, “critique the reasoning of others” (CCSSI, 2010, p. 6).

Research literature on teachers’ and students’ examination of student work shows that student work has proven to be a valuable context for teacher learning as related to making sense of students’ mathematical thinking and assessing and improving teaching. Similarly, studies have shown the value of students examining student work in order to complete error analysis, discuss students’ strategies, or make comparison between multiple instances of student work. Specific to the use of student work in curriculum materials, one group of researchers (Gilbertson et al., 2016) created a framework for analyzing curriculum materials for instances of student work in student textbook tasks. In their framework, they provided explicit criteria about what counts as student work in tasks.

I view my study as exploring one possibility for challenging traditional forms of assessment by using an advancement in curriculum design, the use of embedded student work, to inform assessment task design and explore whether or not student work can serve as a mechanism for assessing SMP3b. In this work, I acknowledge the powerful influence assessments have on teaching and learning as described in the research literature and hypothesize that embedded student work in curriculum-based assessments may further promote the practice of critiquing the reasoning of others and allow teachers to assess students’ abilities to engage in this practice.

CHAPTER 3: METHODOLOGY

Overview

In this chapter, I describe the methods I used in my exploratory validity study of whether or not student work embedded in assessment tasks, or SWAT, could be used as a mechanism for assessing SMP3b. First, I introduce the guiding methodology for my exploratory validity study, Evidence-Centered Assessment Design (ECD), and define key constructs of assessment as *evidentiary argument* and *validity* (Gotwals, Hokayem, Song, & Songer, 2013; Gotwals & Songer, 2013; Mislevy, 2012; Mislevy & Haertel, 2006; Mislevy & Riconscente, 2005; Mislevy, Steinberg, & Almond, 2003). I then describe the methods I used to address my research questions that examined (1) the frequency and nature of SWAT in curriculum-based assessments, (2) how students talked about and made sense of SWAT, and (3) how teachers talked about and understood SWAT and non-SWAT based on the tasks and students' written work on the tasks. I conclude this chapter by connecting the data collection and analysis methods I utilized in my study to the stages of ECD.

Evidence-Centered Assessment Design (ECD)

Evidence-centered assessment design (ECD), pioneered by Mislevy and colleagues (e.g. Mislevy & Haertel, 2006; Mislevy Steinberg, & Almond, 2003), provides tools, concepts, and structures for designing and implementing assessment tasks based on the premise that assessment is a process of examining evidence or making observations in certain circumstances in order to be able to make inferences or claims about what students know or can do. As stated by Mislevy, Steinberg, and Almond (2003):

In assessment, the data are the particular things students say, do, or create in a handful of particular situations, such as essays, diagrams, marks on answer sheets, oral

presentations, and utterances in conversations. Usually our interest lays not so much in these particulars, but in the clues they hold about what students know or can do as cast in more general terms. These are the *claims* we'd like to be able to make about students, on the basis of *observations* in an assessment setting. The nature and the grainsize of assessment claims are driven by the purpose(s) of the assessment. The task of establishing the relevance of assessment data and its value as evidence depends on the chain of reasoning we construct from the evidence to the claims. (p. 11-12)

Therefore, ECD involves assessment as *evidentiary argument*. This means that assessment is “an argument from what we observe students say, do, or make in a few particular circumstances, to inferences about what they know, can do, or have accomplished more generally” (Mislevy, Haertel, 2006). The ECD stages, or layers, provide design and delivery structures for building up arguments linking observations to claims through repeated reasoning (see Figure 3.1).

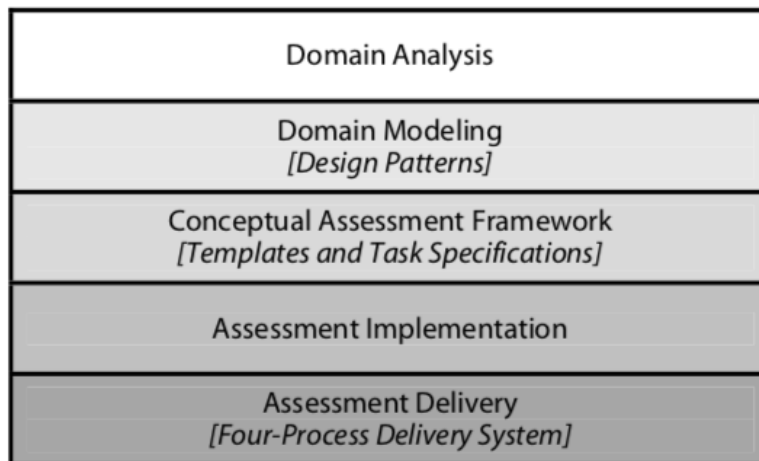


Figure 3.1 ECD Layers (Mislevy & Riconscente, 2005, p. 7)

Through a series of arguments, the design and delivery structures provided in ECD link observations of students to claims about students. Furthermore, “Strong arguments give us confidence in the quality of the inferences and interpretations, or in their validity” (Mislevy, 2012, p. 94). Mislevy defined *validity* as “the degree to which evidence and theory support

interpretations of test scores entailed by proposed uses of a test” (Mislevy, 2012, p. 94). As stated by Gotwals, Hokayem, Song, and Songer (2013), validity “is not just about a given assessment, but also refers to implications for the interpretations and uses of given assessments” (p. 4). Therefore, validity of an assessment is influenced by the strength of the assessment as an argument, using repeated reasoning to link observations and claims, as well as how interpretations of the arguments are then used.

In their study of assessment tasks that fused ecological content and scientific explanation, Gotwals and Songer (2012) used a validity argument to determine the degree to which the tasks provided evidence about students’ knowledge of ecology and scientific explanations. The authors stated, “Developing an argument about an assessment must involve both a clear articulation of the intended knowledge and skills to be measured as well as matching this with empirical evidence of students interacting with the items or tasks” (p. 4). In this way, designing assessment tasks or exploring design features in existing assessment tasks using ECD, require not only a clear idea of what tasks are intended to assess, but also gathered evidence supporting that the tasks do in fact assess the intended purpose.

In my study, I had a conjecture for what the SWAT could assess, SMP3b, as evidenced by my purpose and research questions. I examined explanations of SMP3b in the CCSSM SMPs and research literature on habits of mind related to SMP3b to articulate the knowledge and skills described in SMP3b in more detail (CCSSI, 2010; Koestler, Felton, Bieda, & Otten, 2013; Hull, Miles, & Balka, 2012; NCTM, 2000; 2014; NRC, 2001; Seeley, 2014). I also researched how habits of mind had been previously studied in curriculum-based assessments (Hunsader et al., 2013; 2014). The exploratory components of my study investigated the second part of assessment as argumentation as described by Gotwals and Songer (2012), “matching [intentions]

with empirical evidence of students interacting with the items or tasks” (p. 4). By extending my study beyond determining the potential students have to engage in making sense of someone else’s thinking on curriculum-based assessments to implementing the tasks in clinical interviews with students while also gaining insight into students’ and teachers’ perspectives on SWAT, I sought to better understand what evidence would be required from students in order to demonstrate SMP3b on an assessment task. For this study, I used an exploratory validity approach that mirrored key stages of ECD in order to answer my three research questions. In the sections that follow, I detail my methods for addressing each research question. Then, I show how these methods connect to stages of ECD.

RQ #1: Exploring the Existence of Student Work in Curriculum-Based Assessments

In this section, I detail my methods for answering the following research question:

1. What is the frequency and nature of SWAT in curriculum-based assessments?
 - a. How prevalent are SWAT in curriculum-based assessments?
 - b. How do SWAT vary across curriculum series?
 - c. How do SWAT compare to SWTT in corresponding student textbooks?

To investigate the frequency and nature of SWAT in curriculum-based assessments, I completed text analyses of student textbooks and curriculum-based assessment materials from the following seventh-grade mathematics textbooks intended to be aligned with the CCSSM: *Big Ideas* (Larson & Boswell, 2014), *Connected Mathematics3* (Lappan et al., 2014; I refer to as CMP), *College Preparatory Mathematics* (Dietiker et al., 2013; I refer to as CPM), *Eureka Math* (Great Minds, 2015; I refer to as Eureka), and *Go Math* (Burger, 2014). These analyses focused on instances of SWAT as well of the nature of the SWAT, including the assessment type, CCSSM content, evidence of student thinking, and critique types. In the following paragraphs, I

detail the data collection, data analyses, and inter-rater reliability processes for Research Question 1.

Data Collection

Extending the work of Gilbertson and colleagues (2016) from a focus on a single mathematical topic to an entire grade level and from a focus only on the student textbooks to including the corresponding assessment materials, this study explored the existence and nature of student work in curriculum-based assessments from five seventh-grade mathematics curriculum materials intended to be aligned with the CCSSM. These analyses were compared to analyses conducted (and further extended in this study) as part of an earlier study focused on the existence and nature of student work in the student texts for each of the five seventh-grade mathematics curriculum series (Going, Ray, & Edson, in preparation).

Rationale for Curriculum Selection. The seventh-grade mathematics student textbooks and corresponding assessments selected for this study were chosen for several reasons: (1) the textbooks were designated as aligned with CCSSM, both the Content Standards and the SMPs, by the curriculum writers, (2) the textbooks were representative of a range of middle school mathematics curricula based on publishers and approaches towards teaching mathematics, and (3) the sample was one of convenience with the intention of making comparisons to previously conducted curriculum analyses from earlier studies.

First, all the textbooks were intended to be aligned with the CCSSM and the curriculum writers explicitly highlighted how their textbooks attend to SMPs. Two of the curriculum series, Big Ideas and Go Math, highlighted specific tasks in the student textbooks that attended to each of the SMPs. Specific to SMP3, the writers of Big Ideas detailed how specific problem types – Error Analysis; Different Words, Same Questions; and Which One Doesn’t Belong – were used

to address this practice standard (see Figure 3.2). The Big Ideas authors stated that these problem types “provide students the opportunity to construct arguments and critique the reasoning of others” explicitly referencing SMP3b (Larson & Boswell, 2014, p. vi). Similarly, curriculum writers of Go Math provided a resource in the front of the textbook that provided page numbers and task numbers for tasks that promoted each of the eight SMPs. Specific to SMP3, counterexample tasks, as designated by the curriculum writers, were a common type of task listed for this practice (see Figure 3.3).

12. ERROR ANALYSIS Describe and correct the error in finding the simple interest earned on \$500 at 6% for 18 months.

$I = (500)(0.06)(18)$
 $= \$540$

Figure 3.2 SMP3 Example Task from Big Ideas (Big Ideas, Larson & Boswell, 2014, p. 256)

H.O.T. FOCUS ON HIGHER ORDER THINKING

26. Counterexamples John says that if one side of an inequality is 0, you don't have to reverse the inequality symbol when you multiply or divide both sides by a negative number. Find an inequality that you can use to disprove John's statement. Explain your thinking.

Figure 3.3 SMP3 Example Task from Go Math (Go Math, Burger, 2014, p. 210)

Two of the curriculum series, CMP and CPM, did not highlight specific tasks that promoted the SMPs, but instead discussed how the practices were broadly promoted in their curriculum materials. The writers of CMP described how SMPs “were already embedded in the CMP curriculum.” Specific to SMP3, the authors stated the CMP materials “support a pedagogy that focuses on explaining thinking and understanding the reasoning of others” (CMP, 2018a). Similarly, the curriculum writers of CPM, in their online resources, described how the SMPs were “deeply woven into the daily lessons” and, specific to SMP3, that the structure of the CPM

curriculum ensured that “justifying and critiquing happens every day in a CPM classroom” (Dietiker et al., 2018).

The curriculum writers for the last curriculum series, Eureka, did not provide specific tasks or general information about how their materials addressed the SMPs. Instead, the materials include “Focus Standards of Mathematical Practice” for each module derived from the standards document (CCSSI, 2010; Great Minds, 2015). For example, SMP3 is listed as a Focus SMP for the last two of six modules for the seventh-grade curriculum materials.

A second reason that these five curricula were selected was because they represented a range of middle school mathematics curriculum materials and a variety of approaches towards teaching mathematics. The textbooks were representative of four different publishers: Houghton Mifflin Harcourt (Big Ideas, Go Math), CPM Educational Program, a California nonprofit organization (CPM), Great Minds (Eureka), and Pearson (CMP). The materials also came from curriculum series with varying grade bands: grades PreK-12 (Eureka), grades K-8 (Go Math), grades 6-8 (CMP), and grades 6-12 (Big Ideas, CPM).

The five curriculum series utilized different approaches to engaging students in the process of learning mathematics. Big Ideas used a “research-based strategy of a balanced approach to instruction” which meant students spent an equal amount of time on conceptual understanding through discovery learning and procedural fluency through explicit instruction (Larson & Boswell, 2014, iii). CMP involved an “inquiry-based teaching-learning classroom environment” where students investigated real life situations to make sense of mathematics (CMP, 2018b). Somewhat similar, CPM used “problem-based lessons structured around a core idea” that engaged students in collaborative learning while guided by the teacher (Dietiker et al.,

2018). A key tenet of CPM was “practice with concepts and procedures should be spaced over time” (Dietiker et al., 2018).

The writers of Eureka touted their materials as the first curriculum aligned with the CCSSM, introduced in 2013. The Eureka curriculum was built on progressions of mathematics through modules focused on specific content as detailed in the CCSSM Content Standards (CCSSI, 2010). The writers of Eureka emphasized the alignment of the curricular modules with the CCSSM and the development of mathematics within the seventh-grade textbook and across the sixth, seventh, and eight grade bands, coined *A Story of Ratios* (Great Minds, 2015). Lastly, according to their publisher, the authors of Go Math incorporated “the latest thinking in its comprehensive approach and engages digital natives with cross-platform technology. It helps teachers to differentiate instruction, building and reinforcing foundational math skills that translate from the classroom to real life” (Go Math, 2018). The curriculum materials for Go Math included numerous technological tools included as a part of the curriculum materials along with the student textbook.

Finally, the sample of materials was chosen as a convenience sample based on access to both instructional and assessment materials with the intention of making comparisons to curriculum analyses previously conducted and ongoing for research studies conducted by the CMP research group (Gilbertson et al., 2016; Going, Ray, & Edson, in preparation; Nimitz et al., 2015).

Collecting Curriculum-Based Assessments. For my dissertation study, I explored students’ opportunities to interact with student work in curriculum-based assessments from each of the five seventh-grade curriculum series. The curriculum-based assessments for each curriculum series that I analyzed for this study included the written assessment materials

provided by the curriculum writers of each series. Big Ideas had an Assessment Book that included all of the assessment resources. CMP included written assessments deemed Surveys of Knowledge provided as part of a Teacher Lesson Support DVD. On their online system, CPM provided sample individual and team tests for each of their units as well as a large question bank repository for use on any written assessment. Eureka provided assessment packets for each module that included a Mid-Module and an End-of-Module written assessment. Go Math, similar to Big Ideas, provided an Assessment Resources book.

These written assessments did not account for all the assessment tasks and/or practices that would likely occur in any classroom using the curriculum series investigated in this study. Even so, the written assessment materials designated by curriculum series are important for making sense of what each curriculum series promotes as mathematics and what teachers will be exposed to as important for assessing mathematics. Investigating students' opportunities to engage in making sense of someone else's thinking on curriculum-based assessments is important for understanding how SMP3b could be promoted in classrooms as mediated by curriculum materials.

Refining Criteria for Student Work. Prior to conducting analyses to determine the instances of student work in the curriculum-based assessments, I refined the criteria originally used by Gilbertson and colleagues (2016) and Going, Ray, and Edson (in preparation) for identifying student work tasks in student textbooks as these studies focused on identifying the instances of student work based on all three criteria, not determining whether tasks met the criteria at the level of each criterion. For my dissertation study, it was important for the criteria to be explicit and clear because I conducted analyses at the level of each criterion. Starting from the First Revised Criteria for Student Work used by Going, Ray, and Edson (in preparation, see

Appendix A), I further refined the criteria, providing more detail for each criterion (see Appendices A and B). These refinements were based on initial coding of the curriculum-based assessments using the first set of adapted criteria. Because I was interested in analyzing how tasks met the criteria at the level of each criterion, not just tasks that met all three criteria as used in the earlier studies, the criteria descriptions required additional detail. Additionally, I wanted to make explicit connections between the Criteria for Student Work and SMP3b (see Figure 3.4). The criteria descriptions were refined through iterative reviews of the tasks.

Criteria for Student Work for SWAT		SMP3b
Criterion #1	The mathematical task must include a person based on the following: - Exclude ambiguous “you”, “I”, and “group” - Exclude places or corporations (e.g. school, town, store, company) - Include specific and general people (e.g. Daniel, your friend, boy, girl) - Include professionals (e.g. baker, owner, buyer, manufacturer) - Include groups of people (e.g. team, club, class) A person must be explicitly referred to and not just used to describe a place (e.g. Holly’s basement or Colton’s Gym)	others
Criterion #2	The task must include evidence of a person’s mathematical thinking. Evidence includes the following: - Claims, Conjectures, Statements, Arguments, or Reflections - Methods, Reasoning, or Algorithms - Observations, Measurements, or Diagrams - Actions, when explicitly and intentionally mathematical (e.g. conducted an experiment, designed a game with mathematical components, took a survey with provided purpose, collected/recorded data and evidence is provided, generated a sample, created or used a mathematical object, plotted points) **Attend to verbs and vocabulary** Exclude hypothetical mathematical thinking (e.g. wanting to triple the volume, planning to cut a board) unless evidence of mathematical thinking is included, as described above.	the reasoning of
Criterion #3	The expected activity for the reader of the text depends/relies on the thinking of the person(s) in the task. Activities include the following: - Critiquing/Verifying/Explaining someone’s mathematical thinking - Determining the correctness, fairness, soundness, representativeness, accuracy, or bias of someone’s mathematical thinking - Comparing multiple instances of mathematical thinking Exclude tasks where the reader is asked to: - Complete a task someone has started - Use someone’s mathematical thinking to create a different representation, solve a problem, or apply definitions (e.g. Which number in this equation represents the rate of change?, What type of sampling method was used?)	critique

Figure 3.4 Connections Between Criteria for Student Work for SWAT and SMP3b

In refining Criterion #1, it was important for the people in the tasks to be explicitly described or named as people because the language of SMP3b indicates that students should engage with the mathematical reasoning of “others” (CCSSI, 2010, p.6). I eliminated tasks that included ambiguous people such as tasks that included only “you,” “I,” and “group.” I also eliminated tasks that included places or corporations even when these entities were actors in the tasks. I justified the elimination of these types of tasks because it is not clear that people associated with these places or corporations are actively involved in the tasks. I parallel the idea of a store serving as an actor in a task with a mathematical textbook acting as an instructor in the text. Both of these examples essentially eliminate the people, in other words the humanity, from the mathematics of the task or text.

In refining Criterion #2, which was the most difficult to capture, it was important for the evidence of a person’s mathematical thinking to be explicitly mathematical. Tasks involving probability were the most difficult to code for this criterion because these tasks often involved people engaging in activities. However, these activities were not always explicitly mathematical. For example, “Suzie flipped a coin...” does not indicate that Suzie engaged in an explicitly mathematical activity. However, “Suzie conducted an experiment where she flipped a coin...” indicates that Suzie engaged in the activity of flipping a coin for an explicit mathematical purpose. A number of specific examples are provided in the further revised criterion because mathematical actions were the most difficult to determine. In contrast, claims, methods, or measurements were often much easier to discern as mathematical or not. For Criterion #3, I refined the language to highlight explicit examples and non-examples of critiquing activities. None of the revisions to Criteria #1-3 were in conflict with previous iterations of the criteria. Instead, the refinements provided more precise language and examples/non-examples at the level

of each criterion. Using the student work textbook task examples, or SWTT, from the student textbooks as determined by the earlier studies and the SWAT I analyzed, I conducted analyses to determine the number and nature of student work tasks in curriculum-based assessments.

Data Analysis

Using these further refined criteria which I designated as Criteria for Student Work (see Appendices A and B), I coded the curriculum-based assessment materials associated with each of the five seventh-grade curriculum materials. Beyond the initial review of materials to refine the criteria, three iterations of coding were conducted to determine whether tasks fulfilled the criteria. For the first iteration of coding, I coded all assessment tasks for whether or not the tasks fulfilled Criterion #1. For the second iteration of coding, I coded the tasks that fulfilled Criterion #1 for whether or not the tasks also fulfilled Criterion #2. Lastly, I coded the tasks that fulfilled both Criteria #1 and #2 for whether or not the tasks also fulfilled Criterion #3. For the purposes of this study, I called tasks that fulfilled all three criteria Student Work Assessment Tasks (SWAT)¹.

I completed analyses for the assessment tasks in each curriculum series and across the full set of assessment tasks based on the percentages of tasks fulfilling each criterion. I calculated these percentages based both on the number of tasks per curriculum series as well as the number of tasks fulfilling the previous criterion. For example, in determining how many tasks in CMP fulfilled Criterion #3, I calculated the percentage of tasks that fulfilled Criterion #3 out of the total number of CMP assessment tasks, the total number of CMP assessment tasks that fulfilled

¹The unit of a task was determined by the seriation used by each curriculum series. For a problem in the assessment materials or the student textbook that had multiple components defined by seriation, each component was considered an individual task (e.g. Parts A, B, C, and D of Problem 2 would be considered Tasks 2A, 2B, 2C, and 2D). If a criterion was met prior to a series of tasks (e.g. Use the following for questions 8-9), the criterion was counted for all the tasks in the sequence. However, if a criterion was met within a series of tasks (e.g. Part B of a series of Tasks A-D), the criterion was only applied to the single task (e.g. Part B), unless the criterion was met again in subsequent tasks.

Criteria #1, and the total number of CMP tasks that fulfilled Criterion #2. Due to my coding process in which I only coded tasks that fulfilled both Criteria #1 and #2 for Criterion #3, the number of tasks fulfilling Criterion #2 would be the same as the number of tasks that fulfilled both Criteria #1 and #2. These analyses indicated how many tasks overall fulfilled each criterion as well as how many tasks that included a person and evidence of someone's thinking also required the reader to critique their thinking in some way.

The goal of this work was not only to determine the number of assessment tasks that fulfilled each criterion, but also to make sense of the nature of student work tasks and make comparisons between the student textbook and the curriculum-based assessments for each curriculum series based on both the number and nature of student work tasks. Therefore, additional analyses were completed on the set of SWAT from each curriculum series as well as the SWTT identified by Going, Ray, and Edson (in preparation).

Using the student work tasks (as determined by the Criteria for Student Work; See Appendices A and B) from the student textbooks and the curriculum-based assessment materials, I developed codes and used preexisting codes to identify the assessment types (only applied to SWAT), the CCSSM Content Strands, evidence of student thinking, and the critique types for each task (see Appendix B). Each SWAT received one assessment type code. The assessment types were: diagnostic, periodic, summative, and question bank. These assessment types were determined based on the different types and purposes of the written assessment materials provided across the five curriculum series (see Figure 3.5).

Diagnostic	Assessment tasks included in: - Diagnostic assessments - Placement assessments - Beginning-of-the-year assessments - Pre-assessments - Readiness assessments These tasks are intended for use at the beginning of the year or the beginning of a chapter/unit/module.
Periodic	Assessment tasks included in: - Quizzes - Mid-Module Assessments These tasks are intended for use in the middle of a chapter/unit/module.
Summative	Assessment tasks included in: - Tests - Cumulative Assessments These tasks are intended for use at the end of a chapter/unit/module/semester or the end-of-the-year.
Question Bank	Assessment tasks included in: - Question Banks - Test Banks These tasks are provided as a resource for teachers to pick and choose tasks to use on assessments.

Figure 3.5 Assessment Type Descriptions

Each SWAT and SWTT received one CCSSM content strand code as described in the CCSSM standards document (CCSSI, 2010). Even though units, chapters, or modules could cover multiple CCSSM content strands as described by the curriculum materials, often for assessment and student textbook tasks, the tasks were assigned a specific CCSSM content strand. For tasks that were not designated a CCSSM content strand or were assigned multiple strands, because the size of the tasks was determined by the seriation of the curriculum materials which resulted in small increments of requirements for the reader, I felt that it was appropriate to determine which content strand was the best “fit” for each task from the strands designated by the curriculum series for units, chapters, or modules (see Figure 3.6).

RP	Ratios and Proportional Relationships – Analyze proportional relationships and use them to solve real-world and mathematical problems.
NS	The Number System – Apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.
EE	Expressions and Equations – Use properties of operations to generate equivalent expressions. Solve real-life and mathematical problems using numerical and algebraic expressions and equations.
G	Geometry – Draw, construct and describe geometrical figures and describe the relationships between them. Solve real-life and mathematical problems involving angle measure, area, surface area, and volume.
SP	Statistics and Probability – Use random sampling to draw inferences about a population. Draw informal comparative inferences about two populations. Investigate chance processes and develop, use, and evaluate probability models.

Figure 3.6 CCSSM Content Strand Descriptions (CCSSI, 2010)

Each SWAT and SWTT was coded for types of evidence of student thinking (see Figure 3.7). Each task could receive multiple evidence type codes depending on the evidence(s) of mathematical thinking the reader was required to critique in the task. These codes emerged from reviewing the SWAT and SWTT and noticing different types of evidence of student thinking that were used by the curriculum writers across the textbooks. These evidence type codes do not account for all possible types of evidence that could occur on tasks. Rather, these codes are representative of the evidence types that occurred on the SWAT and SWTT in this study.

Words	The task required the reader to make sense of someone's: - Thoughts - Reasoning - Explanations - Written Solutions - Statements - Predictions/Guesses/Estimates - Noticing - Claims
Symbols	The task required the reader to make sense of someone's: - Computations involving operations and/or variables - Developed Formulas - Developed Expressions/Equations - Operational Representations

Figure 3.7 Evidence of Student Thinking Categories and Corresponding Codes

Figure 3.7 (cont'd)

Visuals	The task required the reader to make sense of: - Diagrams - Tables - Graphs - Drawings - Visual Representations Representations can either be created by a person introduced in the task or serve as evidence of someone's thinking even when not explicitly created by a person.
Actions	The task required the reader to make sense of someone's descriptions of: - Methods/Strategies/Designs - Survey/Sample Plan and/or Results - Measurements - Mathematical Actions **especially important to attend to verbs and vocabulary for this category**

Because I wanted to investigate and determine whether or not tasks provided opportunities for students to engage in critiquing the reasoning of others, I was purposeful in coding for the evidence of student thinking that the reader was required to critique, not just the evidence of student thinking that appeared in tasks. Often, tasks introduced many different instances of students' thinking, but only required the reader to interact with a few. Therefore, the question or set of questions posed to the reader in the task often revealed the evidence types the reader would actually need to critique. For example, a task could provide a person's explanations, computations, visuals, and mathematical methods, but only require the reader to make sense of the person's computations with the prompt, "Is his equation correct?" Although numerous instances of student thinking might appear on many tasks, I wanted to focus on the instances of student thinking the reader actually was required to critique in order to make sense of students' opportunities to engage in critiquing the reasoning of others on student textbook and assessment tasks.

Each SWAT and SWTT was coded for critique type (see Figure 3.8). Each task could receive multiple critique type codes depending on the type(s) of critiquing the reader was

required to engage in on the task. The critique types were often revealed in a question or set of questions that were posed to the reader. Similar to the evidence type codes, the critique type codes emerged from reviewing the SWAT and SWTT and noticing different types of critique types used by the curriculum writers across the textbooks. These critique type codes do not account for all critique types that could occur on tasks. Rather, these codes are representative of the critique types that occurred on the SWAT and SWTT in this study.

ErrorID	The task required the reader to identify someone's error and/or correct someone's error. The existence of an error was explicit in the task.
Eval	The task required the reader to determine correctness, accuracy, validity, truth, bias, viability, fairness, and/or representativeness of student thinking.
Compare	The task required the reader to compare multiple instances of student thinking. This could involve comparing different people's thinking or different types of evidence of thinking from one or many people.
Pref	The task required the reader to make a choice based on preference. This includes determining what makes the most sense, what is easiest, what is efficient, and/or what is preferred.
Insights	The task required the reader to provide insight into some evidence of student thinking. This includes determining the intent, purpose, motivation, reasoning, or meaning of some evidence of student thinking.

Figure 3.8 Critique Categories and Corresponding Codes

Once codes for each assessment task were assigned (assessment type, CCSSM content strand, evidence type, and critique type) using the coding framework (see Appendix B), I conducted analyses across and within the SWAT. First, I looked at the distributions of assessment types, CCSSM content strands, evidence types, and critique types across the full set of SWAT. This allowed me to gain a general picture of the SWAT by the coding categories. Then, I conducted additional analyses at the level of the assessment types. I looked at the distribution of tasks across the three represented assessment types: periodic, summative, and question bank, based on CCSSM content strands, evidence types, and critique types. The results of these analyses were not helpful in terms of answering my research questions, but I was curious if any interesting differences or trends arose from comparison across assessment types.

No major differences or interesting trends emerged. More information about these analyses is provided in Appendix J.

Then, I conducted analyses at the level of the curriculum series, Big Ideas, CMP, CPM, Eureka, and Go Math, based on assessment types, CCSSM content strands, evidence types, and critique types. Lastly, I considered both SWAT and SWTT and made comparisons between the SWAT and SWTT for each curriculum series by CCSSM content strands, evidence types, and critique types. The goal of these analyses was to make sense of students' opportunities to critique someone else's mathematical thinking in curriculum-based assessments as well as make comparisons between these opportunities and those found in the student textbooks.

Inter-Rater Reliability (IRR)

A second coder (a fellow researcher from the CMP research group) was recruited to code sample sets of the assessment tasks from each curriculum series during two rounds of IRR coding. The second coder completed her coding separately. For the first round of IRR coding, I wanted to verify the completeness of the criterion descriptions in the Criteria for Student Work coding framework (see Appendices A and B). Because the coder was previously involved in only considering tasks that fit all three criteria and not at the level of each criterion for the analyses conducted for the earlier studies, she was not involved in the initial construction of the further revised criteria, but she was familiar with the general idea for each criterion. This proved to be extremely valuable in vetting the completeness of the language used for each criterion description.

I selected tasks for the first round of IRR coding based on a number of factors. I wanted the second coder to code at least one assessment from each assessment type (e.g. a periodic assessment) for each curriculum series. Second, I wanted her to code tasks that assessed the

CCSSM content strand expressions and equations from each curriculum series, because this was the content strand used in the assessment tasks for the student and teacher interviews that will be described later in this chapter. Third, I wanted her to code tasks that assessed probability content from each curriculum series, because probability tasks were the most difficult to discern whether or not a person's actions were mathematical. Because the probability tasks were the most difficult for me to code, I wanted to verify that the language I was using for Criterion #2 was reliable when coding these tasks. Based on these requirements, I compiled assessment tasks from at least two chapters/units/modules for each curriculum series, one focused on expressions and equations and one focused on probability, as well as assessment tasks from each assessment type represented in each curriculum series. Of the 4802 total tasks analyzed, the second coder analyzed 1417 of the tasks (29.5%). At the level of each curriculum series, for Big Ideas she coded 279 tasks (26.1% of Big Ideas tasks), for CMP she coded 207 tasks (35.6% of CMP tasks), for CPM she coded 404 tasks (24.9% of CPM tasks), for Eureka she coded 59 tasks (35.3% of Eureka tasks), and for Go Math she coded 468 tasks (34.3% of Go Math tasks).

After the second coder completed criteria coding, she sent me her results which I compiled and used to make comparisons to my coding. The results of IRR coding for the Criteria for Student Work resulted in 99.4% agreement on Criterion #1, 94.3% on Criterion #2, and 97.3% on Criterion #3, with an overall agreement of coding of 97.9%. These percentages were calculated based on comparing differences in coding of tasks to the number of possible tasks for which a criterion could be met. For example, for Criterion #1, there were 1417 possible tasks in which a person could appear, so the IRR percentage was calculated out of 1417 possibilities. For Criterion #2, there were 561 tasks in which people appeared as determined by Criterion #1 (based on consensus coding), so the IRR percentage was calculated out of 561 possibilities.

Similarly, for Criterion #3, there were 112 possible tasks in which a person and evidence of a person's thinking appeared (based on consensus coding), so the IRR percentage was calculated out of 112 possibilities.

We met to discuss and compare differences in criteria coding as well as the language of each criterion in the coding framework and consensus in the coding of tasks was reached. The least initial agreement occurred on Criterion #2 which resulted in editing the language for Criterion #2 to include, “actions, when explicitly *and intentionally* mathematical”. This specific language was added when discussing problematic probability tasks that often involved people engaging in mathematical activity as perceived by the reader, but the actions of the person were not intentionally mathematical on the part of the person in the task. Overall, of the 44 differences in coding, only 10 differences in coding resulted in changes to my initial coding (0.5% of the possible 2090 codes).

No changes to criteria coding were made for the Big Ideas or Eureka tasks. Three changes were made to CMP criteria coding – one for a task with a person that I had overlooked (Criterion #1) and two for tasks that involved a person designing a game with mathematical components that I had not originally counted (Criterion #2). Two changes were made to CPM criteria coding for tasks in which hypothetical student work was included in describing the rules of a game. I had originally coded these tasks to count for Criterion #2, but in reviewing the criteria, these tasks did not count. Three changes were made to Go Math criteria coding for tasks in which codes or passwords were created. I had originally coded these tasks to count for Criterion #2, but it was not clear that code or password creation was explicitly mathematical as deemed in the language for Criterion #2. Two other changes were made to Go Math Criterion #2 coding – one inclusion for a task in which a person had explicit mathematical intent for a

sampling method and one exclusion for a task in which the person was slicing an object, but the intent on the part of the person is not clearly mathematical. Overall, only one change was made to Criterion #1 coding for an omission (0.05% of the possible 2090 IRR criteria codes). Changes in Criterion #2 coding were due to the difficulty of discerning when someone's actions were explicitly and intentionally mathematical (0.4% of the possible 2090 IRR criteria codes). There were no changes in coding for Criterion #3 (0% of the possible 2090 IRR criteria codes).

Beyond refining and verifying my criteria coding, I wanted to verify the completeness of the evidence type and critique type descriptions in the coding framework (see Appendix B). Since assessment types and CCSSM content strands were overwhelmingly determined based on information provided by the curriculum materials as opposed to based predominantly on interpretations of task features, I focused my second iteration of IRR coding on the codes that emerged from the assessment and student textbook tasks – evidence of student thinking and critique types. Of the 127 SWAT that were coded, the second coder analyzed 20 of the tasks (15.7% of the SWAT). I selected tasks for the second coder to code based on the evidence and critique types that occurred in each curriculum series. The tasks that were selected for IRR coding as a whole for each curriculum series accounted for all of the evidence and critique types that occurred for each curriculum series. Selection of the tasks beyond this requirement was random. At the level of each curriculum series, for Big Ideas she coded 2 tasks (50% of Big Idea SWAT), for CMP she coded 4 tasks (20% of CMP SWAT), for CPM she coded 8 tasks (9.5% of CPM SWAT), for Eureka she coded 3 tasks (30% of Eureka SWAT), and for Go Math she coded 3 tasks (33.3% of Go Math SWAT).

After the second coder completed evidence and critique type coding for the selected SWAT, she sent me her results which I compiled and used to make comparisons to my coding.

The results of coding for the evidence and critique types resulted in an overall agreement of 97.8% with an agreement on evidence types of 95% and an agreement on critique types of 100%. These percentages were calculated based on comparing differences in coding to the total possible number of codes that tasks could receive for evidence types (4 possible types per task) and critique types (5 possible types per task).

We met to discuss and compare differences in evidence type coding as well as the language of the evidence and critique types descriptions in the coding framework and consensus in the coding of tasks was reached. Because differences in coding occurred for evidence types, we revisited the evidence type descriptions in the coding framework. We edited the prompt for this coding type to: “Use the descriptions below to code tasks *while attending to the verbs and vocabulary used in the task.*” Additionally, for the Visuals category description we edited the description to include: “*Representations can either be created by a person introduced in the task or serve as evidence of someone’s thinking even when not explicitly created by a person.*” There was a task in the IRR SWAT as well as the full set of SWAT that included a visual that was not explicitly created by the person in the task, but the visual served as evidence of the person’s thinking. We wanted to be sure that these evidence types were included when the reader was tasked with making sense of the visual as a representation of someone’s thinking.

Only one change was made to the evidence type coding based on the IRR process. For a task in CMP that included both a visual representation and a description of a person’s mathematical action and the reader was required to make sense of both, I had not included the Action code. Therefore, for both the evidence and critique types coding, only one change was made in my coding for an omission for an Action evidence type code (1.3% of the possible IRR evidence type codes).

Overall, the IRR process and my intentional focus on problematic types of tasks for coding, specifically tasks related to probability content, resulted in reaffirming the revised descriptions for Criteria #1 and #3 and reaffirming the critique type descriptions. The IRR process and results also prompted me to slightly refine the description for Criterion #2 and refine the evidence type descriptions. I used results from IRR analyses and discussions about the student work criteria, evidence type descriptions, and critique type descriptions when revisiting the full sets SWAT and SWTT. Findings from criteria coding of assessment tasks and coding of the SWAT as well as comparisons of the SWAT to SWTT are detailed in Chapter 4.

RQ #2: Exploring Students' Experiences with and Perspectives on SWAT

In this section, I detail my methods for answering the following research question:

2. How do students talk about and make sense of SWAT?
 - a. What is the nature of students' written work and verbal responses on SWAT and non-SWAT?
 - b. How do students describe their experiences with SWAT?

To explore students' experiences with and perspectives on SWAT, I conducted clinical interviews with six seventh-grade students that involved students completing eight curriculum-based assessment tasks with and without embedded student work and talking about their experiences solving the assessment tasks. I followed up each clinical student interview with a semi-structured interview that included general questions about the students' experiences with the tasks and their perspectives on tasks that require the reader to engage in making sense of someone else's thinking.

The goal of conducting these two types of interviews with each participant and the subsequent data analyses of students' discussions about solving assessment tasks as well as their

written work for these tasks was to be able to understand the nature of students' written work and their verbal reasoning on tasks as well as gain insight into students' experiences solving SWAT. Both of these interview types provided opportunities for me to explore how SMP3b was or was not evident in students' written work and verbal thinking on SWAT. In the following paragraphs, I describe the data collection and analysis methods used for the student interviews.

Data Collection

Based on previous experiences working with CMP teachers as a graduate research assistant on the CMP curriculum project, I was able to solicit CMP teachers for student participation in my study. I reached out to teachers via email for parent permission and student participation in my study during the spring of 2017. Because I was working with students in schools, I was required to follow district protocols for conducting research in classrooms with students. Once permission was granted from two districts after an administrative review of my research study, I was able to solicit student participation from two classrooms, one classroom from each district.

I recruited eight seventh-grade students to participate in my study from two local classrooms where the teacher was using the CMP curriculum series. I secured parent consent and student assent for each student interview conducted with the help of the two teachers that taught these students (see Appendices C and D). Prior to meeting with students in person, I selected assessment tasks to be used in the interviews, prepared clinical interview questions, and prepared semi-structured interview questions. After I conducted interviews with students and with teachers and reviewed student work from all of my student participants, I determined that it would be prudent to eliminate the final two student interviews conducted from my data analyses for this study because they did not differ substantially from the first six interviews and I only had

the clinical portion completed for the last interview. I detail these data collection methods and methodological decisions in the sections below.

Assessment task selection. In order to be able to explore students' experiences engaging with student work on assessment tasks, I selected assessment tasks that both included student work and did not include student work for students to complete in the interview. At the time of the student interviews during the spring semester of 2017, I originally only had four curriculum series selected for my analyses. Eureka was added as a fifth curriculum series due to requests from teachers at a professional development session about student work during the summer of 2017. Therefore, Eureka is represented in the text analyses previously described, but not in the tasks used for the student or teacher interviews.

Because I conducted student interviews at the end of the school year for seventh-grade students, I wanted the assessment tasks to focus on a key mathematical idea in seventh-grade so that the content of the tasks was consistent and the variance in the tasks would be the inclusion or exclusion of student work. Based on discussions with the larger CMP research group and CMP teachers not contributing to my study, I decided to use tasks that focused on one of the key seventh-grade CCSSM content standards, expressions and equations: "Solve real-life and mathematical problems using numerical and algebraic expressions and equations" (CCSSI, 2010, p. 47) (see Appendix F). Due to a lack of varied student work tasks focused on this standard in the assessment materials for each curriculum series and motivated by Shepard's (2000) statement that "good assessment tasks are interchangeable with good instructional tasks" (p.8), I gathered tasks from both instructional and assessment resources.

I selected two tasks from each of the four original curriculum series (Big Ideas, CMP, CPM, and Go Math) with one task that did not fit the Criteria for Student Work and one task that

did fit the Criteria for Student Work (see Appendix B). Table 3.1 below illustrates the variety of the eight tasks, based on the Criteria for Student Work and the evidence and critique types for student work (see Appendix B). As a reminder, all of the selected tasks focused on the CCSSM content strand of expressions and equations. As shown in the summary, Tasks B, C, F, and H were assessment tasks with embedded student work, or SWAT, and Tasks A, D, E, and G were assessment tasks that did not have embedded student work, or non-SWAT. Even so, two of these tasks, Tasks A and G, did fulfill Criterion #1, the inclusion of a person.

Table 3.1

<i>Features of the Eight Assessment Tasks Used in Clinical Interviews with Students</i>					
Tasks	Criterion #1	Criterion #2	Criterion #3	Evidence Types	Critique Types
A	X			--	--
B	X	X	X	Symbols	Error ID
C	X	X	X	Words, Symbols	Eval
D				--	--
E				--	--
F	X	X	X	Words	Eval
G	X			--	--
H	X	X	X	Words, Symbols, Actions	Eval, Compare, Pref

After I had selected the eight tasks that would be used in the student interviews, I constructed my plan for the interviews and determined how the selected tasks would be utilized. Each interview had three main components: (1) introduction, (2) clinical interview, and (3) semi-structured interview. For the introductory component, I asked students to tell me something interesting about themselves, describe their experiences learning math in school, and talk about their experiences taking math assessments or tests. The goal of these questions was to make students feel comfortable talking with me as well as gain insight into each student as a seventh-grade math student and a math assessment taker.

Clinical interview component. Once I had attempted to get to know each student with a few introductory questions, I introduced the clinical component of the interview (Ginsburg, 1981). I audio-recorded each full interview, but I also used an Echo pen for the clinical part of the interview. The Echo pen, which each student used to write any written work, collected another audio file of the clinical portion of the interview as well as video of the student's written work that was saved as a live-PDF document.

In order to have students feel comfortable using the technology, I first asked students to use the pen to "Draw a mathematician." As students were drawing their mathematician, I would ask students to share their thoughts aloud. This drawing task was not used as an explicit part of the clinical interview, but it did allow the students to get used to using the pen and the corresponding Echo-specific paper. This question also allowed students to practice thinking aloud while completing a task that did not involve specific mathematical content. After the mathematician drawing task, students were asked to complete the eight assessment tasks for the clinical part of the interview (see Appendix E).

I provided students with printed copies of each of the tasks as well as access to the Echo pen, writing paper, a scientific calculator, and a graphing calculator. The assessment tasks were not identified by curriculum series or any other feature, such as the inclusion of student work in the task. Also, the tasks were sorted before I assigned a letter to each task so that the student work tasks were not all together in order and tasks from the same curriculum series were not all one after the other. The eight assessment tasks were provided in a packet with one problem per page, so that the students had ample work space to show their written work.

After conducting the first student interview with Jane² in which neither she nor I read each task aloud at the beginning of each task, I then gave students the option for them or me to read the problem aloud. As anticipated, this made the verbal exchange and talking about the tasks aloud seem more natural as a flow of conversation in subsequent student interviews. For all the interviews, when a student would not talk for an extended period of time while completing a task, I would ask a probing question about what she was thinking, what was in her head, or clarification on some component of her written work. The goal of the clinical portion of the interview was to gain insight into students' thinking and processes while solving SWAT and non-SWAT. From the clinical interview, I was able to collect both students' written work on the assessment tasks as well as their verbal reasoning while completing these tasks as data sources for further analyses which provided two different representations of students' thinking on the assessment tasks – written and verbal.

Semi-structured interview component. Once students had solved all eight assessment tasks, I wanted to have students reflect on their experiences solving the assessment tasks and probe ideas related to the inclusion of student work in assessment tasks. The semi-structured portion of the full interview was guided by an interview protocol with specific questions I wanted to ask, but I allowed questions to emerge when appropriate (see Appendix E). Using the techniques for interviewing articulated by Glesne (2006) and Seidman (2012), interview questions focused on general ideas about students' experiences solving the assessment tasks and the similarities and differences the students saw among the various tasks. A few times during the semi-structured part of the interview, students would ask to revise their written work from the clinical portion of the interview. When this happened, I allowed students to revisit their work and

² All student names are pseudonyms.

we would resume the clinical portion of the interview using the Echo pen. Once they concluded their revisions, we would resume the semi-structured portion of the interview.

I also asked students about which tasks were easiest and hardest to solve, why these tasks were easy or hard for them, and whether or not features of the tasks made the student feel that a task was going to be more or less difficult. For one of the last questions of the semi-structured portion of the interview, I introduced the idea of tasks including a person's thinking and requiring the reader to make sense of it. I pointed out the specific tasks (Tasks B, C, F, and H), and asked for students to react to these types of tasks and their general noticing about these tasks and solving them. It was important for me to use the kid-friendly language of "a person's thinking" and "make sense" as opposed to the language of "critique to reasoning of others" used in SMP3b in order for students to understand the ideas I was talking about and feel comfortable expressing their own thoughts. I concluded the interview by allowing students to share any final thoughts about the assessment tasks.

