INTERROGATING LARGE-SCALE SCIENCE ASSESSMENT:
EXPOSING EVIDENCE OF NEXT GENERATION SCIENCE STANDARDS DIMENSIONS

By

Tamara J Heck

A DISSERTATION

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

Curriculum, Instruction and Teacher Education—Doctor of Philosophy

2021

ABSTRACT

INTERROGATING LARGE-SCALE SCIENCE ASSESSMENT:
EXPOSING EVIDENCE OF NEXT GENERATION SCIENCE STANDARDS DIMENSIONS

By

Tamara J Heck

The adoption and implementation of state science standards based on the Next Generation Science Standards (NGSS) have posed significant challenges for the development and interpretation of science Large-Scale Assessments (LSA); specifically, the extent to which the assessment items align with the standards. Previous research has relied on large state or national data sets and classroom assessment data, but has yet to consider students' experiences with the LSA. This research uses qualitative methods to analyze cognitive lab data from Michigan students in Grades 5 and 8 to find evidence of the intended alignment claims of the items designed for the science Michigan Student Test of Educational Progress (M-STEP). The findings indicate that the items elicit the use of the dimensions from students and, in some cases, discriminate among students who chose the keyed response versus those who did not. The evidence of elicitation is often in contrast to the alignment analysis conducted by external reviewers through the Task Annotation Project in Science (TAPS). This work highlights important tensions with which science assessment developers must wrestle and provides recommendations for doing so.

To My Family

ACKNOWLEDGEMENTS

My dissertation study would not have been possible without the support of many people. First, I thank my advisor, Dr. Amelia Wenk Gotwals, whose patience, dedication, and counsel guided me through the dissertation process. I also acknowledge my committee members—Dr. Alicia Alonzo, Dr. Andy Anderson, and Dr. Sandra Crespo—for providing insightful feedback and helping me refine my ideas. My gratitude also extends to Andrew Middlestead, at the Michigan Department of Education, who has supported my work since the beginning. My friends and colleagues—Dr. Mary Starr, Dr. Joi Merritt, and Dr. Angela Kolonich—provided essential counsel, support, advice, and expertise throughout this project. Without each of you, this work would not have been complete. I am grateful for the teachers and colleagues across Michigan whose dedication to science education provided me access to thoughtful students. Additionally, I acknowledge each of the students who so bravely shared their ideas with me.

I am thankful for my parents—my mother who encouraged me to keep smiling as I moved my work forward, and my father whose counsel to confront issues head-on has been integral to my success. Finally, my deepest gratitude goes to my amazing children, Kai and Levi, whose love and tolerance of a mother constantly studying and working has not gone unnoticed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In the modern education system, students face multiple assessments throughout the school year. The purposes of these assessments include formative assessment data collection, programmatic accountability and monitoring, and prediction of success on the collegiate level. The No Child Left Behind Act (No Child Left Behind [NCLB], 2002) initiated the placement of greater emphasis on the use of assessments to track students' educational progress and to serve as an accountability measure of schools. NCLB required states to create "challenging" standards for students at each grade level and assess students from Grades 3 to 8 and once in high school.

In 2015, President Obama signed the Every Student Succeeds Act (Every Student Succeeds Act [ESSA], 2015) into law. With respect to statewide assessments, ESSA continues the tradition begun by NCLB by requiring state-level, high-quality student assessments in mathematics, reading or language arts, and science (ESSA 2A, p. 24) aligned to the state standards and that "provide coherent and timely information about student attainment of such standards and whether the student is performing at the student's grade level" (ESSA 1177-S, pp. 24–25). While ESSA provides more options for state assessment and accountability, it is still the responsibility of the state to endure all students take summative or interim assessments in English Language Arts and Mathematics each year from grades three through grade eight and once in high school. The law also requires all students participate in a science assessment three times in a student's K-12 experience: in the elementary, middle, and high school grade bands.

New Science Standards

In parallel with ESSA, in 2015, Michigan adopted the Michigan K-12 Science Standards (MSS), which are based on the Framework for K-12 Science Education (National Research

Council [NRC], 2012) and the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). The NGSS are written in the form of performance expectations, which are the assessable statements students should be able to do at specific grade levels. The performance expectations were adopted as the MSS with a few Michigan-specific contexts noted throughout the adoption document.

The MSS are three-dimensional (3D), meaning students are required to integrate three separate but interdependent competencies. The first dimension, the Science and Engineering Practices (SEPs), depict the actions, skills, and performances employed by both scientists and engineers in which scientists engage as they investigate and explore the natural and designed world (NRC, 2012). The second dimension is the Disciplinary Core Ideas (DCIs), "which have the power to focus K–12 science curriculum, instruction, and assessments on the most important aspects of science" (nextgenscience.org/three-dimensions). The third dimension, the Crosscutting Concepts (CCCs), traverse all domains of science in an effort to provide coherence across various scientific ideas. When taught with fidelity, the CCCs can shape scientific literate students (NRC, 2012). The amalgamation of these three dimensions promote students' ability to investigate natural phenomena and explore solutions to real-world problems (NRC, 2012; NGSS Lead States, 2013).

Research Questions

This study examines the alignment of the science item clusters using evidence from cognitive labs with 19 students and an external alignment analysis. This study will focus on two clusters: one in fifth-grade Life Science and one in eighth-grade Physical Science. The data collected was used to analyze the assessment claims the writing teams crafted and determine if there is evidence to support using the clusters for large-scale assessments.

The research questions guiding this research are as follows:

1) To what extent do the clusters developed using Michigan Cluster Development process align with the Michigan K-12 Science Standards?

    1a) To what extent do these items elicit and discriminate for the intended dimensions?

A New Science Test

Until Spring 2017, students in Michigan were assessed in science at Grades 4, 7, and 11 using the Michigan Standard Test of Educational Progress (M-STEP) science assessment, which was implemented in 2015 due to a legislative bill that required online state assessments (House Bill No. 5314, 2014). The prior assessments, the Michigan Educational Assessment Program (MEAP), was operationalized in 2006 and continued until 2014. The M-STEP and the MEAP assessed the Grade Level Content Expectations and the High School Content Expectations which were standards that treated science knowledge and inquiry as separate entities.

To ensure alignment with the new standards, assessment development commenced shortly after the adoption of the MSS. To assess the new complex standards, a new assessment was designed to interweave the science content, practices, and cross cutting concepts to determine what students know and can do in science.[1] The new MSS present challenges for the design of assessments. For example, because of the three-dimensional nature of the new MSS, it is nearly impossible to craft a single assessment item that is aligned to all that is contained within a performance expectation. To address this problem, the new State of Michigan science assessment uses item clusters, which are groups of five to eight items that are dependent on a common stimulus based on a scientific phenomenon or engineering problem (The State

---

[1] In this dissertation, I refer to the NGSS performance expectations and the MSS synonymously, often referring to both as "the standards."

3

Assessment Item Collaborative [SAIC], 2015). This process is discussed in more detail in the next chapter.

The Michigan Science Assessment System

This study focuses on just one point in the assessment system, but to clearly describe the conceptual framing for this work, it is important to note that one assessment occasion does not provide sufficient data to meet the requirements of ESSA legislation. Designing an assessment system must be at the fore of science assessment work (NRC, 2014). Stiggins (2007) discusses the importance of an assessment system that includes assessments *of* learning, those that provide a picture of what students learned over a designated period of time, and assessments *for* learning, those designed to help students determine their current understandings in an effort to move learning forward. Others discuss the importance of classroom summative and formative assessment as an essential part of the assessment system (Black & William, 1998; Brookhart, 2014; Pellegrino et al., 2001). A vision for a science assessment system (Figure 1.1) illustrates the ways in which the State of Michigan's Balanced Science Assessment System must include multiple assessment occasions through which data ought to be collected and used to make data-driven decisions.
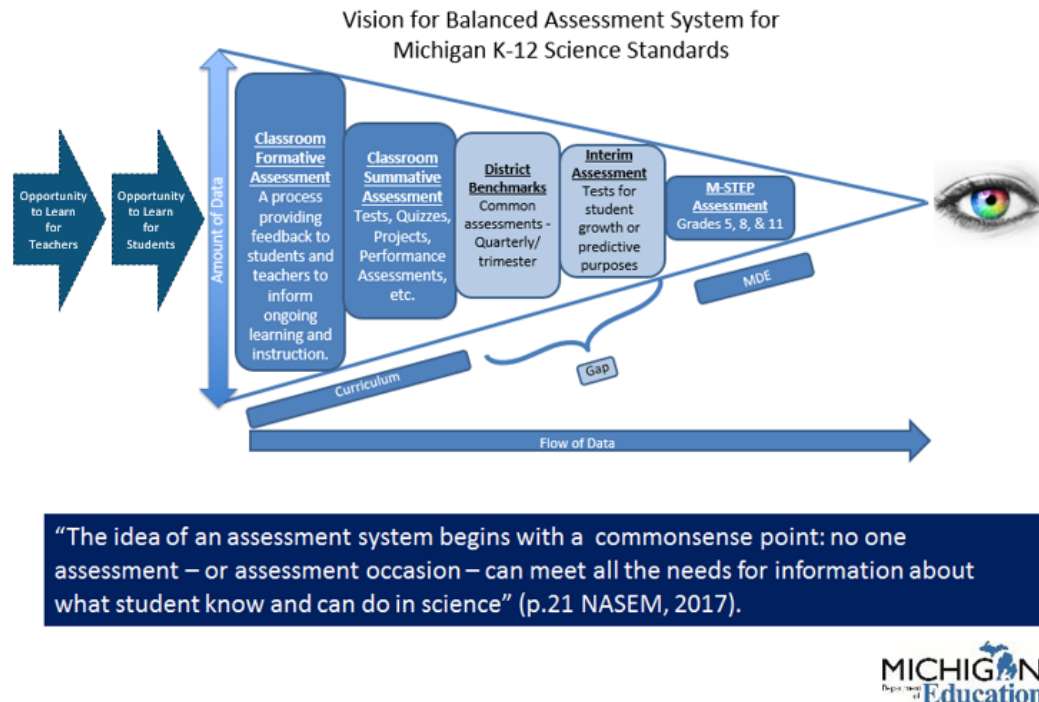
Figure 1.1. Vision for Balanced Assessment System for Michigan K-12 Science Standards.

It is important to situate this research with respect to other forms and purposes of assessment. The large-scale assessment that is the focus of this research resides at the end of the assessment system as depicted in Figure 1.1. When examining the entire assessment system for the MSS, we must realize that while the state assessment is the focus of this particular research, other parts of the assessment system provide much more data on which decisions about students, teachers, and schools should be based.

Nevertheless, the focus of this research is embedded in Michigan's large-scale assessment for science. Therefore, we must examine the forms evidence that ought to be gathered to make inferences about the validity of the data gathered for the purpose determined by the State of Michigan. For example, the Michigan state legislation regarding teacher evaluation mandates that beginning in 2018–2019, 50 percent of teacher evaluations are based on state

assessment data (House Bill 4493, Sec. 1249 2a, 2016, i–ii). The evaluation legislation applies to all tested content areas and grades. Therefore, teachers are facing numerous reforms that require them to shift their practices because of district policies to meet the legislative requirements. As a result, some changes in instruction are delayed until the state assessment looks markedly different regardless of statements that encourage earlier shifts in curriculum and instruction. Clear communication and transparency regarding the processes and decisions about the design and implementation of the new state summative science assessment are essential to creating the impetus for instructional and curricular change necessary for promoting student academic success in science.

The Board on Testing and Assessment (BOTA) report (NRC, 2014) states that the purpose of monitoring assessments can include: determining how much students in a certain school system have learned over the course of a year, comparing student performance in one school system to another, identifying successful instructional techniques, or ascertaining effects of a particular educational policy. The Michigan Department of Education requires that the state assessments provide (a) an important snapshot of student achievement at a state, district, and building level, (b) valuable information to parents on their child's academic achievement, and (c) important data for teachers, schools, and districts to help guide instruction (Michigan Department of Education, 2017). Thus, large-scale assessments can be used in a variety of ways. ESSA legislation requires state assessments to use multiple measures of student academic achievement (B, iii, p. 25) to determine individual student growth over the course of time.

Validity of New Science Assessments

Large-scale assessments used for monitoring and accountability purposes are subject to rigorous validity studies to ensure that inferences that can be made from the data match the

intended purpose of the test. The Standards for Psychometric Testing (NCME/APA/AERA, 2014) recommend that validity studies address five different forms of validity evidence: content validity, response process validity, internal structure, relations to other variables, and consequences of testing. However, large-scale assessment has mainly focused on evidence from item response theory (IRT) or other psychometric analyses (e.g., information about item difficulty and discrimination scores) rather than on all forms of validity evidence. With these new standards (i.e., NGSS and MSS), it is particularly important to construct a validity argument that prioritizes evidence based on test content and response processes in order to determine what students know and can do in science.

Michigan has a long history of designing and implementing valid and reliable large-scale assessments that meet and exceed the federal and state requirements as determined by peer review. In large-scale assessments, validity arguments are essential to convince stakeholders that the assessment provides the information for which it was designed. State assessments are subject to peer-review processes, in which the U.S. Department of Education conducts an analysis of the assessments endorsed by the state to determine if there is alignment between the standards and the assessment. This dissertation provides a deep look into the ways in which these assessments are designed and iteratively tested.

My research addresses the need to determine new and valid ways of assessing at a large-scale these more complex science standards in a manner that meets federal and state assessment legislation in addition to providing all students an opportunity to demonstrate what they know and can do in science. Additionally, my research is framed in an understanding that a coherent assessment system must exist to both use assessment data for the intended purposes and to meet the requirements of federal and state legislation.

Outline of this Dissertation

      Following this chapter, I have a literature review chapter that synthesizes select literature about the *Framework for K-12 Science Education* (NRC, 2012), on which the MSS were based, information about the design of science assessments, and validity of science assessments. In Chapter 3, I provide an overview of the Michigan Science Assessment Design Process. I then present a findings chapter. I finish with a discussion and conclusion chapter where I summarize my findings and discuss the implications of my research.

CHAPTER 2: REVIEW OF THE LITERATURE

What are the Next Generation Science Standards?

The past decade has brought about significant science education reforms. These reforms center on moving classroom science from learning about various science constructs and theories to providing opportunities for students to figure out natural and observable phenomena (NRC, 2006). Students are more engaged when science learning focuses on phenomena and investigating the way in which the world works (NRC, 2012). The *Framework for K-12 Science Education* (*Framework*; NRC, 2012), which lead to the development of the NGSS (NGSS Lead States, 2013), stresses the fact that children are born investigators and are curious about the world around them.

> First, [the *Framework*] is built on the notion of learning as a developmental progression. It is designed to help children continually build on and revise their knowledge and abilities, starting from their curiosity about what they see around them and their initial conceptions about how the world works. The goal is to guide their knowledge toward a more scientifically based and coherent view of the sciences and engineering, as well as of the ways in which they are pursued and their results can be used. (NRC, 2012, pp. 10–11)

To investigate phenomena and solve problems, the *Framework* calls for scientific learning and teaching that integrates three dimensions of scientific knowledge and practice: DCIs, CCCs, and SEPs (see Table 1.1). Additionally, the *Framework* spotlights the importance of Engineering, Technology and the Applications of Science. The DCIs are a "limited set of core science ideas . . . [that] allow for deep exploration" (NRC, 2012, p. 25) in increasingly sophisticated ways across

students' K-12 experience. The criteria for core ideas include topics that are important across disciplines, a resource for learning about more sophisticated ideas, relatable to students and society, and able to grow in sophistication across grades (NRC, 2012). The DCIs are grouped by domains: Physical Science, Life Science, Earth and Space Science, and Engineering.

The SEPs are key practices that scientists and engineers use to develop and test theories about the natural and designed world. Engagement with the SEPs supports students in better understanding the way in which scientific knowledge is developed as well as to promote a deeper understanding of the DCIs (NRC, 2012). The eight SEPs are listed in Table 1.1. Students often use multiple practices together or in succession to make sense of scientific phenomena (Schwarz et al., 2017).

The CCCs are concepts that have broad application across the domains of science (NRC, 2012). The within instruction, the CCCs are often thought of in the metaphorical sense as lenses, bridges, tools, or rules for science (Rivet et al., 2016). The seven CCCs are listed in Table 1.1. The three dimensions (SEPs, DCI, and CCCs) are to be seamlessly embedded within instruction and assessments to provide students an authentic inquiry experience.

Table 1.1

Next Generation Science Standards (NGSS) Three Dimensions

Disciplinary Core Ideas	Science and Engineering Practices	Crosscutting Concepts
PHYSICAL SCIENCES	Asking questions and defining problems	Patterns
Matter and its interactions	Developing and using models	Cause and effect
Motion and stability: Forces and interactions	Planning and carrying out investigations	Scale, proportion, and quantity
Energy	Analyzing and interpreting data	Systems and system models
Waves and their applications in technologies for information transfer	Using mathematics and computational thinking	Energy and matter: Flows, cycles, and conservation
LIFE SCIENCES	Constructing explanations and design solutions	Structure and function
From molecules to organisms: Structures and processes	Engaging in argument from evidence	Stability and change
Ecosystems: Interactions, energy, and dynamics	Obtaining, evaluating, and communicating information	
Heredity: Inheritance and variation of traits		
Biological evolution: Unity and diversity		
EARTH AND SPACE SCIENCES		
Earth's place in the universe		
Earth's systems		
Earth and human activity		

A Need for New Science Assessments

In the past, many large-scale science tests included multiple-choice items that assessed independent pieces of content (Alonzo & Ke, 2016; Blank & Adams, 2018; Pellegrino, 2014). However, to assess the NGSS, assessment tasks must elicit evidence of knowledge-in-use, meaning that students apply science content knowledge while utilizing appropriate SEPs (Harris et al., 2019) and Program for International Student Assessment (PISA; OECD, 2016). For example, students may use modeling (SEP), knowledge of inheritance in organisms (DCI), and apply systems understanding (CCC) to explain scientific phenomena such as why some flowers of the same species may look different (Pellegrino et al., 2013). This type of assessment will look different from most prior assessments (Alonzo & Ke, 2016). Thus, designing assessments to provide information about what students know and can do in science with evidence regarding three-dimensional thinking requires careful design of assessment systems, clearly articulated assessment goals, and innovative assessment design (Gorin & Mislevy, 2013).

Assessment Design for NGSS

Assessment is a form of "reasoning from evidence" in which observations of students' actions and artifacts are used to support inferences about what they know and can do (Pellegrino, Chudowsky, & Glaser, 2001). Students' science knowledge and understanding is a construct, which cannot be directly observed. With respect to assessments, a construct is used to describe a body of content (knowledge, skills, understanding, etc.) that an assessment measures. To develop assessments for the NGSS, the BOTA report (NRC, 2014) recommends using a principled design approach such as Evidence Centered Design (ECD; Mislevy & Haertel, 2006) to provide a framework for developing evidence of construct validity. This design approach has proven to be useful in providing a system for developing assessment claims associated with the NGSS, which

then can be used to design three-dimensional tasks (Debarger et al., 2016). These assessment tasks must elicit knowledge-in-use to bring together the three dimensions to explain specific phenomena or solve problems. Previous iterations of science standards have not required this complex assessment design prior to the NGSS (Pellegrino et al., 2013). The BOTA report (NRC, 2014) recommends, "To adequately cover the three dimensions, assessment tasks will need to contain multiple components, such as a set of interrelated questions" (Conclusion 2-1, p. 63). The SAIC (2015) illustrated this recommendation with sample item cluster prototypes built to assess bundles of performance expectations using a phenomenon-based scenario and multiple two- and three-dimensional items.

Knowing What Students Know (Pellegrino et al., 2001) primed the field of science assessment by using cognitive science, encouraging assessment developers to consider both cognitive learning theory and equity when designing assessments. Building on this work, theories of learning such as sociocognitive and sociocultural learning theory have been used to inform the ways in which assessments are designed in conjunction with curriculum and assessments (Kang & Furtak, 2021; Shepard et al., 2018). While these theories can inform assessment design at the local level, Shepard and colleagues (2018) argue that alignment across districts and within a state can be challenging because it is impossible for any curriculum to cover all the possible intersections of the three dimensions (p. 32). Because the scope of the NGSS is both broad and deep, a principled design approach is necessary to ensure the assessment is designed to gather the evidence to support the claims supported by the assessment (Harris et al., 2019). Assessments also need to foreground sensemaking to provide opportunities for students to show what they know and can do with respect to the three dimensions (Achieve, 2018).

Work from other large-scale assessment projects such as the National Assessment of Educational Progress (NAEP) has influenced the possibilities for what task formats can and should look like. For example, the 2014 NAEP Technology and Engineering Literacy Assessment used engaging, interactive tasks providing students an opportunity to demonstrate their mastery of engineering practices related to problem-solving. While several examples of three-dimensional assessment development for formative and classroom use have been deemed successful (e.g., Anderson et al., 2018; Furtak, 2017), the design and implementation of state-level large-scale science assessment have yet to be the focus of many studies. Nevertheless, the NGSS challenge large-scale assessment design in that the integration of SEPs, DCIs, and CCCs must be assessed. Additionally, the assessments that are created must provide information that can support a validity argument for the stated purposes of the assessment (NRC, 2014).

Importantly, assessment design requires thoughtful consideration of the test format that will give all students the opportunity to demonstrate their ability to integrate the practices, crosscutting concepts, and disciplinary core ideas in the context of investigating phenomena and designing solutions to problems. Additionally, assessment developers should consider multiple student populations with respect to culture, language, ethnicity, gender, and disability to design task formats that are as accessible and fair to as many students as possible (NRC, 2014). Still, considerations for the engagement of diverse populations of students and the variance in their opportunities to learn science present further challenges (Penuel et al., 2019). The complexity of the NGSS requires that we draw on the testing technology used for various testing programs while considering the diverse populations we serve. This challenging task will require that evidence is gathered and synthesized to convince stakeholders that this new assessment design produces a valid and reliable assessment.

Alignment in Large-Scale Science Assessments

The structure of the NGSS makes it difficult to define alignment on large-scale assessments. In the past, large-scale assessments were designed so that one item would assess one standard (NRC, 2014). This one-to-one design provided a simple alignment argument. If a student answered the item correctly, the claim could be made that the student understood the standard to which the item was aligned. NGSS requires the three dimensions be assessed together. Creating an alignment argument becomes more difficult because the assessment no longer has a one-to-one design but requires several items across a cluster to provide alignment to the bundle of standards (Alonzo & Ke, 2016). Therefore, defining alignment for the purpose of large-scale science assessment was a necessary component of this study.

In this study, I define item alignment as the item's ability to elicit evidence that students used the intended dimensions and to discriminate between students who chose the correct response versus those who chose the incorrect response. The intentional design of high-quality state science assessments requires assessments that "assess state science standards in order to provide evidence to support, refute, or qualify state specific claims about students' achievement in science" (Achieve, 2018).

Alignment between a set of content standards and large-scale assessment is integral to the content validity argument that is necessary for an overall validity argument (AERA/APA/NCME, 1999; Ananda, 2003; Impara, 2001; Resnick et al., 2003; Webb, 1997b; Zucker, 2008). Making a claim that the items are representative of the defined construct serves as evidence of students' understanding of the construct (Pellegrino et al., 2001). In other words, the assessment items serve as a structured argument for what students know and can do in science.

The assessment items are designed to elicit evidence from students and that evidence is used to support a claim about the students' knowledge and ability in science.

Validity for NGSS-Aligned Assessments

In 1974, APA standards presented the notion that "validation is a comprehensive effort requiring multiple sources of evidence that support the use of a test for a specific purpose" (Sireci, 2009). Following, the APA (1985) purported, "No test is valid for all purposes or in all situations or for all groups of individuals" (p. 31). To date, Kane's (1992) proposition of an argument-based approach to validation remains in which the validator builds an argument that focuses on defending the use of a test for a particular purpose and is based on empirical evidence to support the particular use. In modern psychometric theory, construct validity—the degree to which a test measures what it claims to measure—serves as an overarching frame for evaluating the strength of assessment arguments (Messick, 1995).

The *Current Standards for Educational and Psychological Testing* (NCME/APA/AERA, 2014) state, "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests…. It is the interpretations of test scores for proposed uses that are evaluated, not the test itself…. It is incorrect to use the unqualified phrase 'the validity of the test'" (p. 11). For any assessment, there are five sources of validity evidence: (a) validity evidence based on test content, (b) validity evidence based on response processes, (c) validity evidence based on internal structure, (d) validity evidence based on relationships to other variables, and (e) validity evidence based on consequences of testing (NCME/APA/AERA, 2014). Additionally, multiple forms of validity are necessary to create a validity argument for an assessment, however, in practice the content and response process validity evidence often takes a second seat to psychometric forms of validity evidence if the

items do not fall within the pre-determined statistical guideposts (NCME/APA/AERA, 2014).

For example, the psychometrics team may determine the P-value of items included on an

assessment has to fall between 0.8 and 0.3 to include an item on an assessment blueprint. Rarely

can the content team argue to include the item due to its content validity evidence alone (Council

of Chief State School Officers (CCSSO) Science Collaborative, personal communication, 2018).

Historically, large-scale assessments have relied heavily on psychometric sources and less on

evidence based on test content and responses processes (Anderson, personal communication,

2019). Because of the complex nature of the NGSS, it is imperative that both content and

response process validity evidence are brought to the fore and carefully examined. The NGSS

are three-dimensional, more complex, and require higher levels of cognition to meet the

performance expectations. Therefore, the traditional data used to validate large-scale assessments

will not provide the evidence necessary to create a validity argument about what students know

and can do in science.

Evidence Based on Test Content

"For educational achievement tests...validity evidence based on test content validity will

represent the foundation of any validity argument" (Sireci & Faulkner-Bond, 2014, p. 106). To

gather validity evidence based on test content, the relationship between the knowledge, skills,

and abilities being measured and the content of a test must be analyzed. In this case, the

construct measured by the State of Michigan science assessment is students' knowledge and

abilities related to the Michigan K-12 Science Standards. I am defining content validity as the

degree to which the content of a test is congruent with testing purposes. Within validity evidence

based on test content, there are four types.

First, the domain definition provides a bridge between the theoretical construct and the concrete content of a domain (Sireci, 1998). In Michigan, a modified Evidence Centered Design (Mislevy & Haertel, 2003) approach was used to define the domain and gain external consensus from a group of independent experts in the field to help develop and evaluate the test specifications (Sireci & Faulkner-Bond, 2014). Second, subject matter experts are employed to determine the extent to which the assessment fully and sufficiently represents the targeted domain. Third, subject matter experts are asked to rate the degree of alignment or the extent to which test items are relevant to aspects of the test specifications. Fourth, the appropriateness of the test development process is considered (Sireci & Faulkner-Bond, 2014).

Evidence Based on Response Process

For any assessment argument of student learning, assessors need data on student actions, such as their response to tasks, to judge the strength of the claim (Pellegrino et al., 2001). This calls for theoretical and empirical analyses of the response processes of the test taker. These analyses can provide evidence concerning the fit between the construct and the detailed nature of the performance or response engaged in by test takers and can be extended to include judges or observers of the test.

Response process data comes from analyses of individual responses. Asking a diverse group of test-takers about their performance strategies or responses to items can provide data to enrich the definition of the construct. Response process information can influence the interpretation of test scores for subgroups. Because assessments often rely on observers or administrators, evidence about the extent to which the processes of observers or judges are consistent with the intended interpretation of scores is important (NCME/APA/AERA, 2014).

Justification for this Research

Given the need for valid large-scale NGSS-aligned assessments, the research presented here focuses on the following research questions:

1) To what extent do the clusters developed using Michigan Cluster Development Process align with the Michigan K-12 Science Standards?

    1a) To what extent do these items elicit and discriminate for the intended dimensions?

What follows in this dissertation includes an overview of the process used in Michigan to develop the science clusters used for this research, the methods used to collect and analyze data, the findings of said data analysis, and a discussion of these findings, including implications for the science assessment field.

CHAPTER 3: MICHIGAN CLUSTER DEVELOPMENT PROCESS

In this section, I describe the decisions made leading to the structure of the new Michigan science assessment. I address the process for forming cluster writing teams, and the decisions leading to the development of clusters for the 2017 Michigan Science Pilot Test. Two of these clusters are the subject of this research project.

Topic Bundles

To determine the blueprint and specifications of the new science assessment for Michigan, the Michigan Science Assessment Advisory Committee (Advisory Committee) was formed. This group, consisting of science education researchers, state assessment specialists, and assessment contractors, was gathered to determine the goals and structure of the new science assessment for the State of Michigan. Researchers brought forward empirical studies regarding assessment design and implementation and state assessment specialists discussed considerations with respect to testing time, budget, and political considerations. The Advisory Committee decided to utilize the "cluster" structure as recommended by the SAIC Assessment Framework (CCSSO, 2015). The idea of clusters builds on the recommendations of the BOTA Report (NRC, 2013) to group items together to coherently assess the complex performance expectations in a manner that forefronts phenomena. Each cluster includes set of five to eight items based on a common stimulus that is written to assess all dimensions of the selected standards.

Assessment contractors discussed the implications of using cluster as the base unit for assessments. The initial focus of the meeting was to determine how the NGSS performance expectations, which Michigan adopted as the Michigan K-12 Science Standards (MSS) would be assessed (example Figure 2.1). It was quickly determined that creating a group of items to assess

each performance expectation would result in too many clusters for a once-a-year state summative assessment. The Advisory Committee agreed that the performance expectations should be "bundled" to facilitate assessment via a single natural phenomenon or engineering problem that is presented within a stimulus (SAIC, 2015). The SAIC (2015) suggested that one approach is bundling the performance expectations to intentionally utilize one of the three dimensions that crossed different performance expectations. For example, performance expectations could be bundled by common Science and Engineering Practices (SEPs). Therefore, two performance expectations with Developing and Using Models as the SEP would be bundled leaving differences in Disciplinary Core Ideas (DCIs) and Cross Cutting Concepts (CCCs). The rationale for this approach was to leverage common dimensions to lessen the task requirements for the students. Similarly, suggestions to bundle performance expectations by DCIs or CCCs were considered.

MS-PS3-1 Energy

Students who demonstrate understanding can:

MS-PS3-1. **Construct and interpret graphical displays of data to describe the relationships of kinetic energy to the mass of an object and to the speed of an object.** [Clarification Statement: Emphasis is on descriptive relationships between kinetic energy and mass separately from kinetic energy and speed. Examples could include riding a bicycle at different speeds, rolling different sizes of rocks downhill, and getting hit by a wiffle ball versus a tennis ball.]

The performance expectation above was developed using the following elements from the NRC document *A Framework for K-12 Science Education*:

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Analyzing and Interpreting Data Analyzing data in 6–8 builds on K–5 and progresses to extending quantitative analysis to investigations, distinguishing between correlation and causation, and basic statistical techniques of data and error analysis. • Construct and interpret graphical displays of data to identify linear and nonlinear relationships.	**PS3.A: Definitions of Energy** • Motion energy is properly called kinetic energy; it is proportional to the mass of the moving object and grows with the square of its speed.	**Scale, Proportion, and Quantity** • Proportional relationships (e.g. speed as the ratio of distance traveled to time taken) among different types of quantities provide information about the magnitude of properties and processes.

Figure 2.1. Example of Performance Expectation from nextgenscience.org.

Ultimately, the Advisory Committee decided to utilize the structure of topic bundles as presented in the MSS for the state assessment. The MSS is structured using the "topic" format of the NGSS (2013). Each topic bundle consists of multiple performance expectations, which are grouped together based on a particular science topic. For example, in middle school, one topic bundle, *Energy*, is categorized in the Physical Science domain and includes five PEs (Figure 2.2).

Energy	
MS-PS3-1	Construct and interpret graphical displays of data to describe the relationships of kinetic energy to the mass of an object and to the speed of an object.
MS-PS3-2	Develop a model to describe that when the arrangement of objects interacting at a distance changes, different amounts of potential energy are stored in the system.
MS-PS3-3	Apply scientific principles to design, construct, and test a device that either minimizes or maximizes thermal energy transfer. *
MS-PS3-4	Plan an investigation to determine the relationships among the energy transferred, the type of matter, the mass, and the change in the average kinetic energy of the particles as measured by the temperature of the sample.
MS-PS3-5	Construct, use, and present arguments to support the claim that when the kinetic energy of an object changes, energy is transferred to or from the object.

Figure 2.2. Middle School Energy Topic Bundle.

Utilizing the topic bundle structure posed promises and challenges. By bundling multiple performance expectations, the shared dimensions could be leveraged to reduce the number of tasks required to assess the whole of the topic bundle. Additionally, utilizing the existing structure of the adopted document for the MSS did not require clarification to stakeholders regarding bundling of performance expectations for the state assessment. The most daunting challenge, however, was determining a process to ensure all the dimensions included in a topic bundle were assessed within a single cluster. Table 2.1 outlines all the dimensions that are part of the middle school *Energy* topic bundle to illustrate this point. Within this topic bundle there are no common SEPs. However, all the performance expectations include the DCI PS3.A, three performance expectations contain PS3.B, and one instance each of PS3.C and ETS1A-B. Moreover, two of the CCCs are represented in two different performance expectations. Therefore, when this cluster is reduced to only one of each unique dimension, there are twelve unique components instead of fifteen. Leveraging this reduction of unique assessable dimensions or elements of dimensions in a large-scale assessment context provides the opportunity to assess more efficiently on various components of the performance expectations. In Michigan, we

consider *The Framework*, the NGSS performance expectations, the assessment boundaries and clarification statements found in the NGSS, and the learning progression information found in Appendices E, F, and G of the NGSS.