Exclusion of the final two student interviews from the data analyses. In this study, I focus on six of the student interviews that were conducted. I omitted the last two interviews because of logistical constraints and the inability to complete the final interview. Because I conducted student interviews at the end of the spring semester, I had to be flexible in the timing of interviews. For one of the classrooms, the teacher was willing for me to meet with students one-on-one in the hallway during their math class as they were reviewing for final exams and she felt that the interview itself was a good way for students to prepare. For the second classroom, I interviewed students during their daily "homework" time in the school library. For about 30 minutes each day, students had an independent study time. For my last interview, which was conducted on the second-to-last day of school, I was only able to complete the clinical portion of

the interview. Due to logistics, I was unable to meet with this student, Jordan, again. I omitted Jordan's interview from the analyzed data set because I was unable to complete the semi-structured portion of the interview.

I also omitted the second-to-last interview conducted from these analyses where I met with a student, Katie. I conducted Katie's interview after I conducted my first teacher interview in which I had the teacher look at the student work of the first six student interviews. After reviewing Katie's work and seeing that her strategies for solving and her written work did not vary substantially from the written work of the other six students, I decided not to include her written work for future teacher interviews. I felt it was already taxing for teachers to make sense of six exemplars of student work in an interview setting. Because Katie's written work was omitted from the teacher interview, I chose to also omit her interview from the analyses of the student interviews.

Six student interviews. Six students from two schools participated in the interviews that were selected for analyses: Anna, Cynthia, David, Ed, Jane, and Susan. Three students were 12 years old (Cynthia, Jane, and Susan) and three students were 13 years old (Anna, David, and Ed) at the time of the interviews. All of the students were in a seventh-grade mathematics class where the teacher taught using CMP except for Jane, who was in an Algebra I class. Even so, Jane had taken the seventh-grade mathematics course the previous year with a teacher that taught using CMP. Students' remarks about their mathematical experiences in elementary school and their experiences on classroom and state level assessments are provided below in Table 3.2. This table details the duration and number of days for each interview as well as students' responses to explicit questions about their experiences in elementary and middle school mathematics classes

and classroom and state level assessments gathered during the first part of the interviews.

Student interviews occurred over one or two days depending on students' schedules.

Table 3.2

Background Information on Student Participants

Student Info	Class Remarks	Assessment Remarks
Anna 44 min. 1 day	Elementary: <i>"In elementary school, people thought of me as the smart one."</i> Middle school: 6th grade was <i>"kind of hard"</i> but <i>"I got the hang of it."</i>	Classroom: <i>"If it's a pre-test or something, I try to think of it like the logical way and see if that would work and then after I get my answer, I check it and see if it's right and then sometimes it was and sometimes it's not and so I just keep trying until I get it."</i> State: <i>"If I get something wrong on the test [in class], I know we're going to be learning it, so I'll get better at it. But on the MEAP and stuff like M-Step, it's just like, you can't change it and it's stuff you already know, so you can't learn it again."</i>
Cynthia 32 min. 2 days	Elementary: <i>"It was difficult. I was getting it slowly"; "multiplication and division were slow to me. It's easy or me now."</i> Middle school: <i>"kind of hard...equations are hard for me"</i>	Classroom: <i>"they don't always go well. If they don't I always redo them and get a better score. But doing homework, I get it done some days."</i> State: <i>"It's actually more easier a little bit ... cuz they're don't actually grading you, so it's a little less stressful I think"; "It's a little harder, but it's different from what we learn. It's more bigger and different tools on there."</i>
David 99 min. 2 days	Elementary: <i>"math has always kind of been easy.";</i> <i>"I think it's hereditary because my grandma is faster than a calculator and my dad is pretty good at math."</i> Middle school: <i>"It's a lot more challenging.";</i> <i>"Trying to work is a little more common.";</i> <i>"The problems require more thinking and it takes me a little bit longer to do."</i>	Classroom: <i>"Taking assessments and tests are honestly my favorite days because it doesn't take me very long to do them and they are pretty easy for me."</i> State: <i>"The M-Step, now, it's quite challenging actually because usually when we do tests and quizzes, we have calculators available, but on the M-Step, there are only certain questions that have calculators. It really requires a lot of thinking and especially with the math M-Step, you're not allowed to go back. I feel like that's a smart idea, because you would go back to a question with a calculator and just use it. I feel like the M-Step is a lot more difficult than regular tests and assessments."</i>
Ed 76 min. 2 days	Elementary: <i>"I coasted until about fourth grade and then I started noticing that the math got harder and I had to pay attention more.";</i> <i>"It was basically just counting and simple math and my mom basically taught me that already."</i> Middle school: <i>"I had a C+ in math, but I fixed that."</i>	Classroom: <i>"I don't want to say easy questions, because when I say that, I mean they're easy for me, not necessarily for everybody else.";</i> <i>"... tests are just some stuff we learned recently so I have that fresh on my mind."</i> State: <i>"On the M-Step, it can be stuff all the way to the very beginning of the year. I'm honestly not that good at remembering things."</i>

Table 3.2 (cont'd)

Jane 57 min. 1 day	Elementary: <i>"It was really easy, but it was fun too."</i> Middle school: <i>"It's still kind of easy, so I got put in Algebra I."</i>	Classroom: <i>"They're not really all that stressful for me."; "The lowest I've gotten is around a C."</i> State: <i>"It was kind of annoying. It doesn't let you go back and look at the answers, but other than that it was fine."; "My math teacher lets us use calculators on quizzes and stuff, but on the M-Step, some of the questions use calculators and some of them, you can't."</i>
Susan 46 min. 1 day	Elementary: <i>"I was not that good at math, but when I hit 5th grade, it started to click."</i> Middle school: <i>"I do get confused around some problems, but then I can just figure it out if I knew a little bit about it and then that helps."</i>	Classroom: <i>"Tests have been pretty easy just because I pay attention more."</i> State: <i>"Those they have from the beginning of the year to the end, but I think those are kind of easy."</i>

For students in Ms. Shirley's class (Anna, David, Ed, and Susan), I conducted interviews with students during their math class time that lasted about 50 minutes. For students in Ms. Christy's class (Cynthia, Jane, and the two interviews that were omitted), I conducted interviews with students during their school-wide study group time that lasted about 30 minutes. However, Jane was able to complete her interview after school, so she only required one day to complete the full interview. For interviews that occurred over two days, I asked students to review their assessment work from the previous day before proceeding with the interview. The interviews lasted an average of 59 minutes with the shortest interview taking 32 minutes and the longest interview taking 99 minutes.

The structure of the student interviews allowed me to gain insight into: (1) students' mathematical learning and assessment experiences, (2) students' engagement with assessment tasks with and without student work, (3) students' descriptions of their experiences with these assessment tasks and their viewpoints on making sense of a person's thinking on assessment tasks. After interviews were conducted, I transcribed the six selected audio recorded interviews as Word documents. I transcribed the audio files collected using my audio recorder, but also had

the Echo pen audio file from the clinical portion of the interview as back-up, as needed. The Echo pen written student work files were used to capture digital copies of students' written work from the interview. These written work copies were compiled by task with students' pseudonyms as labels for each exemplar of written work. I analyzed students' written work on the assessment tasks as well as transcripts of the student interviews in order to answer my research questions for this part of my study.

Data Analysis

I used qualitative methods to conduct thematic analyses of transcriptions of the student interviews (Corbin & Strauss, 2008; Glesne, 2006). Each student interview was divided into three phases previously described: (1) introduction, (2) clinical interview, and (3) semi-structured interview. For two of these phases, the introduction and the semi-structured interview, I constructed summary tables by student and topics. Topics involved the key questions I asked during the interviews. For example, when I asked students about their general experiences solving the tasks, I included General Experience as a row and students' names as columns. For these two summary tables, I summarized main ideas and preserved meaningful quotations from each student for each interview question or topic. The summary table for the introduction component of the student interviews was used to build a mathematics classroom and assessment experiences profile for each student (see Table 3.2). The summary table for the semi-structured interview component of the student interviews was used to answer Research Question #2b.

To answer Research Question #2a, I analyzed students' written work on the assessment tasks as well as their verbal responses during the clinical interview for each task. I analyzed both students' written work on the tasks based on the words, symbols, and visuals students included in their written responses and students' verbal responses based on the words students used to

describe their mathematical thinking. My analysis of students' written work in terms of the student thinking evident in students' written responses showed the types of evidence of student thinking students chose to represent in writing on these assessment tasks. When analyzing students' verbal responses, I specifically focused on the thinking students expressed verbally but did not represent in any way in their written work. This allowed me to capture the nature of another type of representation of student thinking in addition to students' written work as well as allowed me to think about students' responses to assessment tasks allowing for multiple modes of communicating and representing solutions.

The goal of focusing on the student thinking expressed verbally but not in written work was to document all of the student thinking elicited by the different tasks while avoiding redundancy. Therefore, comparison between students' written and verbal work allowed me to capture the full-picture of students' thinking on the tasks while accounting for different representations of student thinking. Taken together, the two types of representations of students' thinking allowed me to analyze the types of evidence of thinking students provided in writing (words, symbols, and visuals) in addition to what students orally provided in their verbal reasoning (words).

To capture the types of evidence of student thinking students provided in their written responses, I reviewed students' written work for the eight tasks iteratively until I had an idea of the range of different types of evidence of thinking students included in their written responses. Once I had developed codes for the types of evidence of student thinking students expressed in their written solutions, I coded students' written work by task and by student. The codes, descriptions, and examples for the evidence of student thinking in students' written responses are provided in Table 3.3 below.

Table 3.3

Evidence of Student Thinking in Students' Written Responses

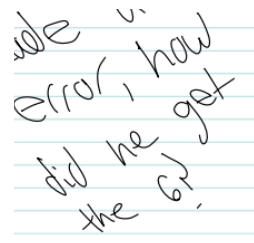
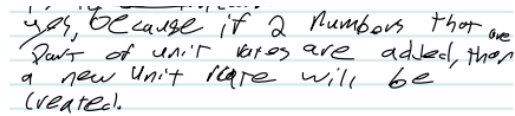
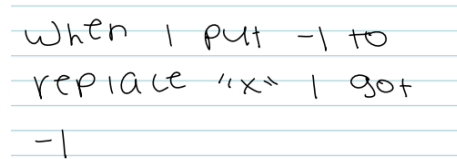
Code	Description	Examples
Drawing	Student included a visual or diagram.	 (Anna, Task D)
Question	Student wrote a question.	 (Cynthia, Task B)
Reasoning	Student included a statement involving a solution or mathematical idea and provided reasoning or justification for the statement expressed using words.	 (David, Task F)
	The use of the Reasoning code assumes the existence of a statement. Do not double code with Statement for the same instance of written work.	
Statement	Student included a statement involving a solution or mathematical idea using words.	 (Susan, Task C)

Table 3.3 (cont'd)

Symbolic Work Student used symbols to detail computations involving operations and/or variables.

For this code, the work extended beyond rewriting the symbols provided in a task to adapting or adding onto the symbols in some way.

(Ed, Task E)

Uncertainty Student expressed uncertainty using words.

side by 3 and I'm 'not sure why.

(Anna, Task H)

My analyses of students' written work only focused on the components of their work that did not require a great deal of interpretation or inference in terms of the types of evidence of student thinking demonstrated. For example, I did not include analysis of markings such as circling of written components or marking out of written work. Many students circled parts of their written work. I could have assumed that this meant the student thought that these written components were important or indicated some sort of answer. However, because that would require an assumption on my part about the purpose of the marking, I did not include circling of written components as a specific type of student thinking. Similarly, a few students marked out parts of their written work. In reviewing this marked out work, I could possibly assume that students revised their thinking. However, because I would have to assume that this was the students' purpose in making the marking, I did not include marked out written work as evidence of revised thinking. Instead, I focused my analyses of written work on words, symbols, or visuals that I could consistently interpret. Because I talked to students about their work as part of the student interviews, I was purposeful in my coding of these tasks to only base my coding

decisions on what students provided on the page in the form of words, symbols, or drawings. I was interested in characterizing the types of evidence of student thinking that existed in students' written work, not the frequency of the types of evidence of student thinking at the task level. Therefore, each task for each student only received at most one instance of each code. Consequently, each of the eight tasks had the possibility of receiving six of the codes for each of the six students.

In order to determine the additional student thinking represented in students' verbal reasoning but not in writing on the tasks, I created a different summary table for the clinical interview phase of the student interviews that included transcription segments from the clinical interview and students' written work, side by side, by task for each student. Then, by making comparisons between the students' verbal remarks, as found in the transcription segments, and students' written work, I was able to determine the ideas and evidence of their thinking from students' verbal remarks that were not evident in any way in their written work. From this process, I created a secondary summary table that detailed, by student and by task, the student thinking that occurred in the clinical part of the interview based on the transcript that was not evident in students' written responses. The contents of this summary table included both direct quotations and summarized ideas for more lengthy components of students' responses.

Once I had collected these instances of student thinking, I coded them using emerging codes based on the mathematical ideas and language students were using to express their thinking verbally. These codes, descriptions, and some examples are provided in Table 3.4. I was curious about the types of evidence of student thinking that existed in students' verbal descriptions in comparison to their written work, so I did not focus on the frequency of the types of evidence of student thinking at the task level. Instead, each task for each student could only

receive one instance of each code. Therefore, each of the eight tasks had the possibility of receiving nine of the emerging codes for each of the six students.

Table 3.4

<i>Evidence of Student Thinking in Students' Verbal Responses Not Found in Written Responses</i>		
Code	Description	Example
Claim	Student stated a mathematical idea without justification (see below).	<i>"The sum is addition."</i> (Cynthia, Task F)
Justification	Student expressed reasons for a mathematical claim, idea, conclusion, answer, or solution. The use of the Justification code assumes the existence of a claim. Do not double code with Claim for the same instance of student thinking.	<i>"Terri and Brian's solutions took a lot of elaborate steps and thinking and stuff, but Jesse's was just like solving it instead of Brian got $2/3 + 1/3 = 1$."</i> (Jane, Task H)
Operation/Method	Student described operations or methods when solving a task.	<i>"120 divided by 60 is 2."</i> (Anna, Task A)
Plausibility	Student expressed suspicion as to the reasonableness of a solution.	<i>"These people better be like small dogs or something because if people weighed that much you should probably see a doctor."</i> (Ed, Task G)
Prior Experience	Student mentioned previous mathematics or classroom experiences when completing a task.	<i>"I've never come across a trick question like that before on the tests I've taken."</i> (Ed, Task E)
Questioning	Student asked questions specific to the task.	<i>"What does it mean to be equivalent? Does that mean which one is not the same?"</i> (Cynthia, Task E)
Revised Thinking	Student expressed a change in thinking or conclusions reached.	<i>"There are like $3x$ and $5x$ and $2x$."</i> Then later, Ed stated, <i>"There's another x right here. So, it would be $11x$."</i> (Ed, Task D)

Table 3.4 (cont'd)

Tools	Student described using a tool for solving a task.	<i>"I do need to use a calculator for that."</i> (David, Task E)
Uncertainty	Student expressed confusion about a task.	<i>"I'm not sure how this problem looks."</i> (Anna, Task C)

I compiled the results of coding for the evidence of student thinking in students' written responses and the evidence of student thinking in students' verbal responses not represented in written responses and made comparisons between my coding results for SWAT and non-SWAT. Because I was making sense of the nature of students' written work based on words, symbols and visuals and students' verbal reasoning based only on words, the types of evidence of student thinking I observed in written form were often similar, although not identical, to those described verbally due to differences in the mediums of written and oral representations. The larger goal of these analyses, as well as the comparisons between coding results for SWAT and non-SWAT, was to be able to characterize and make sense of students' written and verbal representations of their thinking as a means for considering whether or not SWAT could be used to assess SMP3b.

In order to answer Research Question #2b, I reviewed the summary table for the semi-structured interview by comparing within and across student responses. To analyze the summarized main idea and meaningful quotations in the summary table for students' descriptions of SWAT based on their experiences solving SWAT, I used qualitative data analysis methods (Corbin & Strauss, 2008; Glesne, 2006) for developing a coding scheme and codebook based on thematic coding and emergent themes. For these analyses, I specifically focused on parts of the semi-structured interview when students talked explicitly about the SWAT and the summary table row that reflected when students were asked an explicit question about tasks that provided

information about a person's thinking and asked the student to make sense of it. Students were asked their thoughts on these types of tasks, how these tasks compared to the other tasks they solved, and anything that they noticed about the tasks. By analyzing these interview portions using thematic coding, I wanted to capture key ideas about how students perceived SWAT and their experiences solving SWAT in comparison to non-SWAT.

From thematic analyses, two main codes emerged: (1) SWAT as Different and (2) Attributes of SWAT. Four students talked about SWAT as requiring different mathematical processes and/or different final products on the assessment tasks. All six students identified key attributes or features of the SWAT. Within the Attributes to SWAT theme, five sub-codes emerged: (1) Complexity, (2) Connection to Classroom, (3) Imagination, (4) Comparison, and (5) Motivation. The descriptions of these codes are provided in Table 3.5 below along with the number of students out of the total six students that addressed each theme in their semi-structured interview responses.

Table 3.5

Emerging Themes from Students' Experiences with SWAT

Codes	Descriptions	Number of Students
SWAT as Different	Student described SWAT as requiring different processes and/or different final products in comparison to other assessment tasks.	4
Attributes of SWAT	Student highlighted attributes or features of SWAT.	6
Complexity	Student described how SWAT were more complex or more complicated than other assessment tasks.	6
Connection to Classroom	Student described how SWAT were similar to the types of tasks she had seen in class or connected to practices from her mathematics classroom.	3

Table 3.5 (cont'd)

Imagination	Student talked about how SWAT required imagination. This extended beyond thinking to include an explicit mention of the words imagine or imagination.	3
Comparison	Student described how SWAT required comparison between someone's work and her work or someone's thinking and her thinking.	2
Motivation	Student described SWAT as motivating types of tasks.	1

From this process, I was able to look for unifying, emerging themes across subsets or the full set of students in my study as related to ideas that emerged from students' written work and their verbal descriptions of SWAT and non-SWAT. Taken all together, the types of thematic analyses I conducted on the student interview transcripts and their written work on SWAT allowed me to answer my research questions related to (1) students' written work on SWAT and non-SWAT as compared to students' verbal responses and (2) students' experiences and perspectives on SWAT. Findings from the analyses of the student interview transcripts and students' written work are detailed in Chapter 5.

RQ #3: Exploring Teachers' Experiences with and Perspectives on SWAT and Students'

Work on SWAT

In this section, I detail my methods for answering the following research question:

3. How do teachers talk about and understand SWAT and non-SWAT based on the tasks and students' written work on the tasks?
 - a. What do teachers think that SWAT and non-SWAT assess?
 - b. How do teachers describe SWAT?

- c. What evidence of SMP3b do teachers notice in SWAT based on both the written assessment tasks as well as students' written work on the tasks?

To explore teachers' experiences with and perspectives on SWAT and students' written work on SWAT, I conducted semi-structured interviews with teachers that involved teachers talking about and making sense of SWAT and non-SWAT curriculum-based assessment tasks as well as actual students' work on the assessment tasks gathered from the student interviews I described in the previous section. Because teachers are key users of information gathered from curriculum-based assessments, it was important for me to gain insight into their views on embedded student work as a mechanism for assessing SMP3b on curriculum-based assessments. In the following sections, I describe the data collection and analysis methods used for the teacher interviews conducted for this part of my study.

Data Collection

I originally recruited teachers to participate in my study at the CMP Users' Conference during spring of 2017. The CMP Users' Conference is a two-day conference that includes 1-hour sessions about CMP problems, grade levels, big ideas, or teaching strategies presented by the curriculum writers, graduate students, and master CMP teachers. I reached out to both conference presenters and conference attendees. Also, teachers were solicited by email for participation in my study during the spring and summer of 2017. I wanted participants to be teachers who used the seventh-grade CMP3 curriculum materials. First, I recruited seventh-grade teachers, because seventh-grade was the focus of earlier curriculum-based assessment text analyses. Second, I recruited CMP teachers, because based on previous analyses (Gilbertson et al., 2016; Going, Ray, & Edson, in preparation; Nimitz et al., 2015), student work was found throughout the seventh-grade student textbook materials for CMP3. Therefore, CMP teachers

would at least be exposed to tasks with embedded student work and would not be surprised by these types of tasks, as opposed to teachers who used curriculum materials with a more traditional approach towards mathematics teaching where student work might rarely be used in the student textbook. Also, I had worked on the CMP project for four years as a graduate student researcher, so I had access to seventh-grade CMP teachers because of my work on the project.

Six teachers from five different middle schools in Arizona, Illinois, and Michigan participated in my study: Ms. Edwards, Ms. Gables, Ms. Gilbert, Ms. Henderson, Ms. Quinn, and Ms. Shirley³. Some of the background information for the six teachers is summarized below based on information gathered during the first part of the semi-structured interviews (see Table 3.6). Teachers were emailed a consent form that provided an explanation of the research, the participant's rights to participate, and contact information (see Appendix G). Data from this part of the study came from semi-structured interviews with individual teachers about the curriculum-based assessment tasks and students' written work from the student interviews described in the previous research question methods section. Interviews were conducted individually and at the convenience of the participants based on time, location, and access to technology for online interviews.

Table 3.6

Background Information on Teacher Participants

Teacher	Years of Teaching	Grade(s) Taught	Current Grade(s)	Additional Teacher Information
Edwards	18	7th	7th	Certified to teach all subjects Has also taught science, social studies, and communication Has always used CMP (started with CMP1, used CMP2, and now uses CMP3)

³ All teacher names are pseudonyms.

Table 3.6 (cont'd)

Gables	15	5th – 8th	7th, 8th	Math coaching experience Has taught with CMP for 7 years Regularly leads CMP professional development sessions
Gilbert	6	7th, 8th	7th	Has only taught using CMP Started teaching as a long-term sub for 1 year
Henderson	32	6th – 8th	7th, 7th gifted and talented	Has always used CMP (started with MGMP, used CMP1, used CMP2, and now uses CMP3) Regularly leads CMP professional development sessions
Quinn	23	6th – 12 th	7th	Taught high school for 12 years The high school curriculum was written by teachers and involved story lines, similar to CMP Has used CMP for 6th and 7th grade
Shirley	20	6th – 12 th	7th	Experience as administrator at a district central office Taught elementary and secondary mathematics methods courses at local university as a graduate student between teaching and administrative work Experiences as principal at elementary, middle, and high schools

I conducted interviews with the six teachers during the summer and fall of 2017. Two interviews were conducted in person (Quinn and Shirley) and four interviews were conducted online using the Zoom video conferencing platform and Google docs containing supporting materials (Edwards, Gables, Gilbert, and Henderson). The interviews were guided by an interview protocol with specific questions I wanted to ask, but I allowed questions to emerge when appropriate (see Appendix H). Using the techniques for interviewing articulated by Glesne (2006) and Seidman (2012), interview questions focused on the teacher's own understandings of the assessment tasks (see Appendix F) and her analysis of the compiled student work (see Appendix I). All interviews were audio recorded using the Zoom conferencing platform. The interviews lasted an average of 1 hour and 30 minutes with the shortest interview lasting 1 hour

and the longest interview lasting 1 hour and 50 minutes. For each interview, there were six categories of questions that flowed from the beginning to the end: (1) background information, (2) analysis and discussion of assessment tasks, (3) focus on assessment tasks with student work, (4) analyzing students' written work on assessment tasks, (5) revisit assessment tasks with student work, and (6) wrap-up questions.

For the background information, I wanted to gain a snapshot of each teacher's teaching background and their assessment practices. Then, for the second question category, I introduced teachers to the eight assessment tasks described previously in the student interviews part of the study and probed their general reactions, their understandings of what each task was assessing, and similarities and differences the teachers noticed across the tasks or subsets of tasks.

Mirroring the student interviews, assessment tasks were not identified by curriculum materials source or initially whether or not the task included student work as defined by the adapted criteria. For the in-person interviews, teachers were provided with physical copies of the tasks with one task per sheet of paper. For the Zoom interviews, teachers were emailed a link to a Google doc that contained each task on a different page. As the teachers described what each task was assessing, we kept a record of their ideas. For the in-person interviews, either the teacher or I jotted down ideas on the physical sheets of paper that included each task. For the interviews conducted via Zoom, I kept a record of the teacher's ideas on a Google doc that was visible to both me and the teacher. The goal of this was not to keep a record for data analysis, but to have a record to refer back to later in the interview once each teacher had a chance to view students' work on the tasks.

I then explicitly introduced, but did not explain, the concept of *student work* to the teachers and highlighted the tasks that fit the Criteria for Student Work from the eight tasks

(Tasks B, C, F, and H). I asked the teachers what they thought I meant by student work, the advantages and disadvantages of these types of tasks, and whether or not they felt that assessment tasks with student work could assess SMP3b. For the next phase of the interview, I wanted to expose each teacher to students' actual work on the eight assessment tasks. However, I knew that putting six different sets of students' work for eight tasks could be overwhelming and I wanted to make sure the interview was not extremely long or taxing on my participants. Also, because I was interested in making sense of student work embedded in assessment tasks as a mechanism for assessing SMP3b as opposed to teachers' thoughts about the specific assessment tasks, I chose to let teachers tell me which tasks they were most and least interested in looking at students' written work for and reasons why as a way to focus my questioning and respect their time. Once each teacher told me which tasks they were most and least interested in looking at, I selected at least two tasks for us to discuss. I made sure that each teacher talked about at least one task with student work and one task without student work. Because Task F seemed to be different from the three other student work tasks (Tasks B, C, and H) as it involved someone's claim as opposed to someone's symbolic work as pointed out by and often selected by teachers in the first couple of interviews, I made sure that each teacher discussed Task F in this phase of the interview.

For the in-person interviews, I provided teachers with copies of the students' written work. For the interviews conducted via Zoom, teachers clicked a link at the bottom of the assessment tasks Google doc that opened up a PDF document containing the students' written work copies for each task. I prompted each teacher to tell me what they noticed and wondered about when looking at students' work on the selected tasks as well as anything they found surprising. We revisited each teacher's ideas from earlier in the interview about what each task

was assessing to determine if any additions or edits needed to be made to the list of ideas after looking over students' written work. For the next phase of the interview, we focused specifically on the four tasks with embedded student work (Tasks B, C, F, and H). I asked each teacher whether or not they saw evidence that students critiqued the reasoning of others based on students' written work. I concluded the interview by asking each teacher which tasks, if any, she might use as well as how and why she would use them. I also gave each teacher an opportunity to share any final thoughts about the tasks, students' work on the tasks, or the notion of embedding student work in assessment tasks. I told teachers they could keep the assessment tasks documents as well as the students' written work on the tasks documents as resources since student pseudonyms were used to identify students.

The structure of the teacher interviews allowed me to gain insight into: (1) whether or not teachers initially felt that assessment tasks could assess SMP3b, (2) their thoughts about student work embedded in assessment tasks as a mechanism for assessing SMP3b, and (3) what would be required as evidence in students' solutions in order for students to show mastery of SMP3b. After interviews were conducted, I transcribed each interview as a Word document. I analyzed these transcripts in order to answer my research questions for this part of my study.

Data Analysis

I used qualitative methods to conduct thematic analyses of transcriptions of the teacher interviews (Glesne, 2006). Each interview transcription was divided into the six phases previously described: (1) background information, (2) analysis and discussion of assessment tasks, (3) focus on assessment tasks with student work, (4) analyzing students' written work on assessment tasks, (5) revisit assessment tasks with student work, and (6) wrap-up questions. For the first round of thematic analyses, I created summary tables for each interview phase by

teacher and topics. Topics involved the key questions asked or specific tasks discussed during the interviews. For example, when I asked teachers what each task was assessing, the summary table for this phase of the interview included Tasks A, B, C, etc. as rows and teachers' names as columns. For the summary tables, I summarized main ideas and preserved meaningful quotations from each teacher for each interview question or topic. This resulted in six summary tables for the six phases of the teacher interview that included summarized responses and key ideas from the six teachers.

I reviewed the summary tables by comparing within and across teachers' responses. The first summary table of teacher background information helped me gain insight into each teacher's background based on their descriptions of their teaching and assessment. The summary tables for (2) analysis and discussion of assessment tasks, (3) focus on assessment tasks with student work, (4) analyzing students' written work on assessment tasks, and (5) revisit assessment tasks with student work were used to answer my research questions, which involved teachers' thoughts about what the tasks were assessing, teachers' descriptions of SWAT such as the advantages and disadvantages of SWAT, and the evidence of SMP3b teachers noticed in SWAT.

To answer Research Question #3a, I revisited the summary table for teachers' analysis and discussion of assessment tasks. I specifically looked at the rows in the summary table that captured teachers' responses to the question, "What is each task assessing?" In order to make comparisons across tasks and teachers, I identified categories of what tasks assessed based on the summary and key quotations pulled from teachers' descriptions. If I felt I needed more information in order to determine which categories were appropriate for a specific task and teacher, I revisited the interview transcripts and reviewed teachers' original descriptions. Each

teacher description by each task could receive from one to four category codes, but codes were not repeated within tasks. For example, teachers' descriptions of an individual task could not receive multiple Content codes. Table 3.7 below details the category codes and descriptions and provides examples from the teachers' descriptions that illustrate each category of what the assessment tasks assessed.

Table 3.7

<i>Categories of Teachers' Descriptions of What Assessment Tasks Assessed</i>		
Category Code	Category Description	Example
Content	Teacher described mathematical content assessed on a task.	<i>"It could be assessing them writing an equation and then solving an equation."</i> (Edwards, Task A)
Strategies	Teacher elaborated on mathematical content by describing specific strategies or solution pathways students might use to complete a task.	<i>"Kids could make a table. They could do guess and check."</i> (Edwards, Task A)
Related to SMP3b	Teacher mentioned SMP3b, "critique the reasoning of others" (CCSSI, 2010, p. 6), SMP or mathematical practices, or related habits of mind such as error analysis or analyzing other's work. For these instances, preserve the language and/or individual ideas from each teacher (i.e. SMP3b, SMP3, error analysis, critiquing, etc.)	<i>"This is also assessing error analysis because it's still saying whether she did it right or wrong, it's just not telling you that she made an error, so you have to figure it out."</i> (Henderson, Task C)
<i>Other</i>	Teacher described ideas that do not fit into the other three categories. For these instances, preserve the language and/or individual ideas from each teacher.	<i>"So, the piece where it says, 'Explain.', so they have to be able to justify their thinking."</i> (Quinn, Task E)

I created a matrix that included the categories of teachers' descriptions of what the assessment tasks assessed by teacher and by task. Within the boxes of the matrix, I listed the categories discussed by each teacher. In the matrix, I shaded the boxes that corresponded to SWAT, I

bolded ideas that fit with the “Related to CMP3b” code, I italicized ideas that fit with the “Other” category in order to make comparisons between and across tasks based on these attributes. From these analyses, I was able to investigate teachers’ perceptions of what assessment tasks could assess as well as possible differences between SWAT and non-SWAT based on teachers’ understandings of what these tasks were assessing.

To answer Research Question #3b, I completed two different kinds of thematic analyses in order to explore teachers’ descriptions of SWAT. First, I revisited the summary table for the interview questions that focused on assessment tasks with student work. I gathered the summarized teachers’ descriptions of the advantages and disadvantages of SWAT based on their responses to the explicit interview questions: “What might be advantages of using [SWAT] in assessments?” and “What might be the disadvantages of using [SWAT] in assessments?” I looked for common themes of advantages and disadvantages of using student work in assessment expressed by teachers and represented these themes in a chart.

In order to provide a more thorough answer to Research Question #3b that extended beyond reporting on teachers’ answers to direct questions, I conducted additional thematic analyses. I used qualitative data analysis methods (Corbin & Strauss, 2008; Glesne, 2006) for developing a coding scheme and codebook based on thematic coding and emergent themes. I revisited the original teacher transcripts and specifically focused the secondary analyses on portions of the teacher interviews where teachers were discussing the SWAT (Tasks B, C, F, and H) or student work in assessment tasks more generally across all six phases of the teacher interviews. From secondary thematic analyses, four key codes emerged: (1) Interdisciplinary Practices, (2) Connections to Testing, (3) Connections to Curriculum/Classroom, and (4) Teachers as Assessment Writers. The descriptions of these codes are provided in Table 3.8 below

along with the number of teachers out of the total six teachers that addressed each theme in their interview responses.

Table 3.8

Emerging Themes from Teachers' Experiences with and Perspectives on SWAT

Code	Description	Number of Teachers
Interdisciplinary Practices	Teacher discussed how student work related to content or practices promoted in other school subjects besides mathematics.	2
Connections to Testing	Teacher explicitly referenced state tests, standardized tests, or national exams when discussing student work.	4
Connections to Curriculum/Classroom	Teachers explicitly referenced curriculum materials, CMP, or classroom practices when discussing student work.	6
Teachers as Assessment Writers	Teacher talked about adaptations, revisions, or additions she would make to assessment tasks with embedded student work and/or discussed how she would use these tasks in her own teaching or assessment practices.	6

From this process, I was able to look for unifying themes across subsets or the full set of teachers in my study as related to ideas that emerged from teacher talk about student work embedded in assessment tasks. Portions of the teacher interviews that were coded with the four codes corresponding to the four emerging themes were compiled into a different summary table by code and teacher. I reviewed the summary table of key themes by teacher and across teachers.

In order to answer Research Questions #3c, I focused on teachers' interactions with SWAT and traced whether or not teachers used language about SMP3b or related habits of mind to describe and discuss SWAT. I traced teachers' talk about SWAT over three different segments. First, I looked at teachers' general reactions to the assessment tasks based one of the first question I asked teachers during the interview: "What do you think about these tasks?" I

also returned to my analyses conducted for Research #3a and #3b focused on teachers' descriptions of what tasks were assessing and the advantages and disadvantages of student work. These findings revealed different ways in which teachers referenced SMP3b or related habits of mind before I explicitly introduced student work as a possible mechanism for assessing SMP3b. Then, I reviewed and summarized teachers' responses to the question: "Do these four tasks assess this practice?" in reference to the four SWAT tasks from the teacher interview and the practice of SMP3b. These analyses revealed the extent to which teachers felt that student work embedded in assessment tasks could serve as a mechanism for assessing SMP3b based solely on the written tasks.

Lastly, I revisited the summary table for when teachers revisited the assessment tasks for student work and were asked, "Do you see evidence that students critiqued the reasoning of others based on their written student work?" and reviewed teachers' responses. These analyses revealed the extent to which teachers maintained their original thoughts about whether or not student work embedded in assessment tasks could serve as a mechanism for assessing SMP3b based on students' written work on the task and the evidence of student thinking that students included in their written responses. Looking across the analyses conducted for these three segments of the interview in order to answer Research Question #3c, I was able to trace teachers' thoughts on student work as a mechanism for assessing SMP3b based on their reactions to the tasks initially, my hypothesis about student work as possible way to assess SMP3b, and students' written work on the tasks.

Taken all together, the thematic analyses I conducted allowed me to answer my research questions related to teachers' experiences with and perspectives on SWAT. Findings from the analyses of the teacher interview transcriptions are detailed in Chapter 6.

Connecting Study Methods to the Stages of ECD

The previous sections detailed the methodologies for my study on whether or not student work can serve as a mechanism for assessing SMP3b. Within the context of curriculum-based assessments, I detailed the data sources for my study including: (1) curriculum materials, (2) interviews with students, and (3) interviews with teachers. In the following paragraphs, I highlight how the methodologies I utilized in my study align closely with tenets of ECD, an assessment design framework detailed at the beginning of this chapter (Mislevy & Haertel, 2006; Mislevy & Riconscente, 2005; Mislevy, Steinberg, & Almond, 2003). As a reminder, ECD views assessment as a process of examining evidence in certain circumstances in order to make claims about what students know or can do.

While I did not actually design assessment tasks, I did complete specific stages in order to analyze a specific type of assessment task, assessment tasks with embedded student work, that I posited could assess a specific domain, SMP3b. To show how my study aligns closely with ECD, I map stages or components of my study to the stages of ECD which include: *domain analysis* (information about what is being assessed), *domain modeling* (elaborating beyond the preliminary domain information to consider types of proficiency, characteristics, and potential attributes), *conceptual assessment framework* (task specifications), *assessment implementation* (the actual tasks), and *assessment delivery* (administered and completed tasks).

Domain Analysis – SMP3b

According to Mislevy and colleagues (e.g. Mislevy & Haertel, 2006), the first stage of ECD is domain analysis. Domain analysis involves “gathering substantive information about the domain to be assessed” (p. 7). The authors referenced standards documents as a possible resource for this stage. Also, resources gathered during the domain analysis phase “can have important

implications for the assessment, but most of it [will be] neither originally created nor organized in terms of the structures of assessment” (Mislevy, Steinberg, & Almond, 2003, p. 10). This phase involves gathering materials and resources in order to determine a specific domain to be assessed but does not yet extend to creating original content.

Unlike typical assessment task design that involves creating tasks that assess a broad set of skills or proficiencies, my study focused specifically on one component of the eight CCSSM SMPs, SMP3b: “critique the reasoning of others” (CCSSI, 2010, p. 6). In other words, SMP3b served as the domain for the assessment tasks I analyzed in this study.

Domain Modeling –Unpacking and Making Sense of SMP3b

The second stage of ECD is domain modeling. In this stage, assessment designers “lay out what an assessment is meant to measure, and how and why it will do so, without getting tangled in the technical details that will eventually be necessary” (Mislevy & Haertel, 2006, p.8). The domain modeling stage involves gathering additional information about the specific domain in terms of claims about student proficiencies, evidence students could provide, and possible situations or tasks that could elicit this evidence. According to Mislevy, Steinberg, and Almond (2003), “the focus at this stage of design is the evidentiary interrelationships that are being drawn among characteristics of students, of what they say and do, and of task and real-world situations in which they act” (p. 10).

The domain modeling stage was reflected in my study as I began to build up an understanding of SMP3b as a specific habit of mind by diving into the research literature related to this practice (CCSSI, 2010; Koestler, Felton, Bieda, & Otten, 2013; Hull, Miles, & Balka, 2012; Hunsader et al., 2013; 2014; NCTM 2000; 2014; NRC, 2001; Seeley, 2014). Other standards documents and seminal works provided more information about proficiencies closely

related to critiquing someone else's mathematical thinking and what situations would be conducive for eliciting evidence of and assessing SMP3b. The language of the actual practice for SMP3b indicated that critiquing involved responding to others, asking questions, and/or making sense of an argument. The Communication Strand from NCTM's *Principles and Standards for School Mathematics* (2000) discussed how students should be able to "analyze" and "evaluate" the "thinking" and "strategies" of others (p. 268). The National Research Council (NRC) detailed how students should be able to reflect on, explain, or justify their critique of another person's thinking. Therefore, critiquing extends beyond analyzing and evaluating to being able to communicate these practices in a meaningful way using justification, explanation, and/or proof.

Hull, Miles, and Balka (2012) provided a range of three proficiencies for SMP3b starting from the ability to "Understand and discuss other ideas and approaches" (Initial), to "Explain other students' solutions and identify strengths and weaknesses of the solutions" (Intermediate), and to "Compare and contrast various solution strategies, and explain the reasoning of others" (Advanced) (p. 52). These authors highlighted different degrees to which students could provide evidence of their engagement with SMP3b. It should be noted, the authors discussed these practices as evident in classroom interactions, not on written assessments, which is the focus of my study.

I also considered how previous studies had examined the potential for curriculum-based assessments to assess processes and practices (Hunsader et al., 2013; 2014). Hunsader and colleagues (2013; 2014) created a framework for analyzing curriculum-based assessments that focused on particular features of written tasks in order to determine the potential for students to engage with the NCTM process standards. My study extends their work as I focused on SMP3b and I not only analyzed curriculum-based assessments materials but also student textbooks. I also

explored students' and teachers' experiences engaging with SWAT. In making sense of SMP3b by visiting the research literature on this practice and related habits of mind, I began to build an idea of the proficiencies, student evidence, and task features that would be necessary in an assessment task that would assess SMP3b. However, since this was an exploratory validity study, I also viewed the study as a way to gain insight into the proficiencies, student evidence, and task features that would be necessary for an assessment of SMP3b.

Conceptual Assessment Framework – Criteria for Student Work

According to Mislevy and Haertel (2006), the next stage of ECD, the conceptual assessment framework, “concerns technical specifications for the nuts and bolts of assessments” (p. 10). This stage “lays out the blueprint for the operational elements of an assessment” (Mislevy, Steinberg, & Almond, 2003, p. 10). Building from previous iterations of criteria for student work in curriculum materials (Gilbertson et al., 2016; Going, Ray, & Edson, in preparation), I further refined the Criteria for Student Work to analyze the instances of student work in curriculum-based assessments. The three criteria included: (1) the task involves a person, (2) the task presents evidence of the person’s mathematical thinking, and (3) the task requires the reader to engage with the person’s mathematical thinking in some way (see Appendices A or B for a more detailed description of each criterion). Tasks that fit all three criteria were considered student work assessment tasks, or SWAT (see Appendix A for a detailed description of the changes that were made through these iterations). In my study, I viewed the Criteria for Student Work (see Appendices A and B) as a conceptual assessment framework for determining instances of SWAT and thereby, the opportunities for students to engage in critiquing the reasoning of others on assessment tasks.

Assessment Implementation – Selecting Assessment Tasks with and without Student Work

The fourth stage of ECD, assessment implementation, “is about constructing and preparing all of the operational elements specified in the [Conceptual Assessment Framework]” (Mislevy & Haertel; 2006; p. 16). In my study, I did not construct or prepare assessment tasks. Instead, I analyzed curriculum-based assessment materials from five different CCSSM-aligned curriculum series: Big Ideas, CMP, CPM, Eureka, and Go Math for tasks with embedded student work. Using the student work tasks identified in an earlier study from student textbooks (Going, Ray, & Edson, in preparation) and the SWAT from the curriculum-based assessment materials, I developed codes and used preexisting codes to identify the assessment types (only applied to SWAT), the CCSSM Content Strands, evidence of student thinking, and the critique types for each task (see Appendix B). I conducted analyses in order to investigate the frequency and nature of SWAT in curriculum-based assessments.

I also wanted to gain insight into students’ and teachers’ experiences with SWAT. I selected mathematical tasks from the curriculum series (both instructional and assessment resources were used) that focused on one of the seventh-grade CCSSM content standard, expressions and equations: “Solve real-life and mathematical problems using numerical and algebraic expressions and equations” (CCSSI, 2010, p. 47) (see Appendix F). Two problems from each of the four original curriculum series were selected: one task that did not fit the Criteria for Student Work and one task that did fit the Criteria for Student Work that I used in interviews with students and teachers (see Appendix B).

Assessment Delivery – Interviews with Students and Teachers

According to Mislevy and Haertel (2006), the assessment delivery stage of ECD “is where students interact with tasks, their performances are evaluated, and feedback and reports

are produced” (p. 16). This stage involves presenting tasks to students and capturing their work. Since I was conducting an exploratory validity study, I did not have the goal of evaluating students’ performances. Instead, I wanted to be able to observe students interacting with the tasks in order to gain an understanding of whether or not student work embedded in tasks could serve as a mechanism for assessing SMP3b. In this way, I wanted to understand what would be required in the performance if it were to be evaluated.

I conducted clinical interviews with students using the assessment tasks with and without student work described above (see Appendices E and F). I asked students to think aloud as they solved the assessment tasks in order to make their thinking visible (Ericsson & Simon, 1993). In order to explore these tasks further, I also conducted semi-structured interviews with both students and teachers (Glesne, 2006; Seidman, 2012). Students were asked about their experiences solving the tasks and their perspectives on tasks that presented someone else’s thinking and asked them to make sense of it. Teachers were asked about their experiences and perspectives on the assessment tasks and students’ written work on the tasks (see Appendices E, F, H, & I).

Summary

In this chapter, I detailed the methods for my exploratory validity study of student work as a mechanism for assessing SMP3b. I described my guiding design methodology, Evidence-Centered Assessment Design (ECD), and key constructs of assessment as evidentiary argument and validity (Gotwals, Hokayem, Song, & Songer, 2013; Gotwals & Songer, 2013; Mislevy, 2012; Mislevy & Haertel, 2006; Mislevy & Riconscente, 2005; Mislevy, Steinberg, & Almond, 2003). I provided both data collection and analysis methods for each of the three data sources for my study: (1) curriculum materials, (2) interviews with students, and (3) interviews with

teachers. Then, I revisited ECD and described the key stages of the ECD framework and how my study methodologies aligned closely with these stages. I highlighted the ways in which my exploratory validity study differed slightly from the design and delivery processes of ECD intended for validity argument studies. In the following three chapters, I present findings from my exploration of (1) SWAT in curriculum-based assessments, (2) students' experiences with and perspectives on SWAT, and (3) teachers' experiences with and perspectives on SWAT and students' work on SWAT.

CHAPTER 4: RQ #1: SWAT IN CURRICULUM-BASED ASSESSMENTS

Overview

In this chapter, I present findings from text analyses I completed of curriculum-based assessment materials from five CCSSM-aligned seventh-grade curriculum series. The goal of these analyses was to determine the frequency and nature of SWAT in curriculum-based assessments in order to explore students' opportunities to critique someone else's thinking on assessments as mediated by tasks with embedded student work. First, I detail the prevalence of SWAT in the curriculum-based assessments as determined by the Criteria for Student Work (see Appendices A and B) and provide findings from analyses of the features of these tasks (assessment types, CCSSM content strands, types of evidence of student thinking, and critique types). Then, I present detailed findings about the features of the SWAT by curriculum series. I conclude these analyses by making comparisons between the SWAT and student work textbook tasks, or SWTT.