Figure 2.3 illustrates how the dimensions of each performance expectation are assessed in the items of a cluster. Each of the SEPs, DCIs, and CCCs are integrated with one another regardless of the structure of the original performance expectation. For example, item three integrates PS3.A, SEP 6, and CCC 3 in a three-dimensional assessment item. However, none of the performance expectations are written with those three dimensions together. The flexibility in this assessment design allows for the MSS to stay true to The Framework (NRC, 2012) which intends for all the practices and crosscutting concepts to be applied to any core idea depending on the phenomena or problem in question.

Table 2.1

Example Topic Bundle: Middle School Energy

Standard	Science and Engineering Practice	Disciplinary Core Idea	Cross Cutting Concept
MS-PS3-1	4. Analyzing and interpreting data	PS3.A	3. Scale, proportion, and quantity
MS-PS3-2	2. Developing and using models	PS3.A; PS3.C; ETS1.A-B	4. System and System Models
MS-PS3-3	6. Constructing explanations and designing solutions	PS3.A; PS3.B	5. Energy and Matter
MS-PS3-4	3. Planning and carrying out investigations	PS3.A; PS3.B	3. Scale, proportion, and quantity
MS-PS3-5	7. Engaging in argument from evidence	PS3.A; PS3.B	5. Energy and Matter
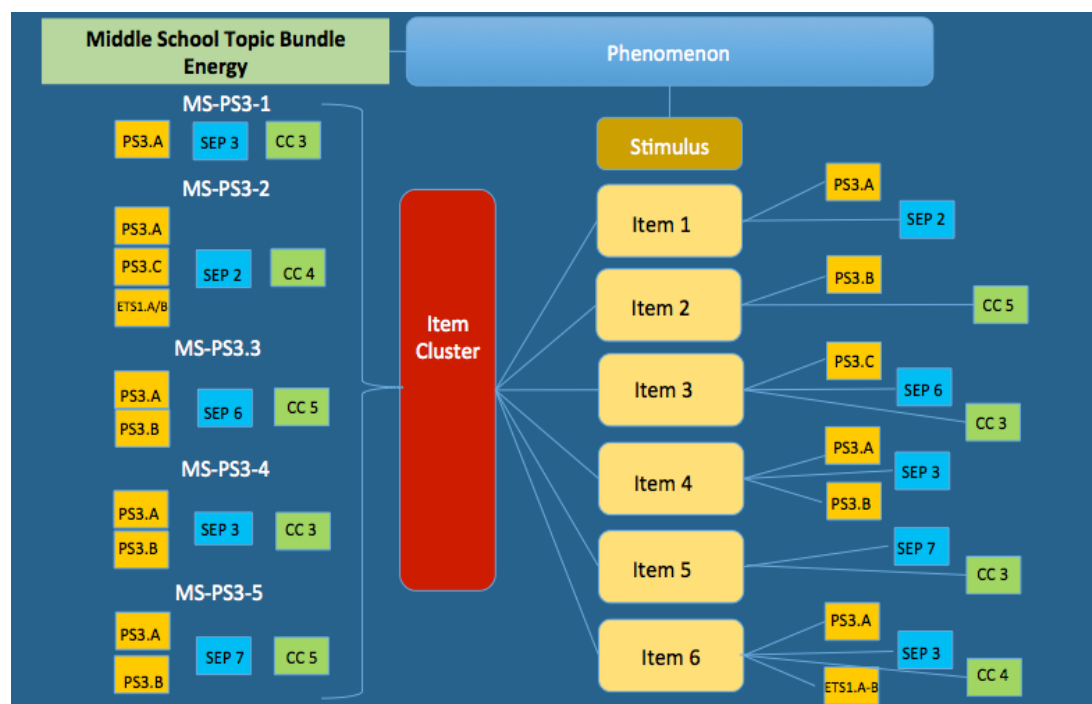
Figure 2.3. Example Topic Bundle to Cluster Map.

Cluster Writer Recruitment

After deciding on the structure of the assessment, we needed to determine a way to

recruit qualified writers to develop the clusters. First, the concept of research-practice

partnerships (Coburn et al., 2013) influenced the decision for both practitioners and researchers

to be involved in the Michigan cluster writing process. For the clusters to represent the true nature of the *Framework* (NRC, 2012), teachers are needed to understand the implications of assessment tasks on students in varying grade levels and contexts. Therefore, practitioner expertise is a crucial element of the process. Additionally, the foundational research and theoretical knowledge possessed by educational researchers is essential to bring forward the intricacies and intent of the Framework (NRC, 2012). By pairing these professionals throughout the process, a wealth of knowledge can be shared and utilized to develop meaningful clusters. Teachers (active classroom science teachers or science curriculum consultants) and researchers (science education graduate students or professors of science education) interested in becoming cluster writers were screened using an application to determine their teaching experience and exposure to NGSS professional development opportunities. The purpose of the screening was to ensure that all participants had prior knowledge of the NGSS and were familiar with the structure of the standards. After screening, teachers and researchers were surveyed to determine their area of expertise (i.e., Earth Science, Physical Science, Life Science, and Engineering) and grade level to determine which topic bundle would best fit their knowledge and expertise.

Following, one teacher and one researcher were paired to create a cluster writing team (writers). Writers were grouped by grade level (Grades 5 and 8) and each grade level group learned how to design one cluster over the course of one week. Table 2.2 summarizes the cluster writer participants and the number of topic bundles addressed during the five 2016 Cluster Workshops. It was intended for teachers and researchers to be paired in every instance, however, over time, the pool of researchers available was less than the number of teachers available. In some cases, teachers who had participated in the cluster workshops during a previous week returned and were paired with a new teacher to act as the mentor. Many of the researchers

participated in multiple weeks of the workshop thereby adding to the percent returning noted in Table 2.2.

Table 2.2

2016 Cluster Writing Participants

Grade Level	Teachers	Researchers	% Returning	Topic Bundles Addressed
Grade 5	10	8	17%	9
Grade 8	12	8	25%	11
Grade 11	21	9	47%	15

Structure of Cluster Writing

Much of the initial thinking around training teachers and researchers to write clusters for the state stemmed from collaboration with the State of Washington. After contacting many states that had adopted the NGSS, Washington was one of the few states where teachers were involved in all aspects of the item development. Educator involvement is an important part of how Michigan designs their assessments as well. After observing Washington's cluster writing process, many of the tools and resources were adopted to suit the needs in Michigan. The Washington State Science Assessment Consultants paved the way for Michigan to engage in developing new science assessments by sharing their work.

Each cluster writing workshop occurred over the course of one week consisting of five days of intense cluster design work (Appendix A). The overall goals were to train teachers and researchers to unpack the NGSS topic bundles to determine what evidence students should be able to provide that would substantiate a claim about what they know and can do in science. The teacher-researcher teams completed one cluster that contained tasks designed to elicit evidence of students three-dimensional thinking centered on investigating phenomenon and designing solutions to problems.

Resources Available to Cluster Writing Teams

In order to develop the clusters, we provided writers with several different resources: an item pool, unpacking process and documents, the *Framework* (NRC, 2012), the NGSS and supporting documentation, and ample feedback loops. Next, I describe each of the resources and processes.

Item Pool

Working together with the Advisory Committee, we offered writers access to item pools from various research projects that developed two- or three-dimensional items in various science domains. The item pool was collected and organized by a team of graduate and undergraduate students and then made available to the cluster writer teams for reference and use in the development process. The item pool was a crucial tool in determining what two- and three-dimensional items look like along with providing examples of phenomena.

Unpacking Process

Understanding all that is contained in the NGSS poses challenges for teachers and researchers. One of the resources utilized to help writing teams think through the information contained in the topic bundles was the unpacking process developed the Next Generation Science Assessment Project (NGSA, 2016; Appendix B). Built on the principles of ECD (Mislevy & Haertel, 2017), the prompts encourage science educators to unpack the content to understand all the complexities involved with a particular topic bundle. Writing teams described evidence one might need from a student to support a claim that the student could demonstrate their understanding of a topic bundle or parts of it. This way, when writing teams developed the tasks, they had the target evidence in mind. This process helps to support a claim about what the ability of the item to elicit specific aspects of what a student knows and can do with respect to a

topic bundle. The writing teams worked through documents that pushed them to think about the underlying elements in each of the three-dimensions, what previous knowledge students may possess, specific vocabulary that is necessary for communication of understanding, the progression of learning that would occur overtime, and more. Such documents included the Framework for K-12 Science Education (NRC, 2011), the Next Generation Science Standards (States, 2013) including Appendices E, F, and G, the American Association for the Advancement of Science Project 2061 Science Assessment Topics (AAAS, 2017), and STEM Teaching Tools (Research and Practice Collaboratory, 2016).

While the unpacking process can seem tedious and overwhelming, especially when unpacking an entire topic bundle, the process offers teachers and researchers the time and tools needed to really dig into the NGSS and understand their complexity. The unpacking process is an essential step in the item writing process because ideas about instruction and assessment stem from understanding what the NGSS entail.

Role of Phenomenon and Stimulus

Defining a phenomenon is a difficult endeavor. Several scholars and researchers have attempted to define a phenomenon for instruction for the three-dimensional science standards. Creating a working definition of phenomenon was an essential part of cluster design. In Michigan, we began with the generalized definition: "a natural event that is observable and repeatable" (Krajcik, personal communication, 2016). However, even this simplistic definition was difficult for teachers and researchers to grasp when it came to microscopic phenomena that are not always directly "observable" or behaviors that are not always "repeatable" due to uncontrollable variables. Therefore, the working definition of phenomenon that grew out of the cluster development process is "something someone can observe and wonder how or why it

happens" (Policy capturing process, Summer 2016). After using the unpacking process, writing teams would brainstorm four to six phenomenon that would apply to their topic bundle. These phenomena were shared among the larger group where discussion would weed out the phenomena that may not be as strong as others. Content and assessment specialists provided feedback on phenomena that may pose equity, bias, or sensitivity issues on a large-scale assessment. Eventually, after collaborative discussion, the writing teams decided on one phenomenon that would best suit their topic bundle.

After deciding the phenomena, writing teams determined the manner in which the phenomenon could be presented to students on a large-scale assessment so that all students would have access to the phenomenon. One way that the writing teams discussed doing this was to identify the phenomenon and then begin with "Once upon a time, students were…." This "story lining" process helped writing teams develop several contexts or stimuli that were relatable and interesting for students. Again, through collaborative feedback and careful consideration, the writing teams determined which stimulus held the most promise for the development of a full cluster.

Item / Task Types

An essential part of large-scale assessment writing is understanding the item types available for the assessment as well as common equity issues developers observe. The state contractor, Data Recognition Corporation (DRC), is currently responsible for training Michigan item writers on "bias and sensitivity" and the item types available within their system. The DRC science consultant led the description of these item types while describing the affordances and constraints of each item type. This was especially important when demonstrating technology-enhanced (TE) item types, which are different from the typical, multiple choice or constructed

response item types that many assessments utilize. Technology-enhanced item types range from multi-select items, where students must choose more than one correct answer to a question, to drag and drop items, where students move graphics or text into a predetermined space to respond to a question.

The equity portion of the training focused on bias and sensitivity issues and Universal Design features that must be considered on large-scale assessments. Writing teams were taught to look for issues around race, gender, regionalism, religion, socioeconomic status, physical disability, and others. Additionally, writing teams were made aware of ways they can intentionally design tasks to be inclusive of all students using common, relatable phenomenon and by including engineering tasks. Writing teams were actively challenged from the beginning of the week to think about how to give students similar experiences through the crafting of the stimulus that does not advantage one child over another.

Draft Stimulus Share

Throughout the week, several opportunities were given for writing teams to discuss the development of the stimuli prepared for the clusters with peers and consultants. The feedback offered within and among teams provided writing teams the opportunity to reflect on the engagement, vocabulary, equity, and necessity of the information within the stimuli.

Item Templates and Alignment Tools

Once the writing teams were ready to write the individual tasks within the cluster, item templates were used and adapted to help guide their process. The item templates included the specifics of task ordering, dimension alignment, and the assessment claim provided by the task design that would supports claims about students understanding of the stated dimensions. The assessment claims written by the writing teams are essential to verifying the alignment of each

cluster. Additionally, writing teams developed their own tools to track alignment. During the second week of writing, one group designed a matrix used to map the items within the cluster and the elements of the dimensions to which they aligned (Appendix C). This Cluster Mapping Tool was utilized by every writing team moving forward as a verification of assessment and alignment for all dimensions of the topic bundle.

Research and Practice Collaboratory Tools

The Research and Practice Collaboratory (Penuel, Bell, et al., 2016) offered many resources and insights for the Michigan item development process. The STEM teaching tools developed by the Research and Practice Collaboratory were also utilized in the process. STEM Teaching Tool #41 – Prompts for integrating Crosscutting Concepts into Assessment and Instruction (Penuel & Van Horne, 2016) and STEM Teaching Tool # 30 - Integrating Science Practices into Assessment Tasks (Van Horne et al., 2016) were valuable resources as the writing teams worked to craft tasks for the clusters. These resources were used both as a reference and as a way to ensure that the SEPs and CCCs were being explicitly assessed.

Peer and Content Review

Built into the initial week of cluster development is dedicated time and protocol for both small and large group content review. After working on their clusters for two and a half days, writing teams would be paired with another team with similar content focus (i.e., the teams working on life science topic bundles would be paired). The writing teams shared their cluster and receive feedback from their partner group. This early feedback session gave writing teams the opportunity to "try out" their stimulus and tasks with another group to determine if there were flaws in either the content or storyline of the cluster. Each group was given one hour to present their work and receive feedback. Specific protocol was followed which allowed for

groups to focus on grade appropriate content, engaging phenomenon, and task alignment to the NGSS. After the peer feedback was received, the writing teams revised their clusters. The following day, the revised clusters were then presented by a facilitator (science education assessment consultant) and reviewed by a larger group. In this content review, a similar protocol was followed as with the peer review by all participants. After feedback was gathered by all writing teams, the teams revised their clusters to prepare for submission.

Policy Capturing Process

A policy capturing process (a method used by researchers to assess how decision makers use information when making evaluative judgements – Zedeck & Kafry, 1977) was utilized to collect data and determine some of the item specifications and requirements for the grade levels and clusters (Aiman-Smith et al., 2002). Throughout the process of cluster design, the writing teams encountered questions that puzzled or stumped them. These questions were added to the "Questions / Decisions" board for later discussion. At the end of each day, the whole group discussed the "Questions / Decisions" made for the day. On the last day of the week, any remaining "Questions / Decisions" were discussed and recorded. For example, the grade 11 writing teams learned that there were several different opinions in the group regarding the use of calculators and formulas. While some writing teams thought that fundamental mathematical equations were fair for the science assessment, applicable under the SEP of Computational and Mathematical Thinking, others expressed their concern for the focus on memorization of a formula at the expense of conceptual understanding. These discussions helped the Michigan Department of Education (MDE) determine some of the item specifications for the new science assessments.

Cluster Refinement

Internal Revisions and Graphics

After the clusters were submitted by the writing teams, the content specialists (state and vendor), worked through each cluster line by line to prepare the clusters for the test engine. Graphic artists from MDE utilized the graphic descriptions provided by the writing teams to create original graphics for the clusters. Once the cluster was prepared in the test engine, a live version of the cluster could be reviewed and interacted with online.

Committee Review Process

Fully developed clusters are required to be reviewed by educator committees with diverse membership that may include state education agency staff, state educators, trained assessment specialists (e.g., district administrators or test coordinators), content specialists, and curriculum developers. Review panels should consider cluster length, readability, format/style, typography, content, vocabulary, sentence complexity, concept load or density, and cohesiveness (SAIC, 2016).

To heed the advice of the SAIC, equity (bias and sensitivity) experts were chosen based on their areas of expertise in visual impairments, English language learners, hearing impairments, urban school settings, and other special education learning situations. For each of the grade levels 5, 8 and 11, five Equity Review Committee members reviewed the three interactive clusters. One goal of equity review is to allow the committee members to see and interact with the cluster as if they were the student. Therefore, student facing clusters were presented through the contractor's test engine in the same manner as the students' experience. First, the Equity Review Committee engaged with the cluster as if they were a student. Next, each member presented written comments as feedback regarding each stimulus and item in the

cluster. All the comments populated a spreadsheet in which the facilitator (science education assessment consultant) could read and review all the feedback. Finally, the facilitator reviewed the feedback with the Equity Review Committee members as a whole and documented consensus notes. Over the course of a day, the Equity Review Committee provided feedback regarding any biases or sensitivity issues identified within the clusters. These consensus notes were then used to make revisions on the clusters following the review process.

The Content Review Committee worked in a similar manner. Science education experts (researchers, teachers, and curriculum coordinators) worked in groups of six to review the three (Grade 5, 8, and 11) grade-level specific clusters in the online test engine. Like the Equity Review Committee, the Content Review Committee engaged with the clusters as if they were students, provided written feedback, and engaged in discourse about the feedback as a whole group with the guidance of a facilitator. Because of the in-depth nature of the three-dimensional clusters, Content Review Committee occurred over the course of two days. The consensus notes from the Content Review Committee also influenced revisions to clusters following the review process.

Internal Revisions

Following the committee reviews, the clusters were sent back to the content specialists (state and vendor) for revisions. These revisions reflected the comments provided by both committees. Graphics, wording, and task types were revised based on the feedback from the committees. Following, the clusters were once more rendered the interactive test engine.

Internal Review

The final layer of review for the clusters was an internal review by the state's English Language Learner specialist, English Language Arts specialists, Mathematics consultants, and

assessment editors. Here, fine grain edits were made to each item within the clusters to ensure accessibility to the largest group of students possible. Following this final review, the clusters were moved to production in the testing engine enabling students to participate in the pilot test.

In summary, the process used by the State of Michigan to develop clusters for the large-scale science assessment provides important contextual information to better understand the research presented here. Next, I present the methods used to gather evidence about the extent to which two of the clusters developed through this process were aligned with the claims of the writers.

CHAPTER 4: RESEARCH DESIGN AND METHODS

Study Overview

The purpose of this qualitative dissertation study is to answer the following research

questions:

1) To what extent do the clusters developed using Michigan Cluster Development

Process align with the Michigan K-12 Science Standards?

    1a) To what extent do these items elicit and discriminate for the intended

    dimensions?

These research questions are important because ultimately, for the Michigan State Science

Assessment, the goal is to provide students opportunities to demonstrate their proficiency in

three-dimensional science. Exploring what dimensions of science understanding students draw

on to respond to specific items will help to better understand what the items are measuring and

what claims can be made about students' science proficiency. In addition, using evidence of

students' engagement with the items can allow insights into how design decisions translate (or

not) into the ability to elicit two- and three-dimensional science understanding from students in a

state-level science assessment.

Study Design

This is a qualitative study in which I used think-aloud interviews (also called cognitive

labs) to understand the extent to which participants used the three dimensions (i.e., disciplinary

core ideas, science and engineering practices, and crosscutting concepts) to respond to the items

designed using the process described in Chapter 3. I compared an external review of the cluster

alignment (Task Annotation Project in Science [TAPS], described below) with the outcomes from cognitive lab data analysis and the analysis of the text.

Participants

Participants for the cognitive labs were identified via convenience sampling (Gall et al., 2007). I used both professional and social networks to seek volunteers for the study. A parental consent form was used (Appendix F) to obtain parental/guardian consent for each participant. This process yielded ten students in Grade 5 and nine students in Grade 8. The students were in the western, central, and eastern areas of southern Michigan (Table 4.1).

Table 4.1

Demographic Characteristics of Grades 5 and 8 Participant Sample

Characteristics	Grade 5 Sample		Grade 8 Sample	
	n	%	n	%
Female	5	50	5	56
Male	5	50	4	44
Grade Level				
Grade 5	3	30	-	-
Grade 6	7	70	-	-
Grade 8	-	-	3	33
Grade 9	-	-	6	67
Race/Ethnicity				
African American	6	60	2	22
Asian	-	-	1	11
Haitian	-	-	1	11
Hispanic	-	-	1	11
Hispanic/Indian	1	10	-	-

Table 4.1 (cont'd)

Mixed Race	1	10	-	-
Caucasian	2	20	4	44
Language				
ELL	-	-	3	33
Non-ELL	10	100	6	67
Region				
Southern east	4	40	-	-
Southern central	6	60	2	22
Southern west	-	-	7	78

Data Collection

The methods used for this study stem from the work of protocol analysis by Ericsson and Simon (1993) and verbal analysis from Chi (1997). Think-aloud protocols for cognitive labs (Conrad et al., n.d.) were a critical part of the data collection and were developed for this research in conjunction with the MDE as part of the validation efforts for the new state-wide science assessment: the Science M-STEP. Think-alouds, or verbal protocols, are a research tool in which participants are asked to complete a task while verbalizing their thinking out loud. The focus that verbal analysis has on learning is appropriate for the context of this research. Chi (1997) argues:

> the goal of the method here is to attempt to figure out what a learner knows (on the basis of what a learner says, does, or manifests in some way, such as pointing or gesturing) and how that knowledge influences the way the learner reasons and solves problems, whether correctly or incorrectly. Thus, the trick is to analyze the learner's utterances (in the case of

verbal data) to capture the knowledge that might underlie those utterances and do so in a

way that is not subjective; therefore, it needs to be quantifiable in some ways. (p. 3)

Think-alouds will never include every thought of the participant but do provide some insight into

their processes for solving tasks (Ericsson & Simon, 1993). The think-aloud protocols that I used

in this study (Appendix D) were designed with some additions to the typical think-aloud process

to elicit students' explanations for their answers and their associated reasoning. For example, the

think-aloud protocols asked the participants to verbalize their responses but then the researcher

asked, "Why did you answer that way?" or "Tell me more about your response." While

interview-like interjections can change performance on the assessment (Beatty & Willis, 2007),

in order to get the most information about how students interacted with the clusters, fusing

protocol analysis with verbal analysis was appropriate.

The cognitive lab data were collected between May and October 2019. Of the nine

clusters developed for the 2017 M-STEP Science Pilot, one Life Science cluster in Grade 5 and

one Physical Science cluster in Grade 8 were chosen as the focus of this dissertation. The

clusters were initially written in the summer of 2016 and developed for the Spring 2017 test

administration period. However, both clusters used for this study were released to the public as

sample clusters, which is why they were chosen for this study. Data collection took place in

schools in three regions of the state (see Table 4.1). Within each, the participants and I were

provided a semi-private location where each participant could focus on the task with little

distraction. For each cognitive lab, the protocol was used as a guide; however, deviations from

the protocol occurred when I thought that more information from the participant was necessary,

resulting in a hybrid procedure between a think-aloud and an interview. I gained permission from

each participant to begin audio recording the cognitive lab and took field notes during the cognitive lab.

Data Processing

The cognitive lab data was stored in password protected digital format. To process the data, I initially transcribed each cognitive lab using speech-to-text software. I did a second round of listening to audio files and made modifications to the transcripts to ensure accuracy of the transcription. Each transcription file was filed by grade level and participant code.

Coding

I used Chi's (1997) steps for verbal analysis in coding and analyzing the data. Specifically, the transcripts were segmented by item and then by utterance, making the utterance the unit of analysis. I am defining an utterance as an idea unit that includes a full idea verbalized by the student.

Each item was designed to be aligned to two or more dimensions of the NGSS. Thus, I examined student responses for evidence of students' using or not using the intended dimensions. To develop codes, I first used the unpacking documents from the Next Generation Science Assessment project (NGSA; Krajcik, n.d.) to clearly define what evidence of each of the dimensions might look like. The rationale for using these unpacking documents stems from the initial design of the items. In Chapter 3, I described the cluster development process, which included unpacking using the resources adapted from the NGSA project, as described in Harris and colleagues (2019). Using the last table in the unpacking documents, *Evidence for Each Component of the Practice / Cross Cutting Concept,* I adapted this verbiage for coding the components of DCIs, SEPs, and CCCs found in the students' responses.

I then engaged in iterative rounds of developing, applying, and refining my coding scheme. I looked for patterns in the ways students were engaging with the dimensions and worked to refine the codes to capture and represent these patterns. In the final round of codes, I developed a coding rule for each SEP and CCC present in the items. These coding rules cut across items and informed how I coded any item aligned with the SEPs or CCCs (Table 4.2). These coding rules allowed me to make claims about dimensions overall rather than having idiosyncratic definitions of dimensions for each item. I used the rules to ensure that the codes were orthogonal, meaning that they were not dependent on each other. For example, a student could get a code for a CCC even if the DCI was not evident in their response. The overarching rules for the SEPs and CCCs are shown in Table 4.2.

Table 4.2

Overarching Coding Rules

Dimension		Overarching Rule
CCC	Cause and Effect	Student states a relationship between two occurrences where one occurrence leads to the other (or needs the other to occur). The language should include linking words such as "because," "and then" (but just having a linking word is not sufficient to get a code of "present" - the linking words have to link the occurrences). If there is a sequence of intermediate events that link the cause and effect, the student states some intermediate events.
SEP	Modeling	Grade 5: Student state connections/interactions between components of the model (where all components are given). For this item, the "arrows" are what is counted as "modeling" because the arrows represent mechanisms by which ….So language for "arrows" could include "leads to" "causes" "and then" .... The description of what the arrow means does not have to be scientifically accurate (e.g., does not need to say "reflect") Grade 8: The limitations portion of the question (Part B) is the focus of the modeling SEP in this item. Limitations: Students must say more than the limitation option they picked. They must explain what is missing in the model that would cause the limitation to be valid or explain why they chose the limitation.

Table 4.2 (cont'd)

SEP	Argumentation	Evidence: Students must indicate that they are using (1) evidence given in the item, (2) evidence from prior knowledge or (3) information from other sources within the item cluster. The evidence does not need to be correct. Reasoning: Students must indicate that they are explaining connections between the evidence and the claim. The reasoning does not need to be scientifically accurate, but it must be clear that they are attempting to make a connection
SEP	Analyzing and Interpreting Data	Students state patterns and relationships in the data and describe why they are meaningful to the investigation question. Language indicating patterns or relationships could include: ● Quantitative or qualitative description of change presented in data (just indicating a "change" happened is not enough) Language for describing why the data is meaningful could include: ● Identifies relationships: Students analyze the data to identify patterns (i.e., similarities and differences), including the changes ● Interpret the data about the changes ● Students use data to determine whether a change occurred ● Students support their interpretation of the data by describing that the change
SEP[a]	Constructing Explanations and Engaging in Argument from Evidence	Evidence: Students must indicate that they are using (1) evidence given in the item, (2) evidence from prior knowledge or (3) information from other sources within the item cluster. The evidence does not need to be correct. Reasoning: Students must indicate that they are making connections between the evidence and the claim. The reasoning does not need to be scientifically accurate, but must be clear that they are attempting to make a connection NOTE: Same as Argument from Evidence in Grade 5 Item 5

[a]Due to the Claim, Evidence, Reasoning item format used for questions assessing these SEPs, the overarching coding rule was also the same.

For the DCIs, I defined how the DCI would be coded item-by-item. Table 4.3 shows

some examples of how I defined the DCI codes. See Appendix E for a complete list of codes.

Table 4.3

DCI Coding Definitions

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.

How the DCI is coded in this item: Student states that light must be present for the plant to be seen AND that light must reflect off the plant (Ref) AND that light must enter the eye after reflecting off the plant (Eye).
The language for "reflect" can include
   ● "directs,"
   ● "bounces off of,"
   ● "goes back,"
   ● etc
Code as 0:
   ● If only the flashlight and seeing the plant is mentioned
   ● If the order or causal mechanism are incorrect
Non-codable:
   ● If the student repeats the answer options or the prompt verbatim

DCI: LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain.

How the DCI is coded in this item: Student states the eyes are sense receptors that take in light information (Sns) AND that light information taken in by the eyes is processed in the brain (Brn).
The language for "sense" can include
   ● "feel,"
   ● "take in,"
   ● "notice"
   ● etc
Code as 0 when.
   ● If only the eyes are mentioned.
   ● If the order or causal mechanism are incorrect.
Non-codable:
   ● If the student repeats the answer options or the prompt verbatim

DCI: LS1.A: Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.

How the DCI is coded in this item: Student states that the pupil regulates the amount of light entering the eye as a function to promote growth, survival, behavior, and reproduction. In this item, this is only seen with some phrases that indicate the function of the pupil is to regulate light due to the body's response system.
   ● For example: "eyes hurt when the lights come on"
   ● "The muscles in the eyes make the change…"
   ● "Pupil needs to open to process light"
   ● "The pupil's diameter doesn't have to open"
Non-codable:
   ● If the student repeats the answer options or the prompt verbatim

DCI: PS1.B.2: The total number of each type of atom is conserved, and thus the mass does not change.
How the DCI is coded in this item: The student must reference to the number of atoms in the final substance Non-codable: ● If the student repeats the answer options or the prompt verbatim

When creating orthogonal coding rules for some items, it was difficult to separate out the dimensions because of the overlap in wording or meaning of the paired dimensions. For example, when the SEP Analyzing and Interpreting Data was paired with the CCC Patterns, there was no way to code the students' responses for one of these dimensions without coding for the other. Therefore, I made the decision to make a combined SEP/CCC code. This happened for three items in Grade 5 Items 1, 3 and 4) and 3 items in Grade 8 (Items 3, 4, and 5). (See Appendix E for full codebook).

Coding Examples

Table 4.4 provides an example of the codebook for a single item that was aligned with a DCI and CCC (see Appendix E for the full codebook). The DCI for this item (PS4.B: An object can be seen when light reflected from its surface enters the eyes) had two potential codes indicating that: (a) a student mentioned reflection and (b) that students mentioned that light has to enter the eyes for something to be seen. In the example student responses, I crossed out the part of the transcript when students just read part of the item. In bold, I include the part of the transcript that provides evidence for the component of the DCI. For example, for the reflection code, the student said, "Because if you shine light on something from a flashlight, it's going to reflect off the plant…" indicating they were using their understanding of reflection to answer the question.

For the crosscutting concept of cause and effect, there were also two potential codes: (a) students explicitly linked between a cause and effect and (b) students explicitly provided an intermediate step between the cause and the effect. An example of a student response coded for a link between a cause and effect is, "the students are able to see where the plant is because of the

flashlight" indicating evidence that the student was linking the flashlight as the cause for being able to see (the effect).

Table 4.4

Grade 5 Item 1 Codebook Sample

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.		
Code	Definition	Example
Ref	Reflection of light off the surface of an object. States that light must be present for the plant to be seen and the light must reflect off the plant	(510) ~~P: Student reads option A - no I don't think so~~ ~~P: Student reads option B and C~~ P: ~~Student reads option D.~~ Well I think that is right. Do I just click…. ~~R: Why do you think D is the right answer?~~ P: **Because if you shine a light on something from a flashlight, It's going to reflect off the plant... Well the thing. And then you can see it.**
Eye	Light enters the eyes for objects to be seen. States that light must enter the eye after reflecting off plant for plant to be seen	(53) P: I think they're able to see the plants now because the light is reflecting off of their eyes To the plant so they can see it. "~~Once the plant produces its own light the students can observe for the plant. Once the plant absorbs all the light from the flash light the students can observe the plant. The light from the flash light is reflected off the student eyes and then back to the plant. The light from the flash light is reflected off the plant and then enters the student eyes."~~ I'm going to say D. ~~R: D? Can you tell me why you answered that way?~~ P: **The light reflects into their eyes and then they can see the plant.**
CCC: Cause and Effect: Cause and effect relationships are routinely identified		
Code	Definition	Example
Lnk	Includes a link between a cause and an effect	(52) I think it's D because while she's pointing at the plant there's a flashlight pointing at the plant. And **the students are able to see where the plant is because of the flashlight.**
Seq	Includes Lnk code and a sequence of intermediate events that link the cause and effect	(51) The light hits the plants and it directs to your eyes. So I think it would be D because the plants into the student's eye because of the flashlight's light that is given to the plant. **It can direct to your eye.**

A science education expert with knowledge of NGSS was recruited to participate in interrater reliability. Early versions of coding had 75% agreement for all items. For all

disagreements, coders met and came to a decision for all the codes and adjusted the codebook to reflect the final agreements. For the final version of coding, 20% of responses were double coded. Interrater reliability was calculated by looking for agreement for coding of present or absent for all codes relevant for a given item. The final interrater reliability was 92.6%. All disagreements were discussed and adjudicated, and examples were entered in the codebook to clarify decisions.

Cognitive Lab Data Analysis

Verbal analysis as defined by Chi (1997) "is a methodology for quantifying the subjective or qualitative coding of the contents of verbal utterances. In verbal analysis, one tabulates, counts, and draws relations between the occurrences of different kinds of utterances to reduce the subjectiveness of qualitative coding" (p. 2). In applying verbal analysis, I first used an analysis question to find and make sense of the patterns within and across items. Which items are discriminating students who chose the correct response from students who chose the incorrect response as evidenced by having codes for specific dimensions? I constructed frequency tables for each item - looking for patterns in codes both within and across dimensions based on whether students selected the correct response. I looked for evidence of elicitation of each dimension. I defined elicitation as the item's ability to provide opportunity for students to use knowledge of a dimension regardless of whether they choose the correct or incorrect response. From the coding pattern indicating elicitation is that the majority of students have a code for the dimension independent of their answer choice or that students who got the answer correct have codes for the dimension while the students who chose the incorrect response did not, linking elicitation and discrimination.