I address the following research questions:

1. What is the frequency and nature of SWAT in curriculum-based assessments?
 - a. How prevalent are SWAT in curriculum-based assessments?
 - b. How do SWAT vary across curriculum series?
 - c. How do SWAT compare to SWTT in corresponding student textbooks?

RQ #1a: Prevalence of SWAT in Curriculum-based Assessments

In the following paragraphs, I present findings from my analyses of tasks in curriculum-based assessments addressing Research Question #1a: How prevalent are SWAT in curriculum-based assessments? I detail the results of the criteria coding and my coding of SWAT based on features of the tasks.

Student Work Criteria Analyses

I analyzed a total of 4802 assessment tasks from curriculum-based assessment materials from five different seventh grade curriculum series: Big Ideas, CMP, CPM, Eureka, and Go Math. I first considered the assessment types for each curriculum series as defined by each series as a way to categorize the assessment tasks (see Figure 4.1). All five series, except for Eureka, had diagnostic assessments either intended for use at the beginning of the year as a course pre-assessment (e.g. Pre-Course Test in Big Ideas) or at the chapter level (e.g. Unit Readiness in CMP). All series except for CPM had periodic assessments that included quizzes at the module, chapter, or unit levels (e.g. Module Quizzes in Go Math). All of the series had summative assessments at the chapter and/or course levels. Two series, CMP and CPM, included question banks. The CMP question bank was for use on the End-of-Year Test as alternative questions. CPM provided a large online repository of questions for use on any type of assessment as determined by the teacher.

	Diagnostic	Periodic	Summative	Bank
Big Ideas	Pre-Course Test	Quizzes	Chapter Tests Cumulative Assessments Alternative Assessments End-of-Course Tests	
CMP	Unit Readiness Assessments	Check Ups Partner Quizzes	Unit Tests End-of-Year Test	End-of-Year Test Bank
CPM	Pre-Assessment		Sample Individual Tests Sample Team Tests	Question Bank
Eureka		Mid-Module Assessments	End-of-Module Assessments	
Go Math	Placement Test Beginning-of-Year Diagnostic Test	Module Quizzes	Unit Tests Unit Performance Tasks Quarterly Benchmark Tests Mid-Year Test End-of-Year Test	

Figure 4.1 Assessment Types Across Curriculum Series

Of the 4802 total assessment tasks analyzed, 1068 tasks were from Big Ideas, 581 tasks from CMP, 1621 tasks from CPM, 167 tasks from Eureka, and 1365 tasks from Go Math. The large range in the number of tasks per curriculum series was due to differences in assessment formats. Both Big Ideas and Go Math had multiple versions of assessments for the same chapters or units which resulted in a high number of assessment tasks. Also, CPM had a large online question bank of tasks in addition to sample tests. CMP and Eureka did not have multiple versions of assessments and did not include a large online repository of tasks. The number of curriculum-based assessment tasks by assessment type and curriculum series is provided in Table 4.1. Each assessment task was analyzed using the Criteria for Student Work (see Appendices A and B). From these analyses, I found that only a small percentage of tasks from each curriculum series in comparison to the total number of assessment tasks fit all three criteria for student work. I also found that curriculum series varied in their inclusion of people in tasks, whether or not these people had evidence of mathematical thinking, and whether or not the reader was tasked with critiquing this thinking in some way (see Table 4.2).

Table 4.1

<i>Number of Assessment Tasks by Assessment Type and Curriculum Series</i>						
Assessment Type	Big Ideas	CMP	CPM	Eureka	Go Math	All Series
Diagnostic	45	83	157	--	130	415
Periodic	274	181	--	79	518	1052
Summative	749	175	136	88	717	1865
Bank	--	142	1328	--	--	1470
Totals (N = 4802)	1068	581	1621	167	1365	

Table 4.2

<i>Percentage of Assessment Tasks by Student Work Criteria and Curriculum Series</i>						
Criteria and Comparisons	Big Ideas (N=1068)	CMP (N=581)	CPM (N=1621)	Eureka (N=167)	Go Math (N=1365)	All Series (N=4802)
Criterion #1	12.4	37.2	26.2	60.5	47.2	31.7
Criterion #2	2.2	7.6	9.7	15.6	6.4	7.0
Criterion #3	0.4	3.4	5.2	6.0	0.7	2.6

Table 4.2 (cont'd)

Criterion #2/ Criterion #1	17.4	20.4	36.4	25.7	13.7	22.2
Criterion #3/ Criterion #1	3.0	9.3	19.5	9.9	1.4	8.3
Criterion #3/ Criterion #2	17.4	45.5	53.5	38.5	10.2	37.6

Across all five curriculum series assessment tasks, the tasks included a person about a third of the time (31.7%), rarely presented evidence of a person's thinking (7%), and rarely required the reader to make sense of someone's mathematical thinking (2.6%). These findings showed that it was very rare for students to have to analyze someone else's mathematical thinking on assessment tasks from the five curriculum series in comparison to the total number of assessment tasks.

Looking across the coding by criteria, Eureka and Go Math had the highest percentages of tasks that included a person (60.5% and 47.2%, respectively). For the tasks that included a person (fulfilled Criterion #1), CPM had the highest percentage of tasks that also included evidence of the person's mathematical thinking (36.4%). None of the other four curriculum series had over a third of the tasks with a person also include evidence of the person's mathematical thinking. Lastly, of the tasks that included a person and presented evidence of the person's thinking, CPM required the reader to make sense of the thinking in some way over half the time (53.5%) compared to over a third for CMP and Eureka (45.5% and 38.5%, respectively). Big Ideas and Go Math had very low percentages of tasks that fulfilled Criterion #3 both overall and for tasks that fulfilled Criteria #1 and #2. Therefore, even when a task included a person and evidence of a person's thinking, students were only required to makes sense of or critique the person's thinking in about half of these tasks for CPM or less for the other four curriculum series.

To summarize these findings by curriculum series, Big Ideas rarely had students encounter any of the criteria for student work on assessment tasks. For Go Math, almost half of the assessment tasks included a person, but students rarely saw evidence of a person's mathematical thinking and, even when they did, they rarely had to make sense of the thinking in some way. Eureka had the highest percentage of assessment tasks with a person. Eureka also presented evidence of a person's thinking for over a quarter of these instances and when presented with evidence, students had to critique it in some way over a third of the time. CMP required the reader to evaluate someone else's mathematical thinking almost half of the time when a person and evidence of the person's mathematical thinking was presented in a task. Lastly, CPM had only about a quarter of the tasks with a person but the highest percentage of tasks that required evaluation when a person and the person's thinking were included in the task.

Overall, 127 tasks out of 4802 total assessment tasks from curriculum-based assessment tasks fit all of the Criteria for Student Work (see Appendices A and B). Of these 127 SWAT, 4 were from Big Ideas (3.1%), 20 were from CMP (15.7%), 84 were from CPM (66.1%), 10 were from Eureka (7.9%), and 9 were from Go Math (7.1%) (see Figure 4.2). As a reminder, the number of tasks provided in the assessment materials for each curriculum series was extremely varied (see Table 4.1). Due to this wide range, the percentages and Figure 4.2 are a little misleading because CPM had so many more tasks overall than the other series. As far as the prevalence of student work tasks in assessments, the percentages of tasks that fulfilled Criterion #3 for each series in Table 4.2 offers a better point of comparison across the curriculum series.

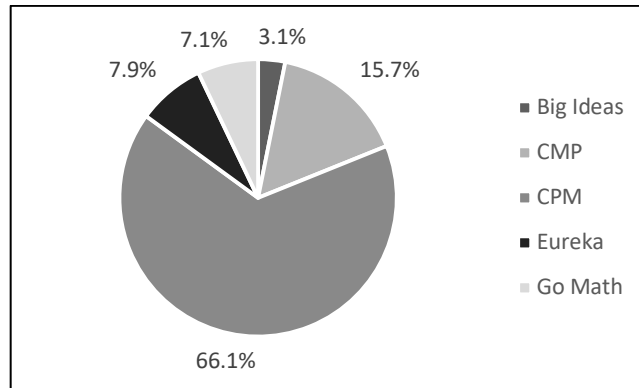


Figure 4.2 SWAT by Curriculum Series

SWAT General Analyses

Assessment Types. Of the 127 SWAT tasks, 0 tasks were diagnostic assessment tasks, 13 tasks were periodic assessment tasks (10.2%), 37 tasks were summative assessment tasks (29.1%), and 77 tasks were question bank tasks (60.6%) (see Figure 4.3).

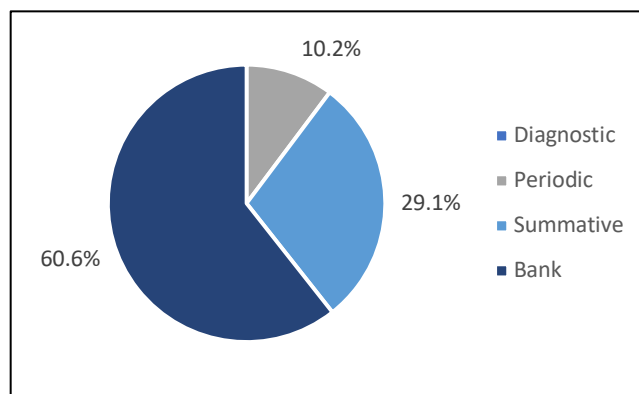


Figure 4.3 SWAT by Assessment Type

CCSSM Content Strands. The 127 SWAT spanned all five of the CCSSM content standard strands for Grade 7. For the main mathematical foci of the assessment tasks, 9 tasks assessed content related to ratios and proportional relationships (7.1%), 54 tasks assessed number system (42.5%), 28 tasks assessed expressions and equations (22%), 17 tasks assessed geometry (13.4%), and 19 tasks assessed statistics and probability (15%) (see Figure 4.4).

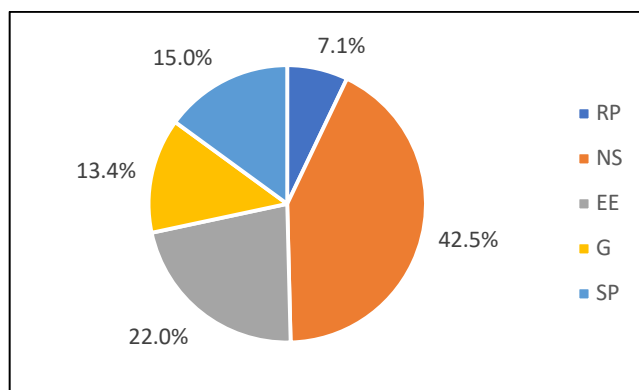


Figure 4.4 SWAT by CCSSM Content Strand

Evidence Types. In the set of 127 SWAT, there were 154 instances of evidence of student thinking that the reader of the tasks was required to critique. Of these 154 instances, 75 were statements, thoughts, or explanations (48.7%), 38 were symbolic computations, developed formulas or expressions, or operational representations (24.7%), 18 were visuals, such as a diagram, graph, or table (11.7%), and 23 were a description of someone's mathematical actions (14.9%) (see Figure 4.5).

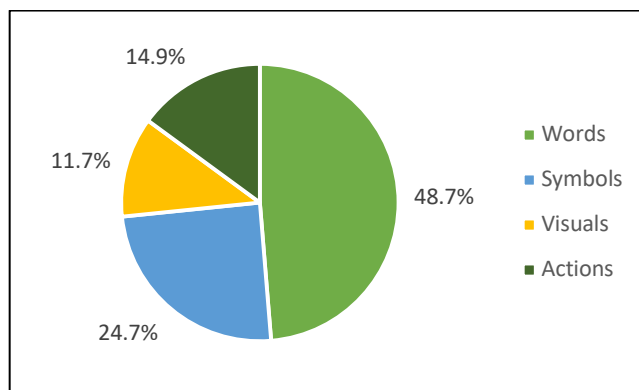


Figure 4.5 SWAT by Evidence Type

Critique Types. In the set of 127 SWAT, the reader was asked to engage in 160 different instances of critiquing spanning four different types of critiques⁴. Of the 160 instances of critiquing, the reader was tasked with completing error identification or correction in 19

⁴ I included the Preference (Pref) category as a type of critique for Figure 4.6 despite the fact that 0 SWAT required the reader to make a choice based on preference. Even so, this type of critique was found in the student work tasks from the student textbooks (SWTT) that will be referenced later in Chapter 4.

instances (11.9%), evaluation in 94 instances (58.8%), comparison in 31 instances (19.4%), and provision of insight into someone's thinking in 16 instances (10%) (see Figure 4.6).

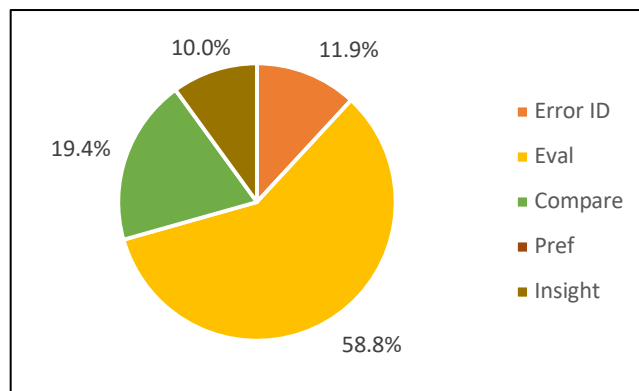


Figure 4.6 SWAT by Critique Type

Overall, analyses of the 127 SWAT indicated that students' opportunities to engage in critiquing someone else's mathematical thinking by interacting with student work on assessments were: (1) mostly found in the curriculum-based assessment materials from CPM and rarely found in Big Ideas, (2) tended to occur on summative assessment tasks or assessment bank tasks and did not occur on diagnostic tasks, (3) often focused on the mathematical content related to the number system and rarely related to ratios and proportional relationships, (4) tended to require the reader to engage in making sense of someone's words and rarely required the reader to make sense of someone's visual representations, and (5) most often required the reader to evaluate someone's mathematical thinking and did not require the reader to make a choice based on preference.

RQ #1b: SWAT by Curriculum Series

Looking across and within curriculum series, I considered how the SWAT compared by task features (assessment types, CCSSM content foci, evidence of student thinking types, and the critique types). In the following paragraphs, I address Research Question #1b: How do SWAT vary across curriculum series? Because CPM made up a majority of the SWAT (66.1%) with

other curricula accounting for only a few SWAT, it was important to understand how features of the SWAT varied at the level of each curriculum series. The following tables and figures detail findings for each curriculum series, Big Ideas, CMP, CPM, Eureka, and Go Math, based on the features of the SWAT including assessment type, CCSSM content strand, evidence of student thinking types, and critique types. In the paragraphs that follow, I elaborate on the findings shown in these tables and figures.

Table 4.3

Number of SWAT by Curriculum Series and Assessment Type

Curriculum Series	Diagnostic	Periodic	Summative	Bank	All Assessments
Big Ideas	0	0	4	--	4
CMP	0	6	7	7	20
CPM	0	--	14	70	84
Eureka	--	2	8	--	10
Go Math	0	5	4	--	9
Totals (N=127)	0	13	37	77	

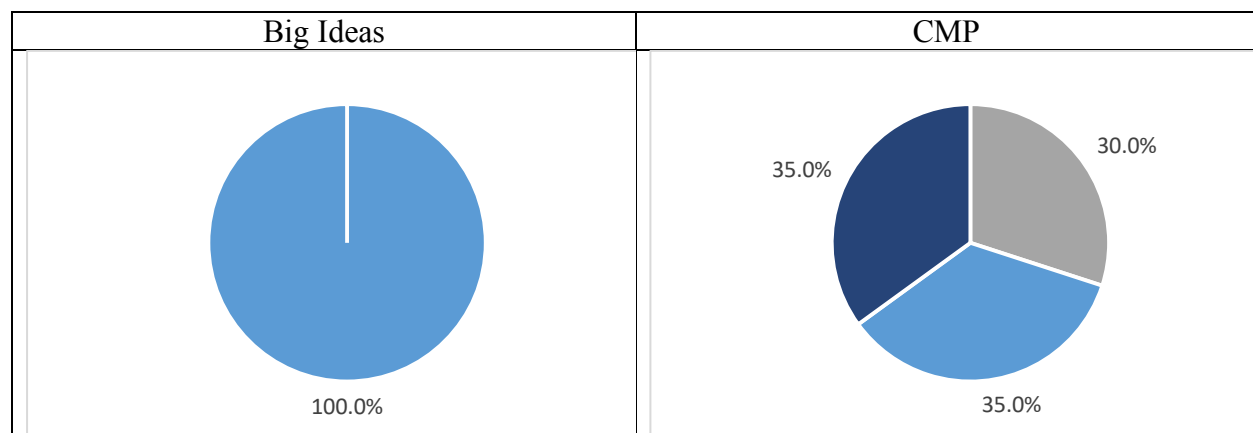


Figure 4.7 Comparison of SWAT by Curriculum Series and Assessment Types

Figure 4.7 (cont'd)

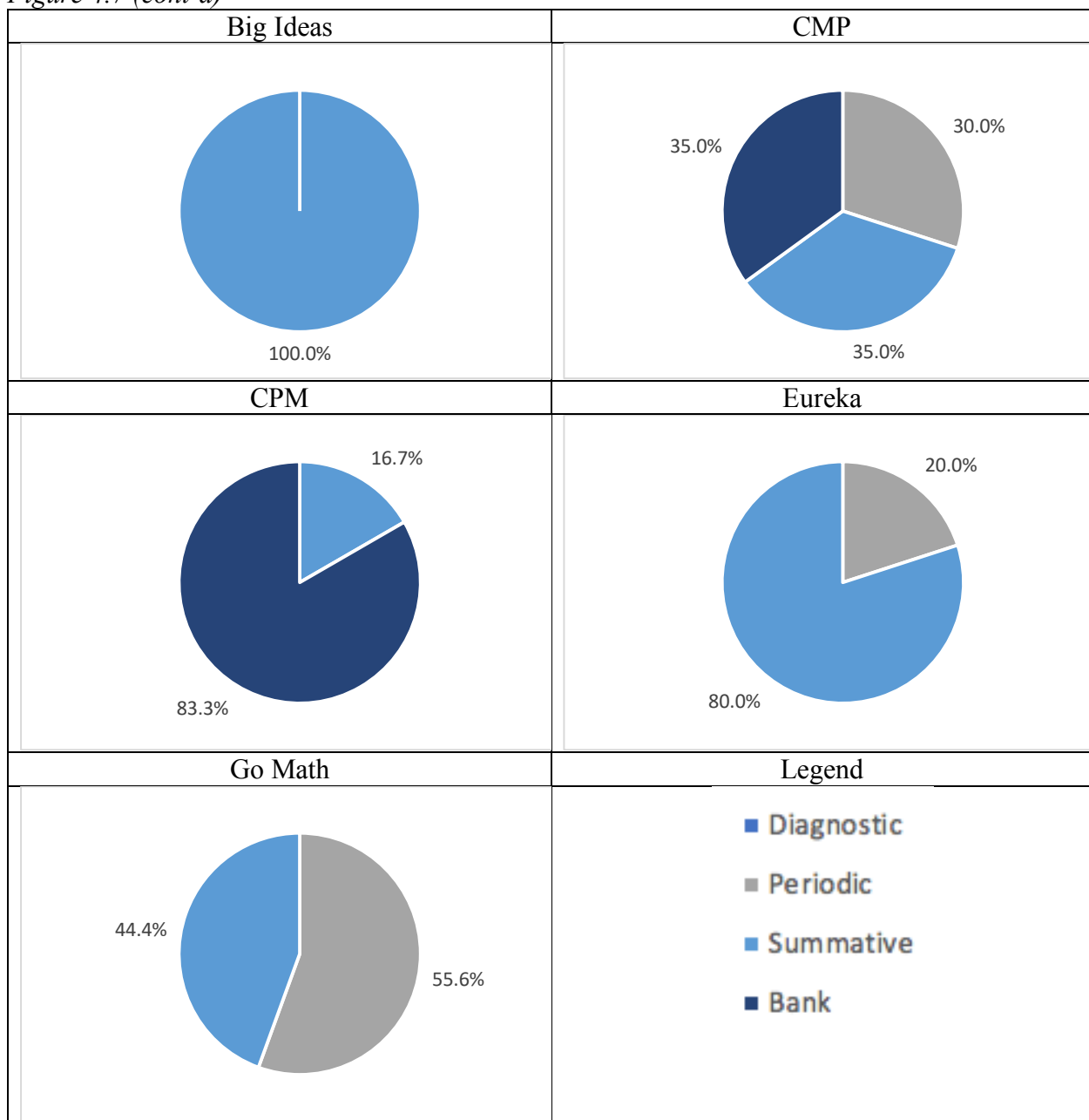


Table 4.4

Number of SWAT by Curriculum Series and CCSSM Content Strand for Grade 7

Curriculum Series	RP	NS	EE	G	SP	All CCSSM
Big Ideas	2	1	0	1	0	4
CMP	5	4	0	2	9	20

Table 4.4 (cont'd)

CPM	1	44	19	13	7	84
Eureka	1	4	3	1	1	10
Go Math	0	1	6	0	2	9
Totals (N=127)	9	54	28	17	19	

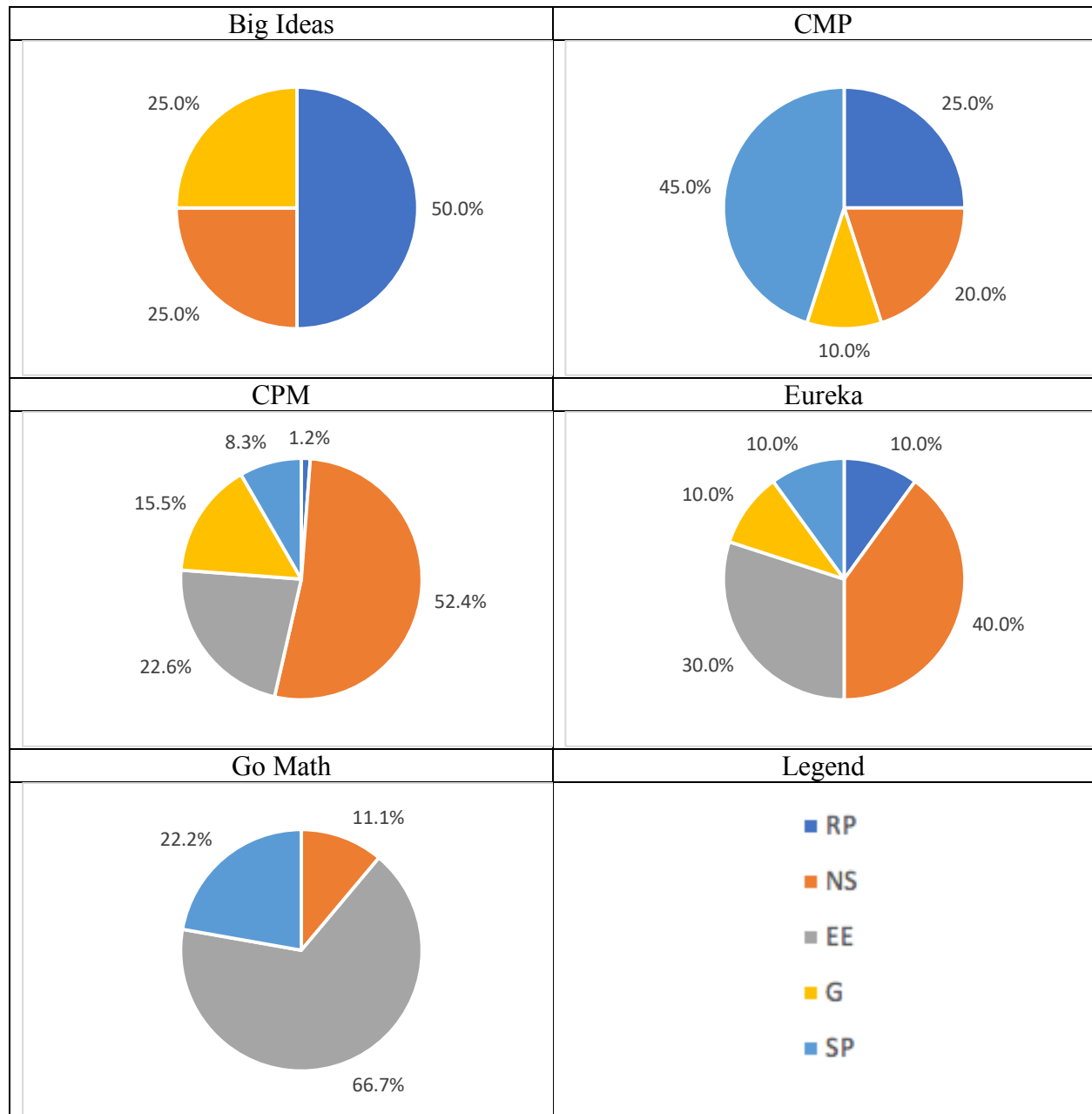


Figure 4.8 Comparison of SWAT by Curriculum Series and CCSSM Content Strands

Table 4.5

<i>Number of Instances of Student Thinking in SWAT by Curriculum Series and Evidence Type</i>					
Curriculum Series	Words	Symbols	Visuals	Actions	All Evidence
Big Ideas	2	2	0	1	5
CMP	12	0	2	8	22
CPM	53	27	15	10	105
Eureka	7	3	1	2	13
Go Math	1	6	0	2	9
Totals (N=154)	75	38	18	23	

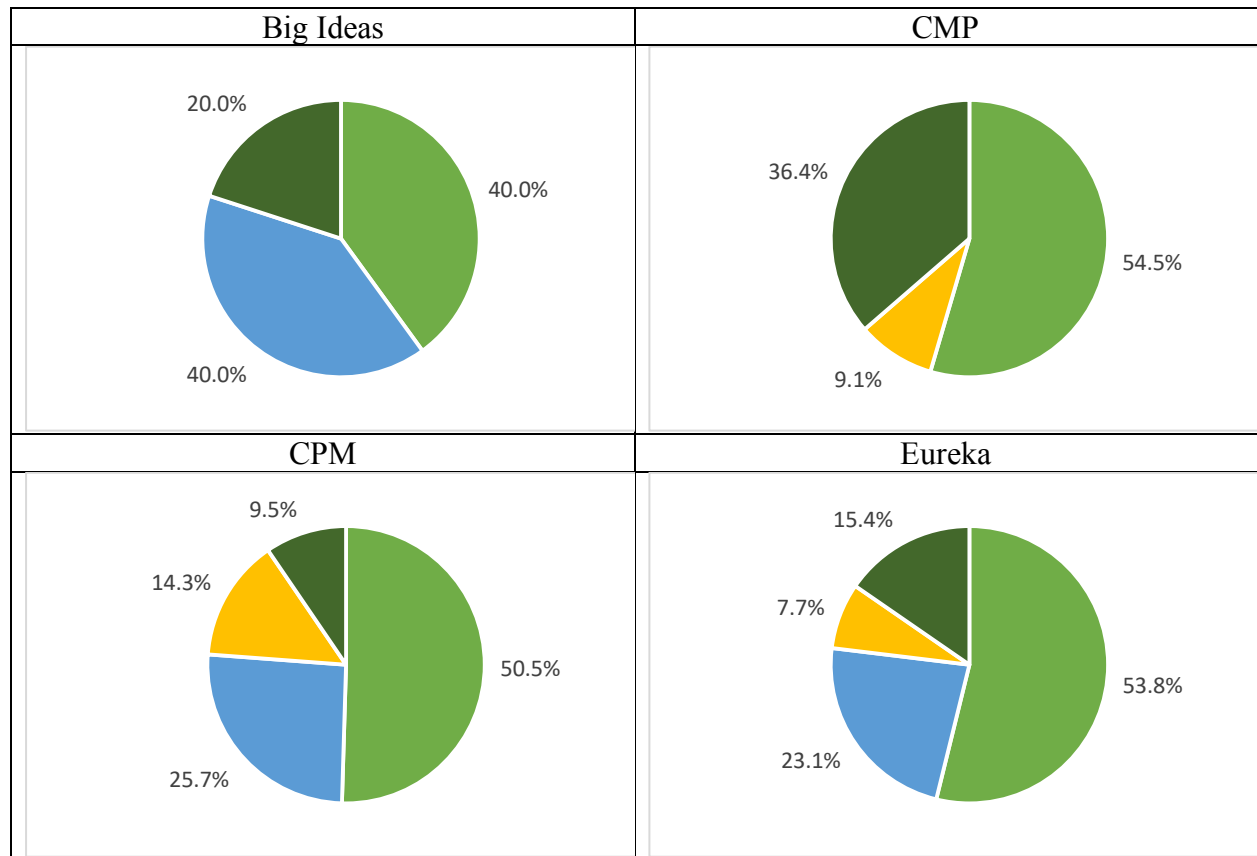


Figure 4.9 Comparison of SWAT by Curriculum Series and Evidence of Student Thinking Types

Figure 4.9 (cont'd)

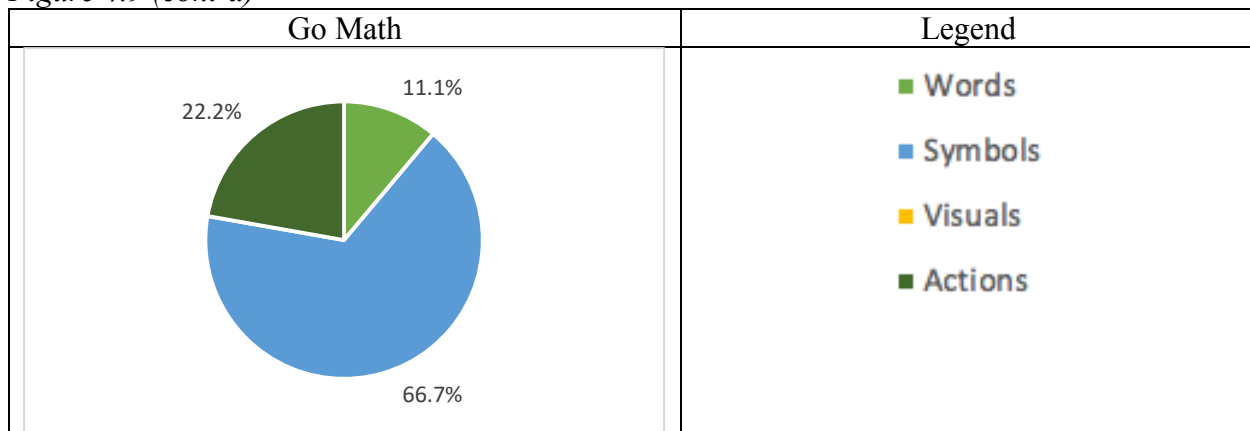


Table 4.6

Number of Instances of Critiques in SWAT by Curriculum Series and Critique Type

Curriculum Series	Error ID	Eval	Compare	Pref	Insights	All Critiques
Big Ideas	2	2	0	0	0	4
CMP	0	19	5	0	1	25
CPM	9	62	20	0	14	105
Eureka	2	7	5	0	1	15
Go Math	6	4	1	0	0	11
Totals (N=160)	19	94	31	0	16	

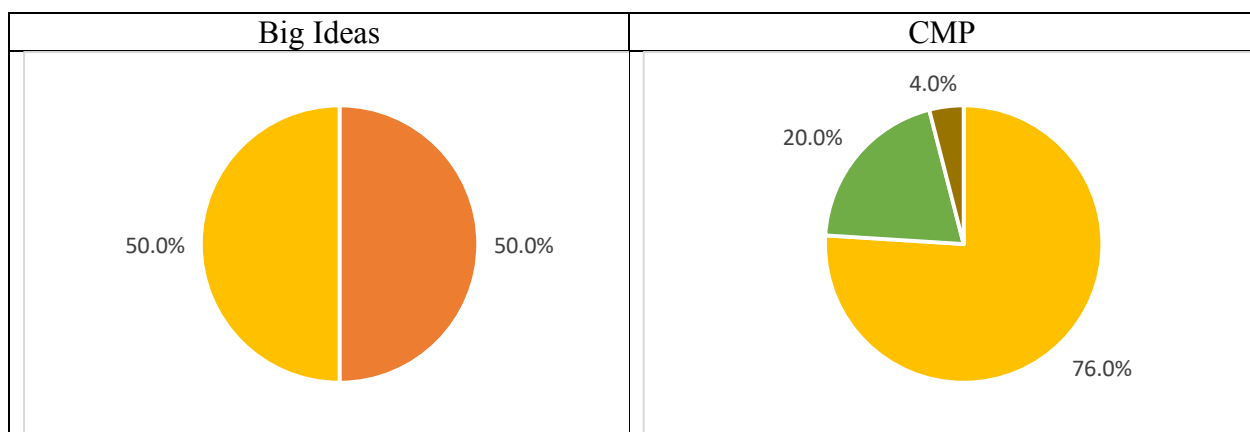
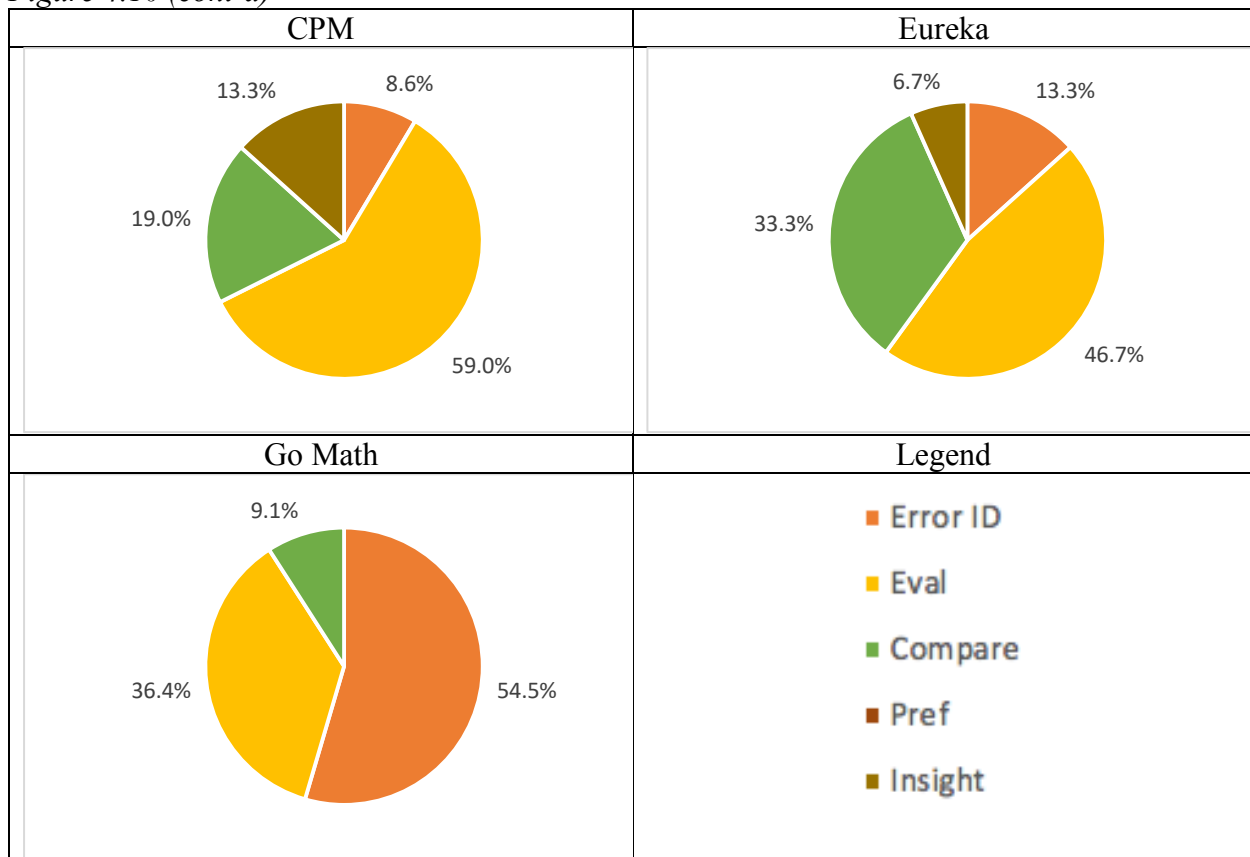


Figure 4.10 Comparison of SWAT by Curriculum Series and Critique Types

Figure 4.10 (cont'd)



Big Ideas

Assessment Types. The 4 SWAT from Big Ideas were all from summative assessments, even though Big Ideas assessment materials also included diagnostic and periodic tasks. Even so, summative assessments made up 70.1% of the total number of assessment tasks from Big Ideas, with periodic assessments (25.7%) and diagnostic assessments (4.2%) accounting for much smaller portions of the Big Ideas assessment tasks (see Table 4.1). So, the only examples of SWAT from Big Ideas were from the most common assessment type found in Big Ideas.

CCSSM Content Strands. The 4 SWAT from Big Ideas were mostly focused on ratios and proportional relationships (50%), with the 2 remaining tasks assessing the number system (25%) and geometry (25%) strands. Big Ideas SWAT did not assess expressions and equations or statistics and probability strands.

Evidence Types. Big Ideas SWAT included the same number of instances of words (40%) and symbolic reasoning (40%) evidence types with mathematical actions accounting for the least represented evidence type (20%). Big Ideas SWAT did not have any visual representations for the evidence type critiqued by the reader.

Critique Types. Big Ideas only required the reader to engage in two critique types in the SWAT. Big Ideas SWAT involved equal parts error identification and evaluation (50% and 50%, respectively) with none of the other critique types represented. Big Ideas was the only curricula in which each SWAT only required one type of critique.

CMP

Assessment Types. Even though the CMP curriculum-based assessment materials contained diagnostic assessment tasks (14.3%), the 20 SWAT from CMP fairly evenly spanned the 3 other assessment tasks – periodic (30%), summative (35%), and question banks (35%). This was not substantially different from the distribution of periodic (31.1%), summative (30.1%), and question bank (24.4%) tasks in the full set of CMP assessment tasks.

CCSSM Content Strands. The 9 CMP SWAT were mostly focused on the statistics and probability strand (45%), with less emphasis on assessing ratios and proportional relationships (25%), the number system (20%), and geometry (10%) strands. None of the SWAT from CMP assessed the expressions and equations strand.

Evidence Types. CMP SWAT did not require the reader to make sense of someone's symbolic reasoning. Instead, the reader was tasked with critiquing someone's words a majority of the instances of evidence types (54.5%), followed by mathematical actions (36.4%). Mirroring the results for the full set of SWAT, the reader was least likely to be required to make sense of someone's visual representations in CMP SWAT (9.1%).

Critique Types. CMP SWAT had a majority of tasks requiring evaluation (76%) followed by 20% of the SWAT requiring comparison. CMP SWAT required insights into someone's mathematical thinking only 4% of the instances of critiques and none of the SWAT required error identification and correction or preference.

CPM

Assessment Types. The CPM curriculum-based assessment materials contained diagnostic tasks (9.7%), but the 84 SWAT from CPM included a majority of question bank (90.9%) and a few summative (9.1%) tasks. When compared to the distribution of summative (8.4%) and question bank (81.9%) assessment types in the full set of CPM assessment tasks, this assessment types result was not very surprising.

CCSSM Content Strands. CPM SWAT mostly assessed the number system strand (52.4%), followed by the expressions and equations strand (22.6%). In CPM, geometry content was the second most assessed CCSSM content strand (15.5%), followed by statistics and probability (8.3%) and ratios and proportional relationships (1.2%).

Evidence Types. CPM required the reader of the task to critique someone's words a majority of the instances of evidence (50.5%). The second most common evidence type was someone's symbolic reasoning (25.7%). Also, visual representations (14.3%) accounted for a larger portion of the evidence types than mathematical actions (9.5%).

Critique Types. CPM SWAT required the reader to engage in evaluation for a majority of the critique instances (59%) followed by comparing (19%), providing insights (13.3%), and identifying and correcting errors (8.6%). This aligned fairly closely with the distribution of critique types for the full set of SWAT.

Eureka

Assessment Types. Periodic and summative tasks were the only assessment tasks represented in the full set of Eureka assessment tasks (47.3% and 52.7%, respectively). For the 10 Eureka SWAT, the majority of tasks were summative tasks (80%) with a couple of periodic tasks (20%) included.

CCSSM Content Strands. Eureka SWAT mostly assessed the number system strand (40%) followed by the expressions and equations strand (30%). From the remaining SWAT in Eureka, ratios and proportional relationships, the number system, and statistics and probability were equally addressed (10% each).

Evidence Types. Eureka SWAT required the reader of the task to critique someone's words a majority of the instances of evidence (53.8%). The second most common evidence type was someone's symbolic reasoning (23.1%) mirroring the results for the full set of SWAT. Eureka SWAT mirrored the results for the full set of SWAT with actions accounting for 15.4% and visuals accounting for 7.7% of the evidence types.

Critique Types. The 15 critique instances in Eureka SWAT required the reader to engage in evaluating someone's evidence of mathematical thinking most often (46.7%) and providing insight least often (6.7%). The remaining critiques involved a considerable portion of comparison (33.3%) and a couple of instances of error identification and correction (13.3%).

Go Math

Assessment Types. Go Math assessment tasks included three different types of assessment tasks – diagnostic (9.5%), periodic (37.9%), and summative (52.5%). However, the 9 SWAT from Go Math did not include any diagnostic tasks, a majority of periodic tasks (55.6%), and a smaller percentage of summative tasks (44.4%).

CCSSM Content Strands. The 9 Go Math SWAT focused on the expressions and equations strand (66.7%) a majority of the tasks, with less attention paid to statistics and probability (22.2%) and the number system (11.1%) strands. Go Math SWAT did not assess ratios and proportional relationships or geometry strands.

Evidence Types. Go Math SWAT required the reader to critique someone's symbolic reasoning a majority of the instances of evidence types (66.7%) with less attention on mathematical actions (22.2%) or someone's mathematical thoughts, statements, or verbal reasoning (11.1%). Go Math SWAT did not have any visual representations for the evidence type critiqued by the reader.

Critique Types. Go Math SWAT required the reader to engage in error identification and correction a majority of the critique instances (54.5%). The two other critiques represented in the Go Math SWAT were evaluation (36.4%) and comparison (9.1%). None of the Go Math SWAT required the reader to provide insights into someone's mathematical thinking or make a choice based on preference.

Comparisons Across Curriculum Series SWAT

Assessment Types. Each curriculum series varied widely in comparison to the others in terms of distributions of assessment types across the SWAT as heavily influenced by differences in distributions of assessment types across the full sets of assessment tasks in each curriculum series (see Table 4.1). Summative assessment tasks made up the majority of SWAT in two curriculum series – Big Ideas and Eureka. The majority of SWAT in CPM were question bank tasks. Go Math SWAT had a majority of periodic assessment tasks. CMP SWAT did not have a majority assessment type. None of the diagnostic assessment tasks across and within each of the curriculum series were represented in the SWAT.

CCSSM Content Strands. The curriculum series SWAT were quite varied by CCSSM content strands assessed. Only two of the series, CPM and Eureka, had all five strands represented in the SWAT. CMP included four strands and both Big Ideas and Go Math only included three. Of the five curriculum series, CPM and Eureka had the most similar distributions of content strands with the number system accounting for most of the tasks (52.4% and 40%, respectively) and ratios and proportional relationships accounting for very few of the tasks (1.2% and 10%, respectively). The three other series, Big Ideas, CMP, and Go Math, all had very different distributions of CCSSM content strands assessed in the SWAT.

Evidence Types. The curriculum series SWAT were fairly varied by the nature of instances of evidence of student thinking the reader was required to critique. Only two curriculum series, CPM and Eureka, included all four evidence types in the SWAT and these two series most closely mirrored the results for the full set of SWAT. Big Ideas, CMP, and Go Math SWAT all included only three evidence types. Big Ideas and Go Math did not have SWAT that required the reader to make sense of someone's visual representations. CMP did not have SWAT that required the reader to make sense of someone's symbolic reasoning. The most prevalent evidence type in Big Ideas, CMP, CPM, and Eureka SWAT was someone's words and these results mirror the results from the full set of SWAT. Go Math SWAT varied significantly from the other curricula with symbolic reasoning accounting for the most number of evidence types and someone's words occurring only once in the Go Math SWAT.

Critique Types. The curriculum series SWAT were fairly varied by the number of instances of critique types required by the reader of the task. Only two series included four of the five possible critique types in the SWAT, CPM and Eureka. CMP and Go Math SWAT both included only three critique types and Big Ideas SWAT only included two. CMP SWAT did not

require the reader to engage in error identification. Go Math SWAT did not require the reader to provide insights into someone's evidence of mathematical thinking. Big Ideas did not require the reader to make comparisons between instances of evidence of mathematical thinking or provide insight into someone's evidence of mathematical thinking. None of the SWAT from any of the curriculum series required the reader to make a choice based on preference. The most common critique type from three of the series, CMP, CPM, and Eureka, was also the majority critique type across the full set of SWAT – evaluation.