Additionally, the items were examined for discrimination—"an index of an item's effectiveness at discriminating those who know the content from those who do not" (Tobin, 2018). In this study, I defined discrimination as the item's ability to separate students who know a particular dimension by choosing the correct answer from those who do not know that dimension evidenced by choosing the wrong answer. From the coding perspective, the pattern in codes is that students who got the item correct were more likely to have the code for that dimension than student who got the item incorrect. For example, if all students who selected the correct response had codes for a DCI, while students who did not get the item right did not have codes of the DCI, this data supports the claim that the item discriminated for the DCI. Therefore, an item was said to discriminate for a particular dimension when the students who chose the keyed response demonstrated evidence of using the targeted dimension and those who chose a non-keyed response did not provide evidence of the targeted dimensions or used a dimension incorrectly. While patterns are not always as neat as this—I looked for trends in the codes to make final claims about elicitation and discrimination by determining that if the majority of students (51% or more) had a code for a particular dimension, then the item elicited or elicited and discriminated based on the definitions described above. Finally, I looked across items to identify characteristics of items that discriminated students who chose the correct response from students who chose the incorrect response as evidenced by having codes for specific dimensions to determine if patterns arose throughout the clusters.

In the findings chapter, I provide information for Grade 5, Items 1–5 and Grade 8, Items 1–5. I did not include Grade 8, Items 6 and 7, because the claims I could make from those items did not add to my evidence due to the items' inability to elicit or discriminate based on any dimension.

TAPS Data Analysis

The data from the Task Annotation Project in Science (TAPS; Achieve, 2019) was used as a secondary data source to provide information about whether items elicited and discriminated for specific dimensions. TAPS employed a diverse set of experts to identify features of three-dimensional assessment tasks across multiple domains and grade levels. Part of the TAPS project was to analyze released state science items and tasks using the Science Task Screener (Achieve, 2018) developed for the project. The Task Screener contains four criteria: (a) Tasks are driven by high-quality scenarios that focus on phenomena or problems, (b) Tasks require sense-making using the three dimensions, (c) Tasks are fair and equitable, and (d) Tasks support their intended targets and purpose. Using the TAPS methodology, released sample statewide summative assessment items from 8 states were reviewed and annotated (Appendix H). Both the Grade 5 and Grade 8 clusters from Michigan were reviewed. Each of the clusters was reviewed by 3 expert reviewers using the Task Screener and facilitated group consensus conversations. The item-level TAPS information was used as the secondary piece of alignment data. Each item was evaluated by the TAPS reviewers regarding the necessity of the claimed dimensions, the extent to which those dimensions were represented in the item, and the role of the dimensions and the item in sensemaking about the phenomenon. (See full list of questions in Appendix I).

The TAPS analysis data came in the form of a spreadsheet that included sections for evaluations of the scenario, individual questions, and the task overall. I focused on the section for the individual questions. Within this section there are three categories: Category A: High-quality phenomena and problem driven; Category B: Sense-making using the three dimensions; and Category C: Connection to assessment purpose. Within each of the categories, several indicators

are applied to the item. For this research, I focused on the indicators in Category B but included information from Category A and C for context (Appendix G).

Category B is divided into four groups of indicators: B1) The task requires students to demonstrate grade appropriate SEP element(s); B2) The task requires students to demonstrate grade appropriate DCI element(s); B3) The task requires students to demonstrate grade appropriate CCC element(s); and B4) The task requires students to integrate multiple dimensions in service of sense-making and problem solving. For each indicator question, the reviewers responded with Yes, No, or N/A and were provided the opportunity to explain their rationale.

The document containing the consensus information was used for this research. After synthesizing the TAPS data, I crafted summary tables for each item (Appendix I) to analyze in comparison to the cognitive lab findings. Each table contained a column labeled *Strengths* to highlight the assets of the item found by the TAPS reviewers and a column labeled *Opportunities for improvement.* These tables were compared to the findings of the cognitive labs for each item. For example, for Grade 5, Item 1, the cognitive lab findings indicated evidence of both the DCI and the CCC. However, the TAPS analysis concluded that the DCI was required by the item, but the CCC was not. In instances where the cognitive lab data and the TAPS data do not agree, I explore the findings to determine why there is disagreement.

Researcher Stance

I come to this work as a white female from the U.S. I was raised in a conservative religious family and was afforded the opportunity to attend parochial schools throughout my K-12 education, including boarding school for my high school years. While education and love of learning has always been fostered within my family, my passion for teaching first manifested through the arts, as a dance teacher. My experiences in traditional classroom education began as

a second career as I moved from the Food and Beverage Industry to Education for familial purposes as a single mother. I moved into my science education career through South Carolina's transition to teaching program, Program of Alternative Certification for Educators (PACE). This program allowed me to maintain full-time employment as a middle school science teacher while earning my teaching credentials in the state of South Carolina.

My teaching context was a small sea island school, which has a rich history in the Civil Rights Movement. As a new teacher, I struggled to connect with my predominately African American students and parents. Over time, I realized that my struggles were due to my lack of awareness about the cultural contexts in which my students were embedded. Over the years, I began to center my students in the classroom and value their voices, experiences, and culture. When I left classroom teaching to pursue a PhD in education, my research interest was to develop effective ways to bridge the research–practice gap. I learned that one way to bridge the research-practice gap is through assessment literacy. Assessment literacy is defined as understanding the process of gathering information about diverse student learning to inform education-related decisions (National Task Force on Assessment Education for Teachers, 2016). For teachers to better understand assessment design decisions, purposes, and intent of various assessments, assessment literacy is necessary for educators. Therefore, by empowering teachers to understand and take part in assessment decisions in Michigan, research regarding assessment becomes available to teachers.

My role as a researcher in this study was that of a participant observer—a researcher who is also an active part of the research context. During cluster development and cognitive labs, I was the primary facilitator of the work. During my research, I was employed by the Michigan Department of Education to develop a state assessment for the new Michigan K-12 Science

Standards that met validity, usability, and budget constraints. Therefore, the need to produce a specific type of assessment product could have influenced design decisions, iterations, timing, and other factors that impact this research. The clusters, which are the focus of this research, were among the first set of those designed for state-level testing in Michigan and across the nation. Because of these uncharted waters, these items were part of that learning endeavor. Therefore, I acknowledge that I was an integral part of the learning community, and I cannot separate myself from the research context and must be conscious of the benefits and drawbacks of engaging as a participant observer throughout the course of the research. I may have made decisions that are good for the assessments but may not have been beneficial to my research agenda. As my aim was to build three-dimensionally aligned clusters that provide an opportunity for all students to demonstrate their understanding of the standards, I acted accordingly.

In the following chapter, I present the findings for the Grade 5 and Grade 8 item clusters using the methods described here.

CHAPTER 5: FINDINGS

This chapter contains the findings based on the methods and data analysis described in Chapter 4. These findings will seek to answer my research questions:

1) To what extent do the clusters developed using Michigan Cluster Development Process align with the Michigan K-12 Science Standards?

1a) To what extent do these items elicit and discriminate for the intended dimensions?

In this chapter, I will discuss the Grade 5 and Grade 8 clusters. As described in Chapter 3, the cluster consists of a stimulus, in one or more parts, and a set of five to eight items associated with the stimulus. The clusters were designed to assess one NGSS topic bundle (a group of performance expectations). I use the cognitive lab data as the primary source of data to discuss item discrimination and elicitation of one or more dimensions and compare these results with the TAPS analysis as a secondary data source. To view the full clusters discussed in this chapter, please see Appendix D. First, I present items that elicited and discriminated students who chose the correct response from students who chose the incorrect response as evidenced by having codes for specific dimensions. Next, I present the items that did not discriminate students who chose the correct response from those who did not. Finally, I present a summary of the coding results for the Grade 5 and Grade 8 clusters respectively.

Elicitation and Discrimination

In this section, I describe the items in the Grade 5 and Grade 8 clusters that elicited knowledge of one or more dimensions and discriminated between students who chose the correct response versus those who chose the incorrect response. There were two items in the Grade 5

cluster (Items 1 and 3) and one item in the Grade 8 cluster (Item 5) that clearly elicited

dimensions and discriminated based on one dimension.

Which statement **best** describes how the students are able
to see the plant?

(a) Once the plant produces its own light, the students
can observe the plant.

(b) Once the plant absorbs all the light from the flashlight,
the students can observe the plant.

(c) The light from the flashlight is reflected toward the
students' eyes and then back to the plant.

(d) The light from the flashlight is reflected off of the plant
and then enters the students' eyes.

Figure 5.1. Grade 5 Item 1.

Item 1 (Figure 5.1) asks students to consider the mechanism for the plant to be seen after

the teacher shined a flashlight on it. Each of the distractors provides common misconceptions

students may have about this mechanism. The correct response, option D, taps into the concepts

of reflection of light and light entering the eyes. This item was designed to assess the DCI,

PS4.B, and the CCC of Cause and Effect.

There were four codes developed for this item (see Table 4.2 and Appendix E). The DCI

codes focused on two components of the DCI: (a) Light must be reflected off of an object to be

seen, and (b) light must enter the eyes for the object to be seen. Examples of the application of

these codes can be seen in Table 4.2. The bolded phrases indicate the section of the transcript

that was considered the utterance for the code. The two CCC codes focused on two components

of cause-and-effect relationships: (a) any indication linking a cause and effect or (b) indication

that the student included an intermediate event between the cause and effect. These codes were

used regardless of whether the cause-and-effect reasoning was scientifically correct.

55

Table 5.1

Grade 5 Item 1 Coding Patterns

Student[a]	DCI		CCC	
	Ref	Eye	Lnk	Seq
51	X	X	X	X
52			X	
53		X	X	X
56	X	X	X	X
59			X	
510	X		X	X
54			X	X
55				
57			X	
58				

[a] *White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.*

Table 5.1 shows the overall coding patterns for student responses to Item 1. All the students who answered the question correctly provided a response that was coded for one or both aspects of cause and effect. Additionally, four of six students who chose the correct response demonstrated their knowledge of one or more aspects of the DCI in their response. Conversely, only two of four students who answered incorrectly had a response that was coded for cause and effect, and none of these students' responses indicated an understanding of the DCI. Based on my criteria for elicitation described in the methods chapter, this item elicits students' understanding of the DCI as is evidenced by codes for the DCI for students who answered the item correctly. The item also elicited evidence of students' cause and effect reasoning shown by

56

students' responses having codes for cause and effect for all students who answered correctly and two of the four who did not.

Table 5.2

Grade 5 Item 1: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking.	The information in the scenario is not necessary to answer the question. The stated CCC is not measured in the item and very little reasoning is required. Overall, the item does not assess what it is intended to assess.

Based on my criteria for discrimination, these findings show that this item discriminates for the DCI because the majority of the students who answered correctly provided evidence of at least one portion of the DCI. The item elicits but does not discriminate based on the CCC because while all the students who answered correctly provided evidence of the CCC, half of the students who answered incorrectly also provided evidence of the CCC. Overall, the patterns suggest that the item elicits and discriminates based on the DCI dimension.

The TAPS analysis agrees with the cognitive lab findings about the DCI, concluding that a substantial part of the DCI is required to answer the question. However, the TAPS analysis concluded that the CCC was not necessary to answer the question (Table 5.2), while my analysis suggests that the CCC is necessary to answer the item correctly (i.e., all students who got the item correct had codes for the CCC) but it is not sufficient to answer the item correctly (i.e., some students who had codes for the CCC without the DCI answered the question incorrectly).

Drag each picture into the model to complete the path that
light and information follow to allow the plant to be seen.
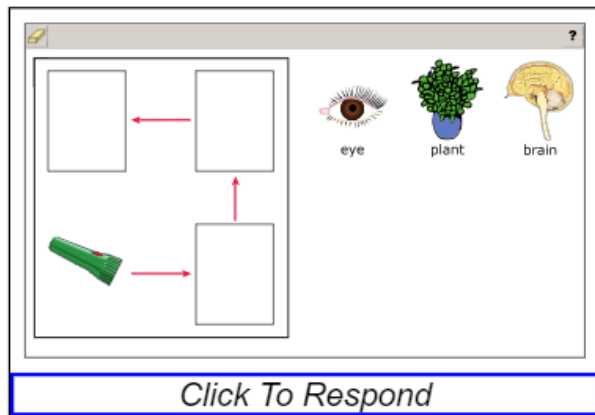
Figure 5.2. Grade 5 Item 3.

Item 2 (Figure 5.2) asks students to model the path of light that would allow the plant to

be seen. The students are provided all the components of the model and must select and move

each component into the appropriate box. The correct response, plant - eye - brain, taps into the

concepts that light reflects off of objects and then enters the eyes, and then information is

processed by the brain in order for us to see. This item was designed to assess all three

dimensions: (a) DCI: PS4.B and LS1.D; (b) SEP: Developing and Using Models; and (c) CCC:

Systems and System Models.

There were seven codes developed for this item (Table 5.3). The four DCI codes were:

(a) light must be reflected off of an object to be seen; (b) light must enter the eyes for the object

to be seen; (c) eyes are sense receptors specialized for light information; and (4) information is

processed by an animal's brain. The three SEP codes for modeling indicate the number of arrows

the student explained in their verbal response. Because the model provided all of the

components, the coding focused on the students' explanation of the arrows between the

components of the model. Examples of the application of these codes can be seen in Table 5.3.

The brackets indicate parts of the student response that were coded for an arrow. As explained in

the Chapter 4, it was not possible to build a unique code for the CCC Systems and System

Models that was separate from the SEP Developing and Using Models. Therefore, a single

SEP/CCC code was used.

Table 5.3

Codes for Grade 5 Item 3

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.		
Code	Definition	Example
Ref	Reflection of light off the surface of an object. States that light must be present for the plant to be seen AND the light must reflect off the plant.	(56) P: First the flashlight goes to the plant and **the light bounces off the plant** into the eyes and then it goes up to the brain so it can process the information.
Eye	Light enters the eyes for objects to be seen. States that light must enter the eye after reflecting off plant for plant to be seen.	(59)P: Well to see the plant you have to have a plant. P: and then once the flashlight turns on, your eyes see it next and then to actually process what it is it goes... Like what's happening in your brain. Because you can't really see stuff when it's in your brain because you can't go through your whole body. R: Okay say more about that P: if you turn on a flashlight it's not going to go into your skin and like through your head into your brain. **It has to go through your eyes** because they're open and they're easier to get into. And then that tracks into your brain so that's why I would say like that it goes before the brain.
DCI: LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain.		
Code	Definition	Example
Sns	Sense receptors specialized for information. States that the eyes are sense receptors that take in light information	No examples in student responses

Table 5.3 (cont'd)

Brn	Information is processed by the animal's brain. States that the light information taken in by the eyes is processed in the brain	(57) P: Because the teacher is trying to reflect the light off the plant and then it got into the students' eyes. **And then the brain now tries to process it so that it can be looked at in the brain** and then you can see

SEP: Modeling + CCC: Systems and Systems Modeling

Students have to explain the arrows, not just point to them. Modeling should not just be pointing to the pictures in order because that isn't evidence that the student is explaining what the arrow represents.

Code	Definition	Example (note that brackets indicate what was coded as one arrow)
1	Includes what one arrow represents in the model	No examples
2	Includes what two arrows represents in the model	(54)P: So basically you see it with your eyes [**and then it goes to your brain**]/ [**and then you see the plant**]. I don't know. Is it the other way around? I don't know if it is the other way around between the eyes and the brain R: ok so what is make you question that P: In order..to like see...cause your brain allows you to see stuff. If your blind you basically can't see stuff. So then something is wrong with your brain and you can't see. R: So you are saying that if you are blind there is something wrong with your brain? P: Isn't there like some parts...cause your [**eyeball is connected to your brain**]. I think it is the other way around Eye, brain, plant *NOTE: Eyeball is connected to your brain is the same "arrow" as "you see it with your eyes and then it goes to your brain"*
3	Includes what three arrows represents in the model	(56) P: First the [**flashlight goes to the plant**] and the [**light bounces off the plant into the eyes**] and [**then it goes up to the brain so it can process the information**]

Table 5.4

Grade 5 Item 3 Coding Patterns

Student[a]	DCI				SEP/CCC		
	Ref	Eye	Sns	Brn	1	2	3
52	X	X					X
55	X	X		X			X
56	X	X		X			X
57	X	X		X			X
59		X		X		X	
510				X		X	
51							X
53						X	
54						X	
58						X	

[a] White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.

Table 5.4 shows the overall coding patterns for student responses to Item 3. All the students who answered the question correctly provided a response that was coded for one or more aspects of the DCI, while none of the students who answered incorrectly provided evidence of the DCI. All students provided some information regarding the SEP/CCC; however, the students who answered correctly explained more of the connections (arrows) in the model than most of the students who answered incorrectly. Based on my criteria for elicitation, these

findings show that this item elicits students' understanding for the DCIs because all the students who answered correctly provided evidence of two or more elements of the DCIs. The item also elicited evidence of students' modeling, illustrated by students having codes for modeling for students who answered the item correctly.

Table 5.5

Grade 5 Item 3: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking.	The information in the scenario is not necessary to answer the question. The SEP is not measured. The stated CCC is not measured. The item requires a visualization of the DCI but does not assess the SEP. Overall, the item does not assess what it is intended to assess.

Based on my criteria for discrimination, these findings show that this item discriminates for the DCIs because all the students who answered the item correctly provided evidence of the DCIs, while students who answered the item incorrectly provided no evidence of the DCIs. However, one element of the DCIs, "different sense receptors specialized for particular kinds of information," was not mentioned in any of the students' responses. Moreover, the item did not meet the criteria for discriminating based on the SEP/CCC because, while all the students who answered correctly provided evidence of the SEP/CCC, all students who answered incorrectly also provided some evidence of the SEP/CCC. Overall, the patterns suggest that the item elicits and discriminates based on the DCI.

The TAPS findings indicated that neither the SEP nor the CCC is required in this item (Table 5.5). However, the cognitive lab data shows that all the students did provide evidence of using the SEP/CCC in their responses. Thus, while it is not possible to distinguish students' use

of the SEP from the CCC, my results suggest that students did use modeling (SEP/CCC) when responding to the item.

Identify and explain the temperature pattern in the graph. Select and move the **best** claim statement, evidence statement, and reasoning statement that can be used to explain the pattern.

Claim	
Evidence	
Reasoning	

Claim Statements:
Energy is transferred from each system to the thermometers.
Energy is transferred from the thermometers to each system.
Energy is not transferred in either of the systems.

Evidence Statements:
The temperature was higher at 50 minutes than at 0 minutes.
The temperature was lower at 50 minutes than at 0 minutes.
The temperature did not change between 0 minutes and 50 minutes in either system.

Reasoning Statements:
The iron absorbed energy from the oxygen during the chemical reaction.
The hand warmer bag absorbed energy from the thermometer.
Energy was released when the iron reacted with the oxygen in the air.
Energy was released when the hand warmer package was opened.

OK

Figure 5.3. Grade 8 Item 5.

Item 5 (Figure 5.3) asks students to choose a claim, evidence, and reasoning that best explain the temperature pattern seen in the stimulus graph (Appendix D). The students must select and move one claim statement, one evidence statement, and one reasoning statement into the boxes to construct their response. The correct response is Claim: Energy is transferred from each system to the thermometers; Evidence: The temperature was higher at 50 minutes than at 0 minutes; and Reasoning: Energy was released when iron reacted with the oxygen in the air. The distractors provide options for students to choose responses that are not consistent with the given data but still provide a logical argument. This item was designed to assess all three dimensions: (a) DCI: PS1.B.3; (b) SEP: Constructing Explanations; and (c) CCC: Energy and Matter.

63

There were three codes developed for this item (Table 5.6). The DCI code focused on one aspect of the DCI PS1.B.3: Some chemical reactions release energy, others store energy. The two SEP codes focus on the evidence and reasoning provided in the students' verbal response. It was not possible to build a unique code for the CCC Energy and Matter that was different from the DCI, so a DCI/CCC code was used. Examples of the application of these codes can be seen in Table 5.6.

Table 5.6

Coding for Grade 8 Item 5

DCI: PS1.B.3: Some chemical reactions release energy, others store energy + CCC: Energy and Matter		
Code	Definition	Example
Re	Students identify that energy is released form the system in the form of heat	(87) the claim that I chose is that the ~~energy is transferred from each system to the two thermometers~~ and now I'm just trying to think of which of the other statements lines up with that. ~~the temperature was higher at 50 minutes and it was a zero~~ **so that means it took longer like in 50 minutes for the temperature to go up**.**(Re)**.
SEP: Constructing Explanations		
Code	Definition	Example
Ev	Evidence from item cluster: Students are drawing on the given data in the stimulus or ideas from prior items or prior experiences/knowledge to support their claim/answer the question.	(85) Evidence statements. ~~Temperature was higher at 50 minutes than at 0 minutes.~~ In which one? **There's two bags. Actually, 50 minutes it is like 90 degrees. 0 minutes it is like 70 degrees**. So yeah, that's not true. That is not true either.
Rsn	Reasoning: students explain how the evidence they stated or chose supports the claim they stated or chose. (the reasoning must go beyond stating that a relationship to the evidence exists but must attempt to explain the relationship (the "why")	(83) the energy was released when the hand warmer package was opened because the **oxygen gets to it as you open the package which allows it to kind of heat up and make that chemical reaction.**

Table 5.7

Grade 5 Item 5 Coding Patterns

Student[a]	DCI/CCC	SEP	
	Ce	Ev	Rsn
83	X	X	X
87	X	X	X
81			
82		X	X
84		X	X
85		X	
86			
88		X	
89			

[a] White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.

Table 5.7 shows the overall coding patterns for student responses to Item 5. The two students who answered the question correctly provided responses in the cognitive labs that were coded for the DCI/CCC, while none of the students who answered incorrectly provided DCI/CCC evidence. All the students who answered correctly provided information regarding both aspects of the SEP; however, some of the students who answered incorrectly also used the SEP. Based on my criteria for elicitation, these findings show that this item elicits students' understanding of the DCI/CCC, shown by the codes for students who selected the correct responses. The item also elicited evidence of students' constructing explanations as is evidenced by codes for the SEP for students who answered the item correctly and those who did not.

Table 5.8

Grade 5 Item 5: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required.	The stated SEP is not measured with the item, rather the reviewers suggested Analyzing and Interpreting Data was being assessed. They argued that selecting options for a CER is not engaging in the cited SEP. The stated CCC is not measured. Overall, the item does assess what it is intended to assess.

Based on my criteria for discrimination, these findings show that this item discriminates for the DCI/CCC because all the students who answered the item correctly provided evidence of the DCI/CCC, while students who answered the item incorrectly provided no evidence of the DCI/CCC. The item does not meet the criteria for discriminating based on the SEP because while all students who answered correctly provided evidence of the SEP, some of the students who answered incorrectly also provided some evidence of the SEP. Overall, the patterns suggest that the item elicits and discriminates based on the DCI/CCC.

The TAPS findings indicated that the targeted SEP was not required by the item (Table 5.8). While there is evidence from the cognitive lab that students were engaging with Claim, Evidence, and Reasoning structures, the item type forces students to do so. However, the TAPS analysis found that an additional SEP was elicited by the item: Analyzing and Interpreting Data. The TAPS findings indicate that the DCI is required by the item. The cognitive lab data also supports this finding. The TAPS analysis determined the CCC was not measured whereas the cognitive labs were coded such that the DCI and CCC were indistinguishable.

For these three items (Grade 5, Items 1 and 3; Grade 8, Item 5), the cognitive lab data provides evidence that the items were able to elicit students' understandings or abilities related to

67

the intended dimensions. In addition, the items were able to discriminate between students who answered correctly versus those who did not based on the DCI or DCI/CCC dimension. While evidence of the SEP and CCC was found in the cognitive lab data, there was not clear discrimination on these dimensions between students who answered the items correctly and those who did not. Therefore, I cannot claim that the items discriminated on any dimension other than the DCI or DCI/CCC.

Non-Discriminating Items

Evidence from the cognitive lab data for Grade 5, Items 2, 4, and 5 suggests that these items elicited some dimensions but did not discriminate on any dimension.

Items that did not elicit evidence of the DCI

Item 2

Eyes collect information about the world in the form of light. Which statement **best** describes how this information is processed?

(a) Light is sensed by the brain and then transferred to the eyes.

(b) Light is processed in the eyes, allowing the object to be seen immediately.

(c) The eyes reflect light back to the object as the information about the object is processed.

(d) The eyes have structures that sense light, and then the information is sent to the brain to be processed.

Figure 5.4. Grade 5 Item 2.

Table 5.9

Grade 5 Item 2 Coding Patterns

Item 2				
Student[a]	DCI		CCC	
	Sns	Brn	Lnk	Seq
54				
55				
56			X	X
58				
51			X	
52			X	
53			X	X
57			X	X
59			X	
510			X	

[a] White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.

Item 2 (Figure 5.4) elicited but did not discriminate for the CCC based on my criteria for elicitation and discrimination. The keyed response, D, foregrounds the eyes as sense receptors that allow light information to be processed by the brain. The cognitive labs showed no evidence of use of the DCI, and only one of four students who chose the correct response provided evidence of the CCC. All of the students who chose the incorrect response (N=6) provided evidence of using the CCC. Some of the students' responses provided insight into their misunderstanding of the term "sense" in the keyed answer option. Two examples from students who chose the correct response are as follows: "I don't really think that your eyes can sense

light. But if light is processed in your eyes that I feel like it would... I think it would more be B because you can't really sense light. You can't sense when the light is going to turn on and when it's going to turn off" (p. 59) and "shut your eyes are probably sensed what the thing is or either knows" (p. 52). These students used the word sense to mean predict or know, which likely informed their choice of an incorrect answer option.

The four students who answered correctly referenced "everything goes to your brains" (p. 54); "my teacher showed an example" (p. 55); "it [brain] produces a picture and then sends it to the eyes" (p. 56); and "you always see things right away" (p. 58). However, these responses provide little insight into the students' understanding of the DCI.

Table 5.10

Grade 5 Item 2: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate.	The information in the scenario is not necessary to answer the question. The DCI is not used in service of sensemaking. The stated CCC is not measured in the item and very little reasoning is required. The item did not require sensemaking because the response is very close to the DCI and could be rote. Overall, the item does not assess what it is intended to assess.

TAPS analysis (Table 5.10) concluded that the DCI is required to answer the question, whereas the cognitive labs found no evidence of the DCI. Additionally, the TAPS analysis concluded that the CCC is not required, where the cognitive labs provided inconclusive data.

Item 4

This question has two parts.

Part A

After a while, the lights come back on in the classroom. How will the students' pupils **most likely** change?

(a) The students' pupil diameters will increase.

(b) The students' pupil diameters will decrease.

Part B

Which statement **best** explains the change in pupil diameter described in Part A?

(a) Pupil diameter increases when there is low light.

(b) Pupil diameter increases when there is bright light.

(c) Pupil diameter decreases when there is low light.

(d) Pupil diameter decreases when there is bright light.

Figure 5.5. Grade 5 Item 4.

Table 5.11

Grade 5 Item 4 Coding Patterns

Item 4				
Student[a]	**DCI**	SEP		CCC
	Stf	Ev	Rsn	Lnk
52		X		X
54		X		X
55		X		X
56	X		X	X
58	X	X		X
59	X	X	X	X
51				X
53			X	X
57	X	X		X
510	X		X	X

[a] White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.

Item 4 (Figure 5.5) elicited but did not discriminate for the DCI, the SEP and the CCC based on my criteria for elicitation and discrimination. The item was an evidence-based selected response item designed for students to choose a response in Part A and then choose a response in Part B that supports their choice in Part A. All of the students' responses to Item 4 provided evidence of the cause-and-effect code for linking, so, based on my criteria for elicitation, I can claim that the item elicits, but does not discriminate on the basis of, the CCC. It also does not discriminate based on the DCI or the SEP because half of students who answered the item correctly and half of those who did not provided evidence of the DCI. Additionally, all but one

student provided evidence of the SEP Because the codes for the SEPs and CCCs were designed to be independent of the DCI, students could use evidence from the item or previous knowledge or experiences regardless of the connection of the evidence to the item stem. Further, if the students linked their reasoning statement to the evidence they provided, this was coded as reasoning for the SEP. Only two students chose the incorrect response for both Part A and Part B. Therefore, a different item type may have provided the opportunity for some students to gain more credit for their knowledge. If the items were designed to provide students with partial credit for Part A and Part B or designed in a way to be two separate items, we would better be able to capture what aspects of the items students are successful with Additionally, four of six students whose response was coded for "SEP-evidence" used prior knowledge or experiences as their evidence instead of the data given in the stimulus. The item was able to elicit of the CCC Cause and Effect. It did not elicit or discriminate for the DCI or SEP.

Table 5.12

Grade 5 Item 4: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP is required to answer the question.	The SEP is different from the identified SEP. It is measured below grade-level and is not used in service of sensemaking because students are expected to read the graph but do not have to apply any ideas from it. The stated DCI is not measured. The stated CCC is not measured. Overall, the item does not assess what it is intended to assess.

The TAPS analysis (Table 5.12) concluded that Analyzing and Interpreting Data was the SEP that was assessed in this item but that it was assessed below grade level. The TAPS analysis does not support any claims about the intended SEP Arguing from Evidence. The TAPS findings also show that the DCI and CCC is not measured by the item. The cognitive lab findings about this item are inconclusive, however, they do suggest that this item elicits (but does not discriminate) for the CCC and the SEP.

Item 5

Based on all of the information, complete the scientific explanation about the effect of light on pupil diameter.

Drag into the chart **two** evidence statements and **one** reasoning statement that support the claim for the effect of light on pupil diameter.

?

Claim	The pupil diameter changes with different amounts of light.
Evidence	
Reasoning	

Evidence Statements

The diameter was largest in the lowest light and smallest in the brightest light.

The diameter was largest in the brightest light and smallest in the lowest light.

The diameter of the pupil increased as the light increased.

The diameter of the pupil decreased as the light increased.

Reasoning Statements

When there is less light, the pupil gets bigger to let in more light.

When the pupil is smaller, it lets in more light so a person can see better in less light.

When there is bright light, the pupil lets in more light so a person can see better.

Click To Respond

Figure 5.6. Grade 5 Item 5.

Table 5.13

Grade 5 Item 5 Coding Patterns

Item 5				
Student[a]	DCI	SEP		CCC
	Stf	Ev	Rsn	Lnk
51				X
52				
56	X		X	X
57				
58	X		X	X
59		X	X	X
53				X
54		X	X	X
55				
510				X

[a] White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.

Item 5 (Figure 5.6) elicited but did not discriminate for the CCC and did not elicit for the DCI or the SEP based on my criteria for elicitation and discrimination. The item used a drag and drop functionality and required the students to choose two evidence statements and one reasoning statement to support a given claim. Of the four students who answered incorrectly, two of them chose both correct evidence statements but the incorrect reasoning statement. The other two students who answered incorrectly choose one correct evidence statement. Like Item 4, a different item type may have provided students the opportunity to gain more credit for their

knowledge. The CCC code is present across most students' responses but does not clearly delineate students who chose the keyed response from those who did not.

Table 5.14

Grade 5 Item 5: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP is required to answer the question. A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking. The students must connect the data and their understanding that light is needed to see. Multiple dimensions are used together. The item measures what is intended.	The SEP is measured below grade-level. The CCC is not measured.

The TAPS analysis concluded that the SEP and DCI were required to answer the question. However, the SEP was measured below grade level. Additionally, the TAPS findings showed that the CCC was not required by the item. The TAPS findings conflict with the findings from the cognitive labs. The TAPS findings concluded that the DCI was needed for students to respond to the item. The cognitive lab evidence does not support this claim in that only two students provided responses coded for the DCI. Furthermore, there is cognitive lab evidence that the CCC is elicited by the item.

Grade 8

The evidence from the cognitive lab data for Grade 8, Items 1, 2, 3, and 4 suggests that these items elicited some dimensions but did not discriminate on any dimension.

Item 1

This question has two parts.

Part A

Use the data table to complete the following statement.

The students can tell that a chemical reaction involving iron
[_____ ▼] because a new substance [_____ ▼] form
overnight.

Part B

Choose one set of properties that **best** supports the completed
statement in Part A.

(a) density and color

(b) color and volume

(c) volume and texture

(d) texture and mass

(e) mass and density

Figure 5.7. Grade 8 Item 1.

78

Table 5.15

Grade 8 Item 1 Coding Patterns

Item 1			
Student[a]	**DCI**	**SEP/CCC**	
	Chp	1	2
83	X	X	X
89			
81			
82		X	X
84		X	X
85		X	X
86			
87		X	X
88		X	

[a] *White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.*

Item 1 (Figure 5.7) elicited but did not discriminate for the SEP/CCC and did not elicit the DCI based on my criteria for elicitation and discrimination. The item was designed as a two-part item in the form of an evidence-based selected response. Part A requires students to choose from two drop down menus to complete the statement correctly. Part B requires students to choose the properties that would support their response in Part A. The keyed response, A, is supported by the table provided in the stimulus. This item provided cognitive lab data that is hard to make sense of because only two students chose the correct response and one displayed knowledge of the SEP/CCC and the DCI, while the other student provided no evidence of

knowledge of either dimension. The majority of students who got the item wrong used the

SEP/CCC.