The analyses of SWAT by curriculum series revealed that the curriculum series SWAT were extremely varied based on the assessment types, influenced by the major differences in the types of assessments included in each curriculum series and the distribution of tasks across these series, and also varied based on CCSSM content strands, evidence types, and critique types. Because the results from the curriculum series analyses required making comparisons between curricula with such a range of numbers of SWAT (e.g. only 4 SWAT from Big Ideas compared to 84 SWAT from CPM) and types of assessment tasks (e.g. Eureka math assessments only included periodic and summative assessments), I concluded my text analyses of the SWAT by making comparisons between tasks in the SWAT and tasks in the corresponding student textbooks. I refer to the tasks found in the student textbook that meet the criteria for student work as student work textbook tasks (SWTT). The goal of these analyses was to compare SWTT to SWAT.

RQ #1c: Comparison of SWAT and SWTT

Based on the opportunities students had to view someone else's mathematical thinking and critique it in some way on the SWTT and the SWAT from each curriculum series, I made comparisons between the student work tasks that appeared in student textbooks, or SWTT, and

the student work tasks that appeared in assessment materials across and within each curriculum series, or SWAT. These comparisons were focused on the key CCSSM content strands emphasized or assessed, evidence of student thinking the reader was required to critique, and critique types. In the following section, I address Research Question #1c: How do SWAT compare to SWTT in corresponding student textbooks? The table and figure below detail the number of SWTT and SWAT and the distribution of SWTT by curriculum series. In the paragraphs that follow, I elaborate on the findings provided in this table and figure.

Table 4.7

<i>Number of SWTT and SWAT by Curriculum Series</i>		
Curriculum Series	SWTT	SWAT
Big Ideas	23	4
CMP	141	20
CPM	112	84
Eureka	63	10
Go Math	70	9
Totals (N=536)	409	127

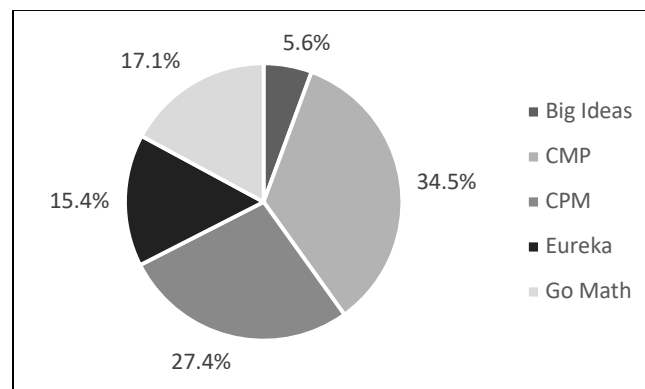


Figure 4.11 SWTT by Curriculum Series

In the analyses conducted on corresponding student textbooks, I found 409 instances of student work. Of these 409 SWTT, 23 were from Big Ideas (5.6%), 141 were from CMP (34.5%), 112 from CPM (27.4%), 63 from Eureka (15.4%), and 70 from Go Math (17.1%). As

compared to the SWAT, with CPM accounting for 66.1% of the SWAT tasks, the full set of SWTT did not have any majority of tasks from a single curriculum series. Instead, CMP was the most represented curriculum series in the SWTT (34.5% as compared to 15.7% in the SWAT), followed closely by CPM (27.4%). The least represented curriculum series, Big Ideas, was consistent for both the SWTT and SWAT (5.6% and 3.1%, respectively). A small difference between the SWTT and SWAT sets was the distribution of Eureka and Go Math. In the SWTT, Go Math was the third most common curriculum series (17.1% as compared to 7.1% in the SWAT) and Eureka was the fourth most common curriculum series (15.4% as compared to 7.9% in the SWAT). The key differences between the curriculum series distributions of the SWTT and SWAT were (1) the lack of a majority in the SWTT as compared to the majority CPM tasks in the SWAT (66.1%) and (2) the most represented curriculum series in the SWTT was CMP accounting for a little over a third of the SWTT (34.5%). However, as a reminder, the number of assessment tasks from each curriculum series accounted for a wide range and I did not capture the total number of textbook tasks in any way as seriation of textbook tasks is not as straightforward as seriation of assessment tasks.

The 409 instances of student work in the student textbook were further analyzed by content strand, evidence of student thinking, and critique types in order to make comparisons between student work occurring in the student textbook and the assessment materials within and across the curriculum series. The following tables and figures detail findings for Big Ideas, CMP, CPM, Eureka, and Go Math based on the CCSSM content strands, evidence types, and critique types of the SWTT. In the paragraphs that follow, I elaborate on the findings shown in the tables and figures including general analyses of the SWTT and a comparison to the SWAT, the findings

for each curriculum series SWAT and SWTT, and comparisons across curriculum series SWAT and SWTT.

Table 4.8

Number of SWTT by Curriculum Series and CCSSM Content Strand for Grade 7

Curriculum Series	RP	NS	EE	G	SP	All CCSSM
Big Ideas	5	5	4	4	5	23
CMP	29	11	26	34	41	141
CPM	29	34	23	12	14	112
Eureka	18	10	7	15	13	63
Go Math	5	14	8	6	37	70
Totals (N=409)	86	74	68	71	110	

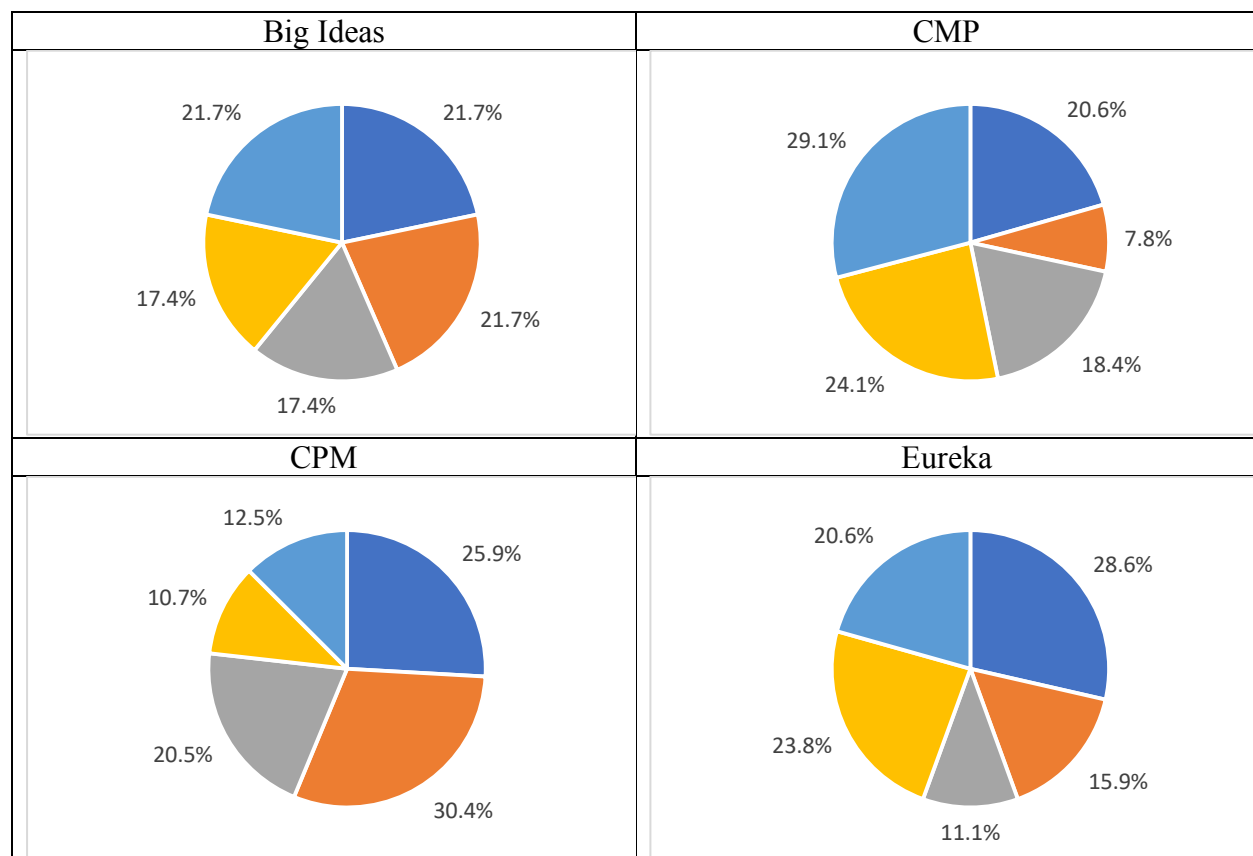


Figure 4.12 Comparison of SWTT by Curriculum Series and CCSSM Content Strand

Figure 4.12 (cont'd)

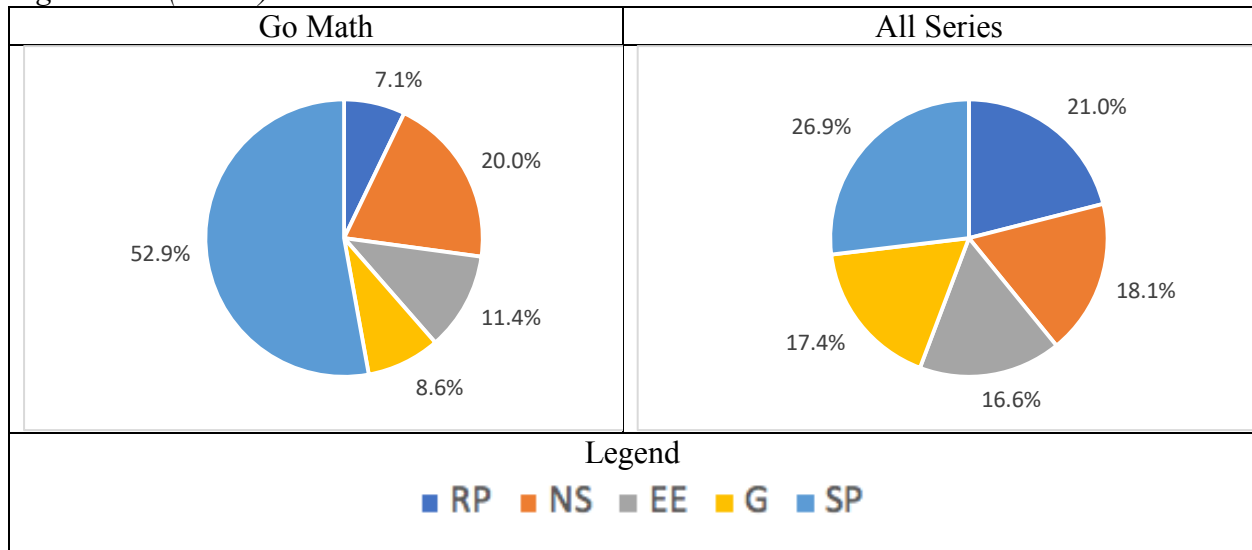


Table 4.9

Number of Instances of Student Thinking in SWTT by Curriculum Series and Evidence Type

Curriculum Series	Words	Symbols	Visuals	Actions	All Evidence
Big Ideas	10	13	1	2	26
CMP	69	56	19	51	195
CPM	38	58	35	27	158
Eureka	41	15	7	12	75
Go Math	36	10	1	25	72
Totals (N=526)	194	152	63	117	

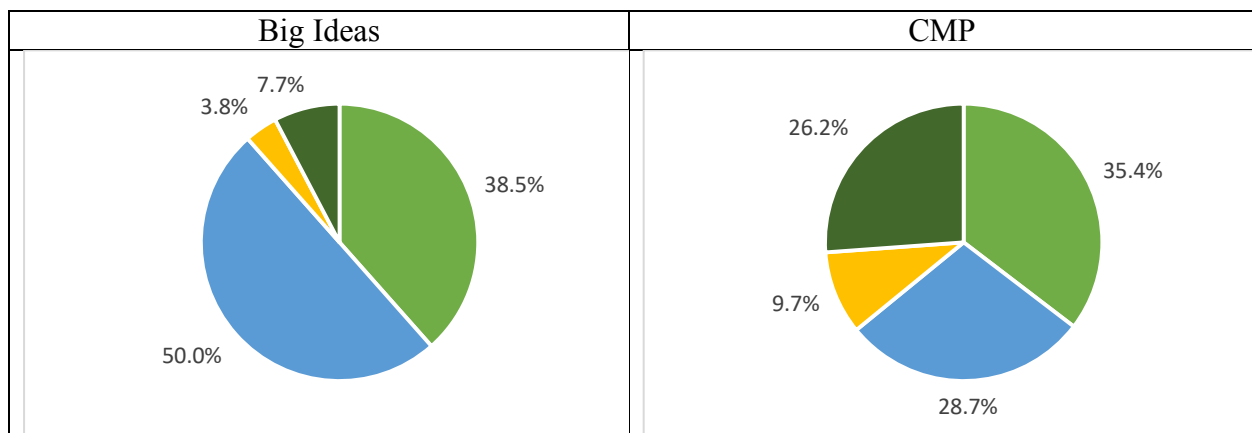


Figure 4.13 Comparison of SWTT by Curriculum Series and Evidence Type

Figure 4.13 (cont'd)

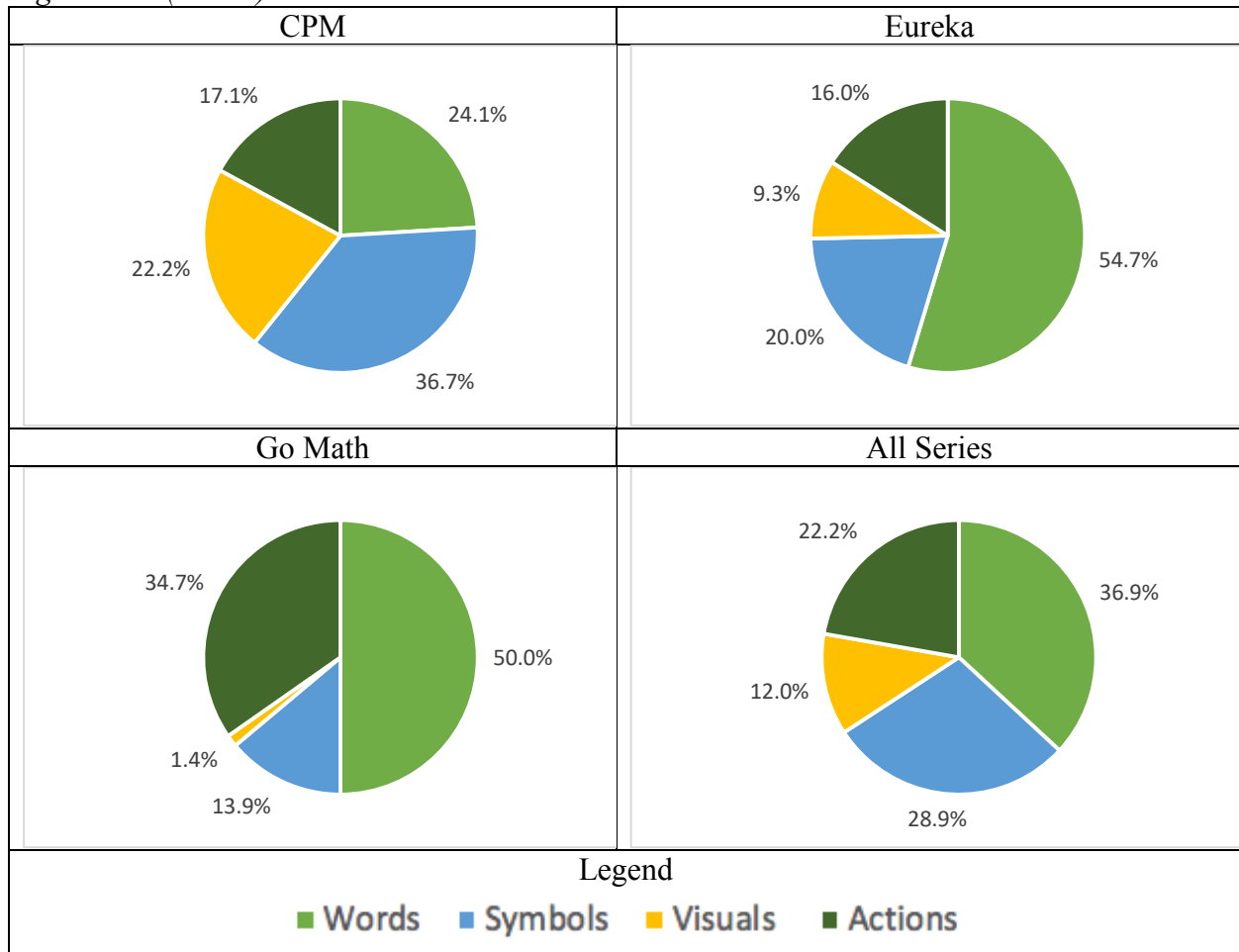


Table 4.10

Number of Instances of Critiques in SWTT by Curriculum Series and Critique Type

Curriculum Series	Error ID	Eval	Compare	Pref	Insights	All Critiques
Big Ideas	11	12	1	0	1	25
CMP	4	98	44	7	28	181
CPM	12	57	19	3	41	132
Eureka	4	49	14	0	10	77
Go Math	18	45	6	0	5	74
Totals (N=489)	49	261	84	10	85	

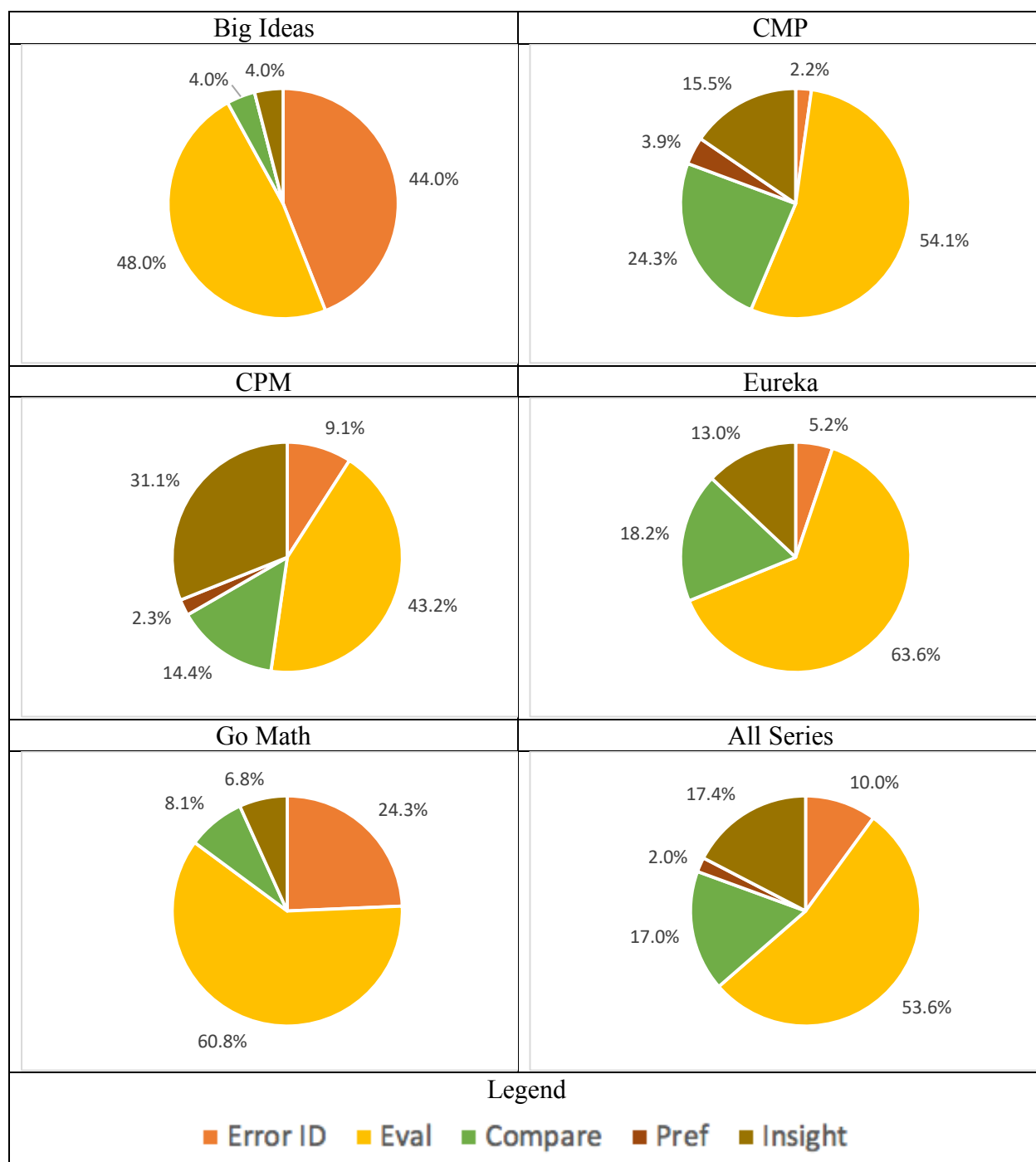


Figure 4.14 Comparison of SWTT by Curriculum Series and Critique Types

SWTT General Analyses and Comparison to SWAT

CCSSM Content Strands. Overall, for the full set of SWTT, the most populous content strand was statistics and probability (26.9%) followed by ratios and proportional relationships (21%), the number system (18.1%), geometry (17.4%), and expressions and equations (16.6%).

This varies substantially from the content strand distribution for the full set of SWAT in which the number system was the most assessed strand (42.5%) followed by expressions and equations (22%), statistics and probability (15%), geometry (13.4%), and ratios and proportional relationships (7.1%). Overall, the full set of SWTT had a more even distribution of content foci as compared to the SWAT, with no single content strand accounting for more than 30% or less than 10% of the total tasks.

Evidence Types. The full set of curriculum series SWTT was not substantially different from the full set of curriculum series SWAT in terms of the evidence types the reader of the tasks was required to critique. Overall in the SWTT and similar to the full set of SWAT, the reader was most likely to be required to critique someone's words (36.9%) followed by symbolic reasoning (28.9%), descriptions of mathematical actions (22.2%), and visual representations (12%). While the distributions vary a bit, the sequence of evidence types from most to least likely was the same for the full sets of SWTT and SWAT. More interesting differences occurred based on comparisons between curriculum series SWAT and SWTT.

Critique Types. Overall in the SWTT, the majority of instances of critiques involved evaluation (53.4%), followed by providing insight (17.4%), comparison (17.2%), error identification and/or correction (10%), and making a choice based on preference (2%). The key differences between the critique types in the sets of SWTT and SWAT was that insight was the second most common and error identification was the fourth. In the full set of SWAT distribution, comparison was the second most common critique type (19.4%) and insight was the fourth most common (10%). None of the SWAT required the reader to make a choice based on preference.

Big Ideas

CCSSM Content Strands. Big Ideas had a fairly balanced distribution of content strands (no more than 30% and no less than 10% for each strand). Big Ideas SWTT had the same number of tasks for ratios and proportional relationships, the number system, and statistics and probability (21.7% for each strand) strands. The two remaining strands, expressions and equations and geometry, also had the same number of tasks (17.4% for both strands). In comparison, Big Ideas SWAT only addressed three of the five content strands.

Evidence Types. The most common evidence type in Big Ideas SWTT was symbolic reasoning accounting for 50% of the instances of evidence. The second most common evidence type was someone's words (38.5%). The remaining evidence types in Big Ideas involved mathematical actions (7.7%) and were least likely to involve visual representations (3.8%). Comparing Big Ideas SWTT to the corresponding SWAT evidence types, symbols and words for both SWTT and SWAT were the most common two evidence types, but Big Ideas SWAT did not include any visual representations of student thinking that the reader was required to critique.

Critique Types. Big Ideas SWAT only involved evaluation and error identification as compared to the addition of comparison and insight in Big Ideas SWTT. Big Ideas SWTT included all critique types except for preference. For Big Ideas SWTT, evaluation was the most common critique type (48%). Error identification and/or correction was the second most common critique type (44%), followed by an equal number of comparison and insight critique types (4% for both).

CMP

CCSSM Content Strands. The CMP SWTT did not have a balanced distribution of content strands, but also did not have any single content strand account for a majority of the

SWTT. In the CMP SWTT, statistics and probability (29%) was the most common content strand followed by geometry (24.1%), ratios and proportional relationships (20.6%), expression and equations (18.4%), and the number system (7.8%). As compared to the CMP SWAT, which did not include any tasks assessing expressions and equations, the only similarity between the CMP SWTT and SWAT was that statistics and probability was the most common content strand for both student work types (29% and 45%, respectively).

Evidence Types. The CMP SWTT included all four of the evidence types. Words accounted for the highest number of instances (35.4%) followed by symbols (28.7%), actions (26.2%), and visuals (9.7%). Words also accounted for the highest number of instances of evidence for CMP SWAT (54.5%), followed by actions (36.4%) and visuals (9.1%). However, none of the CMP SWAT required students to make sense of someone's symbolic reasoning.

Critique Types. CMP SWTT included all five critique types in comparison to the three critique types represented in CMP SWAT. The most common type of critique in CMP SWTT was evaluation (54.1%), followed by comparison (24.3%), insights (15.5%), preference (3.9%), and error identification and/or correction (2.2%). Evaluation was also the most common critique type in CMP SWAT (76%). For CMP SWAT, comparison and insight critique types were also the most common types after evaluation (20% and 4%, respectively).

CPM

CCSSM Content Strands. The CPM SWTT did not have a balanced distribution, but also did not have any single content strand account for a majority of the SWTT. In the CPM SWTT, the number system (30.4%) was the most common content strand followed by ratios and proportional relationships (25.9%), expressions and equations (20.5%), statistics and probability (12.5%), and geometry (10.7%). In contrast, the CPM SWAT not only focused on the number

system for most of the tasks, but also for a majority (52.4%) and ratios and proportional relationships was only assessed in 1 of the 84 CPM SWAT.

Evidence Types. CPM SWTT and SWAT both included all four evidence types, but their distributions were quite different. The most common evidence type in CPM SWTT was symbolic reasoning accounting for 36.7% of the instances of evidence. The second most common evidence type was someone's words (24.1% for CPM). The remaining evidence types in CPM SWTT involved visual representations (22.2%) and were least likely to involve descriptions of mathematical actions (17.1%). In CPM SWAT, words occurred a majority of the instances of evidence (50.5%) followed by symbolic reasoning (25.7%). In both CPM SWAT and SWTT, the reader was least likely to make sense of descriptions of someone's mathematical actions (9.5% and 17.1%, respectively). Overall, CPM SWTT evidence types were more evenly distributed across evidence types than CPM SWAT.

Critique Types. CPM SWAT did not contain preference tasks, but CPM SWTT contained all five critique types. The most common type of critique in CPM SWTT was evaluation (43.2%), followed by insights (31.1%), comparison (14.4%), error identification and/or correction (9.1%), and preference (2.3%). Evaluation was also the most common critique type in CPM SWAT (59%). CPM SWAT had comparison and insight critique types as the most common types after evaluation (19% and 13.3%, respectively).

Eureka

CCSSM Content Strands. Eureka SWTT had a fairly balanced distribution of content strands (no more than 30% and no less than 10% for each strand). However, comparison of the Eureka SWAT and SWTT content strand distributions revealed substantial differences between student work opportunities in the student textbook and the assessment materials. Eureka SWAT

were heavily focused on the number system (40%) followed by expressions and equations (30%) while Eureka SWTT were focused mostly on statistics and probability (28.6%) and geometry (23.8%) with the number system (15.9%) and expressions and equations (11.1%) accounting for the fewest numbers of tasks in Eureka SWAT.

Evidence Types. For both Eureka SWAT and SWTT, the reader was required to make sense of someone else's words a majority of the instances of evidence (53.8% and 54.7%, respectively). This was followed by symbols (23.1% and 20%, respectively) and actions (15.4% and 16%, respectively). The reader was least likely to have to make sense of visuals on Eureka SWAT and SWTT (7.7% and 9.3%, respectively).

Critique Types. Both Eureka SWAT and SWTT contained all critique types except for preference. The Eureka curriculum series was the one series that included the same number of critique type categories in the SWAT and the SWTT. The most common critique type in Eureka SWAT and SWTT was evaluation (46.7% and 63.6% respectively). The second most common critique type for Eureka SWAT and SWTT was comparison (33.3% and 18.2%, respectively). Error identification and insights accounted for the least prevalent critique types for both Eureka SWAT and SWTT other than the lack of any preference tasks, although in opposite order. For Eureka SWAT, error identification accounted for 13.3% of the critique types and insight accounted for 6.7%. For Eureka SWTT, insight accounted for 13% of the critique types and error identification accounted for 5.2%.

Go Math

CCSSM Content Strands. The distribution of student work tasks in Go Math SWTT varied the most as compared to the other curriculum series SWTT distributions from the full set of SWTT. Go Math SWTT focused on the statistics and probability content strand a majority of

the tasks (52.9%) followed by the number system (20%), expression and equations (11.4%), geometry (8.6%), and ratios and proportional relationships (7.1%). This was in stark contrast with the Go Math SWAT in which the expressions and equations strand was assessed a majority of the tasks (66.7%) with neither ratios and proportional relationships nor geometry content strands assessed in any SWAT.

Evidence Types. The evidence of student thinking in Go Math SWTT most often included words (50%), but symbolic reasoning was only the third most common evidence type (13.9%) with actions accounting for the second most common (34.7%) and visual representations for the least common (1.4%). This was significantly different from the evidence types represented in the Go Math SWAT in which no visual representations were included and a majority of the instances of evidence of student thinking included symbolic reasoning (66.7%). Only 11.1% of instances involved someone's words in the Go Math SWAT. This was significantly different from the Go Math SWTT evidence type distribution.

Critique Types. Go Math SWTT included all critique types except for preference. The most common critique types for Go Math SWTT after evaluation (60.8%) were error identification and/or correction (24.3%) and comparison (8.1%). Insight accounted for 6.8% of the critique types. The critique types for Go Math SWAT included error identification (54.5%), evaluation (36.4%), and comparison (9.1%) as compared to the addition of insight in the Go Math SWTT.

Comparisons Across Curriculum Series SWAT and SWTT

CCSSM Content Strands. Comparison of the SWAT and the SWTT by and across curriculum series revealed many interesting differences and a few similarities in terms of CCSSM content strands. For each curriculum series in the SWTT, all five strands were

represented. In the full set of SWAT, two of the curriculum series, Big Ideas and Go Math SWAT, only assessed three of the strands and CMP SWAT only assessed four strands. Only CPM and Eureka SWAT assessed all five strands. Within the set of SWTT, Big Ideas and Eureka had fairly balanced distributions of content strands, CMP and CPM had less balanced distributions, and Go Math was not balanced. For the full set of SWTT, the distribution of content strands was fairly balanced (no more than 30% and no less than 10% for each strand). Go Math was the only set of curriculum series SWTT that had a content strand that accounted for more than a third of the content foci (52.9% for statistics and probability).

The results for the SWTT were substantially different from the SWAT. In the full set of SWAT, the two most common content foci accounted for over 60% of the total number of tasks (42.5% for the number system; 22% for expressions and equations). The five curriculum series SWAT and the full set of SWAT all had a content strand that accounted for over a third of the content foci. So, beyond the student textbook providing more instances of student work tasks than the assessment materials, the content foci for the student work tasks varied substantially in distribution across and within the curriculum series.

Evidence Types. Comparison of evidence types across and within the curriculum series SWAT and SWTT revealed some interesting findings. First of all, all four evidence types occurred in all curriculum series in the SWTT. In contrast, although all evidence types were represented in the full set of SWAT, not all evidence types were represented in each curriculum series. Big Ideas and Go Math SWAT did not include visual representations and CMP SWAT did not require the reader to make sense of someone's symbolic representations. Words was the most common type of evidence that a reader was required to critique in the full sets of SWTT and SWAT as well as three of the five curriculum series SWTT (CMP, Eureka, Go Math) and

three of the five curriculum series SWAT (CMP, CPM, and Eureka). Symbolic reasoning was the most common evidence type for Big Ideas SWTT, CPM SWTT, and Go Math SWAT.

For four of the five curriculum series SWTT and three of the five curriculum series SWAT, visual representations was the least common evidence type, with the exceptions being CMP SWAT with no symbolic representations, CPM SWTT with 22.2% of the evidence types in CPM SWTT, and CPM SWAT with 14.3% of the evidence types in CPM SWAT. The least common evidence type in CPM SWTT and SWAT was mathematical actions (17.1% and 9.5%, respectively). Across all curriculum series SWAT and SWTT, words and symbolic reasoning were the two most common evidence types with the key exceptions of Go Math SWAT and SWTT which also emphasized mathematical actions (22.2% in SWAT; 34.7% in SWTT).

Critique Types. Evaluation was the most common critique type for a majority of the curriculum series SWAT and SWTT and preference was the least common critique type for all of the curriculum series SWAT and SWTT with the exception of CMP SWTT in which error identification and/or correction occurred less often than preference. Second, Go Math and Big Ideas seemed to be the least consistent with the distribution of critique types in comparison to the other curriculum series for both SWAT and SWTT. Lastly, comparing curriculum series SWAT and SWTT, most of the SWTT included more critique types than their corresponding assessment materials. All of the curriculum series SWTT contained at least four of the five identified critique types as compared to the curriculum series SWAT, in which Big Ideas SWAT only involved two critique types and CMP and Go Math SWAT only involved three. Thus, the SWTT offered more variety of critique types as compared to the SWAT.

Comparison of the SWAT and the SWTT by and across curriculum series by CCSSM content strands, evidence of student thinking, and critique types revealed a number of interesting

findings. The main findings included: (1) the SWTT as a whole and across curriculum series accounted for all five CCSSM strands and offered a considerably more balanced coverage of the strands than the SWAT, (2) the sets of SWTT and SWAT and the curriculum series SWTT included all four evidence types, but not all evidence types were represented in each curriculum series SWAT, (3) the most common evidence types for the SWTT and SWAT as the full sets and at the curriculum levels were words and symbolic reasoning, (4) as a whole, the curriculum series SWTT revealed more variety of critique types than the curriculum series SWAT containing at least four of the five identified critique types in each curriculum series as compared to the SWAT in which one series only involved two critique types and two series only involved three, and (5) evaluation was the most common critique type for both the full set of SWAT and SWTT and for a majority of the curriculum series SWAT and SWTT and preference was the least common critique type for all of the curriculum series SWAT and SWTT with the exception of CMP SWTT in which error identification and/or correction occurred less often than preference.

Summary of Chapter Findings

In this chapter, I presented findings from my analyses of curriculum-based assessment materials from five CCSSM-aligned seventh grade mathematics curriculum series. I analyzed 4802 assessment tasks from the curriculum series based on the criteria for student work (see Appendices A and B). Criteria analyses revealed that as a whole, assessment tasks from the five different curricula tended to include a person about a third of the time (31.7%), very rarely presented evidence of a person's thinking (7%) and did not often require the reader to make sense of someone's thinking (2.6%). Even so, when the reader was presented with evidence of someone's thinking, over a third of these tasks required the reader to critique this thinking in

some way (37.6% of tasks fulfilling Criteria #1 and #2). From the criteria analyses, I identified 127 tasks out of 4802 total assessment tasks from curriculum-based assessment tasks (2.6% of the total number of tasks, 37.6% of the tasks that fit Criterion #2) as student work assessment tasks (SWAT).

Analyses of the 127 SWAT based on assessment types, CCSSM content strands, evidence types, and critique types revealed that SWAT were most common in CPM and rarely found in Big Ideas. Also, the SWAT tended to occur on summative assessment tasks or question bank tasks. None of the SWAT were from diagnostic assessments. The full set of SWAT often focused on the CCSSM content strand related to the number system and rarely focused on ratios and proportional relationships. The SWAT also most often required the reader to make sense of someone's words and rarely required the reader to engage in making sense of someone's visual representations. Lastly, the most common critique type represented in the SWAT was evaluation and none of the SWAT required the reader to make a choice based on preference.

Analyses of the SWAT curriculum series by assessment types, CCSSM content strands, evidence of student thinking types, and critique types revealed a number of significant findings. First, the curriculum series SWAT were extremely varied by distributions of assessment types, influenced by the major differences in the types of assessments included in each curriculum series and the distribution of tasks across these series. Second, three of the series, Big Ideas, Go Math, and CMP, varied the most from the full set of SWAT in terms of distributions of content strands, evidence types, and critique types. Third, the two other curriculum series, CPM and Eureka SWAT, were the most similar to the findings for the full set of SWAT in regard to content strands, evidence types, and critique types.

The most significant findings from these text analyses were found when I compared the instances of student work tasks that appeared in the curriculum-based assessments (SWAT) to the student work tasks in the student textbook (SWTT). First, the SWTT as a full set and across curriculum series included all five CCSSM strands and a more balanced coverage of the strands than the full set and individual curriculum series SWAT. Second, not all evidence types were represented in each curriculum series SWAT, but all four evidence types were represented in the set of SWAT, the set of SWTT, and each curriculum series SWTT. Across and within the SWAT and SWTT, the common evidence types were words and symbolic reasoning.

The curriculum series SWTT and the full set of SWTT revealed more variety of critique types than their corresponding SWAT. CMP and CPM contained preference critique types in addition to the four other critique types that also occurred in Big Ideas, Eureka, and Go Math SWTT. In the curriculum series SWAT, one series, Big Ideas, only involved two critique types, and two series, CMP and Go Math, only involved three critique types. Lastly, evaluation was the most common critique type across and within a majority of the SWAT and SWTT and preference was the least common critique type across and within the SWAT and SWTT with the exception of CMP SWTT in which error identification and/or correction occurred less often than preference.

As a whole, SWAT in curriculum-based assessments were not very prevalent across the curriculum series. Even so, the opportunities students did have to engage with student work in the curriculum-based assessment tasks varied across and within curriculum series in terms assessment types, CCSSM content strands, evidence types, and critique types. The most significant differences across CCSSM content strands, types of evidence of student thinking, and critique types were found when making comparisons between SWAT and SWTT, or students'

opportunities to engage with student work in curriculum-based assessments and students' opportunities to engage with student work in student textbooks. I discuss these findings and their implications in Chapter 7.

CHAPTER 5: RQ #2: STUDENTS' EXPERIENCES WITH AND PERSPECTIVES ON SWAT

Overview

In this chapter, I present findings from interviews conducted with students focused on assessment tasks with and without student work (SWAT and non-SWAT). The goal of conducting these interviews was to gain insight into students' experiences with SWAT as compared to non-SWAT. I also wanted to be able to understand the nature of students' written work and verbal descriptions while solving SWAT and non-SWAT as two representations of students' thinking. Broadly, I wanted to explore whether or not students' written and verbal work provided evidence of students' engagement with SMP3b while solving SWAT.

First, I present findings from my analysis of students' written work for the types of evidence of thinking students included in their written work for SWAT and non-SWAT. These analyses provide insight into one type of representation students used to demonstrate their thinking on the curriculum-based assessment tasks. Traditionally, written responses are a predominant way in which students are asked to represent their thinking on assessment tasks. Then, I provide comparisons between students' written work and their verbal responses on the assessment tasks that revealed the evidence of thinking students expressed verbally but did not include in their written work for SWAT and non-SWAT. In other words, these findings offer insight into a second representation type that students used to demonstrate their thinking on the assessment tasks in the clinical interview. Lastly, I summarize students' descriptions of their experiences solving SWAT to highlight students' experiences making sense of someone else's thinking, connecting directly to students' engagement with SMP3b on SWAT.

I address the following research questions:

2. How do students talk about and make sense of SWAT?
 - a. What is the nature of students' written work and verbal responses on SWAT and non-SWAT?
 - a. How do students describe their experiences with SWAT?

RQ #2a: Nature of Students' Written and Verbal Work on SWAT and Non-SWAT

To answer Research Question #2a, which focused on the nature of students' written work and verbal responses on SWAT and non-SWAT, I analyzed portions of the student interview transcripts and students' written work. I explored students' written work for the types of evidence of thinking they included in their written solutions. I coded students' written work based on words, symbols, and drawings students produced for each task (see Table 3.3 for descriptions and examples of codes). Also, I analyzed the differences between students' written work and their verbal responses by coding the types of evidence of student thinking that were evident in students' verbal responses, but not in the written work (see Table 3.4 for descriptions and examples of codes). This allowed me to capture the additional student thinking students demonstrated in the clinical interview but was not represented basely solely on their written work. After I conducted these two types of analyses, I made comparisons between the SWAT and non-SWAT at the levels of tasks and students.

The two sets of findings together detail the types of evidence of thinking students provided in their written work and the types of evidence of thinking they included in their verbal responses for SWAT and non-SWAT. These findings also possibly provide some insight into students' understandings of the types of evidence of their thinking they should include in their written work for SWAT. However, I should note, that due to the clinical nature of the interview in which I asked students to "think aloud" while completing tasks, it is possible that students

expressed thinking verbally that they might typically include in writing in a traditional assessment environment in which they are not asked to “think aloud.” Therefore, these findings provide some insight into the potential benefits of allowing students to use multiple modes (i.e. written, verbal, actions) to represent their thinking that extend beyond the traditional written format, allow for flexibility in expressing different types of evidence of thinking, and provide a more robust picture of students’ thinking on assessment tasks.

Student Thinking in Students’ Written Work

I analyzed students’ written work for the types of evidence of student thinking evident in their written responses to compare student thinking on SWAT and non-SWAT. I reviewed the words, symbols, and drawings that students included as part of their written work and coded each case of written work for each student by task. I was interested in the types of evidence of student thinking that were evident in students’ written work, a traditional way in which students are often required to represent their thinking on mathematics assessments. In Table 5.1 below, I detail the number of instances for each code across all the tasks, the SWAT, and the non-SWAT. I bolded the codes that were most common for SWAT and non-SWAT. For SWAT, Reasoning was the most used code and for non-SWAT, Symbolic Work was the most common code.

Table 5.1

<i>Instances of Student Thinking Codes Evident in Students’ Written Responses</i>			
Code	All Tasks	SWAT	Non-SWAT
Drawing	1	0	1
Question	1	1	0
Reasoning	15	14	1
Statement	9	8	1
Symbolic Work	34	12	22
Uncertainty	1	1	0
All Codes	61	36	25

In Figure 5.1 below, I summarize the student thinking students demonstrated on their written work through their words, symbols, or drawings, by task, student, and type. The four shaded rows indicate SWAT.

Tasks	Anna	Cynthia	David	Ed	Jane	Susan
A	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work
B	Symbolic Work	Statement Question	Statement Symbolic Work	Statement Symbolic Work	Statement Symbolic Work	Statement Symbolic Work
C	Reasoning Symbolic Work	Statement	Reasoning Symbolic Work	Reasoning Symbolic Work	Reasoning Symbolic Work	Statement Symbolic Work
D	Drawing Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work
E	Symbolic Work	Reasoning	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work
F	Reasoning	Reasoning	Reasoning Symbolic Work		Reasoning	Reasoning
G	Symbolic Work	Statement	Symbolic Work	Symbolic Work	Symbolic Work	Symbolic Work
H	Reasoning Uncertainty	Statement	Reasoning	Reasoning Symbolic Work	Reasoning	Reasoning

Figure 5.1 Student Thinking Codes Evident in Students' Written Work Sorted by Task and Student

Figure 5.1 shows that most students included one to two types of evidence of student thinking in their written responses for each task. There was only one student on one task that did not include any type of written response and therefore showed no evidence of their thinking in their written work (Ed, Task F). Tasks B and C had the highest number of types of student thinking (11 each), followed by Task H (7). For Tasks A, E, and G, all students included one type of student thinking in their written work. For Task D, Anna included both a Drawing and Symbolic Work and for Task F, David included both Reasoning and Symbolic Work. Looking across students, students provided a similar number of types of evidence of student thinking across the tasks. Cynthia had the lowest number of types of evidence of student thinking in her written work (9), while Anna and David had the most (11 each). Ed, Jane, and Susan all included 10 types of evidence of student thinking in their written work across all the tasks. I also made comparisons of the types of evidence of thinking students included in their written work based on SWAT and non-SWAT tasks.

In Figure 5.2 below, all student results are combined by task in order to highlight trends for the types of evidence of student thinking students included in their written work by SWAT and non-SWAT. The number next to each type of student thinking indicates the number of students for which the code was applied. The bolded types of evidence of student thinking in each cell represent the most common type of student thinking that occurred in students' written work at the level of each task.

SWAT Tasks	Student Thinking	Non-SWAT Tasks	Student Thinking
B	Symbolic Work (5) Statement (4) Question (1)	A	Symbolic Work (6)
C	Reasoning (4) Symbolic Work (4) Statement (2)	D	Symbolic Work (6) Drawing (1)
F	Reasoning (5) Symbolic Work (1)	E	Symbolic Work (5) Reasoning (1)
H	Reasoning (5) Statement (1) Symbolic Work (1) Uncertainty (1)	G	Symbolic Work (5) Statement (1)

Figure 5.2 Student Thinking Codes Evident in Students' Written Work Sorted by Task and by SWAT and Non-SWAT

For the SWAT, the most common codes for the types of evidence of student thinking found in students' written work were Reasoning, Symbolic Work, and Statement, in descending order. Reasoning was the most common code for Tasks F and H, Symbolic Work was the most common code for Task B, and Reasoning and Symbolic Work occurred the same number of times for Task C.

For the non-SWAT, the most common code was Symbolic Work. This code accounted for 22 of the 25 instances of types of evidence of student thinking for the non-SWAT. The other three instances included one instance each of Drawing, Reasoning, and Statement. Thus, Symbolic Work was also the most common code for the individual non-SWAT tasks.