Table 5.16

Grade 8 Item 1: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required. Overall, the item does assess what it is intended to assess.	The stated CCC is not measured.

The TAPS analysis indicated both the DCI and SEP were necessary to respond to the

item (Table 5.15), but the item does not assess the SEP element at grade level. The TAPS

analysis concluded that the CCC was not required. This is in contrast with the cognitive lab

findings that the suggest the DCI was not elicited. Because the coding for students' responses did

not distinguish between the SEP and CCC it is difficult to determine whether the cognitive lab

data and the TAPS data are in agreement.

Item 2

Complete the statements below.

The final mass of the material on the dish the next day is
[_____ ▼] the initial mass of the material.

This could happen if
[_____ ▼] the
environment.

Figure 5.8. Grade 8 Item 2.

Table 5.17

Grade 8 Item 2 Coding Patterns

Item 2			
Student[a]	**DCI**	**SEP/CCC**	
	C	1	2
81		X	
82		X	
83		X	X
84		X	X
85		X	
86		X	
87		X	
89		X	
88			

[a] *White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.*

Item 2 (Figure 5.8) elicited but did not discriminate for the SEP/CCC and did not elicit the DCI based on my criteria for elicitation and discrimination. The item requires students to choose from options in two drop down menus to explain the result of the experiment presented in the stimulus. All students who answered the item correctly provided some evidence of SEP/CCC knowledge, and the student who answered the item incorrectly did not. While this could be considered clear data to support that the item discriminates on the SEP/CCC dimension, only one student chose the incorrect response. Therefore, there is not enough data to determine if the item discriminates on that dimension.

Table 5.18

Grade 8 Item 2: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required, however not at grade level. Overall, the item does assess what it is intended to assess.	The SEP is not engaged at grade level. The stated CCC is not measured.

The TAPS findings disagree with the cognitive lab findings. The TAPS findings concluded that a substantial portion of the DCI was required to answer the question. Additionally, TAPS found that the CCC is not needed to answer the question.

Item 3

This question has two parts.

The students decide to develop a model to explain why the iron appeared to change overnight. To complete the model the students need to include:

- iron atoms from the hand warmer
- oxygen atoms from the air
- the final substance they observed

Part A

Complete the student model by dragging the appropriate number of atoms from the key into the model.

| Initial Iron Atoms | Oxygen in the Air | Final Substance | Key |

Click To Respond

Part B

Select one limitation of the model shown in Part A.

(a) The model does not show conservation of matter.

(b) The model does not show the color change of the final substance.

(c) The model does not show how the atoms are organized in the final substance.

Figure 5.9. Grade 8 Item 3.

83

Table 5.19

Grade 8 Item 3 Coding Patterns

Item 3		
Student[a]	DCI/CCC	SEP
	Atc	Lim
81	X	
82	X	X
85	X	X
86	X	
87	X	X
88	X	X
83	X	
84		X
89	X	

[a] *White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.*

Item 3 (Figure 5.9) elicited but did not discriminate for the DCI/CCC and SEP based on my criteria for elicitation and discrimination. Part A of Item 3 uses a drag and drop item type to allow students to complete an atomic-level model of the chemical reaction taking place between iron and oxygen. Part B requires students to think about modeling as a practice and choose a limitation of the model they completed in Part A. Students who chose the correct response used the DCI and CCC. Students who chose the incorrect response also provided some evidence of DCI or CCC knowledge.

Table 5.20

Grade 5 Item 3: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required. Overall, the item does assess what it is intended to assess.	The application of the DCI is at a low level. The stated CCC is not measured.

The TAPS findings indicate that the DCI is required by the item but at a low level (Table 5.17). Additionally, the TAPS determined that the SEP was required to answer the question, however, the CCC is not measured by the item.

Item 4

Figure 5.10. Grade 8 Item 4.

Table 5.21

Grade 8 Item 4 Coding Patterns

Item 4			
Student[a]	DCI	SEP/CCC	
	E	1	2
81		X	
82			
83	X	X	X
85		X	X
86		X	
88		X	
89		X	
84		X	
87		X	

[a] *White shading indicates students who chose the correct response. Grey shading indicates students who chose the incorrect response.*

Item 4 (Figure 5.10) elicited but did not discriminate for the SEP/CCC and did not elicit the DCI based on my criteria for elicitation and discrimination. The item is a hot spot item type. Here, the students select two sentences from the similarities column and two from the differences column to compare the data provided in the stimulus. All but one student provided evidence of the SEP/CCC regardless of whether they chose the correct response or the incorrect response. Additionally, only one student provided any knowledge of the DCI.

Table 5.22

Grade 8 Item 4: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. The SEP and CCC are both engaged in this item. Multiple dimensions are used together and sensemaking or problem solving is required.	The SEP and CCC are not engaged at the appropriate grade level. The DCI is not measured. Overall, the item does not assess what it is intended to assess.

The lack of DCI evidence in the cognitive labs is aligned with the TAPS findings (Table 5.20). While the TAPS analysis found that both the SEP and the CCC were measured by the item, they were assessed below grade level.

The pattern tables created for each item provided an overarching look at how the students' verbalized thinking was mapped onto the dimensions intended to be assessed. Many of the patterns were hard to make sense of and provided no clear pattern for what dimensions students were using to choose the correct response versus the incorrect response. Therefore, it is not possible to determine if the items discussed in this section are providing the intended discrimination. There may be some factors that impacted the items' ability to discriminate among students. These factors will be discussed in the next section and the next chapter.

Item Cluster Analysis

Table 5.23

Grade 5 Summary Table

	Item 1	Item 2	Item 3	Item 4	Item 5
PS4.B	E/D	-	E/D	-	-
LS1.D	-	-	E/D	-	-
LS1.A	-	-	-	E	-
Modeling (SEP/CCC)	-	-	E	-	-
Argumentation	-	-	-	E	-
Cause and Effect	E	-	-	E	E

E = Elicits, D = Discriminates, - = no evidence

The cluster writing team was given the task to write a cluster that assessed each element of the Life Science topic bundle "Structure, Function, and Information Processing" (NGSS, Lead State, 2013) at Grade 4. Some of the requirements for this cluster were that each item was to be two-dimensionally aligned and at least one item was to be three-dimensionally aligned.

For the Grade 5 cluster, all the targeted DCIs across the cluster were elicited and discriminated by at least 1 item. Additionally, all the SEPs and CCCs were elicited but did not discriminate among students who answered correctly and those who did not. There was also evidence of embedded dimensions in three of the five items (Table 5.23).

Table 5.24

Grade 8 Summary Table

	Item 1	Item 2	Item 3	Item 4	Item 5
PS1.B.1	-	-	-	-	-
PS1.B.2	-	-	E	-	-
PS1.B.3	-	-	-	E	E/D
ETS1.B	-	-	-	-	-
ETS1.C.1	-	-	-	-	-
Analyzing Data	E	E/D	-	E	-
Modeling (SEP/CCC)	-	-	E	-	-
Constructing Explanations	-	-	-	E	E
Designing Solutions	-	-	-	-	-
Patterns	-	-	-	E	E
Energy and Matter	-	-	-	-	-

E = Elicits, D = Discriminates, - = no evidence

For the Grade 8 cluster, only one of the targeted DCIs was elicited and discriminated between students got the item correct and those who did not. Additionally, only one SEP was elicited and discriminated between students who got the item correct and those who did not. All

other targeted dimensions, with the exception of DCI PS1.B.1, were elicited but did not discriminate among students who answered correctly and those who did not.

Summary

The data and analysis presented in this chapter serve to answer the research questions:

1) To what extent do the clusters developed using Michigan Cluster Development Process align with the Michigan K-12 Science Standards?

    1a) To what extent do these items elicit and discriminate for the intended dimensions?

As noted in Chapter 4, I defined elicitation as the item's ability to provide opportunity for students to use knowledge of a dimension regardless of whether they choose the correct or incorrect response. From the coding pattern indicating elicitation is that the majority of students have a code for the dimension independent of their answer choice or that students who got the answer correct have codes for the dimension while the students who chose the incorrect response did not, linking elicitation and discrimination. I defined discrimination as the item's ability to separate students who know a particular dimension by choosing the correct answer from those who do not know that dimension evidenced by choosing the wrong answer. From the coding perspective, the pattern in codes is that students who got the item correct were more likely to have the code for that dimension than student who got the item incorrect.

The Grade 5 and Grade 8 Item Clusters proved to have some value for eliciting and discriminating students based on the DCI dimensions. Three of twelve total items provided clear cognitive lab data to support this claim. Other items across the clusters were able to elicit students' knowledge of the three dimensions but did not meet the discrimination criteria.

CHAPTER 6: DISCUSSION AND CONCLUSION

Discussion of Findings

The purpose of this research was to determine the extent to which the State of Michigan science item clusters elicit evidence of each of the three dimensions of the Michigan K-12 Science Standards. This chapter includes a discussion of major findings related to the Grade 5 and 8 clusters, including how alignment is defined for NGSS and how grade-level sophistication is considered. Also included is a discussion of the exclusion of the Grade 11 data. This chapter concludes with limitations of the study, implications for large-scale assessment, areas for future research, and a summary.

This chapter contains discussion and future research possibilities to help answer the research questions:

1) To what extent do the clusters developed using Michigan Cluster Development process align with the Michigan K-12 Science Standards?

    1a) To what extent do these items elicit and discriminate for the intended dimensions?

Overall Findings

The Grade 5 cluster findings revealed that all of the intended dimensions were elicited across the items in the cluster. As seen in Table 5.5, most of the elements were elicited in some way. However, only two of the items (Items 1 and 3) discriminated based on the intended DCI alignment. The Grade 8 cluster findings (Table 5.18) revealed that all but one DCI were elicited by the items. Two items, Item 4 and Item 5, failed to elicit the DCI. One item (Item 5) discriminated based on the DCI.

Overall, across Grade 5 and Grade 8, items seemed to be most able to elicit DCIs. Items that elicited CCCs had closely related DCIs or SEPs, making it difficult to tease them apart and make claims about discrimination based on CCCs. Because of this close alignment, items also were not able to discriminate based on the SEPs or CCCs between students who chose the correct response from those who did not. Based on these overall findings, I will now examine three discussion points: (a) Tensions associated with alignment for NGSS large-scale assessments; (b) Challenges of large-scale assessment of high school NGSS and rationale for excluding Grade 11 data in this study; (c) Limitations of this study; and (d) The implications of this study for various aspects of large-scale assessment processes.

Alignment Tensions

As a reminder, I define item alignment as the item's ability to elicit evidence that students used the intended dimensions and to discriminate between students who chose the correct response versus those who chose the incorrect response. There are two main alignment tensions that were identified in this research: embedded dimensions and dimensional density. The following sections will describe each tension, summarize the evidence of the tension found in the data analysis, and provide considerations and recommendations for dealing with the tensions from a large-scale assessment development perspective.

<div align="center">

Tension 1: Embedded Dimensions

</div>

In this section, I present the first main tension that arose when analyzing these items, which I have termed embedded dimensions. I define embedded dimensions as a set of dimensions used together in the development of an item where two dimensions are so closely related it is difficult to separate them out when examining students' responses to the items. This idea of embedded dimensions came up in two main ways: (a) in the cognitive lab data; and (b) in

the structure of the codes. After discussing each of these instances, I make sense of these two categories by examining the actual language of the dimensions and by using the TAPS analysis. I will describe each of these ways that have been revealed through this study and then discuss potential reasons.

Embedded Dimensions in Cognitive Lab Data

For several items in Grade 5, the cognitive lab data illuminated the issue of embedded dimensions. This was evident because every time a students' response was coded for one dimension it was also coded for another dimension – suggesting that the two dimensions were correlated in some way. For example, in Grade 5, Item 1, when a student's response was coded for one more aspect of the DCI, it was also coded for both aspects of the CCC. This suggests that there was some type of relationship between the DCI and CCC. This pattern also occurred for Grade 5, Item 3, where students' responses that were coded for a DCI also were coded with the SEP/CCC code and for Grade 5, Items 4 and 5, where students responses that were coded for the DCI were also coded with the SEP and the CCC. These patterns indicate some dependency among the dimensions when students were answering the questions. I suggest that these patterns may be explained using the concept of embedded dimensions. However, these patterns were not present in the Grade 8 data. One reason that I may have found these patterns of embeddedness in Grade 5, but not in Grade 8 is because of the way in which dimensions were chosen during item design. I discuss this possibility after examining how embeddedness occurred in the coding structure.

Embedded Dimensions in Coding Structure

The other place where issues of embeddedness occurred was when attempting to develop a coding scheme for examining the cognitive lab data. As described in my methods chapter, I

attempted to develop codes for each dimension that were orthogonal to each other, meaning that coding for one dimension was independent from coding for the other dimensions. However, there were two instances when this orthogonal coding was not possible because two dimensions were indistinguishable and required a combined code. One example was in Grade 5, Item 3 which was designed to assess a DCI, the SEP of "Developing and Using Models" and the CCC "Systems and System Models." When working to unpack these two dimensions (see Table 6.1), it became clear that any evidence that a student would provide of the SEP would also count as evidence for the CCC. There was no way to disentangle these dimensions and thus, I coded student responses as either having both the SEP and CCC (i.e., SEP/CCC code) or neither of the two. This is evidence of embedded dimensions because it was impossible to separate out the dimensions when looking at students' response.

Table 6.1

Grade 5 Item 3: Embedded Dimensions

SEP: Developing and Using Models: Develop and/or use models to describe and/or predict phenomena.	CCC: Systems and System Models: Describe a system in terms of its components and their interactions.
Develop and/or use a model to describe	Describe a system
The components (i.e., images) in the model (as presented in the item)	Components (given in the item)
The arrows in the model (as presented in the item)	Interactions (the arrows in the item)

Embedded Dimensions in Language of the Dimensions

In this section, I look at the language of the dimensions to help explain embedded dimensions and examine how the external analysis (TAPS) was unable to capture embeddedness

94

in their review of the items. As described above, there were instances when both students'

responses to certain items and the coding of certain items revealed a relationship between certain

dimensions to which items were aligned. To further examine this phenomenon of the embedded

dimensions trends, I examined the language of the dimensions to look for relationships between

the dimensions to which the items were aligned. As discussed above, patterns in student

responses to Grade 5, Item 1 appeared to show embeddedness. For this item, the intended

alignment is to DCI: PS4.B: An object can be seen when light reflected from its surface enters

the eyes; and CCC: Cause and Effect: Cause and effect relationships are routinely identified. I

looked for a relationship in the language of these two dimensions (see Table 6.2). In this

example, I found that the language of the DCI included cause and effect relationships intended

by the CCC. The DCI states, "An object can be seen when light reflected from its surface enters

the eyes." Closer examination of this DCI reveals that [An object can be seen] is the effect,

[when light is reflected from its surface] is the cause, and [enters the eyes] is an intermediate

event. This analysis assisted me in determining the relationship between the intended alignment,

the item design, and the responses that students may provide. Therefore, when students answered

the item correctly, it was difficult to tease apart students' use of cause-and-effect thinking as

separate from the cause-and-effect relationship set forth in the DCI.

95

Table 6.2

Grade 5 Item 1: Embedded Dimensions

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.	CCC: Cause and Effect: Cause and effect relationships are routinely identified
An object can be seen	Effect
when light reflected from its surface	Cause
enters the eyes	Sequence of Events

TAPS Analysis and Embedded Dimensions

When comparing the TAPS reviews (see Appendix H and the findings chapter) for items that I claim have include embedded dimensions, it is clear that there were different perspectives on alignment determinations. This relationship between dimensions (i.e., embeddedness) is likely the cause of the disagreement between the cognitive lab findings and the TAPS conclusions (Appendix H). For example, returning to Grade 5, Item 1, which I found elicited both the DCI and CCC (but only discriminated on the DCI), the TAPS analysis found that the DCI was required to answer the question but that the CCC was not. These results seem to disagree with a portion of the cognitive lab data in that the CCC was coded for every student who answered the item correctly. One reason why the TAPS review may have made this determination is because of the embedded nature of the CCC within the DCI. To answer the item correctly, there is no indicator of cause-and-effect reasoning separate from DCI understanding, the knowledge that the item elicits is the same for the DCI and CCC. Therefore, distinguishing the CCC from the DCI is impossible. As a result, we would not expect to see unique evidence of the CCC and DCI in

students' correct responses. However, if we consider students' incorrect responses, there were instances where the CCC was elicited, indicating that while this dimension did not discriminate students who got the item correct from those who did not, the item did provide students the opportunity to use the CCC (i.e., cause and effect reasoning) when interacting with the item.

There were several other instances where the TAPS analysis found that an item was not aligned to one dimension, but I found that, in fact, the item elicited that dimension, but that it was embedded in another dimension. Thus, it is important to clarify the role of embedded dimensions in large-scale assessment development.