Additional Student Thinking in Students' Verbal Responses

In order to make sense of the additional student thinking students provided in the clinical interview in the form of their verbal responses on SWAT (Tasks B, C, F, and H) and non-SWAT

(Tasks A, D, E, and G), I compared students' written work to their verbal responses on the tasks. I analyzed the mathematical thinking that was evident in students' verbal responses but did not appear in their written work. These analyses allowed me to explore a second type of representation that students used to demonstrate their mathematical thinking – their verbal reasoning – without revisiting mathematical thinking evident in students' written work, in order to avoid redundancy. These analyses provided a fuller picture of students' mathematical thinking on both SWAT and non-SWAT. Instances of additional student thinking were identified by comparing students' written work to the transcript portions for each task in the clinical interview. After I identified the instances of mathematical thinking represented in students' verbal responses but not in their written work, I coded the instances using emerging codes at the level of each task.

As a reminder, I was interested in the types of evidence of student thinking that existed in students' verbal responses in comparison to their written work, not the frequency of these types for individual students. Therefore, in my coding scheme, each task for each student could only receive one instance of each code. In Table 5.2 below, I detail the number of instances for each code across all the tasks, the SWAT, and the non-SWAT. I bolded the codes that were most common for SWAT and non-SWAT. For SWAT, Justification was the most common code and for non-SWAT, Operation/Method was the most common code.

Table 5.2

<i>Instances of Student Thinking Codes Evident in Students' Verbal but not Written Responses</i>			
Code	All Tasks	SWAT	Non-SWAT
Claim	12	10	2
Justification	17	11	6
Operation/Method	26	8	18
Plausibility	2	0	2
Previous Experience	6	2	4
Questioning	6	4	2

Table 5.2 (cont'd)

Revised Thinking	6	3	3
Tools	7	5	2
Uncertainty	11	9	2
All Codes	93	52	41

In Figure 5.3 below, I summarize the types of evidence of mathematical ideas students verbally expressed but did not write down in their written work by task, student, and type. The four shaded rows indicate SWAT.

Tasks	Anna	Cynthia	David	Ed	Jane	Susan
A	Operation/Method Revised Thinking	Justification Operation/Method	Operation/Method	Operation/Method	Operation/Method	Operation/Method
B	Justification	Tools Uncertainty	Justification Revised Thinking	Operation/Method	Claim Justification	Claim Operation/Method Revised Thinking
C	Claim Uncertainty	Claim Prior Experience Uncertainty	Claim Tools Uncertainty	Claim Justification Operation/Method	Claim Justification Operation/Method	Operation/Method
D	Prior Experience	Operation/Method Prior Experience Uncertainty	Operation/Method Revised Thinking	Operation/Method Revised Thinking	Operation/Method	Justification Operation/Method
E	Justification Prior Experience Uncertainty	Justification Questioning Tools	Operation/Method Tools	Justification Operation/Method Prior Experience Questioning	Operation/Method	Justification Operation/Method
F		Claim Questioning	Claim Operation/Method Questioning Revised Thinking Tools Uncertainty	Prior Experience Uncertainty	Claim Justification	Uncertainty
G	Operation/Method	Operation/Method	Plausibility	Operation/Method Plausibility	Claim	Claim
H	Justification	Justification Tools Questioning	Justification Operation/Method Questioning Tools Uncertainty	Justification Operation/Method	Justification	Uncertainty

Figure 5.3 Student Thinking Codes Evident in Students' Verbal but not Written Responses Sorted by Task and Student

As Figure 5.3 shows, few additional types of evidence of students' thinking occurred verbally for Tasks A and G. In comparison, many additional types of evidence of students' thinking occurred for Tasks C and E. Alternatively, David and Cynthia demonstrated a lot of their thinking verbally that was not evident in their written work as compared to Anna who did not demonstrate a lot of additional thinking verbally as compared to her written work. While these findings provide interesting trends at the levels of individual tasks and students, in order to

answer my research question #2a, it was essential for me to compare the completed coding of evidence of student thinking evident in students' verbal responses but not in their written work for SWAT and non-SWAT tasks.

In Figure 5.4 below, I combine all the types of evidence of student thinking results by task in order to highlight trends for the types of evidence of student thinking that appeared in students' verbal responses but not their written work by SWAT and non-SWAT. The number next to each type of student thinking indicates the number of students out of six total for which the code was applied. The bolded student thinking types for each task highlight the most prevalent type of student thinking that occurred in students' verbal responses but not their written work at the level of each task.

SWAT Tasks	Student Thinking	Non-SWAT Tasks	Student Thinking
B	Justification (3) Claim (2) Operation/Method (2) Revised Thinking (2) Tools (1) Uncertainty (1)	A	Operation/Method (6) Justification (1) Revised Thinking (1)
C	Claim (5) Operation/Method (3) Uncertainty (3) Justification (2) Prior Experience (1) Tools (1)	D	Operation/Method (5) Prior Experience (2) Revised Thinking (2) Justification (1) Uncertainty (1)
F	Claim (3) Uncertainty (3) Questioning (2) Justification (1) Operation/Method (1) Prior Experience (1) Revised Thinking (1) Tools (1)	E	Justification (4) Operation/Method (4) Prior Experience (2) Questioning (2) Tools (2) Uncertainty (1)
H	Justification (5) Operation/Method (2) Questioning (2) Tools (2) Uncertainty (2)	G	Operation/Method (3) Claim (2) Plausibility (2)

Figure 5.4 Student Thinking Codes Evident in Students' Verbal but not Written Responses Sorted by Task and by SWAT and Non-SWAT

For the SWAT, the most common codes for types of evidence of student thinking that were evident in students' verbal responses but not in their written work across the full set of SWAT were, in order, Justification, Claim, and Uncertainty. These findings are upheld based on reviewing the data at the level of individual SWAT. Justification was the most common code for

Tasks B and H, Claim was the most common code for Task C, and both Claim and Uncertainty were the most common codes for Task F.

For the non-SWAT, the most common codes for evidence of student thinking that was evident in students' verbal responses but not in their written work across the non-SWAT were, in order, Operation/Method and Justification. Again, these findings are reinforced at the level of the individual non-SWAT. Operation/Method was the most common code for Tasks A, C, and G and both Operation/Method and Justification were the most common codes for Task E.

My analyses of students' written and verbal work allowed me to investigate two different types of representations of student thinking that students in my study used to express their mathematical thinking on the eight curriculum-based assessment tasks. Analyses of students' written work allowed me to characterize the types of evidence of thinking students used to demonstrate their thinking and provided insight into the types of evidence of thinking that students might traditionally express in writing on assessment tasks. Analyses of students' verbal reasoning allowed me to characterize the types of evidence of thinking students expressed verbally during the clinical interview and provided insight into an additional mode of representation of thinking that students could possibly use to demonstrate their mathematical thinking on assessment tasks. Taken together, these two different types of representations of evidence of student thinking – written and verbal – could possibly present a more robust picture of students' thinking on assessment tasks. More specifically, for assessment tasks intended to assess students' abilities to analyze someone else's mathematical thinking, multiple modes of representations could allow students more opportunities to demonstrate SMP3b on written assessment tasks. These ideas are further discussed in Chapter 7 and explicitly connected and compared to teachers' perspectives on SWAT as a possible mechanism for assessing SMP3b.

RQ #2b: Students' Descriptions of Experiences with SWAT

In order to answer Research Question #2b focused on students' descriptions of their experiences with SWAT, I analyzed the summary table for the semi-structured portion of the student interview for key themes. More specifically, I concentrated on summarized ideas and students' responses focused on the SWAT and related to a specific set of questions about the SWAT included in the interview (see Appendix E, Questions #7a, 7b, and 7c). From these analyses, two main types of themes emerged as students (1) noticed that SWAT required different processes and final written work products than the non-SWAT, and (2) identified key features or attributes of SWAT. I detail these themes in the paragraphs that follow.

SWAT Required Different Thinking Processes and Final Products

Four of the six students talked about how the four SWAT in the interview required them to think differently and/or produce different written solutions than the non-SWAT tasks or other math assessment tasks, more generally. Cynthia discussed how the SWAT required her to think about the problem differently. When asked about her experiences solving assessment tasks that required her to make sense of someone else's thinking, Cynthia expressed her preference for non-SWAT stating, "I don't really like these kind[s] of tasks because you have to find their equations and what they said and see whether they are right or wrong. I would rather just get the equation and just do it." Cynthia favored "simple questions" where she was only required to "do it" as opposed to tasks that required her to evaluate someone else's work.

Anna, Ed, and Jane not only talked about the SWAT requiring different thinking in order to complete the assessment tasks, but also discussed how these tasks required them to produce different types of written work for their assessment solutions. Anna made comparisons between SWAT and "just math" or "actual math" assessment tasks. She stated, "For those [non-SWAT],

it's mostly just all math instead of changing your thinking for these [SWAT]." She elaborated, "I had to change my thinking and think how they would think and what they did. I had to ... think different than I usually do." She further detailed, "I think different than how the people, the examples in the paper kind of showed" which required her to change her own thinking in order to make sense of the way "they showed it."

Beyond having to think differently on the SWAT, Anna also talked about differences in what she was expected to produce as written work for the SWAT as compared to "just math" assessment tasks. She stated, "I have to write an actual response" as compared to a non-SWAT in which, "I would just have to solve it." In reflecting on her experiences solving SWAT and non-SWAT, Anna concluded, "I'm more good at actual math instead of writing responses to their answers." Anna saw a clear distinction between the SWAT, tasks that required her to "write an actual response", and non-SWAT, tasks that were "actual math" tasks based on both the mathematical thinking she had to complete when solving these two types of tasks as well as the expectations for the types of written work she should produce.

Similar to Anna, Jane talked about how the SWAT were "not really about what you do, it's more about what someone else did." She stated, "It's kind of like trying to think like someone else ... because it's like analyzing other people's work and seeing if they are right or wrong." She noted, "It's not asking, 'Why do you think you're right? Explain.' It's more, 'Are they right or wrong? Explain why they are right or wrong.'" In other words, instead of defending her own solution to a mathematical problem, Jane detailed how SWAT required her to review someone else's mathematical work to determine correctness and then defend her reasoning for her conclusions about the correctness of the work. Jane's observations reveal an awareness of the

SWAT requiring different thinking processes for solving problems as well as different results for written solutions based on different questions asked of the reader.

Instead of discussing how the SWAT might require him to think like someone else as discussed by Anna and Jane, Ed talked about how the SWAT required him to extend his own thinking. He stated, “I feel like it’s a way to push your thinking, not just solve this problem. Another person solved this problem, evaluate if they were correct or not.” He elaborated by making comparisons to non-SWAT:

If they just tell you to do the problem, then you just do the problem and find the answers.

But, if they tell you that a person did this problem and you have to explain if they are right or not, then you have to do the problem and then, you have to think about it even more to be able to answer the second question.

Ed described how SWAT not only required you to do the problem like any other kind of problem, but you also have to complete a second part of evaluating someone else’s work on the problem. Ed’s description of the SWAT also revealed that he felt that the SWAT required him to “explain if they are right or wrong”, which has implications for what his written responses to these types of tasks would entail in comparison to tasks where he only had to “find the answer.” Students’ descriptions of the SWAT revealed that most of the students viewed the SWAT as different from the other assessment tasks based on what mathematical thinking the task required as well as the written work students were asked to present.

Attributes of SWAT

Based on students’ reflections about their experiences solving assessment tasks in the clinical interview, a number of ideas related to the attributes or features of SWAT emerged: (1) the complexity of SWAT in comparison to other assessment tasks, (2) how SWAT reflected

students' experiences in their mathematics classrooms, (3) the use of imagination when solving SWAT, (4) the need for comparison when solving SWAT, and (5) the possibility of helping someone else as a motivating feature of SWAT for students. In the paragraphs below, I detail students' responses related to these main ideas concerning features or attributes of SWAT.

Complexity. All six students discussed how the SWAT were more complex or more complicated than the other assessment tasks they solved. In Figure 5.5 below, I detail the tasks that students felt were easiest and hardest of the eight. I have maintained the order in which students stated the ease or difficulty of the tasks and I have also bolded SWAT.

Task Type	Anna	Cynthia	David	Ed	Jane	Susan
Easiest	B , A, G	A, E, B	A, B , G, D	F	A, G, D, E	A, G, B , F
Hardest	C , E	C , D, H	F , H , C	H , C	F	D, C , E, H

Figure 5.5 Students' Designated Easiest and Hardest Tasks

Tasks C and H were considered the hardest tasks by a majority of the students. Based on the task coding framework previously discussed, these two tasks both required the students to evaluate the correctness of at least one student's work. Task H also required the students to make comparisons among multiple instances of student work and make a choice based on preference (see Chapter 3, RQ #2 for more information about the SWAT task features). Susan described why she thought Task H was difficult stating, "I had to read through all of them to see which I thought was correct." Similarly, Ed said, "Like H, I know people that would look at that problem and would give up immediately, wouldn't even read the words, just look at the math and quit."

Two students, David and Jane, also considered F to be one of the hardest tasks. David felt that Task F was a "complicated problem" and Jane noted that the task "was more in-depth" and required her to consider "all the different possibilities." She continued, "You have to think about all the instances about adding linear expressions to see if there is anything that contradicts or goes against your friend's explanation" revealing why she felt that this task was the hardest.

Task B was not considered one of the hardest problems by any of the students. A number of students discussed how the task revealed that Paco had made an error in his work, so the students already knew this and did not have to determine if he had made error just where he had made an error, which they discussed was an easier task. Even so, when reflecting on the four SWAT, students talked about these tasks as more complex. Anna felt that the SWAT were “more elaborate” than the other tasks because, “I have to write an actual response.” Echoing Susan and Ed’s reaction to Task H, Cynthia felt overwhelmed by what was provided in the SWAT saying, “I see a whole bunch of stuff” when comparing SWAT to “simple problems” that just required the reader to “do it.”

When asked his thoughts on SWAT, Ed stated, “You have to think about it even more” and Jane said, “It’s basically those, but just explain why they are right or wrong” expressing the idea that explanation component of the tasks was one of the main differences between the SWAT and the non-SWAT.

Connection to Classroom. Three students discussed how the SWAT were similar to experiences or tasks they had encountered in their mathematics classrooms. When discussing Task B, Susan stated, “This one we do in class.” In her experience, she had seen similar types of tasks involving looking at someone’s symbolic work and finding an error. Similarly, when discussing Task C, Ed stated, “We’ve done problems exactly like this where it’s like someone already tried to put in the values of x and you have to determine whether they were right or not.” Ed connected how the task required him to determine whether the work was correct or not to problems he had previously seen in which he had to evaluate someone’s work.

When discussing her experiences with the SWAT, Jane stated:

Well, they're kind of like what we do in math class a lot. In math class, our teacher doesn't make us memorize stuff, so we discuss stuff a lot. For basically the whole class, we spend it going over the warm-up and she really asks us to ask questions and push our thinking a lot, think about why it's right or wrong.

Jane detailed experiences from her math class and connected practices in class to the SWAT. She discussed how questioning and evaluation were major parts of her mathematics classroom experience.

Imagination. Three students discussed imagination when reflecting on their experiences with the SWAT. When describing Task F, Anna stated, "This one is imagination, it's in your head." In order to solve Task F, Anna felt that she had to use her imagination and think about the task "in [her] head" in order to complete the task. Reflecting on his experiences solving SWAT, David said, "I have an imagination, so I have to sit there and imagine somebody sitting there working on a problem." When he encountered a problem that showed someone's thinking, he had to "pause and think about it." Referencing a feature of the criteria for student work (see Appendix B) in his discussion of SWAT, David said, "For some reason, every single time I hear a name, I have to think of a person." He elaborated:

In a math problem, I think of a person ... just a person in a school trying to figure out this problem looking down at their paper. I'm always behind them watching them trying to figure out the problem. And so, it gives me a minute to think about it.

He talked about how, on tasks that included a person, he had to slow down and actually visualize the person in the task.

Similarly, Ed expressed, "To me, it's like these problems about your friend, you can imagine yourself going up to them and going, "Hey dude, can I show you how to do this?" rather

than just being “Solve the problem.” He discussed how the tasks “about your friend” or other tasks that included people involved imagining showing someone how to do the problem instead of just solving the problem for yourself.

Comparison. Two students, Anna and Susan, talked about how the SWAT required them to make comparisons between their work and the work provided in the tasks. In reference to Task B, Anna said that the task involved, “comparing my work to his” to determine where Paco made his error. When discussing all the SWAT, Anna stated that for these tasks, “you would have to compare your work to theirs” in order to complete the assessment tasks. Similarly, in describing her process for solving SWAT, Susan said, “I compare my thinking to their thinking to see if I’m correct or if they are correct.”

Motivation. Lastly, one student, Ed, talked pretty extensively about how SWAT could be motivating for some students. Ed said, “I know some very kind people that aren’t that into doing work so maybe stuff like that would help motivate them.” He elaborated:

I think some people could think about it as their friend did a problem, but you notice that they did it wrong, so you’re trying to help them. So, that might motivate them to try to do the problem because there are some people that just don’t want to do it because it looks difficult.

Ed felt that tasks that required students to make sense of someone else’s work and determine correctness could be motivating for students because they would be helping the person in the problem. In talking about himself, Ed stated, “I care about my friends and if they got something wrong, I would want to help them with it.” Ed personally felt that helping others correct their mistakes on a mathematics problem was motivating.

He also contrasted tasks with a person, which he called “story problems” (he identified Tasks A, B, C, F, G, and H as “story problems”), to tasks without a person. He said, “If it’s just a problem, it might be difficult for [students] to look at because there is no one to connect to.” Thus, Ed thought that tasks with people in them allowed students to connect to the people in the tasks. More specifically, tasks that asked students to correct an error in someone’s work, a type of SWAT, could be motivating for students because these tasks asked them to help the person in the problem and, according to Ed, helping others can be motivating.

Summary of Chapter Findings

In this chapter, I presented findings from analyses of students’ written work on assessment tasks and student interviews I conducted to explore students’ experiences with and perspectives on SWAT. The findings from the analysis of students’ verbal reasoning on tasks and their written work and the analysis of students’ written work revealed insights into (1) the types of representations students used to demonstrate their mathematical thinking, (2) the evidence types students used within various representations (e.g. written and verbal), and (3) the types of evidence of thinking students demonstrated on tasks verbally but were not evident in writing on the page. This third point, as previously stated, is not a very robust one due to the clinical interview methods used in this study as opposed multiple assessment settings involving multiple representations of student thinking.

Even so, comparing SWAT and non-SWAT revealed students often verbalized lots of thinking not captured in writing for SWAT. Students included different types of evidence of thinking in their written work for SWAT and non-SWAT. Also, the types of evidence of thinking students verbalized but did not include in their written work differed for SWAT and non-SWAT. As a whole, analyses of the two different types of representations of student thinking captured in

the clinical interviews – written and verbal – revealed possible benefits of allowing for multiple modes of representations of thinking on assessment tasks in order to capture a more robust picture of students’ mathematical thinking on written assessment tasks.

The findings from my analyses of students’ discussions of their experiences solving tasks that required them to make sense of someone else’s thinking (SWAT) revealed that students noticed SWAT required different processes and/or final written work products when compared to non-SWAT. Students also identified key features of attributes of SWAT including: (1) the complexity of SWAT in comparison to non-SWAT, (2) the SWAT were similar to tasks students had seen in their classrooms, (3) the SWAT encouraged students to use their imagination, (4) the SWAT required students to make comparisons between instances of student thinking, and (5) the SWAT could be motivating tasks for some students. I discuss these findings and their implications in Chapter 7.

CHAPTER 6: RQ #3: TEACHERS' EXPERIENCES WITH AND PERSPECTIVES ON SWAT AND STUDENTS' WORK ON SWAT

Overview

In this chapter, I present findings from teacher interviews focused on teachers' experiences with and perspectives on curriculum-based assessment tasks with and without embedded student work as well as actual students' work on the tasks. The goal of these interviews was to gather information about teachers' viewpoints on student work as a mechanism for assessing SMP3b and insights into what evidence of student thinking would be required for students to include in their written responses to demonstrate SMP3b on an assessment task. First, I detail teachers' thoughts on what each task was assessing. Then, I discuss how the teachers described SWAT. I summarize teachers' perspectives on the advantages and disadvantages of SWAT based on explicit interview questions and teachers' responses. Also, I present themes that emerged from a broader examination of teachers' descriptions of SWAT and general discussions of student work embedded in assessments tasks.

Lastly, I track how teachers' ideas about student work as related to SMP3b evolved or were reinforced across the interview phases. In doing so, I detail whether or not teachers discussed SMP3b or related habits of mind when reviewing assessment tasks initially – before I made any explicit mention of SMP3b. I refer back to my analyses conducted for Research Questions #3a and #3b in which teachers discussed what tasks were assessing and the advantages and disadvantages of using student work in assessment tasks, which revealed how teachers referred to SMP3b or related habits of mind prior to my mentioning this practice. Next, I report on teachers' ideas about whether or not student work can assess SMP3b, an explicit question I asked after exploring teachers' thoughts on SWAT. Finally, I summarize teachers' perspectives

on student work as a mechanism for assessing SMP3b after they reviewed students' written work on SWAT tasks.

I address the following research questions:

3. How do teachers talk about and understand SWAT and non-SWAT based on the tasks and students' written work on the tasks?
 - a. What do teachers think that SWAT and non-SWAT assess?
 - b. How do teachers describe SWAT?
 - c. What evidence of SMP3b do teachers notice in SWAT based on both the written assessment tasks as well as students' written work on the tasks?

RQ #3a: What SWAT and Non-SWAT Assessed

During the teacher interviews, I asked teachers to tell me what each of the eight assessment tasks assessed (see Appendix F for Assessment Tasks). I asked, "Looking at each task individually, what is each task assessing?" I asked this question towards the beginning of the interview when I had not yet mentioned student work or SMP3b. I coded teachers' responses to this question at the level of each task using four category codes: Content, Strategies, Related to SMP3b, and Other (see Table 3.5 for code descriptions). In Figure 6.1 below, I have summarized the category codes for what tasks were assessing by task and by teacher. In the figure, bolded items are included in the "Related to SMP3b" category and italicized items are included in the "Other" category. For these two categories, I preserved the language and/or individual ideas that teachers discussed in order to highlight the different ideas that emerged for these categories based on teachers' descriptions. The shaded rows are SWAT and the non-shaded rows are non-SWAT.

Tasks	Teachers					
	Edwards	Gables	Gilbert	Henderson	Quinn	Shirley
A	Content Strategies	Content Strategies	Content Strategies	Content	Content Strategies	Content Strategies
B	Content SMP3	Content Strategies Error analysis	Content Strategies Error ID/Correct	Content Strategies Error ID	Content Error analysis	Content Critique
C	Content SMP3	Content Strategies	Content	Content Error Analysis	Content	Content <i>Perseverance</i>
D	Content Strategies	Content	Content Strategies	Content	Content	Content
E	Content	Content Strategies	Content Strategies	Content Strategies	Content Strategies <i>Justification</i>	Content Strategies
F	Content SMP3	Content Strategies Analyzing someone's claim	Content Strategies	Content Strategies <i>Generalization</i>	Content Strategies <i>Proof</i>	Content <i>Generalization</i>
G	Content <i>Language</i>	Content	Content Strategies	Content <i>Reading</i>	Content <i>Language</i>	Content Strategies <i>Perseverance</i>
H	Content SMP3	Content Strategies	Content Analyzing student work	Content <i>Perseverance</i>	Content Strategies Make sense of someone's work Choice	Content Critique <i>Perseverance</i>

Figure 6.1 Summary of Teachers' Descriptions of What Interview Assessment Tasks Assessed

Teachers mostly described tasks as assessing content and strategies. All teachers for every task discussed the mathematical content that the task assessed. While not as prevalent as content, teachers often described specific strategies or solution pathways students might use to complete a particular task. Teachers also talked about some of the tasks assessing things other than content and strategies. Ideas related to SMP3b were evident on SWAT, but not on non-SWAT. Ideas that did not fit into the first three categories (content, strategies, and related to SMP3b) also emerged for a few of the tasks.

As shown in Figure 6.1, teachers often identified that the SWAT tasks assessed SMP3b or related habits of mind. This was most evident on Tasks B and H and less so on Tasks C and F. Ms. Edwards explicitly used the term “mathematical practice” and quoted SMP3, “construct

viable arguments and critique the reasoning of others” when describing what Tasks B, C, F, and H assessed (CCSSI, 2010, p.6). The other teachers did not use the term “mathematical practice” when describing ideas related to SMP3b, but these five teachers did discuss similar habits of mind such as analysis of someone’s work, critiquing, sense-making, error identification, and choice. All six teachers identified at least two tasks as assessing ideas related to SMP3b. These findings show that the teachers in my study felt that assessment tasks could assess habits of mind, such as SMP3b.

Two teachers talked about tasks assessing perseverance. Both Ms. Henderson and Ms. Shirley felt that Task H assessed perseverance. Ms. Shirley stated, “the different ways of doing it would require some perseverance” when reviewing the three student strategies the reader would encounter on Task H. Ms. Shirley expressed that Tasks C and G also assessed perseverance. For Task E, Ms. Quinn felt that the task assessed students’ abilities to provide a justification or “how to justify.”

For Task F, three teachers, Ms. Henderson, Ms. Quinn, and Ms. Shirley, felt that the task assessed generalization or proof. Ms. Quinn stated that Task F assessed students’ understanding of “sufficient proof.” Lastly, for Task G, three teachers, Ms. Edwards, Ms. Henderson, and Ms. Quinn, talked about the task assessing language or reading. Ms. Henderson stated, “this is more about reading than math because it's really playing a game with these things that are pretty meaningless” when describing what Task G assessed.

RQ #3b: Teachers’ Descriptions of SWAT

Advantages and Disadvantages of SWAT

In the teacher interviews, teachers were explicitly asked about the advantages and disadvantages of the use of embedded student work in assessment tasks prior to responding to

any questions related to the use of student work as a mechanism for assessing SMP3b. Teachers' descriptions of the advantages and disadvantages of student work embedded in assessment tasks were gathered and compared across teachers. In Figure 6.2 below, I have summarized the main points that teachers made about the advantages and disadvantages of SWAT. I describe each of these points in the paragraphs that follow.

Advantages of SWAT	Disadvantages of SWAT
• Promote robust student understanding • Expose students to different strategies • Connect to the Standards for Mathematical Practice (SMPs) • Task Specific ○ Task B – error analysis ○ Task C – error analysis; credibility of symbols ○ Task F – generalizability	• Difficulty of assessing students' understanding of someone else's work on a written assessment • Unfair if not mirrored in classroom practice • Inability to ask questions • Possibly overwhelming for students (most specifically Task H)

Figure 6.2 Advantages and Disadvantages of SWAT

Advantages of SWAT. Teachers described a number of advantages of using tasks with embedded student work on assessments. First, Ms. Edwards and Ms. Gables described how SWAT promote robust student understanding. Ms. Edwards said, “It's that level of understanding. You're truly getting a good level of understanding of that concept if [students] can see that there's multiple ways to get there.” She talked about how using student work on an assessment can help the teacher determine if students “know more than one way to solve a problem” which would show that they “truly understand the concept.” Ms. Gables stated,

I think just the advantage here is bumping up the expectations and bringing kids to a higher level. You know, like that wealth of knowledge thing and ... Bloom's Taxonomy in bumping up the level of rigor. I think that is a definite advantage.

She expressed how SWAT were more rigorous than many of the non-SWAT tasks because students had to think differently. She also said that SWAT “would give you a deeper insight into

what kids know compared to [Task A] I think is the lower level.” So, SWAT promote robust student understanding of mathematics, but possibly provide opportunities for teachers to gain deeper insight into students’ understandings of mathematics.

Another advantage of SWAT expressed by two teachers was that these tasks expose students to different strategies. Ms. Gilbert said that students “can see a different strategy that might be different from what they would have used and see how it works.” So, the use of student work on an assessment could show students different strategies and how they work. Ms. Gables suggested, “If kids had not seen one of these solution strategies before, they might be open to thinking more about it or using it in the future. So, I think it would ... open their minds up to other possibilities.” Ms. Gables’ observations suggest that SWAT can not only expose students to new strategies, but also encourage students to use different strategies in the future.

Three teachers expressed how SWAT connect to the Standards for Mathematical Practice (SMPs) (CCSSI, 2010). Ms. Henderson said that it was advantageous to use SWAT “because one of the mathematical practices is can you confront someone else’s thinking and critique it.” Similarly, Ms. Quinn simply stated, “It’s one of our practice standards.” In other words, a clear advantage of SWAT to Ms. Quinn was that, in her view, these tasks “fit” a practice standard. Ms. Shirley echoed the other two teachers, “You know, you’re supposed to be making arguments and critiquing the reasoning of others and that’s engaging them in the mathematical practices.”

Teachers also identified advantages of SWAT at the level of individual tasks. For Tasks B and C, which involved identifying where an error occurred in someone’s work and determining the correctness of a method, respectively, Ms. Gilbert stated, “It definitely highlights common errors that other students are making and they can now go through and see what that error is and try to correct it.” So, Ms. Gilbert viewed Tasks B and C as exposing students to

common errors and providing opportunities for students to identify and correct errors. For Task C, Ms. Quinn also expressed that an advantage of this task was that it gave credibility to symbols as a way of communicating mathematics. She stated:

[Task C is] just interpreting the symbols, which I think gives credibility to symbols. It gives that idea of communicating mathematics this way and being able to get information from it and draw conclusions from it. I think it's pretty powerful! I don't think that if you just have to do your own work, I don't think that that pushes a kid as hard in terms of being able to communicate with symbols than looking at someone else's symbolic work.

In her view, Task C promoted symbolic representations as a valuable way of communicating mathematical ideas. Lastly, Ms. Gables discussed how Task F “[gave students] the opportunity to generalize ideas because I think that’s important.”

Disadvantages of SWAT. The teachers in my study also expressed disadvantages of SWAT. Three teachers discussed the difficulty of assessing students’ understanding of someone else’s work on a written assessment. Ms. Henderson expressed, “It’s hard to test how someone undoes or does someone else’s thinking. It’s easier to assess how someone is thinking about their own thinking!” Similarly, Ms. Shirley anticipated, “There would be lots of students who might do that and I still don’t know whether they’ve memorized steps of whether they understand what’s going on.” Both of these teachers expressed that it would be difficult to gain insight into what students actually understand about someone else’s mathematical thinking on a written assessment task. Ms. Gables expressed that she did not think SWAT had a disadvantage, but said, “I guess, for people who are more traditional, it would take more time to grade.” So, perhaps SWAT would require additional time to grade in comparison to traditional written assessment tasks.

Three teachers felt that a disadvantage of SWAT was that using these types of tasks on an assessment would be unfair to students if they were not exposed to student work during their mathematics classes. Ms. Edwards argued:

If you don't have students share out their work and have those deep discussions about what different students are doing or that error analysis, I think that would make it very challenging for an assessment. Because, again, they might be looking at, well, this is how it has to be solved, and then they're not seeing that bigger picture of, there's different ways to go about solving the problem. I think that would be the biggest disadvantage, that idea of not having that prior knowledge themselves of looking at other students' work.

Similarly, framing assessments as expectations, Ms. Gables stated:

If the kids weren't prepared for something like this, it would be really challenging for them. So, I think the opportunities that kids have in the classroom have to match the expectations. You can't just be going through problems without these conversations before a test and then giving them this task.

Lastly, Ms. Quinn expressed, "It depends on how kids are engaging in the classroom about similar types of tasks to see whether this would be a fair assessment." These three teachers expressed that it was essential for classroom practices and expectations on assessment tasks to mirror one another.

Ms. Quinn highlighted a unique disadvantage not discussed by the other teachers. She felt that a key disadvantage of SWAT in comparison to the ways her students interact with other student's work in class was that on a written assessment, students are unable to ask questions of the person who created the mathematical work. She stated:

Even when kids have a lot of practice doing that in the classroom, looking at this, when I've used questions like this in assessments, there's a little bit of anxiety about it because I don't know what they did wrong, and it says there's an error there and I don't know what they did wrong. And it stinks because the kid can't ask a question to the other kid. Therefore, while teachers can engage students in similar types of practices in the classroom and on written assessment tasks, the nature of SWAT does not provide the option for students to interact with the person who is responsible for the mathematical thinking presented in the assessment task. Even so, Ms. Quinn expressed that these types of tasks "add another level of complexity" because students do not have the ability to question the person in the task.

Lastly, two teachers talked about how SWAT could possibly be overwhelming for students. These two teachers specifically referenced Task H when describing this disadvantage. Ms. Edwards said, "Having too many student samples might also be an issue for some students. It might be time consuming depending on how many questions you have on your assessment." Similarly, Ms. Gilbert stated, "I know I would have students look at that and shut down just because it's so much [to read]." So, for SWAT, a disadvantage expressed by teachers was that too many instances of student work on an assessment task could be overwhelming for some students.

All six teachers in my study expressed advantages and disadvantages of SWAT both as a genre of assessment tasks and at the level of individual tasks in response to direct questions about advantages and disadvantages of SWAT. In the next section, I provide findings about the key themes that arose from teachers' discussions of SWAT and the use of student work in assessment tasks.

Key Themes about SWAT and Using Student Work in Assessment Tasks

In order to explore broader themes related to how teachers made sense of the SWAT used in the interviews (Tasks B, C, F, and H) as well as the use of student work in assessment tasks more generally, I revisited portions of the teacher interviews where these ideas were discussed and analyzed the transcript portions for emerging themes. The four main themes that emerged as teachers discussed student work in assessment were: (1) interdisciplinary practice, (2) connections to testing, (3) connections to curriculum and/or classroom, and (4) teachers as assessment writers. In the following paragraphs, I elaborate on these themes and highlight key points articulated by the teachers in my study.

Interdisciplinary Practice. Two teachers, Ms. Edwards and Ms. Quinn, discussed how student work embedded in assessment tasks related to content or practices their own students encountered in other school subjects. When discussing Task B, Ms. Edwards described the task as an error analysis task and detailed how students would be required to identify where the student in the task made his mistake and provide valid reasons for their solutions, which Ms. Edwards named “justifying.” She continued, “We try to use that term with kids like justify your reasoning, provide evidence to support it, and try to connect it back to what they’re doing in science and communications.” She also discussed how justification related to what her students were doing in their writing course work. She said, “when they write a paper, an argumentative paper, they have to provide evidence as to why they feel this way.” Extending beyond saying ‘show your work’, Ms. Edwards highlighted how teachers in her school, across various school subjects, were trying to be purposeful in using common language about justifying or providing evidence for solutions that connects to but is broader than the language provided in SMPs, specifically SMP3.

Ms. Quinn also discussed how justification was tied to students analyzing student work and how justification related to students' experiences in writing. She detailed how she and her colleagues were working on supporting students' developing ideas about what justify means across the middle school grade levels by developing rubrics that would scaffold students through their middle school experiences. She stated, "kids in sixth grade are beginning to write argument papers, so [justification] weaves so beautifully, why aren't we working on this together?" In making connections between justification in mathematics and argumentation in writing, Ms. Quinn saw an opportunity for her and her colleagues to work together to help students have an understanding of justification that extended across school subjects. Beyond writing, Ms. Quinn also highlighted how justification connected to science and social studies. Both Ms. Edwards and Ms. Quinn detailed the importance of justification in how students made sense of someone else's work on assessment tasks and also connected justification to practices promoted in other school subjects.

Connections to Testing. Four teachers discussed how SWAT and the use of student work in assessment tasks related to their experiences with state testing, standardized assessments, or national exams such as the SAT or ACT. Two teachers, Ms. Gables and Ms. Gilbert, talked about how tasks with student work were similar to the tasks that their students would be required to solve on state tests. When discussing initial reactions to Tasks B, C, F, and H, which she called analysis questions, Ms. Gables stated, "that's the kind of thing that kids are going to be expected to do when they take the state test." Similarly, Ms. Gilbert identified Tasks B and H as tasks found on the state test "where you've got to check other students' work." In preparing for the state test, Ms. Gilbert detailed:

To prepare for the [state test] in years past, we have gone through the books and pulled those types of questions, like [Task] H out of the ACE ... and then did a review of that before the [state test] where they had to look at other students' explanations and try to think through their thought process.

Therefore, both Ms. Gables and Ms. Gilbert saw assessment tasks with embedded student work as similar to some of the types of tasks their students would be required to complete on their state test. Ms. Gilbert also detailed how, in order to prepare for the state test, student work tasks from the student textbooks were used as review tasks to help students make sense of someone else's mathematical work and think about how students would respond on the state test.

Two other teachers, Ms. Edwards and Ms. Henderson, also talked about connections to testing as related to SWAT. They discussed the difficulty of preparing for standardized assessments that included SWAT as well as the difficulty of writing assessment tasks, especially standardized assessment tasks, that assess mathematical practices, or more specifically a particular practice, students' abilities to make sense of someone else's mathematical thinking. Ms. Edwards stated, "If you're not clear, and this is why standardized testing can be very challenging because you don't know what type of questions they're going to ask and what pieces... students may not clearly understand the question." She felt that the prompt for this task, which required students to choose which student strategy made the most sense to them, would not elicit responses from students that would show that they actually engaged in making sense of all the student strategies in the task. And, if this type of task were to appear on a standardized assessment, Ms. Edwards thought it would be difficult for students to know what to include in their written responses for their solutions to be considered sufficient. This teacher highlighted that whether or not students engaged in the practice of making sense of the student strategies on

Task H might not always be apparent based on the evidence that students provide in their written work for the task.

When discussing her final thoughts on the assessment tasks and student work embedded in assessment tasks, Ms. Henderson expressed how standardized assessments, in her opinion, did not currently assess the depth of what students do in mathematics classrooms. She stated, “we say we value all this collaboration with kids, but we only test them individually at a computer and take away some of their tools” and “we nationally have this big weight on these SAT and ACT kinds of tests that are really closed and narrow.” She felt that standardized assessments did not give students an opportunity to really show all they knew mathematically, but she also acknowledged, “it’s really hard to write [assessments] that get at the complexity of student thought.” She further stated, “I don’t know that we can ever write assessment tasks that really cover the depth of what kids know.” With these statements, Ms. Henderson highlighted the limitations of written assessments as a means for eliciting students’ understandings of broader mathematical practices beyond rote skills or basic knowledge and she questioned the individualized and restrictive nature of current standardized testing environments that are highly consequential for students.

Overall, four of the six teachers saw connections between student work in assessment tasks and large-scale testing. Two teachers, Ms. Gables and Ms. Gilbert, talked about how SWAT looked like the types of tasks that their students would encounter on their state tests. When discussing Task H, Ms. Edwards felt that the use of this type of a task on a standardized assessment would be difficult, because the instructions for students were unclear as far as what evidence was required in their written responses in order to sufficiently answer the question. Lastly, Ms. Henderson wrestled with whether or not written assessments could assess complex

mathematical practices and criticized current high-stakes assessments for their narrowness and stark contrast from the types of classroom experiences currently being promoted, such as student collaboration.

Connections to Curriculum/Classroom. All six of the teachers in my study discussed student work in assessment tasks as connected to the CMP curriculum and/or their classroom practices in a number of different ways. Three main ideas that emerged from teachers' descriptions of connections between SWAT and the CMP curriculum and/or classroom were (1) SWAT aligned with what teachers had found in the CMP curriculum materials, (2) connections between SWAT and the role of collaboration in the CMP classroom, and (3) an emphasis on explanation as a feature of SWAT and teachers' classroom practices.

CMP Curriculum Materials. Two teachers, Ms. Edwards and Ms. Gilbert talked about how the student work tasks were similar to the types of tasks found in the CMP3 curriculum materials. Ms. Edwards said, "I'm noticing in CMP3 ... there's a lot of these types of tasks where they have two different students do the work and then they have to either determine if one student is accurate or not and do you agree or disagree with them." Similarly, when initially discussing Tasks B and C, Ms. Gilbert stated, "they have to analyze student work and I mean, those types of problems are throughout CMP too." She elaborated on the use of student work in CMP, "I don't think any of our assessments have that. We'll do it in class or a lot of the ACE⁵, that's when those types of problems show up more." Both Ms. Edwards and Ms. Gilbert discussed how tasks with embedded student work were a lot like the types of tasks they had previously encountered in the CMP curriculum materials. It is interesting to note that Ms. Gilbert

⁵ ACE problems in CMP are Application, Connection, and Extension exercises. ACE problems are intended for use as additional learning experiences on homework, entrance activities, or exit slips.

also remarked that tasks with embedded student work did not appear on the assessment tasks from CMP that she had previously used.

Collaboration and the CMP Classroom. Five teachers discussed the importance of engaging in making sense of student work in class if analysis of student work also appeared on assessments and because student collaboration and analyzing each other's work are key components of CMP classrooms, as described by the teachers. Ms. Edwards said, "if you don't have students share out their work and have those deep discussions about what different students are doing or that error analysis, I think that would make it very challenging for an assessment." Similarly, when discussing assessment tasks with embedded student work, Ms. Gables stated:

If the kids weren't prepared for something like this, it would be very challenging for them. So, I think, the opportunities that kids have in the classroom have to match the expectations. You can't just be going through problems without these conversations before a test and then give them this task.

Similarly, Ms. Quinn stated, "it depends on how kids are engaging in the classroom about similar types of tasks to see whether [Task B] would be a fair assessment." Ms. Edwards, Ms. Gables, and Ms. Quinn all discussed the importance of students' preparation for student work tasks by engaging in making sense of student work in class. The teachers talked about both making sure students are sharing and making sense of each other's work during class as well as using tasks with embedded student work in class.

Alternatively, Ms. Henderson, the teacher in my study with the most years of teaching experience and the most years of experience teaching and leading professional development with the CMP curriculum, had a slightly different way of connecting student work in assessment tasks with her classroom practices. She said:

If I'm doing my job as an implementor of CMP as it's intended, those conversations are going to come out when we're going over day-to-day problems in our Summarize and in our Explore... it would be through a more involved process than just having [student work] be thrown in there.

Ms. Henderson talked a lot about students making sense of other students' work in a "natural" way. She felt that student work embedded into tasks was not as authentic as students actually engaging in making sense of each other's work in class or on an assessment. She stated:

I would say that happens more in the learning setting ... if I'm working in a small group and everyone is tackling the problem and we come together and kind of look, oh wait a minute, this doesn't make sense to me and you can talk it out.

Ms. Henderson acknowledged the dilemma of translating having students confront and critique someone else's thinking during authentic classroom experiences to similar experiences on written assessments. She brainstormed, "if you put something on a partner quiz where you had the two partners tackle a problem and then confront others, that might be a way to get at it," but she upheld the idea that natural, authentic exposure to other students' thinking was more ideal than embedded student work in written tasks used in class or on an assessment.

Echoing a similar idea about natural mathematical experiences, Ms. Shirley expressed frustration about creating authentic mathematical classroom experiences stating, "That's very difficult to do in the climate we're in to have students do that, you know in a natural, more natural way in the classroom." In reviewing Task H, she was admiring how the student would be required to determine what was warranted in the task and which of the solution strategies was most elegant, "a very important part of mathematics." She felt that the current emphasis on expressions and solving in seventh grade prevented students from thinking more broadly about

family of functions, making connections across mathematical ideas, engaging with other students' ideas, and thinking about the viability or elegance of various approaches towards problems. Similarly, she expressed how an exploratory and collaborative classroom was not always possible due the current approach promoted for teaching the mathematical content.

In connecting student work in assessment tasks to classroom practices, Ms. Quinn highlighted questioning as a key component of making sense of someone else's work, making a claim, and justifying it. She detailed this process:

There's a lot of intellectual work that has to happen in order to examine someone's work, especially if you can't ask questions and so kids in class when they present things, there are opportunities for kids to say, now why did you do this? Or how are you thinking about this? Or what did you visualize? Or what thinking did you do from this step to this step because I have no idea how you did that?

She talked about how the SWAT, if used on a written assessment, would not allow students to ask questions of the students in the task, an important part of critiquing others' work and justifying claims in her classroom. She elaborated on the difficulty of translating interpreting others' work in class to interpreting others' work on a written assessment task. Ms. Quinn said:

Even when kids have a lot of practice doing that in the classroom, looking at this, when I've used questions like this in assessments, there's a little bit of anxiety about it because I don't, I don't know what they did wrong, and it says there's an error there and I don't know what they did wrong. And it stinks because the kid can't ask a question to the other kid. But again, that just means it's another level of complexity in terms of the thought process, in terms of the understanding that kids have, because they have to not just have that understanding, but be able to look at somebody else's.

So, while Ms. Quinn promoted asking questions as an important activity in her classroom as students share and make sense of each other's work in class, she highlighted how tasks with embedded student work do not allow students to ask questions of the student in the task. Instead of viewing this as a negative consequence, Ms. Quinn described students' inability to ask questions as "another level of complexity" on these types of assessment tasks.

Emphasis on Explanation. A final way in which three of the six teachers made connections between student work in assessment tasks and their classroom practices was an emphasis on evidence and explanation when making sense of someone else's work. Ms. Edwards discussed how teachers in her building were focused on getting students to provide evidence and explaining why in mathematics. In reference to Task B, Ms. Edwards said students should detail "not just here's the mistake, but why is that the mistake. Then it's showing your understanding of the concept, as well." Extending beyond identifying an error, Ms. Edwards wanted her students to be able to explain the error in order to show their understanding of the mathematics.

When discussing Task F, Ms. Henderson talked about a common classroom phrase that her students used on a regular basis. "My kids would say, in order to have us believe you, you have to show us the money, can you give us some examples of these things to prove or disprove it?" She detailed how "show us the money" meant more to her students than the simple prompt "Explain." She continued, "My students have become routine. If they make a claim about something, this is why, this is what my evidence showed. I think that's part of that culture of expectation." Ms. Henderson set the expectation that mathematical claims, in this case claims about students' mathematical thinking, required justification that would be convincing to other students.

Ms. Quinn described how students should be involved in creating a collaborative understanding of what it means to explain and what a sufficient explanation would include. She talked about how discussions about explanation and justification should not occur just so that students can provide correct answers on assessments, but “it should be every single day and that conversation about what it means to explain, if we’re going to assess kids on it, then we have to give them the skills to do that.” In introducing her own students to building up an understanding of explanation and justification, Ms. Quinn described a collaborative process:

So, we started to brainstorm, so taking kids’ work, putting it up in front of others, and saying, okay, what is really powerful about their explanation? What details do we need to make sure we include? So, we started making a list. It feels like, in having that posted, kids that say, I don’t know what I can do, I seemed to be able to call them out on it, you should be able to do at least some of these things.

Ms. Quinn realized the importance of providing students with the skills to be able to construct an explanation in order to show sufficient evidence about making sense of someone else’s thinking. Also, Ms. Quinn talked about how she and her students collaboratively created a shared definition of what a powerful or sufficient explanation included and how this definition could be used as a resource for students struggling to provide appropriate justifications for their claims. In reference to both the CMP curriculum series and/or their own classrooms, all six teachers discussed the ways in which student work in assessment tasks connected to their curriculum and assessment practices.

Teachers as Assessment Writers. A final theme that emerged from conversations with teachers about SWAT was that all six teachers felt it would be necessary to adapt, revise, or add onto the existing SWAT in order to use the tasks in their own classrooms and/or they discussed

how they would use tasks in their own classrooms for teaching or assessment purposes. All six teachers talked about adapting at least one of the SWAT in order to get students to provide more evidence to support their solutions and show that they engaged in making sense of someone else's thinking. For Task B, Ms. Edwards felt that it would be important to "add why" because "the idea of the math practice of critiquing others is so that you can figure out where the mistakes are and why those mistakes happened, why the mistake is made because then you learn from that yourself." She described how it was important for students to be able to identify errors, but it was even more important for students to be able to explain why the errors were errors in order to show their understanding of the mathematics.

Three teachers discussed rewording Task C from "How did she do?" to a more descriptive prompt. Ms. Edwards made connections to her own assessment practices of green, yellow, and red light levels of understanding. She suggested adding onto the existing prompt with, "If you had to give her a grade, what do you feel that she would receive?" She felt that this addition would give students a bit more information about what to include in their responses. She also thought that asking students to give a student a grade based on their work would connect to the levels of understanding students used on a daily basis and were familiar with from seeing these levels on their own assessment feedback. Also wanting to give students more detail about expectations for the task, Ms. Gilbert stated, "I might reword it to say, if there's an error, what is it and explain what the error is" and Ms. Henderson said, "It should be, did she solve this correctly?" Both Ms. Gilbert and Ms. Henderson felt that students would be confused by the prompt "How did she do?" and responses to this question would not provide them with enough evidence of students' understandings.

Four teachers talked about rewording the prompt for Task F from “Explain.” to something that provided students with more guidance of their expectations for an explanation. Ms. Edwards suggested, “Explain by providing an example to negate it or to support it.” while Ms. Gilbert simply replaced “Explain.” with “Prove with an example.” Ms. Henderson stated, “Rather than “Explain.” I would say, “What’s your evidence for how you decided?” or something along with that.” Ms. Shirley added onto “Explain.” with “Explain and justify your reasoning with examples, definitions and/or evidence.” All four of these teachers felt that a simple, one-word prompt was insufficient guidance for students in crafting their responses on the assessment task. Elaborating on the inadequacy of “Explain.” as a prompt for students to provide an explanation or justification for their work, Ms. Henderson stated:

I often change “Explain.” on quizzes to like, “What makes you say that?” or “What's your evidence?” I change “Explain.” to a bigger phrase. I try to not make it the same phrase every time because I think that gets to be old, but, yeah, it's, you know, “Is your strategy evident to others?” “Could someone else follow your reasoning?” That kind of thing to prompt it to be more.

She felt it was important to provide students with more guidance on assessment tasks that required explanation in order to elicit more robust explanations from students that showed more of their mathematical thinking.

Four teachers discussed adaptations they would make to Task H either because they felt the task would be overwhelming for students as it provided three different student strategies and explanations on a single task, or they felt that students could complete the problem without actually having to make sense of all three students’ work. Ms. Edwards suggested, “maybe having fewer student sample pieces would be better” on an assessment so that students would

find Task H more accessible and would not be overwhelmed by the amount of symbolic reasoning and text on a single page. Two teachers, when discussing Task H, expressed uncertainty about whether or not students would actually engage in making sense of all three sets of student work on the task when answering the posed questions. On Task H when students were asked to “Explain which method makes the most sense to you.”, Ms. Gilbert said, “they’re not going to analyze student work as much because if it’s their method, whatever it might be, they’re going to pick that one compared to themselves.” Similarly, Ms. Gables pondered:

I wonder if kids would really have to deeply analyze all three of these because they could just say, yeah, they're all correct and Jesse's solution makes the most sense to me because that's the one they feel most familiar with. So, even if they didn't really understand Terri's or Brian's solutions, they could avoid that. They could avoid dealing with them. So, I guess it looks really challenging at first, but I'm just, I wonder if a kid could just get around any challenging part of this the way the question is worded. And I don't know how to word the question differently. I'd have to think about it, to get them to really look at each one.

Ms. Shirley also shared the concern raised by Ms. Gilbert and Ms. Gables and thought of an adaptation to Task H in order to avoid students only making sense of a single instance of student work. She said:

I would just change H, so “Which one makes the most sense to you and why?” and then “Choose one that didn’t make as much sense to you and then talk about steps they did do, and why those worked.” Something like that because the strategies are different enough.

Ms. Shirley felt that requiring students to describe and justify a non-preferred strategy would reveal more about students’ mathematical understanding than only requiring students to explain

their preferred strategy. With her adaptations, she sought to actually require students to engage in making sense of someone else's reasoning instead of just identifying which student's thinking matched their own reasoning.

Beyond making adaptations to the prompts for the four examples of SWAT introduced in the interview, three teachers also discussed ways in which they would use the assessment tasks other than on a written assessment. Ms. Edwards discussed using Task H in order to place students in appropriate algebra courses. She stated:

Not all of our students take algebra in 8th grade, so this would be a type of problem that would help indicate whether they are ready to move forward to the next level of algebra, they can take the high school level algebra class. It would be a factor to use for that.

The two other teachers, Ms. Gables and Ms. Quinn, talked about using the SWAT in class to spark important conversations. Ms. Gables discussed using Task F as an exit ticket to collect student responses and then “start the discussion by selecting varying levels of thought on that and projecting them for the kids to see to have a discussion.” Ms. Gables also felt that Task H would be fruitful to use in class for kids to come “back together to talk about the ideas.”

Ms. Quinn felt that all of the SWAT would be useful to have discussions in class about student strategies and sufficient evidence to support claims. In reference to Task B, she discussed:

I can imagine this being a conversation in class and kids talking about, ohhh, a mistake and then everyone, we are so thankful that you helped us see that this is a mistake, this is something we're going to make sure we're careful about. This question on test would seem very easy. Oh yeah, remember when Antonio said he made this mistake and oh yeah, we all talked about it, kids would just be able to write about it.

Similarly, Ms. Quinn felt that Task C would be a conversation starter in class where students explore how to interpret someone else's symbolic work and generating examples of responses to share with the full class. Beyond the mathematical ideas, Ms. Quinn talked about how Task F provided an opportunity for students to have a powerful conversation about evidence and justification that could extend beyond Task F. She stated:

So that whole idea of what counts as proof, and how do you know. And what is convincing for you that might be not convincing for someone else, what else would you need to tell them in order to be convincing feels like it could be a very powerful conversation that could extend to anything else where they need to explain and justify.

She also discussed how Task F could be used and revisited throughout the year as students built up more and more of their understanding of linear expressions and operational fluency with linear expressions.

Lastly, Ms. Quinn also felt that Task H could be used as a written assessment with a different prompt provided or it could also be used in class in order to brainstorm ideas about critiquing the reasoning of others. She discussed this process through a series of questions:

This could be a let's read through this, so what do you do when you're looking at someone else's work? Let's teach you how to critique someone else's work. At what part does it break down for you? What do you do when it starts to break down?

Ms. Quinn felt that all of the SWAT introduced in the interview could be used for in-class purposes in order to start mathematical conversations as well as deepen students' understandings of justification, explanation, and ideas about proof in mathematics.

All six teachers in my study discussed adapting, revising, or adding onto the existing prompts for the SWAT in order to provide students with more direction on the type of evidence

required in their solutions. Three teachers also discussed the ways in which they would use the SWAT other than on a written assessment. One teacher discussed using Task H as a placement task for high school algebra. Two teachers talked about using SWAT in order to spark classroom conversations about mathematical strategies. Lastly, one teacher emphasized how Tasks F and H could provide opportunities for students to have conversations about and deepen their understanding of critiquing the work of others in mathematics.

My analyses at the levels of individual interview questions and broader themes as teachers discussed SWAT and student work in assessment tasks revealed teachers' perspectives on SWAT. In order to explore more in detail whether or not student work could serve as a mechanism for assessing SMP3b, I investigated whether or not ideas related to SMP3b emerged from teachers' discussions of SWAT and students' work on SWAT.

RQ #3c: Evidence of SMP3b in Teachers' Discussions of SWAT and Students' Work on SWAT

In the following paragraphs, I trace how teachers talked about SMP3b or related habits of mind when discussing the SWAT from their first introduction to the SWAT, to an explicit discussion of student work as a possible mechanism for assessing SMP3b, and, lastly, to the teachers' review of actual students' work on the SWAT for evidence of engagement with SMP3b. Figure 6.3 below provides a snapshot of teachers' thoughts about SWAT as related to SMP3b through different segments of the teacher interviews. These analyses revealed that while all the teachers saw some kind of evidence of SMP3b in the SWAT both when initially introduced to the SWAT and when asked whether these tasks assessed the practice, teachers felt that evidence of students' actual engagement in SMP3b was lacking on students' written work.

Interview Segment	Teachers					
	Edwards	Gables	Gilbert	Henderson	Quinn	Shirley
Teachers' Initial Reactions to SWAT	" math practices "; "construct viable arguments and critique the reasoning of others "	" analysis tasks "	" analyze student work " (Tasks B, C, and H)	" error analysis " (Tasks B and C)	"look at someone else's thinking and kind of critique it "	--
Do SWAT Assess SMP3b?	" Yes. "	" Yes. "; "I just think there would have to be some criteria you've established for each of these ... how would I see that in a students' response? "	" Yes. "	"At least on the surface level "; "Is it really analyzing that I can find someone else's error if I know how to do it myself?" (in reference to Task C)	" Yes, but it should not just be in an assessment. It should be every single day."	"I can assess their aptitude for addressing the issue, for sure." Tasks B and C: "I'm pretty much just assessing whether they can do it right. "
Do Teachers See Evidence of SMP3b in Students' Work?						
Task B	" None of them are really saying why it's the error."	Identified 3 students that " kind of explained ", so "...in that sense, they're critiquing the reasoning of another person, I think."	"I would not say that a lot of them did. They pretty much solved it on their own way and then looked at his answer to compare."	" It's really not , it's saying, where did their thinking not match with mine."	"This feels like identifying, but not critiquing. "	" Yes, and justifying. I mean, they're checking the answer. "
Task C	"I don't think Susan explains how she did and neither does Cynthia."	Four students provided " some kind of critique to Roneasha's work but the level of understanding or explanation is different for each "	"I do not see that."; " I'm not sure what that's going to look like. "	" To some extent, yes. "	"...doesn't feel like they're critiquing. I wonder what the explanation would be? "	

Figure 6.3 Snapshots of Evidence of SMP3b in Teachers' Thoughts about SWAT and Students' Work Across Interview Segments

Figure 6.3 (cont'd)

Task F	“I don’t know that this was as much critiquing other people. I think it’s just showing their understanding of it.”	“No, ... I almost feel like these were where kids more had to construct their own argument.”	“You’ve got to prove it to me.”	“I don’t see real evidence.”	She wanted to ask students, “When you write, who is your audience?”	Teacher would rewrite the task to say: “Explain and justify ... with examples, definitions, and/or evidence.”
Task H	“I think Susan really talked about each of the three students, you can tell from just her written piece, that she looked at all three sample solutions”; “David is kind of showing because he’s got those check marks next to the people”; Three students “really tried to critique the other students’ work, like look at them.”	Generally, yeah, I think that their work is showing that they are critiquing the work of other students.”	“They’re looking for which person did it the easiest and not really critiquing their method.”; “It does say, “Explain which one makes the most sense to you.”, so they’re going to pick the one that fits their style.”	“Yes, to some extent.”; “It was really in the ways they confronted it. If I’m really fragile, I have less skills to confront someone else’s. If I’ve got that kind of medium level, I’m at least going to try and solve it myself and see if I agree or disagree, and then we had a couple of people in each of these groups that really tried to follow the other person’s pathways until it fell apart, but I think that’s a rarer thing.”	“So, visually, I thought this would be great, you’d get a lot of information from it because explaining which method makes sense to you, but I should have known because they choose the method that makes the most sense to them, it is easier and faster in their mind. Being able to explain the one that is different from yours feels like it would’ve given a lot more information.”	“Yeah.”; “I am seeing that play through, absolutely.”

Teachers’ Initial Reactions to SWAT

Before introducing the possibility of SWAT as a mechanism to assess SMP3b, I asked teachers to view the eight different assessment tasks and provide their initial reactions, ideas about what the tasks were assessing, and the advantages and disadvantages of using SWAT on

assessments. These different parts of the interview provided insight into teachers' thoughts about SWAT and allowed me to see whether or not ideas about SMP3b emerged from our discussions.

Initial Thoughts. When introduced to the assessment tasks and asked, “What do you think about these tasks?”, five teachers discussed SMP3 or related habits of mind when discussing SWAT. Ms. Edwards identified the SWAT as related to the “math practices” and she quoted SMP3, “construct viable arguments and critique the reasoning of others” (CCSSI, 2010, p.6). Ms. Gables identified the SWAT as “analysis tasks.” She detailed how on these types of tasks, “[students] have to examine solutions or work and determine if the students’ methods and solutions are correct.” Ms. Gilbert stated that Tasks B, C, and H required students to “analyze student work.” Ms. Henderson identified Tasks B and C as “error analysis” tasks. Lastly, Ms. Quinn identified SWAT as tasks that required students to “look at someone else’s thinking and kind of critique it.” Ms. Shirley was the only teacher that did not mention SMP3 or related habits of mind. Instead, she initially described all eight assessment tasks as traditional problems.

What SWAT Were Assessing. When viewing the tasks individually and responding to my question about what each task was assessing, teachers often used SMP3 or related habits of mind to describe what the SWAT were assessing. Findings for Research Question #3a, summarized in Figure 6.1, show that all six teachers for Task B, two teachers for Tasks C and F, and four teachers for Task H, identified these tasks as assessing SMP3b or related habits of mind. It is important to note that Ms. Shirley identified Tasks B and H as “assessing [a student’s] ability to make sense of and critique the work of others” when asked what these tasks assessed. Therefore, prior to being introduced to ideas about student work embedded in assessment tasks and whether or not student work can serve as a mechanism for assessing SMP3b, all six teachers had mentioned, in some way, ideas related to SMP3b when discussing SWAT.

Revisiting Advantages of SWAT. After I had introduced student work as a feature of four of the assessment tasks, I asked teachers about the advantages and disadvantages of using tasks with embedded student work on assessments. Findings from Research Question #3b, summarized in Figure 6.2, show that three teachers talked about how SWAT connect to the Standards for Mathematical Practice (SMPs) (CCSSI, 2010). Ms. Henderson, Ms. Quinn, and Ms. Shirley described the connection they viewed between SWAT and SMP3b as an advantage. Ms. Shirley stated, “You know, you’re supposed to be making arguments and critiquing the reasoning of others and that’s engaging them in the mathematical practices.” Therefore, before asking teachers explicitly whether or not SWAT could assess SMP3b, all six teachers had already discussed ideas directly or closely related to SMP3b when talking about SWAT.

Do SWAT Assess SMP3b?

After informing teachers that, “Some teachers and curriculum writers have hypothesized that assessment tasks with student work can assess this practice (SMP3b),” I asked them, “Do these four tasks (the SWAT) assess this practice?” Four teachers, Ms. Edwards, Ms. Gables, Ms. Gilbert, and Ms. Quinn, said “Yes.” Ms. Gables elaborated on her response saying, “How would I see that in a student’s response?” She agreed that the SWAT assessed SMP3b, but she questioned exactly how it would work in terms of what students would include in their solutions. Ms. Quinn also elaborated on her response stating, “... but it should not just be in an assessment. It should be every single day.” She felt that SMP3b should be a component of the mathematics classroom on a daily basis.

Ms. Henderson said that the SWAT assessed SMP3b “at least on the surface.” She felt that more authentic opportunities for students to engage in and teachers to assess this practice could occur in classrooms with students generating the student work as opposed to a fictionalized

person embedded into a written task. Ms. Quinn stated that with the SWAT, “I can assess their aptitude for addressing the issue, for sure,” but she qualified that for Tasks B and C, saying the tasks “just assess[ed] whether they can do it right.” Overall, all six teachers, to varying degrees, expressed that SWAT could assess SMP3b to some extent.

Do Teachers See Evidence of SMP3b in Students’ Work?

Once teachers had a chance to look at students’ written work on the SWAT, I asked teachers, “Do you see evidence that students critiqued the reasoning of others based on their written student work?” Many teachers’ thoughts about SWAT as a mechanism for assessing SMP3b shifted. As teachers talked about students’ written work on the SWAT, Ms. Gables question from the previous section, “How would I see that in a student’s response?” seemed extremely relevant.

For Task B, four teachers felt that students did not show evidence of SMP3b. Ms. Quinn shifted her thoughts about Task B assessing SMP3b. She stated, “This feels like identifying, but not critiquing.” Ms. Gables identified three student responses in which the student “kind of explained.” She continued, “... in that sense, they’re critiquing the reasoning of another person, I think.” Ms. Shirley viewed Tasks B and C together and said that she saw evidence that students critiqued the reasoning of others. She stated, “Yes, and justifying. I mean, they’re checking the answer.” In this way, as students checked the person’s answers in the tasks, Ms. Shirley saw this as evidence that students critiqued the person’s reasoning. For Task B, most teachers did not see evidence of SMP3b. Also, Ms. Quinn shifted her view that Task B assessed SMP3b. Ms. Gables also expressed uncertainty about Task B assessing SMP3b.

For Task C, two more teachers joined Ms. Gables in questioning what evidence of student thinking related to SMP3b students would need to provide on a written assessment

response. After discussing that she did not see evidence of SMP3b on Task C, Ms. Gilbert stated, “I’m not sure what that’s going to look like.” Also, Ms. Quinn expressed that students’ responses, “[didn’t] feel like they’re critiquing. I wonder what the explanation would be?” Also, responses from Ms. Edwards, Ms. Gables, and Ms. Henderson revealed that explanation was a component of the evidence of SMP3b that teachers were looking for in responses and students were demonstrating different levels of understanding based on their explanations. Ms. Gables felt that four of the students provided “some kind of critique,” but “the level of understanding or explanation is different for each.”

For Task F, most of the teachers expressed that they did not see evidence of SMP3b in students’ responses either by explicitly saying so or suggesting edits to the task in order to elicit more evidence. Three teachers, Ms. Gables, Ms. Gilbert, and Ms. Henderson, did not see evidence of SMP3b. Ms. Henderson said, “I don’t see real evidence” and Ms. Gilbert stated, “You’ve got to prove it to me.” Ms. Quinn and Ms. Shirley suggested edits to Task F in order to support students in sharing their mathematical thinking on a written assessment task. Ms. Quinn wanted to ask students, “When you write, who is your audience?” Ms. Shirley wanted to rewrite the task to say, “Explain and justify ... with examples, definitions, and/or evidence.” Two teachers expressed doubt that this SWAT could assess SMP3b. Ms. Edwards said, “I don’t know that this was as much critiquing other people. It’s just showing their understanding of it.” Similarly, Ms. Gables stated, “I almost feel like these were where kids more had to construct their own arguments.”

For Task H, a task that presented three different strategies for solving and asked the reader to look at the three solutions, determine the correctness of the solutions, and explain which method made the most sense, teachers revealed different ideas about whether or not

students' solutions showed evidence of SMP3b and what teachers counted as evidence. Ms. Shirley expressed that there was evidence of SMP3b in students' responses. She stated, "I am seeing that play through, absolutely." Ms. Gables also said, "Generally, yeah, I think their work is showing that they are critiquing the work of other students." Ms. Henderson also saw evidence of SMP3b, "to some extent." Her description of how different types of students would go about solving this task revealed different levels of engagement in making sense of someone else's thinking. She expressed how it was rare for students to "follow the other person's pathways until it fell apart," and that students were more likely to "try and solve it myself and see if I agree or disagree."

Ms. Edwards expressed doubt as to whether or not there was evidence of SMP3b in students' responses. Ms. Edwards detailed how three students "really tried to critique the other students' work, like look at them." She based this on the fact that the students either "talked about each of the three students" or had "check marks next to the people." Ms. Edwards looked for evidence that students "looked at all three sample solutions." Both Ms. Gilbert and Ms. Quinn expressed doubt about the prompt for the task requiring students to critique. For the part of the task that asked the reader to explain which method made the most sense, Ms. Gilbert said, "They're looking for which person did it the easiest and not really critiquing their method."

Similarly, Ms. Quinn said:

I should have known they chose the method that makes the most sense to them, it is easier and faster in their mind. Being able to explain the one that is different from yours feels like it would've given a lot more information.

Ms. Quinn posited that requiring students to describe a method that differs from their preferred method would provide more evidence of SMP3b on an assessment task.

After viewing students' written work on the SWAT, teachers expressed doubt about whether or not the tasks could assess SMP3b and many teachers were puzzled about what would count as sufficient evidence that students would need to include in their written responses in order to demonstrate that they engaged in critiquing someone else's reasoning. Two teachers, Ms. Gables and Ms. Henderson, discussed different levels of engagement with SMP3b demonstrated in students' written work. Similarly, these findings, and previous findings mentioned in this chapter, show that a few teachers wanted to adapt task instructions in order to elicit additional evidence from students to gain a better understanding of students' engagement with SMP3b.

Summary of Chapter Findings

In this chapter, I presented findings from analyses of teacher interviews focused on teachers' experiences with and perspectives on curriculum-based assessment tasks with and without embedded student work as well as students' written work on these tasks. Analyses of teachers' discussions of what the assessment tasks were assessing revealed that teachers thought that assessment tasks could assess SMP3b and related habits of mind. Also, teachers expressed advantages and disadvantages of SWAT both at the general level and the use of specific tasks. Connected to the focus of this study, three teachers discussed student work as connected to the SMPs as an advantage for the use of student work embedded in assessment tasks (CCSSI, 2010).

When I analyzed teachers' descriptions of SWAT and the use of student work in assessment tasks, four main themes emerged. First, two teachers discussed how student work related to content or practices promoted in other school subjects besides mathematics. Second, four teachers referenced state or standardized test when discussing student work. Third, all six teachers made explicit connections between student work and the CMP curriculum materials or

CMP classroom practices. Lastly, all six teachers talked about making adaptations, revisions, or additions to SWAT and/or discussed how they would use tasks with embedded student work in their own teaching or assessment practices.

In terms of the evidence of SMP3b that teachers noticed in the SWAT and students' work on the SWAT, all six teachers discussed SMP3b or related habits of mind when first introduced to the SWAT. Also, all six teachers, to varying degrees, expressed that they felt SWAT could assess SMP3b. However, once they looked at students' written work on the SWAT, teachers began to question whether or not SWAT could serve as a mechanism for assessing SMP3b. Also, many teachers began to question the expectations for evidence that students would have to provide on these tasks in order to sufficiently demonstrate SMP3b. I discuss these findings and their implications in Chapter 7.

CHAPTER 7: DISCUSSION

Overview

In this chapter, I revisit the findings detailed in the previous three chapters and present the big picture as informed by each research question and key takeaways from the results of my analyses of curriculum materials, student interviews, and teacher interviews in which I explored whether or not student work embedded in assessment tasks can be used as a mechanism for assessing SMP3b. The takeaways from my study of student work embedded in assessment tasks as a possible mechanism for assessing SMP3b relate to: (1) the potential of SWAT and (2) the limitations of SWAT. For each takeaway, I highlight how the results of my study parallel, differ from, or extend the research literature on curriculum-based assessments, SMP3b, and student work and how my study contributes to the field of mathematics education.

Across these takeaways, I discuss implications for the development of and research on curriculum materials, assessment tasks, and, more broadly, types of mathematics assessments used to assess habits of mind. Then, I revisit validity as a tenet of ECD and discuss my findings in terms of the validity of SWAT as a mechanism for assessing SMP3b. I conclude this chapter by summarizing the implications of this study for teachers, curriculum and assessment writers, standards writers, and researchers in mathematics education as related to written assessment tasks as mechanisms for assessing dynamic practices that involve communication and interaction with others and consider next steps for my own research.

Findings Big Picture

My analyses of curriculum materials, students' written work on curriculum-based assessment tasks, and students' and teachers' descriptions of their experiences with SWAT resulted in many interesting findings. Looking across my findings, I noticed trends related to

whether or not SWAT could serve as a mechanism for assessing SMP3b. Namely, although (1) SWAT existed in curriculum-based assessments, (2) students recorded evidence of their thinking in writing on SWAT, and (3) teachers initially felt that SWAT aligned with classroom practices and could assess SMP3b, closer examination of my study findings revealed: (1) differences between curriculum-based SWAT and SWTT (RQ #1), (2) differences between students' written and verbal evidence of thinking (RQ #2), and (3) teachers' uncertainty about SWAT assessing SMP3b (RQ #3).

In the following paragraphs, I describe how my analyses revealed that while students had opportunities to “read” and “respond to the arguments of others” on SWAT, students did not have the opportunity to and/or teachers did not see evidence that students “decide[d] whether [arguments] make sense and ask[ed] useful questions to clarify or improve the arguments” (CCSSI, 2010, p. 6). For each research question, I discuss my findings and how they illustrate the potential for SWAT to serve as a mechanism for assessing SMP3b “on the surface level” as Ms. Henderson would say, but that closer examination of SWAT based on analyses of curriculum materials, students' experiences, and teachers' experiences revealed that assessing dynamic processes that involve communication with others is problematic when translated to static, written assessment tasks.

RQ #1: SWAT in Curriculum-Based Assessments

Although SWAT existed in curriculum-based assessments, closer examination revealed differences between curriculum-based SWAT and SWTT. Coding of five CCSSM-aligned seventh-grade curriculum-based assessment materials revealed that SWAT occurred in all five curriculum series (see Figure 4.2). Because the Criteria for Student Work corresponds directly with the language of SMP3b (see Figure 3.4), these analyses showed that students had the

opportunity to “critique the reasoning of others” in the curriculum-based assessment tasks across all of the series analyzed (CCSSI, 2010, p.6). However, comparisons between SWAT and SWTT at the level of each curriculum series revealed differences between students’ opportunities to critique someone else’s mathematical thinking in terms of the evidence and critique types evident on SWAT and SWTT (see Figures 4.9, 4.10, 4.13, and 4.14). For example, all four evidence types (words, symbols, visuals, and actions) occurred in all curriculum series SWTT. However, not all evidence types were represented in each curriculum series SWAT. Big Ideas and Go Math SWAT did not include visual representations and CMP SWAT did not require the reader to make sense of someone’s symbolic representations. In terms of the critique types, all of the curriculum series SWTT contained at least four of the five identified critique types (Error ID, Eval, Compare, Pref, and Insight). This contrasts with the critique types represented in the curriculum series SWAT in which Big Ideas SWAT only involved two critique types and CMP and Go Math SWAT only involved three. Therefore, SWTT across the curriculum series offered more variety of evidence and critique types for students to engage in as compared to the SWAT.

Differences in evidence and critique types included in SWTT as compared to SWAT can be problematic if the textbook tasks are communicating different ideas about what it means to critique someone else’s mathematical thinking as compared to the assessment tasks. If the assessment tasks are communicating a more limited perspective on the types of evidence and critique types students should engage in when analyzing someone else’s work, the more diverse evidence and critique types found in the SWTT may not be utilized by teachers or seen by students if the assessment tasks are deemed more consequential or “what counts” for analyzing someone else’s work. These analyses revealed that while students did have opportunities to engage in reading and responding to the mathematical work of others in the curriculum-based

assessments, the same types of tasks found in the student textbook were more diverse in terms of the evidence and critique types students were required to complete. Therefore, there was a disconnect between students' opportunities to critique someone else's mathematical thinking in the student textbook as compared to on curriculum-based assessment tasks.

RQ #2: Students' Experiences and Perspectives on SWAT

Although students recorded evidence of their thinking on SWAT, closer examination revealed differences between students' written and verbal evidence of thinking. Analyses of students' written work on SWAT revealed that students recorded evidence of their thinking on SWAT in writing for all of the instances except for Ed on Task F (see Figures 5.1 and 5.2). Therefore, most students were able to provide evidence of their thinking in writing on SWAT for most of the tasks (see Appendix I for students' actual written work on SWAT and non-SWAT). However, closer examination of students' written work as compared to their verbal responses not represented in any way in their written work revealed that students often provided additional evidence of their thinking in their verbal reasoning.

For my analyses, I purposely used different codes for the two types of data (see Tables 5.1 and 5.2) due to the difference in mediums (spoken word versus written words, symbols, and diagrams), but direct connections between types of evidence for these two analyses existed for many of the codes. For example, a claim found in a student transcript was quite similar to a statement in a student's written work. Figure 7.1 below details pairings I observed between the two sets of evidence type codes. Notice, four codes used for analyzing the interview transcripts did not arise in a similar way in the written work codes: plausibility, prior experience, revised thinking, and tools. Also, drawing was a unique code for the written work analyses.

Student Thinking Codes Used to Analyze Written Work	Student Thinking Codes Used to Analyze Transcripts in Comparison to Written Work
Statement	Claim
Reasoning	Justification
Symbolic Work	Operation/Method
	Plausibility
	Prior Experience
Question	Questioning
	Revised Thinking
	Tools
Uncertainty	Uncertainty
Drawing	

Figure 7.1 Student Thinking Codes for Written Work and Transcripts Pairings

Figure 7.1 shows the possible limitations of written work for capturing certain types of evidence of student thinking that students were often able to provide verbally. This figure, in addition to multiple comparisons between students' written and verbal work, also demonstrates how much more information can be gained about students' thinking when considering more than one mode of students' representations of their thinking. Taken together, both written and verbal work, these two types of representations of student thinking could allow students more opportunities to demonstrate their mathematical thinking, in particular, their ability to demonstrate SMP3b on a written assessment task.

RQ #3: Teachers' Experiences and Perspectives on SWAT and Students' Work on SWAT

Although teachers initially felt that SWAT aligned with classroom practices and could assess SMP3b, closer examination revealed teachers' uncertainty about SWAT assessing SMP3b. When first introduced to SWAT in the teacher interview, teachers discussed these tasks as connected to their curricular and/or classroom practices. Some teachers made explicit connections to CMP materials, the role of collaboration in their classroom, and the importance of students' providing explanations when analyzing someone else's mathematical work (see Table 3.8). Three teachers explicitly talked about how SWAT connected to the Standards for

Mathematical Practice (SMPs) (CCSSI, 2010) (see Figure 6.2). Lastly, when asked, ‘Do these [SWAT] assess [SMP3b]?’ , most teachers responded affirmatively (see Figure 6.3). One teacher, Ms. Gables, expressed a question about what students would provide on these types of tasks and one teacher, Ms. Henderson, stated that she felt SWAT could assess SMP3b “at least on the surface.”

However, once I asked teachers to review students’ actual written work on the tasks, teachers were less certain that SWAT could assess SMP3b (see Figure 6.3). After teachers had an opportunity to browse through students’ actual written work on both SWAT and non-SWAT, I asked teachers to review each SWAT individually and answer, ‘Do you see evidence that students critiqued the reasoning of others based on their written student work?’ Two teachers, Ms. Gilbert and Ms. Quinn, joined Ms. Gables in questioning what type of evidence they would see from students in order to demonstrate SMP3b. Ms. Gilbert stated, “I’m not sure what that’s going to look like.” Similarly, Ms. Quinn questioned, “I wonder what the explanation would be?” on Task C.

Comparison across teachers’ responses to, ‘Do you see evidence that students critiqued the reasoning of others based on their written student work?’ revealed differences in expectations for what students would need to provide as evidence of engagement in SMP3b in their written work, often differing at the task level. For example, on Task H, Ms. Edwards wanted to see evidence that students had “really tried to critique...like look at them.” She wanted to see evidence in students’ written responses that they had looked at all three instances of student thinking provided in the task. In contrast, Ms. Shirley was content with the evidence students were providing on Task H in terms of evidence that students critiqued the reasoning of others. In reflecting on Task H and whether or not she saw evidence of SMP3b, Ms. Shirley stated, “I am

seeing that play through, absolutely.” Overall, teachers were less confident that SWAT could assess SMP3b after reviewing students’ actual written work on these tasks and reflecting on whether or not they were seeing evidence that students critiqued the reasoning of others based on their written responses.

Overall, findings across research questions revealed that while SWAT provided opportunities for students to read and respond to someone else’s mathematical thinking, teachers were not seeing sufficient evidence that students critiqued the reasoning of others based on their written work. Extending beyond individual research questions, I observed a number of takeaways from my study about the potential and limitations of SWAT as a mechanism for assessing SMP3b. In the following sections, I discuss these takeaways and make connections to the research literature on curriculum-based assessments, SMP3b, and student work. I also discuss how my study contributes to the field of mathematics education and identify implications for stakeholders in mathematics education.

Potential of SWAT

My previous curriculum research experiences taught me that curriculum writers included tasks with embedded student work in seventh-grade textbooks for all five of the curriculum series used in this study (CMP, 2018; Gilbertson et al., 2016; Going, Ray, & Edson, in preparation). In this study, I explored how these curriculum writers used student work in their assessment materials and examined if student work could be used to assess SMP3b. From my review of literature on assessments, I knew that Hunsader and colleagues (2013; 2014) had found limited evidence of curriculum-based assessments assessing NCTM process standards, but for the evidence that did exist, there was wide variability across curriculum series. Furthermore, Shepard (2000) expressed that assessment methods often lag behind instructional methods.

While the results of my study reinforced Shepard's (2000) observation and echoed the results of Hunsader and colleagues' (2013; 2014) assessment analyses, my findings highlighted the potential of SWAT as a possible mechanism to assess SMP3b. In the following paragraphs, I discuss the potential of SWAT to (1) characterize SMP3b in the context of curriculum tasks, (2) align assessment and instruction, (3) elicit student thinking, (4) motivate students to solve tasks and (5) provide benefits for students.

Characterize SMP3b in the Context of Curriculum Tasks

My exploration of curriculum-based assessments from each of the five curriculum series using the Criteria for Student Work (see Appendices A and B) at the level of each criterion revealed the opportunities students had to, first, encounter a person in a task, second, be exposed to a person's mathematical thinking, and, third, analyze the person's thinking in some way. These analyses not only identified the tasks where students had the opportunity to engage in SMP3b, as supported by the direct mappings between the criteria for student work and the content of SMP3b "critique the reasoning of others," but also detailed students' opportunities to engage with the smaller aspects of the practice – "others" (Criterion #1) and "the reasoning of others" (Criteria #1 and #2) (CCSSI, 2010, p. 6). Additional analyses of both assessment and student textbook tasks that fulfilled all three criteria, the SWAT and SWTT, revealed the nature of SWAT based on assessment types, and the nature of SWAT and SWTT based on CCSSM content, evidence types (or "the reasoning of others"), and critique types ("critique").

The methods I used to collect instances of SWAT by coding at the level of each criterion provided a robust conceptualization of SMP3b by clearly defining "what counts" for each criterion as mapped onto smaller parts of the practice. Also, the iterative coding revealed potential tasks that could be minimally adapted or added to, as compared to tasks that did not fit

any criteria, in order for tasks to fulfill all three criteria for student work. Tasks that fulfilled all three criteria for student work have the potential to assess SMP3b as they required the reader of the task to engage in some form of critique of a person's reasoning.

For example, the task provided in Figure 7.2 is an example of an assessment task from CMP that fulfilled Criterion #1 ("the principal", "students") and Criterion #2 ("The principal concluded the number of students n would be related to a session price p by the equation $n = 350 - 2p$ "). However, the task failed Criterion #3. The reader was not required to engage in an activity that involved critiquing or explaining someone's mathematical thinking. Instead, the reader was asked to use someone's mathematical thinking to determine an appropriate different representation of that thinking (e.g. translating an equation to a graphical representation).

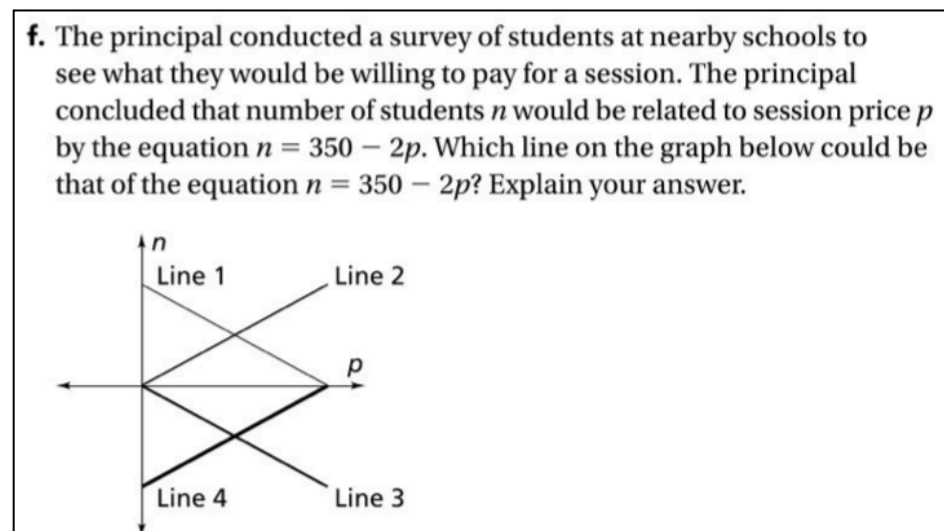


Figure 7.2 Assessment Task that Fails Criterion #3 (CMP, EOY Assessment Test Bank, p. 17)

However, because the task already fulfilled Criteria #1 and #2, there exists an opportunity for the task to be adapted to also fulfill Criterion #3. Students could be asked whether or not the principal's equation makes sense (Evaluation) or how they think the principal arrived at this equation (Insight). Therefore, the methods used in my study and the results of my detailed analyses at the level of the individual criteria provide a possible starting point for teachers and

curriculum writers to consider how tasks that already fulfill Criterion #1 or both Criteria #1 and #2 could be revisited and adapted to be SWAT in order for assessment materials to include more opportunities for students to make sense of someone else's mathematical thinking, thinking that is already embedded in assessment tasks.

The methods I used to analyze students' opportunities to engage with SMP3b as well the nature of these opportunities on assessment and textbook tasks extended the work of Hunsader and colleagues (2013; 2014), to not only capture the potential for practices to be assessed in curriculum-based assessments and promoted in textbooks, but to also provide insight into the nature of these tasks. The evidence and critique types I used in my coding contribute to research on mathematical practices and habits of mind (CCSSI, 2010; NCTM, 2010; 2014; NRC, 2001; Seeley, 2014) by providing examples of different types of "reasoning of others" and "critiques" that were evident in curriculum materials and could be used in tasks to promote and/or assess SMP3b (CCSSI, 2010, p. 6).

The results of evidence and critique type coding for SWAT revealed evidence and critique types varied by curriculum series indicating different emphases on what it means to critique someone else's mathematical thinking on assessment tasks by curriculum series. At the level of the curriculum series, each series communicated different ideas about what types of evidence of thinking students should use to represent their mathematical thinking to others and what evidence of thinking students should be able to critique. Similarly, each curriculum series communicated different ideas about the types of critiques students should be able to demonstrate on a written task. Therefore, the methods used in my study and the results of my analyses could be used by curriculum and assessment writers to inform future assessment and textbook task

design, by teachers to vet materials for use, and by researchers to conduct analyses of tasks based on the frequency and nature of opportunities students have to critique someone else's reasoning.

While the methods I used for criteria and SWAT/SWTT coding contribute to the field in terms of identifying the potential for student work to be embedded in curriculum tasks and conceptualizing SMP3b and the nature of SMP3b in terms of evidence of student thinking and critique types, there exist opportunities for future work around identifying and characterizing SMP3b in tasks. The evidence and critique types I identified in the SWAT provide a way to capture and compare different examples of “the reasoning” and “critique,” but I did not explore conceptualizations of “others” as seen in the SWAT or SWTT, which is Criterion #1 of the Criteria for Student Work (see Appendices A and B). Analyses of whose reasoning students are asked to critique in SWAT/SWTT would provide insight into what curriculum writers communicate to students, as mediated through tasks, about whose mathematical reasoning students should critique or are able to critique. Specifically, for SWAT, analyses of “others” would show what curriculum writers communicate about whose mathematical thinking students should be able to demonstrate their abilities to critique on an assessment task.

To extend my conceptualizations of evidence and critique types, I would not only want to compare the types of evidence and critiques students were required to analyze (evidence) and complete (critiques) on assessment tasks to those found in textbook tasks, as I did in this study, I would also want to investigate how the evidence and critique types identified in assessment and textbook tasks compare to the types of evidence and critiques students encounter during daily classroom practices. In this way, I want to consider additional evidence and critique types not evident in SWAT or SWTT that could possibly be adapted for use on assessment and textbook

tasks, as informed by how students critique another person's reasoning in mathematics classrooms.

Align Assessment and Instruction

Comparison of SWAT and SWTT coding of features revealed differences between the types of CCSSM content strands, evidence types, and critique types students encountered on assessment tasks as compared to textbook tasks. Differences between SWAT and SWTT based on these features of tasks (CCSSM content strand, evidence types, and critique types) could be problematic as the assessment tasks may emphasize that critiquing other's mathematical thinking only (1) occurs for certain mathematical topics, (2) involves a few types of representations of student thinking, and (3) requires a few types of critiques in comparison to the student textbook. These findings aligned with Shepard's (2000) observation that assessment methods often lag behind instructional methods.

The methods I used and the results of my study highlighted the potential for SWAT to be used to align assessment and instruction based on students' opportunities to critique someone else's thinking on tasks. For example, SWTT coding by critique types revealed that, while evaluation was the dominant critique type found across the full sets of SWTT and SWAT, students had opportunities to engage in additional types of critiques in the curriculum series SWTT as compared to the SWAT. Preference was a critique type that occurred on SWTT (CPM and CMP) but not SWAT. Preference tasks required the reader to make a choice based on the reader's preference. An example of a preference task is shown below for Problem 7-98b from CMP SWTT where the reader is asked, "Which equation is easiest to solve?" In order to answer this question, the reader must make a choice based on which equation is easiest. Note, this task

was also coded as a comparison task since the reader had to compare multiple instances of student thinking.

7-98. Shiloh knows that it takes 5 cups of flour to make 2 recipes of cookies. He has 13 cups of flour left in his cupboard. He wants to know how many recipes he can make. He wrote the equivalent ratios below.

$$\frac{5 \text{ cups}}{2 \text{ recipes}} = \frac{13 \text{ cups}}{x \text{ recipes}}$$

This equation looked difficult to solve, so he decided to ask his friend Carolyn for help. Carolyn said,

"If it takes 5 cups of flour to make 2 recipes, then it takes $2 \frac{1}{2}$ cups of flour for each recipe. This is the multiplier. Can't you just use the multiplier to write an equation?"

Carolyn then wrote the equation $F = 2 \frac{1}{2} R$ where F stands for cups of flour and R stands for the number of recipes. As Carolyn wrote the equation below, she said, *"Then I would solve this equation to find out the number of recipes."*

$$13 = 2 \frac{1}{2} R$$

a. Compare Shiloh and Carolyn's equations. Are they correct? Why or why not?

b. Which equation is easiest to solve? Solve it.

Figure 7.3 Preference Critique Type Example (CPM, Problem 7-98b, p.421)

Because assessments are highly consequential (Hamilton, Stecher, & Yuan, 2012; Swan & Burkhardt; 2012), even when students are provided diverse opportunities to critique someone else's mathematical thinking in the student textbook, a lack of variety in critique types provided in assessment tasks may communicate to teachers and students that it is really only important for students to be able to critique someone's reasoning certain kinds of ways and based on certain types of evidence of student thinking. Therefore, it is important for teachers using curriculum materials and writers creating curriculum materials to attend to the opportunities, and the features of these opportunities, students are provided in materials to engage with mathematical practices in the student textbook and on assessment tasks.

Findings from comparisons of SWAT and SWTT suggested next steps. The variety in the nature of student work embedded in textbook tasks as compared to assessments means that perhaps the evidence and critique types used in SWTT could inform assessment task design and selection. Teachers could use methods and findings from this study to inform their own assessment practices. Shepard (2000) stated that "good assessment tasks are interchangeable with good instructional tasks," so teachers could use SWAT and SWTT interchangeably in their

practice (p. 8). While curriculum writers could use the methods and/or results of my analyses to inform the design of their materials, mathematics education researchers could also use the methods detailed in my study to further investigate the use of student work in different curriculum materials or as a model to investigate students' opportunities to engage in other important mathematical habits of mind in curriculum materials. For my own purposes as a researcher on the CMP project, I think that results from the analyses of CMP SWAT and SWTT could be used to inform future iterations of CMP assessment and textbook task design.

Elicit Student Thinking

As evidenced by analyses conducted on data gathered from student interviews, SWAT elicited student thinking from students both verbally and in written form (see Figures 5.1-5.4). The description of SMP3b states that students should be able to “read the arguments of others” and “respond to the arguments of others” (CCSSI, 2010, p.6)). Analyses of students' written work and verbal descriptions for types of evidence of student thinking provided robust evidence that students both “read” and “responded to” the mathematical thinking of others on SWAT. Therefore, SWAT has the potential to elicit student thinking that demonstrates students' engagement with SMP3b.

While research has focused on the benefits for students of learning from examples (e.g. Rittle-Johnson, Star, Durkin, 2009), my findings suggest that there is potential for student work, an exemplar of a worked example, to be used not just for student learning, but also to assess students. To extend this study, I would investigate what evidence of student thinking is essential for students to include in their solutions to demonstrate SMP3b. Since writers of the CCSSM SMPs (CCSSI, 2010) did not provide information about how individual practices can be assessed, they only asserted that practices could be assessed, additional work is needed to

determine the types of evidence of student thinking that would be necessary for students to demonstrate SMP3b. I elaborate on this in a later section when I consider the lack of explicit expectations for what evidence of student thinking students should include in their written solutions as a limitation of SWAT. When discussing this limitation, I expand on the implications and next steps related to how students demonstrate their thinking on written assessments that are intended to assess SMP3b or other habits of mind.

Motivate Students to Solve Tasks

Not only did analyses of students' written and verbal responses indicate that SWAT has the potential to elicit student thinking, one student in my study described SWAT as tasks that students would be motivated to complete. Ed expressed how the possibility of helping someone else in an assessment task was motivating for him and could possibly motivate other students that may not typically enjoy doing mathematics. Ed's comment revealed that the desire to help others may be an added incentive for students to engage with SWAT and demonstrate their thinking on these tasks.

Provide Benefits for Students

When asked about the advantages of student work embedded in assessment tasks, teachers in my study discussed a number of advantages for SWAT. The teachers identified the potential of SWAT to promote robust student understanding, expose students to different strategies, and promote the Standards for Mathematical Practice (SMPs) as well as a number of advantages of individual SWAT tasks (see Figure 6.2). Also, a theme that emerged from teachers' description of SWAT was that these tasks promoted interdisciplinary practices. Therefore, teachers in my study felt that SWAT were beneficial for students based on identified advantages and connections to other school subjects.

As teachers discussed the benefits of SWAT for students, SMP3b and related habits of mind were evident in their descriptions. Teachers affirmed that SWAT could assess SMP3b and discussed how SWAT promoted interdisciplinary practices of justification, explanation, and providing evidence for claims (NCTM, 2000; NRC, 2001; Seeley, 2014). Therefore, not only did teachers view SWAT as having the potential to assess SMP3b, but also SWAT had the potential to promote and assess important practices students were required to demonstrate in other school subjects.

The role of critiquing across subject matters and disciplines, as identified by teachers in my study, has implications for how teachers, curriculum writers, and assessment writers consider assessing SMP3b on written assessments. These stakeholders should consider: How are students expected to demonstrate justification and explanation in other school subjects? What evidence of their thinking are students expected to include in their written responses? How could assessment techniques in other school subjects be adapted for use in mathematics in order to assess SMP3b? Utilizing an interdisciplinary approach towards designing assessment tasks in mathematics that assess SMP3b or other habits of mind would not only affirm and align with students' experiences in other school subjects but would also utilize advancements in assessment design that may already be developed in other disciplines.

Limitations of SWAT

While findings from across my research questions suggested the potential use of SWAT for assessing SMP3b, the results of my study also revealed the limitations of the use of SWAT for assessing SMP3b. While SWAT may have the potential to assess SMP3b, students' and teachers' experiences with and perspectives on SWAT revealed that additional work is needed to determine how students should demonstrate how they “decide whether [arguments] make sense

and ask useful questions to clarify or improve the arguments” (CCSSI, 2010, p.6) on SWAT. Also, my analyses showed the standards writers’ decision to not describe how practices can or should be assessed can be problematic for designing tasks that assess mathematical practices. In the following paragraphs, I discuss the limitations of SWAT based on (1) the expectations for evidence communicated in SWAT and (2) the student thinking not captured and the components of SMP3b not assessed on SWAT.

Expectations for Evidence

In reviewing students’ verbal and written responses on the assessment tasks to answer my second research question, I became curious about the relationship between the evidence of student thinking that students communicated in their written work and the evidence of student thinking that tasks required based on the expectations for evidence expressed in the task prompts. A limitation of SWAT seemed to be related to what students thought they should communicate about their thinking in written form. Returning to the actual assessment tasks used in the clinical component of the student interviews, I considered the expectations for students in each task to get a sense of what assessment tasks were communicating about expectations for evidence.

Expectations involved what the student would have to do to complete the task and record their work often indicated by the presence of a question or explicit instructions or directions. Expectations for what a student should include in a written response can come from a number of sources: task instructions, assessment norms, and/or teacher expectations. I specifically focused on expectations provided in written task instructions. In reviewing the task prompts, I noticed two types of expectations provided in the written tasks. For many of the tasks, students were required to respond to a question. For other tasks, students were asked to complete some action, most often, “Explain.” The tasks offered different arrangements of these expectations.

In Figure 7.4 below, I detail the expectations provided in each task in the order in which the expectations are presented. Bolded rows indicated SWAT. For example, for Task H, students are first directed to “Look at the three solutions below.” Then, they are asked, “Are they correct?” Lastly, students are directed to “Explain which method makes the most sense to you.” In other words, Task H presented three different explicit expectations for completing this task as indicated by the question and directions included in the task prompt.

Task	Expectations Provided in Assessment Tasks
A	Question: “How many hours did the technician work?”
B	Question: “Where did Paco make an error in his calculation?” Direction: “Explain.”
C	Question: “How did she do?” Direction: “Explain completely.”
D	Direction: “Write an expression ...”
E	Question: “Which of the following expressions is not equivalent to the others?” Direction: “Explain.”
F	Question: “Is your friend correct?” Direction: “Explain.”
G	Direction: “Write an expression ...” Question: “... how much does Royce weigh?”
H	Direction: “Look at the three solutions below.” Question: “Are they correct?” Direction: “Explain which method makes the most sense to you.”

Figure 7.4 Student Response Expectations Provided in Assessment Tasks

These expectations offered insight into the evidence of student thinking students were expected to include, as indicated by the task prompts, in their written responses in order to meet these expectations. However, a number of the questions and directions in the tasks did not offer a clear indication of what evidence was sufficient or necessary to complete the task. I compared the expectations for SWAT and non-SWAT to the evidences of student thinking that was most often not included in the written response but evident in students’ verbal descriptions for these two assessment types. From these comparisons between the expectations for SWAT and non-SWAT to the evidence of student thinking types for SWAT and non-SWAT detailed in Figure

5.4, I found that the direction of “Explain” (Tasks B, C, E, F, and H) in a task often resulted in students making claims or justifications in their verbal responses that were not reflected in their written work. This occurred for all the SWAT, and also included the non-SWAT, Task E.

For example, in Figure 7.5 below, Anna on Task B provided the following written work:

Paco solved the equation $-3x - 5 = 1$.
His work is shown below.

Step 1: $-3x - 5 = 1$

Step 2: $-3x = 6$

Step 3: $x = 2$

Where did Paco make an error in his calculation? Explain.

$$\begin{array}{r} -3x - 5 = 1 \\ +5 \quad +5 \\ \hline 0 \quad 6 \\ -3x = 6 \quad \rightarrow 2 \\ \hline -3 \quad -3 \end{array}$$

Figure 7.5 Task B and Anna's Written Work on Task B

Her written work was coded as symbolic work. However, in her verbal response on this task, she said, “He messed up at Step 3 because he had to divide by -3.” and “It was Step 3 because he didn’t make it a negative.” She provided justification in her verbal response that was not evident in any way, as a statement or reasoning, in her written work.

Even when students did provide statements or reasoning in their written work, students often had claims or justification that were evident in their verbal responses that extended beyond what they included in writing. For example, in Figure 7.6 below, Jane on Task H provided the following written work:

Students in Mr. Rickman's class are asked to solve the equation $\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) = 3$. Look at the three solutions below. Are they correct? Explain which method makes the most sense to you.

Jess's Solution

I began by distributing the numbers outside the parentheses on the left side.

$$\begin{aligned}\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) &= 3 \\ 4x - 6 + 2x - 3 &= 3 \\ 6x - 9 &= 3 \\ 6x &= 12 \\ x &= 2\end{aligned}$$

They all are correct, but Jess's solution makes the most sense because it is the way I would do it and it's the most straightforward.

Terri's Solution

I began by multiplying each side of the equation by 3.

$$\begin{aligned}3\left[\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9)\right] &= 3(3) \\ 2(6x - 9) + 1(6x - 9) &= 9 \\ 12x - 18 + 6x - 9 &= 9 \\ 18x - 27 &= 9 \\ 18x &= 36 \\ x &= 2\end{aligned}$$

Brian's Solution

I knew that $\frac{2}{3} + \frac{1}{3} = 1$, so I simplified the right side to $6x - 9$.

$$\begin{aligned}\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) &= 3 \\ 1(6x - 9) &= 3 \\ 6x - 9 &= 3 \\ 6x &= 12 \\ x &= 2\end{aligned}$$

Figure 7.6 Task H and Jane's Written Work for Task H

Jane's written work was coded as reasoning. However, in her verbal response, she offered additional justification when she discussed the solutions provided by Terri and Brian. She said, "Terri and Brian's solutions took a lot of elaborate steps and thinking and stuff, but Jesse's was just like solving it instead of like Brian got $2/3 + 1/3 = 1$." Her verbal response indicated more information about why she chose Jess's solution in comparison to the other two solutions.

These examples highlight, and my analyses of evidence of student thinking in students' verbal and written responses showed, that the simple direction of "Explain" does not provide explicit directions for students as to what should be included in an explanation or what exactly it is that students should be explaining. Is an explanation checking your work, justifying your solution path, or simply detailing or elaborating on your methods? Even for Task H, where students are provided a bit more detail and are asked to, "Explain which method makes the most sense to you," it is unclear, based on the task instructions, how the student should go about meeting this expectation. Should the student explain a method, explain why a method makes sense, or explain how she made sense of the method?

For the non-SWAT, other than Task E, expectations for students included answering questions, “How many ...?” (Task A), “...how much...?” (Task G), and completing the direction, “Write an expression ...” (Tasks D and G). For these tasks, the most prevalent evidence type that students expressed verbally but was not captured in writing was operation/method. However, the questions and directions included in the written task instructions for these three non-SWAT (Tasks A, D, and G) do not indicate that students were required to show their reasoning or methods in any way in their written work beyond providing an answer to a question or writing an expression. In their written work for these tasks, all students did provide symbolic work and one student also included a drawing for Task D.

So, while students’ omissions of reasoning along with operation/methods in their written responses for Tasks A, D, and G might not be problematic for meeting the expectations for these tasks as indicated by the task instructions, students’ omissions of statements and/or reasoning for Tasks B, C, E, F, and H when providing a written explanation might prove to be problematic. This is because the expectation for the task indicated by the instruction “Explain” does not provide sufficient information about what students should include. My analyses of students’ written and verbal responses showed that for tasks that included an “Explain” prompt, while students did provide statements and/or reasoning in their written work, they also often verbally articulated claims and/or justifications that provided more insight into their thinking.

The research literature on SMP3b and related habits of mind emphasizes the importance of justification, explanation, and providing evidence (CCSSI, 2010; Koestler, Felton, Bieda, and Otten, 2013; Hull, Miles, and Balka, 2012; NCTM, 2000; NRC, 2001; Seeley, 2014). Students are expected to not only be able to critique someone else’s thinking, but also “justify and explain ideas in order to make their reasoning clear” (NRC, 2001, p. 130). However, this literature does

not provide a lot of information about how students should be able to communicate their reasoning about another person's thinking in writing based on student work embedded in a task.

Teachers, curriculum writers, and assessment writers should consider the expectations for evidence that are communicated in tasks used to promote and assess SMP3b. Teachers in my study talked about adapting, revising, or adding on to SWAT in order to encourage students to show evidence of their thinking in their written solutions and demonstrate that they engaged in making sense of someone else's thinking. For example, Figure 7.7 details Ms. Shirley's brainstorming of revised expectations for Task F in which she attempted to provide students with clearer instructions about what evidence of student thinking she wanted students to provide on this task.

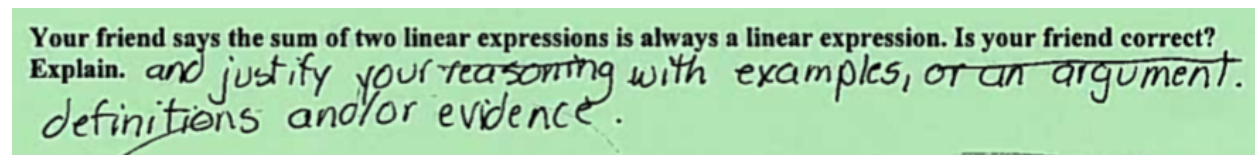


Figure 7.7 Ms. Shirley's Revised Expectations for Task F

For SWAT, students are required to make sense of someone else's thinking and are often required to provide an explanation. Therefore, it is important to consider the expectations the written assessment tasks convey to students about what evidence of student thinking should be included in the written responses. As evidenced in my study, the expectations for the types of evidence that students should provide in their written responses based on the task prompts for SWAT were not clear about what students should include in their written work. Also, the analyses of the evidence of student thinking for SWAT revealed that students often expressed evidence of their thinking verbally in the form of claims and justifications that were not evident in their written work. Consequently, a limitation of SWAT is expectations for evidence these tasks communicate about the types of evidence of thinking students should include in their written responses.

Evidence of Thinking Not Captured and Components of SMP3b Not Assessed on SWAT

The prevalence of types of evidence of student thinking that were evident in students' verbal responses but not their written work for both SWAT and non-SWAT revealed that the features of written assessment tasks and the environments in which students solve these tasks present limitations. In other words, the traditional nature and task-solving environment of written assessment tasks require that students not only solve the assessment task, but also appropriately record or document their thinking, as determined by the task instructions, assessment norms, and/or teacher expectations, to be reviewed by teachers without access to the students' task solving experience.

It is not surprising that students' written work did not present a full-picture representation of the thinking students talked about for both SWAT and non-SWAT tasks as evidenced by the types of evidence student thinking students provided verbally, but not in their written work. In much the same way, the student interview transcripts also did not capture the entirety of students' task solving experience since the transcripts only detailed students' verbal responses and did not capture students' gestures, timing, or movements. Even so, my analyses of the instances of evidence of student thinking that occurred in students' written and verbal responses revealed interesting trends about the evidence types students expressed in their written work as compared to the evidence types students expressed verbally but omitted from their written responses for SWAT and non-SWAT. The evidence of their thinking that students detailed verbally but was not represented in their written work reveals the limitations of written assessments for capturing evidence of students' thinking only in writing.

When first presented with SWAT, teachers in my study were fairly confident that these tasks could be used to assess SMP3b. After reviewing students' written work, teachers were not

convinced that students provided sufficient evidence that they critiqued someone else's mathematical thinking in their written work. As stated above, once teachers noticed that students were not provided sufficient information about the evidence they should provide in their written responses, many teachers began to revise the existing expectations provided in the task to make it clearer to students what they expected in their written solutions. Another possibility to consider as far as supporting students in providing evidence of their thinking and their abilities to critique someone else's mathematical thinking is to consider allowing students to use multiple modes of communicating their thinking. Analyses of students' written work as compared to their verbal reasoning revealed that students provided a lot of evidence of their thinking that was not captured in their written solutions (see Figures 5.1 – 5.4; see Figure 7.1). The examples of students' written work as compared to their verbal responses illustrated in Figures 7.3 and 7.4 highlight how more insight into students' thinking could be evident when considering both modes of thinking, written and verbal, instead of only reviewing students' written work.

For example, Ms. Edwards revealed that for Task H, she wanted to see evidence that students had “really tried to critique...like look at them.” She wanted to see evidence in students' written responses that they had looked at all three instances of student thinking provided in the task. In reviewing Jane's written work (see Figure 7.4), Ms. Edwards did not see evidence that Jane reviewed all three instances of student thinking provided in the task. However, if Ms. Edwards also had access to Jane's verbal reasoning on this task, she would have heard Jane discuss all three instances of student thinking. In her verbal response, she said, “Terri and Brian's solutions took a lot of elaborate steps and thinking and stuff, but Jesse's was just like solving it instead of like Brian got $\frac{2}{3} + \frac{1}{3} = 1$.” Therefore, Jane's written work and verbal reasoning together would have provided Ms. Edwards a more robust representation of Jane's thinking as

well as evidence that Jane did look at all three instances of student thinking, something Ms. Edwards was looking for as evidence that students engaged in SMP3b on Task H. Attending to both students' written and verbal representations of their thinking could have allowed teachers in my study to gain additional insight into whether or not students provided evidence of SMP3b orally, if not evident in writing.

Since teachers were not seeing evidence of SMP3b in students' written work, but students are most often required to show evidence of their thinking in writing on assessments, I returned to literature on SMP3b to determine if the literature provides insight into how this practice could be assessed. The research literature on SMP3b and related habits of mind (CCSSI, 2010; Koestler, Felton, Bieda, and Otten, 2013; Hull, Miles, and Balka, 2012; NCTM, 2000; NRC, 2001; Seeley, 2014) described SMP3b as a dynamic process that involves interacting with and communicating with others. The standards writers described SMP3b as involving student actions of "decid[ing]," "ask[ing]," and "communicating" (CCSSI, 2010, p. 6). Also, each level of Hull, Miles, and Balka's (2012) proficiencies for SMP3b involve student actions: "discuss," "explain," and "compare and contrast" (p. 52). The authors' descriptions of these practices and proficiencies do not provide insight into written products students might create to demonstrate SMP3b.

The results of my analyses showed a limitation of SWAT that is a consequence of the limitations of traditional written assessments. How can written assessments assess students' abilities to engage in dynamic processes involved in "critiquing the reasoning of others" (CCSSI, 2010, p. 6) as detailed by the research literature? My findings showed that students engaged in lots of thinking that was not evident in their written responses. However, the standards writers indicated that SMPs can be assessed and two testing consortia are currently producing written assessments that align with CCSSM and assess mathematical practices (Chandler, Fortune,

Lovett, & Scherrer, 2016; Herman & Linn, 2013; Hull, Balka, Miles, 2013; Schoenfeld, 2013). Therefore, it is important for teachers, curriculum writers, and assessment writers to consider how students' evidence of their thinking is currently captured on SWAT and possibilities for designing SWAT and assessment formats that capture more evidence of students' thinking.

My text analyses indicated that students had the opportunity to complete five different types of critiques across the textbook and assessment tasks: error identification and/or correction, evaluation, comparison between multiple instances of mathematical thinking, making a choice based on preference and providing insight into someone's thinking (see Figure 3.8 for descriptions). However, the description of SMP3 provided by the standards writers reveals other components of SMP3b that were not evident in the types of critiques that students were asked to complete on SWAT and SWTT. As stated earlier, the research literature on SMP3b and related habits of mind (CCSSI, 2010; Koestler, Felton, Bieda, and Otten, 2013; Hull, Miles, and Balka, 2012; NCTM, 2000; NRC, 2001; Seeley, 2014) described SMP3b as a dynamic process that involves interacting with and communicating with others. The standards writers suggested that students should "listen ... to the arguments of others," "ask useful questions to clarify or improve the arguments," and "communicate [conclusions] to others" (CCSSI, 2010, p.6). These descriptions of SMP3b indicate other types of critiques not evident in the SWAT or the SWTT, which I describe as Questioning and Dialogue/Interaction.

Questioning was not a critique type represented in SWAT or SWTT. Students were not explicitly prompted to ask questions or provide questions in their written responses as part of a task. Even so, students demonstrated questioning as a type of evidence of their thinking when they solved assessment tasks in both written work and verbal responses. Also, one teacher, Ms.

Quinn, identified students' inability to ask the person in the task questions as a disadvantage of SWAT.

Due to the nature of written tasks, dialogue or interactions were not evident in SWAT or SWTT because the reader of the task cannot interact with the person in the task in a collaborative way. Students completing tasks with embedded student work only have access to the evidence of thinking presented in tasks. They do not have the opportunity to gather more evidence of thinking from a person or to interact back-and-forth. Teachers in my study talked about the importance of dialogue and interactions in promoting SMP3b in their mathematics classrooms. They expressed the importance of collaboration when students analyzed each other's work in their mathematics classrooms. In relation to that, the research literature on the benefits of student work for teacher learning often situates teachers' examination of student work as a collaborative learning experience in which teachers reflect on and discuss student work as a professional development component in a teacher learning community (e.g. Silver & Suh, 2014). While collaboration, interaction, and dialogue are important for promoting SMP3b based on how SMP3b is defined in the research literature and teachers' descriptions of the practice, SWAT did not allow students to interact with the person that created the mathematical thinking in the task in order to build up an argument together.

The lack of questioning and interactions or dialogue in the SWAT revealed that a limitation of SWAT is that they are unable to assess all components of SMP3b, as defined by the standards writers. Implications for teachers, curriculum writers, and assessments writers in addressing this limitation include revisiting SWAT and considering: How can questioning be promoted in SWAT and SWTT? How can dialogue and interaction be promoted in SWAT? Are there components of SMP3b that written assessments cannot assess? As I personally reflect on

these questions, I anticipate that SWAT could include expectations that students generate questions for the person in the task based on the person's written work. However, I would argue that although this might achieve embedding Questioning critique types into SWAT, students would not be able to "ask useful questions to clarify or improve the arguments" as described by the standards writers (CCSSI, 2010, p. 6). In order for questioning to be used to improve arguments, the person in the task would need to be able to respond, which is not possible for written assessments. Similarly, dialogue and interaction are not possible critique types for SWAT because the nature of written tasks does not allow for the reader to interact with, question, communicate, and build up an argument with the person in the task.

Therefore, written assessment tasks may not be the best mechanism for assessing certain components of SMP3b such as Questioning and Dialogue/Interaction. Instead, perhaps alternative assessment methods that involve groups of students interacting around a task, generating evidence of their thinking, sharing their thinking with one another, and critiquing each other's work would be more conducive for assessing certain components of SMP3b. Even so, because assessment materials provided in curriculum materials communicate to teachers what mathematics is important for students to know and do (Hunsader, 2013; 2014), students' opportunities to engage in and be assessed on SMP3b and other habits of mind should be carefully considered when curriculum and assessment writers design written assessment tasks.

Revisiting Validity

Considering my findings across research questions and the takeaways that emerged related to the potential and limitations of SWAT as a mechanism for assessing SMP3b in my exploratory validity study, I now return to the construct of validity and ask: Are SWAT valid assessment tasks for assessing SMP3b? As a reminder, Mislevy (2012) stated that, "Strong

arguments give us confidence in the quality of the inferences and interpretations, or in their validity” (p. 94). In reflecting on the exploratory argument I made in this study connecting the claim that SWAT could serve as a mechanism for assessing SMP3b to my observations of students’ and teachers’ experiences with SWAT, I would suggest that while SWAT have the potential to assess SMP3b, additional work is needed before these tasks can be used on traditional written assessments. Currently, the Criteria for Student Work (see Appendices A and B) serves as a Conceptual Assessment Framework to vet tasks based on their inclusion of “critique”, “the reasoning of”, and “others” (CCSSI, 2010, p. 6). However, my analyses of SWAT did not capture the expectations for what students would need to provide or demonstrate in writing as evidence of SMP3b.

In other words, my analyses allowed me to capture the potential for students to demonstrate SMP3b, not necessarily clear expectations for students to demonstrate SMP3b in writing on written assessment tasks. Findings from students’ and teachers’ experiences with SWAT revealed that expectations for evidence was a limitation of SWAT, as most tasks depended on the vague direction of “Explain” for eliciting students’ thinking. My initial attempt to characterize the expectations provided to students on SWAT (see Figure 7.4) is one way in which I can build a more robust argument about the validity of SWAT for assessing SMP3b on written assessment tasks. Therefore, one way to build a more robust argument, and consequently a more valid assessment task, would be to consider the expectations for evidence that are included in SWAT and, just as teachers in my study revised tasks to include, provide students with clear expectations for how they should demonstrate SMP3b in writing.

Another way to build a more robust argument linking claims about SWAT assessing SMP3b to observations of students’ and teachers’ experiences with SWAT is to consider whether

or not written assessment tasks are the most appropriate format for assessing a practice that is routinely described as dynamic in nature. The standards writers as well as authors of seminal works describing related habits of mind, discuss SMP3b as involving communication, interaction, and engagement with others. Is a static task on a page that includes an “other” that is unable to respond to the reader the best mechanism for assessing SMP3b? Perhaps SWAT could serve as a basis for conversation between multiple readers and a teacher could observe these students’ experiences engaging with the thinking of an “other”? Even so, it is problematic for standards writers to assert that practices, processes, and related habits of mind can be assessed without providing insight into their intentions about how these practices should be assessed. Similarly, two assessment consortia are crafting tasks that assess practices in traditional formats. How are teachers expected to prepare students to demonstrate SMP3b in writing on an assessment task if not with tasks similar to SWAT?

Therefore, while SWAT may possibly be considered valid assessments when using traditional modes of assessing mathematics, I would assert, and I suggest a number of my teacher participants would agree, that there are more authentic and more valid ways in which to assess students’ abilities to demonstrate SMP3b that involve real communication with peers, interactions with others, and student-created instances of student work. In this context, the types of evidence of student thinking captured through observations of students’ interactions with one another would more closely align with how SMP3b is described by standards’ writers.

Summary of Implications, Next Steps, and Final Thoughts

Overall, the explorations I completed in this study showed that SWAT have the potential to assess SMP3b, but additional assessment task design work is needed for these tasks to be adapted, improved, and implemented in ways that support students in providing evidence of their

thinking and support teachers in interpreting students' written work for evidence of SMP3b. The methods used in my study for analyzing curriculum-based assessment materials for the existence of SWAT and the results of my analyses could be used by teachers, curriculum and assessment writers, and researchers to determine potential ways in which tasks could be adapted to be student work tasks based on the criteria for student work. Similarly, my text analyses methods and analyses results could be used by curriculum writers to inform future assessment and textbook task design, by teachers to vet materials for use, and by researchers to conduct analyses of tasks based on the frequency and nature of opportunities students have to critique other's reasoning or as a model to investigate students' opportunities to engage in other important mathematical habits of mind in curriculum materials.

Students' and teachers' experiences with SWAT detailed in my study have implications for teachers, curriculum and assessment writers, standards writers, and researchers and their efforts to assess SMP3b on written assessments. These stakeholders should consider: What are the expectations for evidence communicated in written assessment tasks? How can students be encouraged to show evidence of their thinking in written form on written assessment tasks? What components of SMP3b cannot be assessed on a written assessment task? What are the most conducive assessment types for assessing SMP3b? How could assessment techniques from other school subjects, that also promote explanation, justification, and critiquing, be adapted for use on written assessments or other types of assessments in order to assess SMP3b? Answers to these questions would help minimize the limitations of SWAT and strengthen the potential of SWAT as a mechanism for assessing SMP3b.

In my own work, I want to use the disconnect between the evidence of thinking students discussed verbally and what they produced in writing to inform future coding of SWAT. In my

current coding framework for criteria and features of SWAT, I do not account for the expectations for evidence of thinking that students are prompted to include based on the questions or directions provided in the tasks. I want to revisit the coding framework and include a feature of tasks that would allow me to capture and compare expectations of the SWAT. In this way, I could explore what SWAT communicate about the evidence of their thinking students should include in their written responses. Also, beyond the use of written assessment tasks, I want to explore ways in which teachers assess practices, such as SMP3b, using interactive experiences that require students to communicate, question, and collaboratively make sense of each other's thinking. Teachers' promotion and assessment of SMP3b in their classrooms could provide insight into how SWAT could possibly be incorporated into classroom experiences or be adapted to more closely reflect classroom and assessment experiences.

APPENDICES

Appendix A: Criteria for Student Work Revision Iterations

SWTT Initial Development → Revised Version		Connection to SMP3b	SWAT Dissertation
Gilbertson et al. (2016) Criteria	First Revised Criteria (Going, Ray, & Edson, in preparation)		Further Revised Criteria
1	The mathematical task must mention at least one person (the <i>character</i>) to which the work is attributed.	others	The mathematical task must include a person based on the following: - Exclude ambiguous “you,” “I,” and “group” - Exclude places or corporations (e.g. school, town, store, company) - Include specific and general people (e.g. Daniel, your friend, boy, girl) - Include professionals (e.g. baker, owner, buyer, manufacturer) - Include groups of people (e.g. team, club, class) A person must be explicitly referred to and not just used to describe a place (e.g. Holly’s basement or Colton’s Gym)
2	The task must include a character’s thinking or actions or prompt the reader to determine the character’s thinking or actions. Thinking might include a written mathematical claim, a conjecture, a strategy, some form of reasoning, an observation or measurement, an algorithm, or a reflection on a mathematical idea.	the reasoning of	The task must include evidence of a person’s mathematical thinking. Evidence includes the following: - Claims, Conjectures, Statements, Arguments, or Reflections - Methods, Reasoning, or Algorithms - Observations, Measurements, or Diagrams - Actions, when explicitly and intentionally mathematical (e.g. conducted an experiment, designed a game with mathematical components, took a survey with provided purpose, collected/recorded data and evidence provided, generated a sample, created or used a mathematical object, plotted points) **Attend to verbs and vocabulary** Exclude hypothetical mathematical thinking (e.g. wanting to triple the volume, planning to cut a board) unless evidence of mathematical thinking is included, as described above.
3	There must be an expected activity for the reader of the text. These activities might include analyzing, critiquing, or reflecting on the mathematical thinking/actions of the character in the written materials.	critique	The expected activity for the reader of the text depends/relies on the thinking of the person(s) in the task. Activities include the following: - Critiquing/Verifying/Explaining someone’s mathematical thinking - Determining the correctness, fairness, soundness, representativeness, accuracy, or bias of someone’s mathematical thinking - Comparing multiple instances of mathematical thinking Exclude tasks where the reader is asked to: - Complete a task someone has started - Use someone’s mathematical thinking to create a different representation, solve a problem, or apply definitions (e.g. Which number in this equation represents the rate of change?, What type of sampling method was used?)

Figure A.1 Criteria for Student Work Iterations

Appendix B: Student Work in Curriculum Materials Analysis Codebook

Process of Coding:

1. **Task Size:** The unit of a task is determined by the seriation used by each curriculum. For a problem in the assessment materials or the student textbook that has multiple components defined by seriation, each component is considered an individual task (e.g. Parts A, B, C, and D of Problem 2 would be considered Tasks 2A, 2B, 2C, and 2D).
2. **Student Work Coding:** Using the Criteria for Student Work, each task will be coded first to determine if it fits Criterion #1. Tasks that fit Criterion #1 are then coded to determine if they also fit Criterion #2. Lastly, tasks that fit both Criteria #1 and #2 are coded to determine if they fit Criterion #3. Tasks that fit all three criteria are considered Student Work Tasks. If a criterion is met prior to a series of tasks (e.g. Use the following for questions 8-9), the criterion should be counted for all the tasks in the sequence. However, if a criterion is met within a series of tasks (e.g. Part B of a series of Tasks A-D), the criterion should only be applied to the one task (e.g. Part B), unless the criterion is met again in subsequent tasks. In the curriculum-based assessment materials, identify tasks that meet all three criteria as Student Work Assessment Tasks (SWAT). In the student textbooks, identify these tasks as Student Work Textbook Tasks (SWTT).
3. **Assessment Type Coding:** Using the Assessment Type Descriptions, each assessment task will be categorized by the type of assessment in which the task is provided in the assessment materials. Each task will receive only one designation based on the descriptions provided.
4. **SWAT and SWTT Coding:** Using the CCSSM Content Strands Descriptions, the Evidence of Student Thinking Categories and Corresponding Codes, and Critique Categories and Corresponding Codes, each SWAT and SWTT will be coded.
 - a. Each task will receive only one CCSSM content strand code based on information provided by the curriculum materials as well as the mathematical content of the task. The CCSSM content strand code should be the strand that “fits” most closely with what the materials designate as the mathematical content as well as the mathematical content actually provided in the task.
 - b. Each task will be coded for the type of evidence of student thinking the reader is required to engage with in/evaluate/analyze/make sense of in the task. A task can have multiple instances of evidence of student thinking. Do not apply code(s) based on what evidence of student thinking appears in the task. Rather, apply code(s) based on the types of evidence the task requires the reader to critique. Evidence must be explicit and fit in an appropriate category (e.g. a “method” could take on many forms including symbolic work, descriptions of the method, a visual). For tasks that designate specific student thinking that the reader should engage with/evaluate/analyze/make sense of, include only the codes that “fit” the provided evidence (e.g. Are her formulas correct?; Is the sampling plan biased?). For tasks that do not designate specific student thinking that the reader should engage with/evaluate/analyze/make sense of (e.g. What do you think?; How did he do?), include all instances of evidence that the reader would engage with on the task.
 - c. Each task will be analyzed for the type of critique the reader is required to complete in the task. A task can have multiple instances of critiques as defined in

the categories. Do not apply code(s) based on the type of critique a reader could complete. Rather, apply code(s) based on the type of critique required in order to complete the task as designated in the task.

Criteria for Student Work

Use the criterion descriptions below to code tasks. Analyze all tasks for Criterion #1. Only analyze tasks fitting Criterion #1 for Criterion #2. Only analyze tasks fitting both Criterion #1 and Criterion #2 for Criterion #3. Tasks that fit all three criteria are Student Work Tasks.

Criterion #1	The mathematical task must include a person based on the following: - Exclude ambiguous “you”, “I”, and “group” - Exclude places or corporations (e.g. school, town, store, company) - Include specific and general people (e.g. Daniel, your friend, boy, girl) - Include professionals (e.g. baker, owner, buyer, manufacturer) - Include groups of people (e.g. team, club, class) A person must be explicitly referred to and not just used to describe a place (e.g. Holly’s basement or Colton’s Gym)
Criterion #2	The task must include evidence of a person’s mathematical thinking. Evidence includes the following: - Claims, Conjectures, Statements, Arguments, or Reflections - Methods, Reasoning, or Algorithms - Observations, Measurements, or Diagrams - Actions, when explicitly and intentionally mathematical (e.g. conducted an experiment, designed a game with mathematical components, took a survey with provided purpose, collected/recorded data and evidence is provided, generated a sample, created or used a mathematical object, plotted points) **Attend to verbs and vocabulary** Exclude hypothetical mathematical thinking (e.g. wanting to triple the volume, planning to cut a board) unless evidence of mathematical thinking is included, as described above.
Criterion #3	The expected activity for the reader of the text depends/relies on the thinking of the person(s) in the task. Activities include the following: - Critiquing/Verifying/Explaining someone’s mathematical thinking - Determining the correctness, fairness, soundness, representativeness, accuracy, or bias of someone’s mathematical thinking - Comparing multiple instances of mathematical thinking Exclude tasks where the reader is asked to: - Complete a task someone has started - Use someone’s mathematical thinking to create a different representation, solve a problem, or apply definitions (e.g. Which number in this equation represents the rate of change?; What type of sampling method was used?)

Figure A.2 Criteria for Student Work

Assessment Type Descriptions

Use the descriptions below to code tasks. Each task will receive only one assessment type code.

Diagnostic	Assessment tasks included in: - Diagnostic assessments - Placement assessments - Beginning-of-the-year assessments - Pre-assessments - Readiness assessments These tasks are intended for use at the beginning of the year or the beginning of a chapter/unit/module.
Periodic	Assessment tasks included in: - Quizzes - Mid-Module Assessments These tasks are intended for use in the middle of a chapter/unit/module.
Summative	Assessment tasks included in: - Tests - Cumulative Assessments These tasks are intended for use at the end of a chapter/unit/module/semester or the end-of-the-year.
Question Bank	Assessment tasks included in: - Question Banks - Test Banks These tasks are provided as a resource for teachers to pick and choose tasks to use on assessments.

Figure A.3 Assessment Type Descriptions

CCSSM Content Strand Descriptions

Use the descriptions and standards information provided in the curriculum materials to code tasks. Each task will receive only one CCSSM content strand code.

RP	Ratios and Proportional Relationships – Analyze proportional relationships and use them to solve real-world and mathematical problems.
NS	The Number System – Apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.
EE	Expressions and Equations – Use properties of operations to generate equivalent expressions. Solve real-life and mathematical problems using numerical and algebraic expressions and equations.
G	Geometry – Draw, construct and describe geometrical figures and describe the relationships between them. Solve real-life and mathematical problems involving angle measure, area, surface area, and volume.
SP	Statistics and Probability – Use random sampling to draw inferences about a population. Draw informal comparative inferences about two populations. Investigate chance processes and develop, use, and evaluate probability models.

Figure A.4 CCSSM Content Strand Descriptions

Descriptions of the CCSSM content strands are from the *Common Core State Standards for Mathematics* for Grade 7. Additional language for each CCSSM content strand for Grade 7 can be found in:

Common Core State Standards Initiative (CCSSI). *Common Core State Standards for Mathematics*. Washington, D.C.: National Governors Associations Center for Best Practices and the Council of Chief State School Officers, 2010. <http://www.corestandards.org>

Evidence of Student Thinking Categories and Corresponding Codes

Use the descriptions below to code tasks while attending to the verbs and vocabulary used in the task. Each task can receive multiple evidence type codes.

Ask yourself: “*What evidence(s) of mathematical thinking is the reader required to critique?*”

Words	The task requires the reader to make sense of someone’s: - Thoughts - Reasoning - Explanations - Written Solutions - Statements - Predictions/Guesses/Estimates - Noticing - Claims
Symbols	The task requires the reader to make sense of someone’s: - Computations involving operations and/or variables - Developed Formulas - Developed Expressions/Equations - Operational Representations
Visuals	The task requires the reader to make sense of: - Diagrams - Tables - Graphs - Drawings - Visual Representations Representations can either be created by a person introduced in the task or serve as evidence of someone’s thinking even when not explicitly created by a person.
Actions	The task requires the reader to make sense of someone’s descriptions of: - Methods/Strategies/Designs - Survey/Sample Plan and/or Results - Measurements - Mathematical Actions **especially important to attend to verbs and vocabulary for this category**

Figure A.5 Evidence of Student Thinking Categories and Corresponding Codes

Critique Categories and Corresponding Codes

Use the descriptions below to code tasks. Each task can receive multiple critique type codes.

Ask yourself: *“What type(s) of critiquing does the task require the reader engage in?”*

ErrorID	The task requires the reader to identify someone’s error and/or correct someone’s error. The existence of an error is explicit in the task.
Eval	The task requires the reader to determine correctness, accuracy, validity, truth, bias, viability, fairness, and/or representativeness of student thinking.
Compare	The task requires the reader to compare multiple instances of student thinking. This can involve comparing different people’s thinking or different types of evidence of thinking from one or many people.
Pref	The task requires the reader to make a choice based on preference. This includes determining what makes the most sense, what is easiest, what is efficient, and/or what is preferred.
Insights	The task requires the reader to provide insight into some evidence of student thinking. This includes determining the intent, purpose, motivation, reasoning, or meaning of some evidence of student thinking.

Figure A.6 Critique Categories and Corresponding Codes

Appendix C: Research Participant Information and Assent Form

RESEARCH PARTICIPANT INFORMATION AND ASSENT FORM

You are being asked to participate in a research study. Researchers are required to provide a consent form to your parent(s) and an assent form to you to inform you about the research study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision.

You should feel free to ask the researchers any questions you may have.

Study Title: Use of Student Work in Curriculum-based Assessment Materials: Student Perspectives, Teacher Perspectives, and Comparisons to Student Work in Textbooks
Researcher: Corey Drake, Associate Professor and Amy Ray, Ph.D. Candidate
Department and Institution: Teacher Education, Michigan State University
Address and Contact Information: 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu

1. EXPLANATION OF THE RESEARCH and WHAT YOU WILL DO

You are being asked to participate in a research study of assessment tasks from middle school mathematics curriculum materials. You have been selected as a possible participant in this study because you are a student in a 7th grade classroom and your teacher uses the CMP curriculum. You will be asked to participate in a one-on-one interview with the researcher where you will solve a series of assessment tasks and then answer questions about these tasks.

2. POTENTIAL BENEFITS

You may not directly benefit from your participation in this study. However, your participation in this study may contribute to the understanding of how students understand and complete assessment tasks from the written materials. Your participation may also contribute to improvements in CMP3 assessment materials and possible teacher supports for using these assessments.

3. POTENTIAL RISKS

You may feel uncomfortable about talking out loud and being audio-recorded while you are solving a series of assessment tasks and answering questions about these tasks.

4. YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW

Participation in this research project is completely voluntary. You have the right to say no. You may change your mind at any time and withdraw. You may choose not to answer specific questions or to stop participating at any time.

5. PRIVACY AND CONFIDENTIALITY

- You will participate in a private, audio-recorded conversation with the secondary researcher. Your written student work will be recorded and collected.
- Although we will make every effort to keep your data confidential there are certain times, such as a court order, where we may have to disclose your data (audio-recording and any collected documents). Your voice will be audible in the audio-recording of the conversation with the secondary researcher. Unless required by law, only members of the research team and the Institutional Review Board will have access to this audio-recording.
- Consent and assent forms will be stored in a locked file cabinet in a locked office accessible only the primary investigator for three years after the close of the project. All electronic data collected in the study, including the audio-recording of your conversation with the secondary researcher, your written work on the assessment tasks, and transcribed student interview responses, will be scanned (when applicable) and stored on the secondary researcher's password-protected personal computer. Any written documents scanned for storage will then be destroyed. All data will be stored for at least three years.
- The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain anonymous.
- The appropriate administrator at your campus will be consulted and notified that a student on campus is participating in a study about assessment tasks.

6. CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you have concerns or questions about this study, such as scientific issues or how to do any part of it, please contact: Corey Drake, 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu or the graduate student researcher Amy Ray, 469-855-0781, rayamy1@msu.edu.

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Rd, Suite 136, Lansing, MI 48910.

7. DOCUMENTATION OF INFORMED ASSENT

I understand that even if my parent(s) signed a consent form, it is my choice to participate in this study. I understand that my choice will not change my grade.

By signing my name below, I volunteer to participate in Ms. Ray's research study. I understand that possible risks are small and are typical of other activities in math class.

My printed name _____

My signature (cursive) _____

Today's date _____

My age _____

Appendix D: Research Participant Information and Consent Form

RESEARCH PARTICIPANT INFORMATION AND CONSENT FORM

Your student is being asked to participate in a research study. Researchers are required to provide a consent form to inform you about the research study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision for your student. You should feel free to ask the researchers any questions you may have.

Study Title: Use of Student Work in Curriculum-based Assessment Materials: Student Perspectives, Teacher Perspectives, and Comparisons to Student Work in Student Textbooks

Researcher: Corey Drake, Associate Professor and Amy Ray, Ph.D. Candidate

Department and Institution: Teacher Education, Michigan State University

Address and Contact Information: 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu

1. EXPLANATION OF THE RESEARCH and WHAT YOUR STUDENT WILL DO

Your student is being asked to participate in a research study of assessment tasks from middle school mathematics curriculum materials. Your student has been selected as a possible participant in this study because he/she is in 7th grade and the CMP curriculum is used in your student's mathematics classroom. Your student will be asked to participate in a one-on-one interview with the researcher where your student will solve a series of assessment tasks and then discuss his/her experiences solving the tasks.

2. POTENTIAL BENEFITS

Your student may not directly benefit from his/her participation in this study. However, his/her participation in this study may contribute to the understanding of how students understand and solve assessment tasks from the written materials. Your student's participation may also contribute to improvements in CMP3 assessment materials and possible teacher supports for using these assessments.

3. POTENTIAL RISKS

Your student may feel uncomfortable about talking out loud and being audio-recorded while he/she is completing assessment tasks and answering questions about the tasks.

4. YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW

Participation in this research project is completely voluntary. You and your student have the right to say no. You and your student may change your mind at any time and withdraw. You and your student may choose not to answer specific questions or to stop participating at any time.

5. PRIVACY AND CONFIDENTIALITY

- Your student will participate in a private, audio-recorded conversation with the secondary researcher.
- Although we will make every effort to keep your student's data confidential there are certain times, such as a court order, where we may have to disclose your student's data (audio-recording and any collected documents). Your student's voice will be audible in the audio-recording of the conversation with the secondary researcher. Unless required by law, only members of the research team and the Institutional Review Board will have access to this audio-recording.
- Consent and assent forms will be stored in a locked file cabinet in a locked office accessible only the primary investigator for three years after the close of the project. All electronic data collected in the study, including the audio-recording of your student's conversation with the secondary researcher, copies of your student's written assessment work, and transcribed student interview responses, will be scanned (when applicable) and stored on the secondary researcher's password-protected personal computer. Any written documents scanned for storage will then be destroyed. All data will be stored for at least three years.
- The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain anonymous.
- The appropriate administrator at your student's campus will be consulted and notified that a student on campus is participating in a study about assessment tasks.

6. CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you have concerns or questions about this study, such as scientific issues or how to do any part of it, please contact: Corey Drake, 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu or the graduate student researcher Amy Ray, 469-855-0781, rayamy1@msu.edu.

If you have questions or concerns about your student's role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Rd, Suite 136, Lansing, MI 48910.

7. DOCUMENTATION OF INFORMED CONSENT.

Your signature below means that you voluntarily agree to have your student participate in this research study.

Signature

Date

Appendix E: Clinical and Semi-Structured Interview Questions for Student Interview

Introduction: Today, you will be asked to complete eight assessment tasks about solving problems using expressions and equations. As you are solving the tasks, I would like to you to not only use the space provided to show your written work, but also will ask you to “Think Aloud” as you complete your work. In other words, whatever thinking you are doing in your head, I would ask you to say it aloud. After you have completed all the assessment tasks, I will ask you a few questions about the tasks and your experience completing the task. Do you have any questions before we begin?

Background Information/Getting to Know

1. Can you tell me something interesting about yourself?
2. How would you describe your experiences learning math in school? [Probe elementary and middle school experiences]
3. Tell me about your experiences taking math assessments/tests.

Clinical Task Solving Interview

Here are the 8 assessment tasks I would like you to complete. You should use the pages provided to show any written work you do. You can also use a calculator if you need one. I will ask that you “Think Aloud” as you are solving each task. This is probably different from how you have solved assessment tasks in the past, so I will remind you of this idea as you are working! I will not be able to help you solve any of the tasks, but I can help you with the directions. I am interested in your thinking as you are solving tasks, not whether you get the answer right or wrong! Do you have any questions before you start?

As the student is solving the task, remind them to share their thinking if there is an extended period of silence. Feel free to ask questions to probe students’ thinking.

Questions about the Task Solving Tasks and Experience

1. Tell me about your general experience solving these tasks.
2. Do you see any similarities between any of the assessment tasks? [Could be the whole set or a subset]
3. Do you see any differences between any of the assessment tasks? [Could be the whole set or a subset]
4. [Keep track of which ones the student has mentioned] For the ones not mentioned, can you tell me your thoughts on these tasks?
5. Which tasks were the easiest for you to complete?
 - a. Why were these the easiest?
 - b. Was there anything in the way they wrote the problem down that made it easier?
6. Which tasks were the hardest for you to complete?
 - a. Why were these the hardest?
 - b. Was there anything in the way they wrote the problem down that made it harder?
7. You might have noticed that some of these questions provide information about a person’s thinking and ask you to make sense of it.
 - a. What do you think about these types of problems?
 - b. How are these problems similar or different from the other problems?
 - c. What do notice about these problems?

8. Do you have any final thoughts about any of the assessment tasks that you would like to share?

Appendix F: Clinical Interview Assessment Tasks

Go Math, Unit 3 Test A (Assessment Book, p. 86)

A technician charges an initial fee of \$300 plus an hourly fee of \$60. Mr. Jenks paid the technician \$480. How many hours did the technician work?

Figure A.7 Task A

Go Math, Expressions and Equations Module 6 Quiz B (Assessment Book, p. 35)

Paco solved the equation $-3x - 5 = 1$.
His work is shown below.

Step 1: $-3x - 5 = 1$

Step 2: $-3x = 6$

Step 3: $x = 2$

Where did Paco make an error in his calculation? Explain.

Figure A.8 Task B

CPM, Online Assessment Bank Repository #11801

Roneasha's solution to an equation is shown below. How did she do? Explain completely.

$$\begin{aligned} 3(x + -2) + 4 &= -1 + 3(1 - x) \\ 3x - 6 + 4 &= -1 + 3 - 3x \\ 3x - 2 &= 2 - 3x \\ \text{It's the reverse so} \\ \boxed{x = -1} \end{aligned}$$

Figure A.9 Task C

Big Ideas, Chapter 3 Test A #10 (Assessment Book, p. 31)

Write an expression in simplest form that represents the perimeter of the polygon.

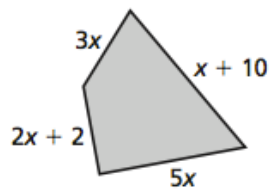


Figure A.10 Task D

CMP, Moving Straight Ahead Unit Test #4 (p. 1)

Multiple Choice Which of the following expressions is not equivalent to the others? Explain.

A. $6(x - 1) + 5$

B. $6x - 1$

C. $6(1 - x) + 5$

D. $5 + 6x - 6$

Figure A.11 Task E

Big Ideas, Chapter 3 (Student Textbook, p. 91)

Your friend says the sum of two linear expressions is always a linear expression. Is your friend correct? Explain.

Figure A.12 Task F

CPM, Chapter 6 Sample Test Task #9

Write an expression to represent Royce's weight if she is three pounds more than twice Ted's weight. If Ted weighs 30 pounds, how much does Royce weigh?

Figure A.13 Task G

CMP, Moving Straight Ahead ACE Problem (p. 83)

Students in Mr. Rickman's class are asked to solve the equation $\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) = 3$. Look at the three solutions below. Are they correct? Explain which method makes the most sense to you.

Jess's Solution

I began by distributing the numbers outside the parentheses on the left side.

$$\begin{aligned}\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) &= 3 \\ 4x - 6 + 2x - 3 &= 3 \\ 6x - 9 &= 3 \\ 6x &= 12 \\ x &= 2\end{aligned}$$

Terri's Solution

I began by multiplying each side of the equation by 3.

$$\begin{aligned}3\left[\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9)\right] &= 3(3) \\ 2(6x - 9) + 1(6x - 9) &= 9 \\ 12x - 18 + 6x - 9 &= 9 \\ 18x - 27 &= 9 \\ 18x &= 36 \\ x &= 2\end{aligned}$$

Brian's Solution

I knew that $\frac{2}{3} + \frac{1}{3} = 1$, so I simplified the right side to $6x - 9$.

$$\begin{aligned}\frac{2}{3}(6x - 9) + \frac{1}{3}(6x - 9) &= 3 \\ 1(6x - 9) &= 3 \\ 6x - 9 &= 3 \\ 6x &= 12 \\ x &= 2\end{aligned}$$

Figure A.14 Task H

Appendix G: Research Participant Information and Consent Form

RESEARCH PARTICIPANT INFORMATION AND CONSENT FORM

You are being asked to participate in a research study. Researchers are required to provide a consent form to inform you about the research study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision. You should feel free to ask the researchers any questions you may have.

Study Title: Use of Student Work in Curriculum-based Assessment Materials: Student Perspectives, Teacher Perspectives, and Comparisons to Student Work in Student Textbooks

Researcher: Corey Drake, Associate Professor and Amy Ray, Ph.D. Candidate

Department and Institution: Teacher Education, Michigan State University

Address and Contact Information: 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu

1. EXPLANATION OF THE RESEARCH and WHAT YOU WILL DO

You are being asked to participate in a research study of assessment tasks from middle school mathematics curriculum materials. You have been selected as a possible participant in this study because you teach 7th grade and use the CMP curriculum. You will be asked to participate in a one-on-one interview with the researcher about assessment tasks and students' written work on assessment tasks.

2. POTENTIAL BENEFITS

You may not directly benefit from your participation in this study. However, your participation in this study may contribute to the understanding of how teachers understand and use assessment tasks from the written materials. Your participation may also contribute to improvements in CMP3 assessment materials and possible teacher supports for using these assessments.

3. POTENTIAL RISKS

You may feel uncomfortable about talking out loud and being audio-recorded while you are answering questions about assessment tasks and students' written work on assessment tasks.

4. YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW

Participation in this research project is completely voluntary. You have the right to say no. You may change your mind at any time and withdraw. You may choose not to answer specific questions or to stop participating at any time.

5. PRIVACY AND CONFIDENTIALITY

- You will participate in a private, audio-recorded conversation with the secondary researcher.
- Although we will make every effort to keep your data confidential there are certain times, such as a court order, where we may have to disclose your data (audio-recording and any collected documents). Your voice will be audible in the audio-recording of the conversation with the secondary researcher. Unless required by law, only members of the research team and the Institutional Review Board will have access to this audio-recording.
- Consent forms will be stored in a locked file cabinet in a locked office accessible only the primary investigator for three years after the close of the project. All electronic data collected in the study, including the audio-recording of your conversation with the secondary researcher, and transcribed teacher interview responses, will be scanned (when applicable) and stored on the secondary researcher's password-protected personal computer. Any written documents scanned for storage will then be destroyed. All data will be stored for at least three years.
- The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain anonymous.
- The appropriate administrator at your campus will be consulted and notified that a teacher on campus is participating in a study about assessment tasks.

6. CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you have concerns or questions about this study, such as scientific issues or how to do any part of it, please contact: Corey Drake, 620 Farm Lane, 118A Erickson Hall, East Lansing, MI, 48824, 517-355-1713, cdrake@msu.edu or the graduate student researcher Amy Ray, 469-855-0781, rayamy1@msu.edu.

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Rd, Suite 136, Lansing, MI 48910.

7. DOCUMENTATION OF INFORMED CONSENT.

Your signature below means that you voluntarily agree to participate in this research study.

Signature

Date

Appendix H: Semi-Structured Interview Questions for Teacher Interview

Introduction: Today, we will be discussing assessment tasks and students' written work on assessment tasks. I will provide you with copies of the assessment tasks and (later in the interview) the corresponding student work on the assessment tasks. I will ask you questions related to your background, your thoughts on the provided assessment tasks and your understanding of students' work on the assessment tasks. Do you have any questions before we begin?

Background Information

1. To begin, can you please tell me about your teaching background? [Grade levels, subjects, curriculum materials, years of teaching, etc.]
2. What subjects and grade levels are you teaching this year?
3. How would you describe your assessment practices?
 - a. What kind of assessments do you use?
 - b. Where do you find assessment materials?
 - c. What role does assessment play in your classroom?

Analysis and Discussion of Assessment Tasks

Teacher will be provided copies of the assessment tasks students previously solved. These assessment tasks focused on the CCSSM Standard: 7.EE – Solve real-life and mathematical problems using numerical and algebraic expressions and equations.

1. What do you think about these tasks?
 - a. What do you notice about these tasks?
 - b. What do you wonder about these tasks?
2. Looking at each task individually, what is each task assessing? What in the written task tells you this? [Have the teacher look through Tasks A-H and jot down their response on the provided lines.]
3. Do you see any similarities between any of the assessment tasks?
4. Do you see any differences between any of the assessment tasks? [Could be the whole set or a subset]

Focus on Assessment Tasks with Student Work

We are going to look at a subset of questions in this set as compared to the others. I am interested in learning more about a particular type of assessment question. The subset of questions here all have student work. [Tasks B, C, F, & H]

1. In looking at this subset, what do you think I mean by “student work”?
 - a. What counts as “student work” in Tasks B, C, F, & H?
2. What might be the advantages of using these types of problems in assessments?
3. What might be the disadvantages of using these types of problems in assessments?
4. One part of the CCSSM Practice Standards promotes students' *critiquing the reasoning of others* (part of Mathematical Practice #3). Some teachers and curriculum writers have hypothesized that assessment tasks with student work can assess this practice.
 - a. What are your thoughts on this idea?
 - b. Do these four tasks assess this practice?

Analyzing Students' Written Work on Assessment Tasks and the Assessment Tasks Themselves

1. Which assessment tasks are you interested in seeing written student work for? Which tasks are you not interested in seeing written student work for?
 - a. Why?

[Choose 1-2 Tasks the teacher is interested in and 1-2 Tasks the teacher is not interested in and share written student work one at a time. Be sure that this collection includes “student work” and non-“student work” task examples. Make sure to include Task F even if the teacher does not mention this task.]

2. In looking at the students' written work...
 - a. Does anything surprise you?
 - b. What do you notice?
 - c. What do you wonder about?
3. In looking at the students' written work and the ideas you jotted down earlier about what the task was assessing, would you add to or edit anything you have written? [Provide teacher with a different color pen and ask them to jot down their additional or revised ideas.]

Revisit Assessment Tasks with Student Work

1. Revisiting the four tasks with “student work”, do you see evidence that students critiqued the reasoning of others based on their written student work?
 - a. Do you see evidence in Tasks B, C, F, or H?

Wrap-up Questions

1. Of these eight assessment tasks, which would you use, if any?
 - a. How? [ex. In-class, on assessment, etc.]
 - b. Why?
2. Do you have any final thoughts about the assessment tasks, the written student work, or the use of student work in assessment tasks?

Appendix I: Students' Written Work from Clinical Interview Assessment Tasks

A Jane $\begin{array}{r} 480 \\ -300 \\ \hline 180 \end{array}$ $60 \overline{)180}^3$ 3 hours	Susan $\begin{array}{r} \$300 \\ 180 \\ \hline \$480 \end{array}$ $\begin{array}{r} 60 \\ \times 3 \\ \hline 180 \end{array}$ 3 hours	Cynthia $\begin{array}{r} 300 \\ 60 \\ \hline 1800 \\ 60 \\ \hline 420 \\ 60 \\ \hline 1470 \\ 60 \\ \hline 530 \end{array}$
David $\begin{array}{r} 480 \\ -300 \\ \hline 180 \end{array}$ $180 \div 60 = 3$	Anna $\begin{array}{r} 300 \quad 60 \\ 360 \end{array}$ $\begin{array}{r} 480 \\ 360 \\ \hline 120 \end{array}$ 2 hours	Ed 3 hours $\begin{array}{r} 300 \\ 3 \times 480 \\ -300 \\ \hline 180 \end{array}$ $60 \times 3 = 180$

Figure A.15 Student Work on Task A

B Jane $\begin{array}{r} -3x - 5 = 1 \\ +5 +5 \\ \hline -3x = 6 \\ \div -3 \\ \hline x = -2 \end{array}$ Between step 2 and 3	Susan $x = -2$ $-3x - 5 = 1$ -3×6 Paul did not put the negative before the 2 in "x=2"	Cynthia $-3x - 5 = 1$ step 2 may have made a error, how did he get the 6?
David step 3, because $-3 \cdot 2 \neq 6$ $x = -2$	Anna $\begin{array}{r} -3x - 5 = 1 \\ +5 +5 \\ \hline 0 \quad 6 \\ -3x = 6 \quad -2 \\ \hline -3 \quad -3 \end{array}$	Ed $\begin{array}{r} -3x - 5 = 1 \\ -3 + 2 = -6 \end{array}$ step 2 is wrong because $-3 \times 2 \neq 6$, it equals -6.

Figure A.16 Student Work on Task B

C Jane $3(x+(-6))+4=-1+3-3x$ $3x-2=2-3x$ $3x+3x=2+2$ $6x=4$ $x=\frac{4}{6}$ $x=\frac{2}{3}$ Roneasha got the answer wrong because she skipped a step at the end, which made her get confused on the negatives negatives and if you were to solve and check using her answer, you would get $-3\frac{2}{3}-2=2\frac{2}{3}$ which is not the same thing at all. ($-5=5$).	Susan $3(x+2)+4=-1+3(1-x)$ $3x-6+4=-1+3-3x$ $3x-2=2-3x$ $3-1=3-2=-1=2-3-1$ $-3-2=-$ $x=-1$ when I put -1 to replace "x" I got -1	Cynthia $x=-1$ stands for the ending answer.
David $3(x+2)+4=-1+3(1-1)$ The problem is not equivalent so she would not have been able to get this question.	Anna $3(x+2)+4=-1+3(1-x)$ $3x-1x+4=-1+3-1x$ $-3x+4=-2-1x$ $1x=-1x$ I think she did good because I solved the equation and may have messed up but the answer to her equation in the end was the reverse so I think she is right.	Ed She did badly. None of the equations are equal if $x=-1$.

Figure A.17 Student Work on Task C

D Jane $3x+5x+x+10+2x+2$ $=3x+5x+x+2x+10+2$ $=11x+12$	Susan $(2x+2)+(3x)+(x+10)+(5x)$ $(3x)+(5x)+(4x)+(x+10)$	Cynthia $(3x \times 5x + 10 \times 2x + 2)$
David $11x+12$ $11x+12$	Anna $3x+2x+2+5x+x+10$ 	Ed $11x+12$

Figure A.18 Student Work on Task D

E Jane A. $6x-6+5$ $6x-1$ B. $6x-1$ C. $6-6x+5 = 6x+11$ (C)	Susan B. in C $6x(6-x+5)$ $6x-1$ $6(1-5)+5=$ $6-30=24+$ $6+5=30-1=29$ $\frac{29}{29}$ $5+6+5=6$ $5+30=35-6=29$ A. $6(5-1)+5$ $30-1=29+5=34$	Cynthia -1 1 11 -1 C. is not the same as A,B,D by not having the same answer.
David $6=2.5$ $6(2.5)=5$ A=14 $6+5=14$ $6+2.5=1$ B=14 $15-1=14$ $6(1-2.5)+5$ C=-4 $6+5=14$ $5+6+2.5=6$ D=14 $15+5=6$ $20-6=14$ (C)	Anna $6(x-1)+5$ $6x-1$ $6 \cdot -1+5$ $5x$ $-6+5$ -1 $6(1-x)+5$ $5+6x-6$ $6 \cdot -1+5$ $-6+5$ -1 B	Ed $T=1$ $6(1-1)+5$ $6 \cdot 0+5$ $0+5=5$ $6(1-1)+5$ $6 \cdot 0+5$ $0+5=5$ $5+6+1-6$ 11

Figure A.19 Student Work on Task E

F Jane linear expression = $y = mx + b$ Ces. When you add two values, you do not get any powers higher than 1, which means you won't get any quadratics or exponential functions or anything with a weird curve. Also, you can't get an inverse variation or something because that's in a whole different form and the basic form's not going to change. If you were multiplying however, that's a whole different story.	Susan no, yes, both expressions are linear so when added together would be linear.	Cynthia no, because it may be up linear (meaning being straight).
David 59 114 Linear ✓ 8-16 24 11-23 34 14-30 44 17-37 54 9-8 17 Linear ✓ 10-9 19 11-10 20 12-11 21 13-12 22 14-13 23 15-14 24 16-15 25 17-16 26 18-17 27 19-18 28 20-19 29 21-20 30 22-21 31 23-22 32 24-23 33 25-24 34 26-25 35 27-26 36 28-27 37 29-28 38 30-29 39 31-30 40 32-31 41 33-32 42 34-33 43 35-34 44 36-35 45 37-36 46 38-37 47 39-38 48 40-39 49 41-40 50 42-41 51 43-42 52 44-43 53 45-44 54 46-45 55 47-46 56 48-47 57 49-48 58 50-49 59 no, because the more complicated the equation, the less likely it is to be linear. yes, because if 2 numbers that are sums of unit ones are added, then a new unit there will be created.	Anna I think that she is correct because if you add up the two linear expressions and combine like terms there would still be things that would represent a linear expression.	Ed

Figure A.20 Student Work on Task F

G Jane $R = 3 + 2T$ $R = 3 + 2 \cdot 30$ $R = 3 + 60$ $R = 63$	Susan $R = T + 3 \text{ pounds}$ $R = 33 \text{ pounds}$	Cynthia 60 pounds Maybe by doubling it by 10.
David $\begin{array}{r} 60 \\ + 3 \\ \hline 63 \end{array} \text{ Pounds}$ $\begin{array}{r} 30 \cdot 2 + 3 \\ 60 + 3 \\ \hline 63 \end{array}$	Anna $\begin{array}{r} 30 \\ \times 2 \\ \hline 60 \end{array}$ 63 lbs	Ed $30 \times 2 + 3 = 63$

Figure A.21 Student Work on Task G

H Jane They all are correct, but Jess's solution makes the most sense because it is the way I would do it and it's the most straightforward.	Susan Not correct Terri is multiplying by 2 the inside of the parentheses 30 lbs the number (3) on the side of outside of the parentheses Brian said he could get "60-9" from "3 + 1" but if it was "3 + 1" then he can't get "60-9" I think Jess is correct because she is multiplying 30 and 2 by the inside of the equation	Cynthia Jess's solution more easier and simply to look at.
David Jess ✓ Terri ✓ Brian ✓ Jess, because she worked out the problem mathematically with the correct formulas and didn't over or over simplify it. Terri over-simplified it, so it took more time. Brian did it easily, but could have easily messed up at any stage.	Anna I think Jess's solution made the most sense because I would also start by distributing. Also the other 2 did not make sense to me that much because in Terri's solution he multiplies each side by 3 and I'm not sure why.	Ed $\begin{array}{r} 60 \div 2 \frac{1}{2} \\ = 9 \\ \times 3 \\ \hline 27 \end{array}$ They are all correct. Brian's solution made the most sense to me because it was easy and simple. I also liked how he noted the trick to do this. Brian's solution is closest to what I would have done.

Figure A.22 Student Work on Task H

Appendix J: SWAT Analyses by Assessment Type Additional Findings

SWAT by Assessment Type

I considered how the 127 SWAT varied across assessment types based on the CCSSM content strand foci, the evidence of student thinking provided, and the type of critique required in the task to determine if any interesting differences or trends emerged across assessment types. As a reminder, of the 127 SWAT, 0 tasks were diagnostic assessment tasks, 13 tasks were periodic assessment tasks (10.2%), 37 tasks were summative assessment tasks (29.1%), and 77 tasks were question bank tasks (60.6%). Thus, SWAT were mostly found in question bank, which were only represented by CMP and CPM. The following tables and figures detail findings for periodic, summative, and question bank assessment types based on the CCSSM content strands, evidence types, and critique types.

Table A.1

Number of SWAT by Assessment Type and CCSSM Content Strand for Grade 7

Assessment Type	RP	NS	EE	G	SP	All CCSSM
Periodic	4	3	4	1	1	13
Summative	4	8	6	6	13	37
Bank	1	44	18	9	5	77
Totals (N=127)	9	54	28	17	19	

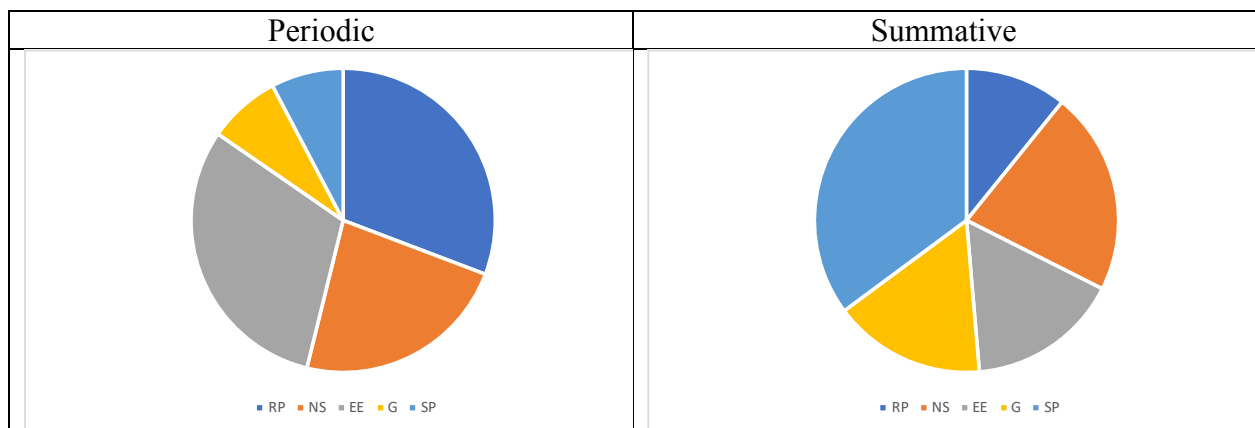


Figure A.23 Comparison of SWAT Assessment Types by CCSSM Content Strands for Grade 7

Figure A.23 (cont'd)

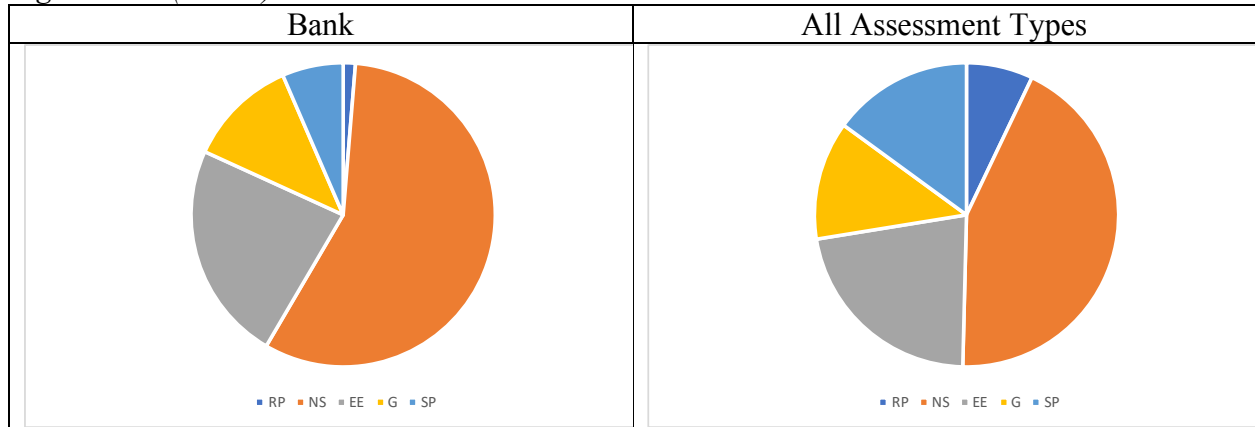


Table A.2

<i>Number of Instances of Student Thinking in SWAT by Assessment Type and Evidence Type</i>					
Assessment Type	Words	Symbols	Visuals	Actions	All Evidence
Periodic	6	4	2	4	16
Summative	19	9	5	11	44
Bank	50	25	11	8	94
Totals (N=154)	75	38	18	23	

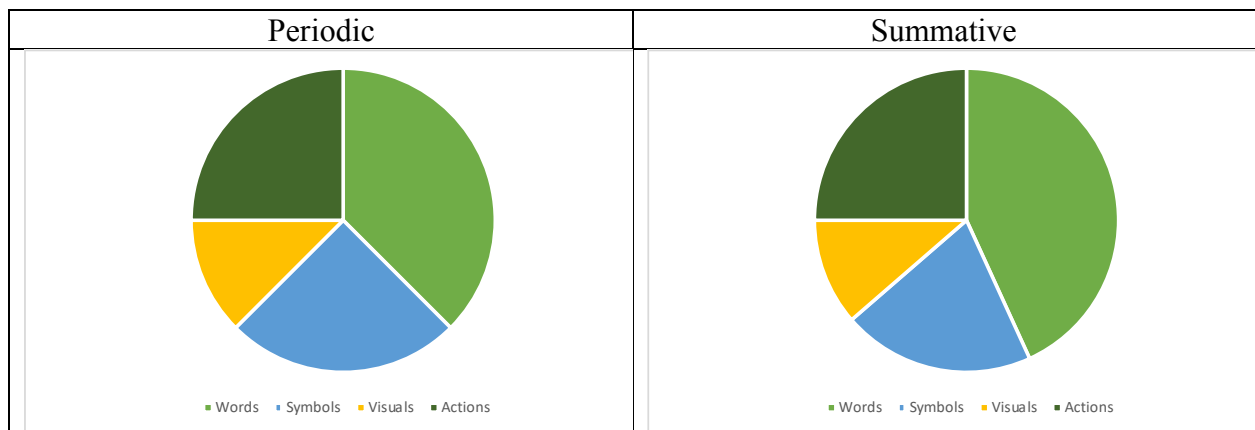


Figure A.24 Comparison of SWAT Assessment Types by Evidence of Student Thinking Types

Figure A.24 (cont'd)

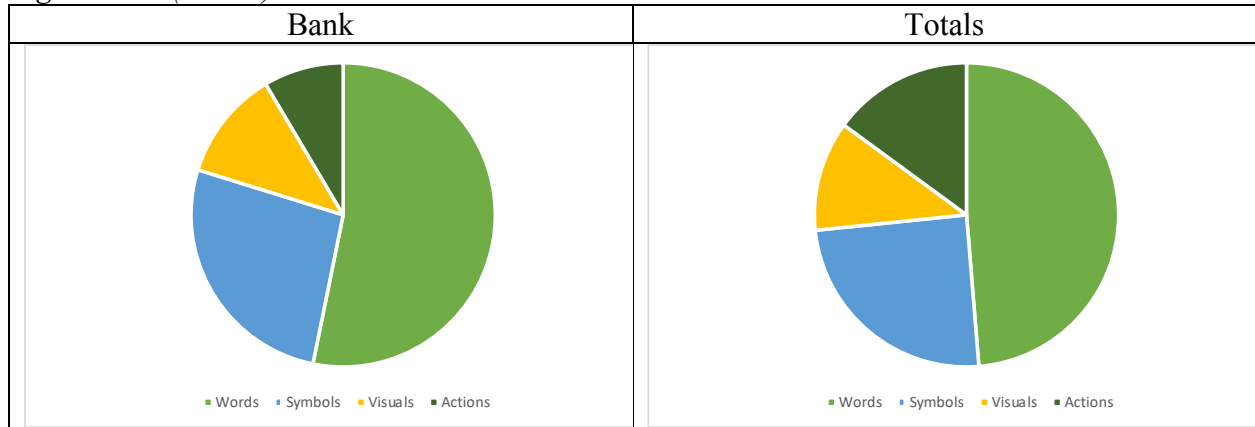


Table A.3

Number of Instances of Critiques in SWAT by Assessment Type and Critique Type

Assessment Type	Error ID	Eval	Compare	Pref	Insights	All Critiques
Periodic	4	7	3	0	2	16
Summative	6	28	8	0	3	45
Bank	9	59	20	0	11	99
Totals (N=160)	19	94	31	0	16	

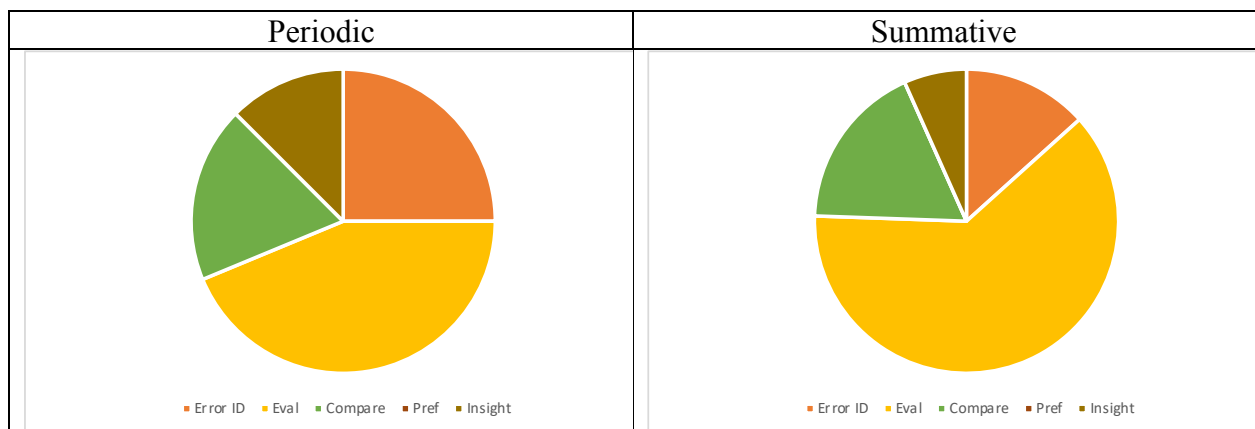
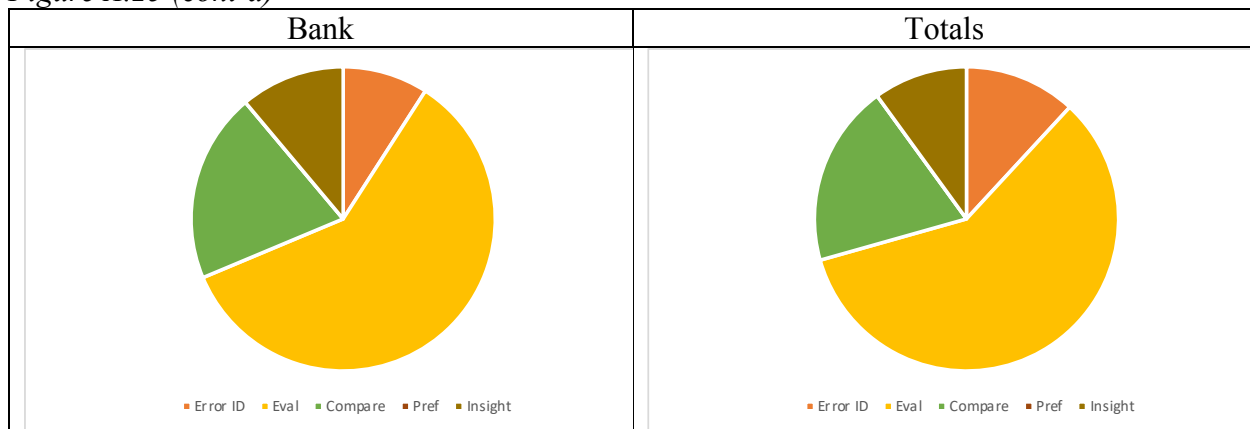


Figure A.25 Comparison of Assessment Types in SWAT by Critique Types

Figure A.25 (cont'd)



Periodic Assessments

For the 13 total periodic tasks, ratios and proportional relationships and expressions and equations were the most common content strands each accounting for 30.8% of the periodic SWAT tasks. Geometry (7.7%) and statistics and probability (7.7%) strands were the least common. The most common evidence type for periodic tasks was in the form of someone's thoughts, claims, or explanations (37.5%). The two second most common evidence types were symbolic representations and mathematical actions (25% each). The reader was least likely to be tasked with making sense of someone's visual representations (12.5%). The most common critique type was evaluation (44.8%). Error identification was the second most common critique type (25%). The least common type of critique found in the tasks, other than preference, was providing insight into some evidence of student thinking (12.5%).

Summative Assessments

For the 37 summative tasks, statistics and probability (35.1%) was the most common content strand accounting for over a third of the summative SWAT tasks, followed by the number system (21.6%). The least common strand was ratios and proportional relationships (10.8%). The most common evidence type was someone's words (37.5%). The second most common evidence type was descriptions of someone's mathematical actions (25%). The reader

was least likely to be tasked with making sense of someone's visual representations (11.4%). The critique type of evaluation occurred a majority of the instances for summative tasks (62.2%). The second most common critique type was comparison (17.8%). The least common type of critique found in the tasks, other than preference, was providing insight into some evidence of student thinking (6.7%).

Question Bank Assessments

The 77 question bank tasks heavily assessed content focused on the number system (57.1%), accounting for a majority of the question bank tasks. The second most common content focus for the question bank tasks was expressions and equations (23.4%). Only 1 question bank task assessed content related to ratios and proportional relationships (1.3%). The most common evidence type was someone's words (53.2%). The second most common evidence type was symbolic representations (26.6% for question bank). The reader was least likely to be tasked with making sense of someone's mathematical actions (8.5%). The critique type of evaluation occurred a majority of the instances for question bank tasks (59.6%). The second most common critique was comparison (20%). Error identification was the least prevalent critique type (9.1%) other than preference.

Summary of SWAT by Assessment Type Findings

From the analyses of the SWAT assessment types by content strands, evidence types, and critique types, few significant findings emerged from comparing the SWAT by the different assessment types. The main findings included: (1) periodic and summative assessment tasks were more balanced in assessing across the content strands with no single strand having a majority as compared to the question banks, (2) evidence type distributions varied little across the assessment types other than the majority of question bank tasks requiring the reader to make

sense of someone's words, and (3) critique type distributions varied little across the assessment types with one of the few exceptions being a greater emphasis on error identification in the periodic assessments.

REFERENCES

REFERENCES

- An, S., & Wu, Z. (2012). enhancing mathematics teachers' knowledge of students' thinking from assessing and analyzing misconceptions in homework. *International Journal of Science and Mathematics Education*, 10(3), 717-753.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2), 181-214.
- Booth, J., Lange, K., Koedinger, K., & Newton, K. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, 25, 24-34.
- Boston, M. D. (2014). Assessing Instructional Quality in Mathematics Classrooms Through Collections of Students' Work. In *Transforming Mathematics Instruction* (pp. 501-523). Springer International Publishing.
- Burger, E. (2014). *Go math*. Orlando, Fla: Houghton Mifflin Harcourt.
- Cameron, M., Loesing, J., Rorvig, V., & Chval, K. B. (2009). Using student work to learn about teaching. *Teaching Children Mathematics*, 15(8), 488-493.
- Chandler, K., Fortune, N., Lovett, J. N., & Scherrer, J. (2016). What should Common Core assessments measure?. *Phi Delta Kappan*, 97(5), 60-63.
- Common Core State Standards Initiative (CCSSI). (2010). *Common Core State Standards for Mathematics*. Washington, D.C.: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from <http://www.corestandards.org>
- Connected Mathematics Project (CMP). (2018a). *CMP and the Common Core Standards for Mathematical Practice*. Retrieved from <http://connectedmath.msu.edu>
- Connected Mathematics Project (CMP). (2018b). *The Student Work in Curriculum Materials Research Project*. Retrieved from <http://connectedmath.msu.edu>
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research, 3e*. Thousand Oaks, CA: SAGE.
- Dietiker, L., Kassarjian, M., & Nikula, M. (2013). *Core connections 2, course 2: College Preparatory Mathematics*. Elk Grove, CA: CPM Educational Program.

- Dietiker, L., Kassarian, M., & Nikula, M. (2018). CCSS Standards for Mathematical Practice in CPM core connections courses. In *Core connections, course 2: College Preparatory Mathematics*. Retrieved from <http://ebooks.cpm.org>
- Driscoll, M., & Moyer, J. (2001). Using students' work as a lens on algebraic thinking. *Mathematics Teaching in the Middle School*, 6(5), 282-287.
- Ely, R. E., & Cohen, J. S. (2010). put the right spin on student work. *Mathematics Teaching in the Middle School*, 16(4), 208-215.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, Mass: Bradford Books.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (2005). Research on middle school renewal: Teacher views on collaborative review of student work. *Middle School Journal*, 37(2), 56-60.
- Ghousseini, H., & Sleep, L. (2011). Making practice studyable. *Zdm*, 43(1), 147-160.
- Gilbertson, N.J., Edson A.J., Grant, Y., Lawrence, K.A., Nimitz J., Phillips, E.D., & Ray, A. (2016). An Analytic Framework for Examining Curriculum-Generated Student Work, presented at the American Educational Research Association Annual Conference, Washington, D.C., 2016.
- Ginsburg, H. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics*, 1(3), 4-11.
- Glesne, C. (2006). *Becoming qualitative researchers: An introduction, 4th edition*. Boston: Pearson.
- Go Math. (2018). *Meeting rising standards with focus, coherence, and rigor*. Retrieved from <http://hmhco.com>
- Going, T., Ray, A., & Edson, A. J. (in preparation). Analysis of student work tasks embedded in 7th grade CCSSM-aligned mathematics curricula.
- Gotwals, A. W., Hokayem, H., Song, T., & Songer, N. B. (2013). The role of disciplinary core ideas and practices in the complexity of large-scale assessment items. *Electronic Journal of Science Education*. 17(1), 1-25.
- Gotwals, A. W., & Songer, N. B. (2013). Validity evidence for learning Progression-Based assessment items that fuse core disciplinary ideas and science practices. *Journal of Research in Science Teaching*, 50(5), 597-626.
- Great Minds (2015). *Eureka Math: A story of ratios*. Washington D.C.: Great Minds.

- Hamilton, L., Stecher, B., & Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. *Education Inquiry*, 3(2). Retrieved from <http://www.education-inquiry.net/index.php/edui/article/view/22025>
- Herbel-Eisenmann, B. A., & Phillips, E. D. (2005). Using student work to develop teachers' knowledge of algebra. *Mathematics Teaching in the Middle School*, 11(2), 62-66.
- Herman, J., & Linn, R. (2013). On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia. CRESST Report 823. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.
- Hull, T. H., Balka, D. S., & Miles, R. H. (2013). Mathematical rigor in the Common Core. *Principal Leadership*, 14(2), 50-55.
- Hull, T. H., Miles, R. H., & Balka, D. S. (2012). *The common core mathematics standards: Transforming practice through team leadership*. Thousand Oaks, Calif: Corwin Press.
- Hunsader, P. D., Thompson, D. R., & Zorin, B. (2013). Engaging elementary students with mathematical processes during assessment: What opportunities exist in tests accompanying published curricula?. *International Journal for Mathematics Teaching & Learning*.
- Hunsader, P. D., Thompson, D. R., Zorin, B., Mohn, A. L., Zakrzewski, J., Karadeniz, I., Fisher, E.C., & MacDonald, G. (2014). Assessments accompanying published textbooks: The extent to which mathematical processes are evident. *ZDM*, 46(5), 797-813.
- Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education*, 7(3), 203-235.
- Kersaint, G., & Chappell, M. F. (2004). What do you see? A case for examining students' work. *The Mathematics Teacher*, 97(2), 102-105.
- Koestler, C., Felton, M. D., Bieda, K., Otten, S. (2013). *Connecting the NCTM process standards and the CCSSM practices*. National Council of Teachers of Mathematics.
- Lannin, J., Townsend, B., & Barker, D. (2006). The reflective cycle of student error analysis. *For the Learning of Mathematics*, 26(3), 33-38.
- Lappan, G., Phillips, E. D., Fey, J. T., & Friel, S. N. (2014). *Stretching and shrinking: Connected Mathematics 3*. Boston, MA: Pearson.
- Larson, R., & Boswell, L., (2014). *Big Ideas Math Course 2: A bridge to success*. Boston, MA: Houghton Mifflin Harcourt.
- Little, J. W., Gearhart, M., Curry, M., & Kafka, J. (2003). Looking at student work for teacher learning, teacher community, and school reform. *The Phi Delta Kappan*, 85(3), 184-192.

- McDonald, J. P. (2002). Teachers studying student work: Why and how? *The Phi Delta Kappan*, 84(2), 120-127.
- Mislevy, R. J. (2012). The case for informal argument. *Measurement: Interdisciplinary Research and Perspectives*, 10(1-2), 93-96.
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of Evidence-Centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6-20.
- Mislevy, R. J., & Riconscente, M. M. (2005). Evidence-centered assessment design: Layers, structures, and terminology (Principled Assessment Designs for Inquiry Technical Report 9). *Menlo Park, CA: SRI International*.
- Mislevy, R. J., Steinberg, L. S., Almond, R. G., National Center for Research on Evaluation, Standards, and Student Testing, & University of California, Los Angeles. Center for the Study of Evaluation. (2003). *On the structure of educational assessments, CSE technical report*. Distributed by ERIC Clearinghouse.
- National Council of Teachers of Mathematics. (2014). *Principles to action: Ensuring mathematical success for all*. Reston, VA: NCTM.
- National Research Council (NRC). *Adding it up: Helping children learn mathematics*. Edited by Kilpatrick, J., Swafford, J., Findell, B. Washington, DC: National Academies Press, 2001.
- Nimitz, J.L., Gilbertson, N.J., Lawrence, K.A., Ray, A., Edson, A.J., Grant, Y., & Phillips, E. Student work as a context for student learning. In Bartell, T. G., Bieda, K. N., Putnam, R. T., Bradfield, K., & Dominguez, H. (Eds.). (2015). *Proceedings of the 37th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. East Lansing, MI: Michigan State University.
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? an experimental study on learning to solve equations. *Journal of Educational Psychology*, 99(3), 561-574.
- Rittle-Johnson, B., & Star, J. R. (2011). The Power of Comparison in Learning and Instruction: Learning Outcomes Supported by Different Types of Comparisons. *Psychology of Learning and Motivation*, 55, 199-225.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, 101(4), 836-852.
- Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures?: Developing procedural flexibility. *British Journal of Educational Psychology*, 82(3), 436-455.

- Rowland, T. (2008). The purpose, design and use of examples in the teaching of elementary mathematics. *Educational Studies in Mathematics*, 69(2), 149-163.
- Ryken, A. E. (2009). Multiple representations as sites for teacher reflection about mathematics learning. *Journal of Mathematics Teacher Education*, 12(5), 347-364.
- Sandholtz, J. H. (2005). Analyzing teaching through student work. *Teacher Education Quarterly*, 32(3), 107-122.
- Schoenfeld, A. H. (2007). Issues and tensions in the assessment of mathematical proficiency. In A. Schoenfeld (Ed.), *Assessing Mathematical Proficiency* (pp. 3–15). Cambridge: Cambridge University Press.
- Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, 10(1/2), 9-34.
- Seeley, C. L. (2014). Smarter than we think. *Sausalito, CA: Math Solutions*.
- Seidman, I. (2012). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Silver, E. A., & Suh, H. (2014). Professional development for secondary school mathematics teachers using student work: Some challenges and promising possibilities. In *Transforming Mathematics Instruction* (pp. 283-309). Springer International Publishing.
- Slavit, D., & Nelson, T. H. (2010). Collaborative teacher inquiry as a tool for building theory on the development and use of rich mathematical tasks. *Journal of Mathematics Teacher Education*, 13(3), 201-221.
- Smith, M. S., Hughes, E. K., Engle, R. A., & Stein, M. K. (2009). Orchestrating discussions. *Mathematics Teaching in the Middle School*, 14(9), 548-556.
- Star, J., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology*, 40, 41-54.
- Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, 102(4), 408-426.
- Swan, M., & Burkhardt, H. (2012). A designer speaks: Designing assessments and performance in mathematics. *Educational Designer*, 2(5), 1-41.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (Expand 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.