Discussion of Embedded Dimensions

When there are embedded dimensions, determining whether an item is eliciting unique evidence for both dimensions is difficult. As mentioned earlier, traditional large-scale assessment items only had to measure one idea at a time—science content was often assessed separate from "inquiry skills" like analyzing data (Alonzo & Ke, 2016). The NGSS requires multiple dimensions be used together to figure out phenomena and solve problems. So, an item that claims to measure two dimensions but provides the same evidence for both dimensions may be problematic to validate. So, in the case of Grade 5, Item 1, which is designed to measure a DCI that includes a cause-and-effect statement and the CCC of cause and effect, how do you know if the student is using the DCI, the CCC, or both?

Alonzo and Ke (2016) point out:

Thus, it is not enough to ask whether a particular assessment includes NGSS content and practices (i.e., to match up the assessment framework and/or items from a particular assessment with the disciplinary core ideas and practices from the NGSS). Unless

97

students are asked to coordinate the two in explicit and meaningful ways, the assessment

does not integrate content and practices as intended by the Framework/NGSS" (p. 137).

This demonstrates the conundrum with NGSS-aligned large-scale assessment. We do not want to

claim that a test is aligned just by checking off the dimensions used and not considering the way

in which the dimensions are coordinated. However, when dimensions that perhaps "should" be

used together in responding to an item have an embedded nature, the validation of the items is

challenging.

There are two ways of dealing with this tension of embedded dimensions. On one side of

the tension, alignment on large-scale science assessment could be defined to mean that each item

can claim alignment to multiple dimensions even if the dimensions are so closely related the

evidence is indistinguishable. In this study, item writers did not consider issues of embeddedness

and the thought that they were able to develop multidimensional items even though the evidence

for each dimension was not different. Because this was the first round of item writing with the

MSS, the issue of embedded dimensions was not yet illuminated for writers to consider. If large-

scale assessment were to come down on this side of the tension, it would allow large-scale

assessment developers to design items that are multidimensional without concern for the

embeddedness of those dimensions.

On the other side of the tension, alignment on large-scale science assessment could be

defined to mean that claims about students' achievement on the basis of the items should be

supported with unique evidence for each dimension. If the same evidence counts for multiple

dimensions, then the item cannot claim to be multidimensional. This appears to be the approach

taken by the external content reviewers conducting the TAPS analysis. The tension was apparent

in the disagreement between the results of my data analysis and the TAPS analysis. For example,

Grade 5 item 1 was designed to be aligned with both the DCI and the CCC and I found that it elicited both dimensions, but only discriminated for the DCI. However, the TAPS analysis concluded that the CCC was not assessed by this item.

The findings of this research illustrate some of the ways item evidence can "count" for more than one dimension. For example, some students were engaged in cause and effect thinking but, due to the embedded dimensions of the CCC and the DCI, it is difficult to make a claim about each dimension separately. On large-scale assessments, developing items with unique evidence of each of the dimensions may require the development of more sophisticated item types. For example, using constructed response (CR) items could provide opportunities for students to demonstrate how and when they are intertwining the dimensions. The drawback is that scoring CR items takes human, temporal, and financial resources that are not often allocated to state science assessments. Therefore, until the resource allocation changes, large-scale assessment developers must partner with psychometricians to develop validity arguments that support alignment to multiple dimensions even if the dimensions are so closely related the evidence is indistinguishable.

Recommendations

While this may not cohere with the assessment philosophy of some developers, my recommendation is to reduce the footprint of large-scale assessments and provide more resources to develop other parts of the assessment system that can more readily accommodate assessment items that provide unique evidence of multiple dimensions and build teachers' capacity for assessment development, analysis, and the associated instructional shifts to build on students' undeveloped scientific understandings. If the ultimate goal remains to improve student learning outcomes, then spending precious resources on the classroom and formative assessment end of

the Balanced Assessment System for Science (Figure 1.1) would be most beneficial (e.g.; Black & Wiliam, 1998; Decristan, et al., 2015).

Tension 2: Dimensional Density

The second alignment tension focuses on dimensional density, which I define as the complexity of each of the dimensions and the extent to which that complexity can be measured in state-level assessments. Because each of the dimensions is distinct, dimensional density can be thought of separately for each of the dimensions. It became clear through the cognitive labs, examining the language of the dimensions, and comparison to the TAPS analysis that there are features of the NGSS and MSS that are difficult to assess within the constraints of large-scale assessment. Specifically, there are considerations in each of the dimensions (SEP, DCI, and CCC) that may pose difficulties for large-scale assessment. In this concept of dimensional density, I think of the SEPs and the DCIs as adding "mass" to the assessment challenge. The CCCs add "volume" to the assessment challenge. Therefore, in the following sections, I am using the concept of dimensional density to describe these features and discuss their implications for science assessment.

Dimensional Density of SEPs

The items did not always reflect the sophistication in SEPs that was expected according to the standards. For example, Grade 8 Item 2 was aligned with Analyzing and Interpreting Data, which at this grade level means students should be interacting with large data sets, using data to identify causal and correlational relationships, or look across data sets to determine similarities and differences in findings (Appendix F, NGSS Lead States, 2012). However, in this assessment, students only needed to determine that the value provided for the mass of the substance increased.

The NGSS were developed based on learning progressions (NRC, 2012), which indicate the how the sophistication of the SEPs grows throughout the K-12 educational experience. In the Grades 5 and 8 cluster, many of the items were determined to be "below grade-level" by the TAPs analysis. For example, the TAPS analysis found the SEP, Analyzing and Interpreting Data, in both Item 2 and Item 4 in Grade 8 to be below grade level. Ideally, each item would elicit and discriminate based on the level of knowledge, skills, and abilities aligned with expectations for their grade level. However, designing a forced choice assessment item to meet these criteria is considerably more challenging than creating items with lower SEP sophistication. This adds "mass" to the assessment design challenge.

If assessment items are not crafted to elicit the SEPs at the appropriate grade-level sophistication, students do not have the opportunity to demonstrate what they know and can do in science, and the assessment cannot support the claims about students' SEPs. While it is not impossible, building and scoring items that measure complex reasoning using SEPs will require more research and resources than is available at the state level.

Dimensional Density of DCIs

The DCIs in the NGSS were crafted with varied depth and breadth. Some DCIs are very specific and narrow. For example, PS4.B: An object can be seen when light reflected from its surface enters the eyes, is very specific in nature and only applies to a few phenomena. On the other hand, LS1.A: Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction, is a very broad DCI that can be applied to any number of organisms and phenomena. This aspect of dimensional density also adds "mass" when crafting assessment items. These two DCIs are very different in grain size but may receive the same amount of attention on an assessment. Therefore, a question like Grade 5

Item 1 may be able to elicit and discriminate information about the DCI (PS4.B), but a question like Grade 5 Item 4 cannot cover all aspects of the DCI (LS1.A).

This same issue was also present when examining the macro to micro mechanistic reasoning that is required by the DCIs in the middle school chemical reactions topic bundle. The Framework calls for students in middle school to be able "to relate patterns to the nature of microscopic and atomic-level structure – for example, they may note that chemical molecules contain particular ratios of different atoms" (NRC, 2011, p. 86). While the options in the force-choice responses included the word, "atoms," there is no clear indication that the students understood that the atomic make-up of substances determined their characteristic properties as is required by the DCI. For example, Student 83 said there was "more stuff inside of there" and Student 84 said "because then if it combines with matter, it could possibly get heavier." Both students illustrate that their thinking is still on the macroscopic level, and they are not providing evidence of relating to microscopic or atomic-level structures. Therefore, the items analyzed for Grade 8 provided little to no opportunity for the students to engage with this DCI at grade level.

Dimensional Density of CCCs

The ubiquitous nature of the CCCs produces challenges for capturing students' knowledge and use of CCCs on large-scale assessments. The CCCs are to be woven throughout science learning and have been described as a bridge across disciplines, a lens to investigate phenomena, the "grammar rules" for science, and many other metaphors (Fick, 2017). While the usefulness of the CCCs is not in question, our ability to measure students' use of CCCs is. The "volume" at which the CCCs play a role in developing science assessments presents the third aspect of dimensional density. In this study, even without canonical scientific understanding,

students were able to use the CCCs. However, using the CCCs and choosing the wrong answer was not information that this large-scale assessment could capture outside of the cognitive labs.

For example, the Grade 8 cluster analysis revealed while Scale, Proportion, and Quantity is not identified as a CCC in the Chemical Reactions topic bundle (Grade 8), understanding this CCC is important when moving between macroscopic observations of bulk quantities of substances and explaining their properties using microscopic reasoning with respect to atomic make-up (Chesnutt et al., 2018). Therefore, there seems to be at least one CCC (and I would assume there are more) that are essential for students to know and understand to succeed in science that are not identified as part of the assessed topic bundle. This poses a challenge for large-scale assessment as the construct needs to be defined to measure it.

Recommendations for Dealing with Dimensional Density

Developers need to wrestle with whether all dimensions need to be at grade-level or whether certain dimensions be below grade level to either highlight another dimension or to serve as an on-ramp for students' interaction with the task. If an entire assessment or cluster is too difficult for a portion of the student population, then the assessment or cluster is useless in providing information about what students know and can do in science. Using on-ramp items at the beginning of clusters can provide access to a larger number of students. Designers may also choose to use a less sophisticated form of an SEP to foreground another dimension in an item, such as a DCI, so that the evidence elicited by the item is focused on the DCI. These can be considered legitimate reasons for lowering the sophistication of one or more dimensions. One consequence of this may be that students do not have the opportunity to show the sophistication of their knowledge on SEPs across the assessment, yet the assessment designers can still make claims about students' ability to do the SEPs in their specific grade band. Another stance might

be that SEPs provide opportunities for students at all levels. Therefore, if scaffolding is needed to provide a range of difficulty on a large-scale assessment, the sophistication of the DCI or CCC should be varied, not the SEPs. Since the SEPs are written to mirror the practices that scientists and engineers do in their professions, it is very difficult to develop forced choice items that require students to engage in the complex processes involved with the Science and Engineering Practices. Both on-ramping items and foregrounding dimensions in some items are legitimate reasons to design an item with below grade level dimensions.

One consideration for the "on grade-level argument" is the number of items across a task that have to be considered "on grade level" for the task to be deemed so. The Achieve Criterion indicates the "vast majority of items need to be grade level appropriate" (Achieve, 2017). Therefore, assessment developers are required to make a judgement call regarding the number of items across an assessment that require "on grade-level" knowledge and skills within the multidimensional argument they are making.

Yet another consideration is the scaffolded nature of the items and tasks. The idea of scaffolding comes into play when thinking about the cognitive complexity of the task (Achieve, 2019). Highly scaffolded items and tasks reduce the extent to which students are "doing science." For example, the items designed to elicit evidence of the SEPs Arguing from Evidence and Constructing a Scientific Explanation. Item 5 in both grades were designed to support claim, evidence, and reasoning responses from students. However, many of the students interacted with the statements as if they were interacting with a multi-select item type or true/false interactions instead of using the given phrases to create a scientific explanation. Even though the evidence from the cognitive lab found that both of these items elicit and discriminate for their respective DCIs, the TAPs findings indicated that the SEPs were below grade level. Therefore, what level

of scaffolding is appropriate for large-scale assessments for these SEPs to be considered "on grade level?" With the constraints of technology enhanced item types, there may be little more designers can do. However, this again highlights a need for constructed response items. Constructed response items can provide students the opportunity to show their abilities when it comes to both Arguing from Evidence and Constructing Scientific Explanations. Without constructed response items, the nature of technology-enhanced item types makes grade level appropriate SEPs difficult to attain.

Assessment developers could argue that the high school DCIs are so complex that it is impossible to assess them on a large-scale assessment. Therefore, assessment designers need to determine which portions of the DCIs will serve as a proxy for all the standards. This determination would provide a more focused set of ideas which the items could assess given the restrictions of large-scale assessment.

Another stance may be that determining which pieces of the DCIs should be assessed on large-scale assessment is akin to reducing the standards or choosing priority standards, which is a slippery slope and should be avoided at all costs. Indirectly narrowing the curriculum due to messaging from the large-scale assessment can have dire consequences for classroom instruction.

Exclusion of Grade 11

This research study was originally designed to examine cognitive lab data from Grades 5, 8 and 11. The data were collected for all three grade levels. After preliminary analysis was complete, three main issues, two of which relate to dimensional density, came to light with respect to the Grade 11 data, resulting in the decision to exclude the full analysis from this study: (a) The DCIs were so complex that the design of the forced choice items provided no evidence of DCI elicitation; (b) The restrictive nature of the available item types and design resulted in all

SEPs eliciting below-grade level sophistication; and (c) All the participants reported they had never had the opportunity to learn Earth Science in their high school coursework.

The DCIs become increasingly sophisticated throughout the K-12 progression. While this sophistication is important for students to understand the major disciplinary ideas in science, the depth of the DCIs poses a problem for large-scale assessment design and implementation. The grade 11 topic bundle initially examined was Earth Systems, which contains eight DCI elements. An example of one element is as follows:

> ESS2.C: The abundance of liquid water on Earth's surface and its unique combination of physical and chemical properties are central to the planet's dynamics. These properties include water's exceptional capacity to absorb, store, and release large amounts of energy, transmit sunlight, and expand upon freezing, dissolve and transport materials, and lower the viscosities and melting points of rocks. (NGSS Lead States, 2013)

Constructing one or two forced choice items to provide adequate evidence of students' understanding of this DCI is difficult, let alone for eight of these DCIs paired with two more dimensions. The preliminary cognitive lab data analysis revealed that none of the students (N=11) provided evidence of engaging any of the DCIs when responding to the items. Appendix I shows the full Grade 11 Cluster.

Like Grades 5 and 8, the Grade 11 cluster resulted in restrictive item types that did not provide students the opportunity to engage with the SEPs at grade level. Within the cluster, students are required to use the following SEPs: Developing and using Models, Planning and Carrying Out Investigations, Analyzing and Interpreting Data, and Engaging in Argument from Evidence. As is the case with the DCIs, the SEPs are expected to grow in sophistication

throughout the K-12 science experience. One example of how students are required to engage in an SEP is as follows:

> Planning and Carrying Out Investigations: Plan and conduct an investigation individually and collaboratively to produce data to serve as the basis for evidence and in the design: decide on types, how much, and accuracy of data needed to produce reliable measurements and consider limitations on the precision of the data (e.g., number of trials, cost, risk, time), and refine the design accordingly. (NGSS Lead States, 2013)

It is nearly impossible to construct an item, or a series of items, that allows students to demonstrate this practice given the current time and technology constraints on large scale assessment programs. Small pieces of the practice could be included on forced-choice assessments, which is what the item writers attempted to develop for this cluster. However, the cognitive lab data revealed that the students were not provided the opportunity to use this or any of the three other SEPs at a high school level.

The third issue presented in the Grade 11 data was that of Opportunity to Learn (Moss et al., 2008). At the end of each student's interview, I asked them a series of questions related to their high school science pathway (Appendix I). Among these were questions to learn which science courses they had been offered and which courses they had taken or planned to take throughout their high school career. All eleven students reported that they had not taken an Earth Science course, while only three of them reported having learned about the phenomenon related to Atmospheric Changes Over Time in any science course. This poses a problem for the reliability of the cognitive lab data. Without a range of learning experience with the science content, the data collected can be called into question.

As a result of this research, my recommendation is once again that we decrease the footprint of large-scale assessment, move valuable resources and assets into building teachers' capacity to build classroom assessment systems that allow for students to dig deeply into the NGSS dimensions and provide opportunities for teachers to analyze those assessment to build on students' existing knowledge. Additionally, the opportunity to learn Earth Science specifically in high school is something that Michigan schools need to continue to improve. In the recent past, most schools offered Biology, Chemistry, and Physics as the main science courses. Now, with the adoption of the MSS, schools are starting to weave Earth Science into more traditional high school science courses. Without the opportunity to learn Earth Science, the participants fell short of standards and assessment expectations, and this lack of knowledge and skills will affect their ability to be consumers of scientific information throughout their lives.

Limitations of the Study

The clusters used for this study were among those that resulted from the first year of development in 2016. Since then, the field of science assessment has grown with a wealth of information, tools, research, and criteria to help assessment developers better create science assessments (e.g., Achieve, 2018; Campbell et al., 2020; Clark, et al., 2017; Harris et al., 2019; Penuel et al., 2019). As these clusters are now five years old, many of the processes used to develop them have been iteratively improved over time. Additionally, due to the security of the state assessment, the only clusters allowed to be used for this study were those that were released to the public. A study that focused on operational (and therefore secure) clusters, may have different outcomes due to the layers of iteration the operational clusters undergo.

Another limitation of this work was the way the cognitive lab data was collected. Initially, the study was designed to mimic the cognitive lab protocols used by the Office of

Assessment and Accountability at the Michigan Department of Education (Appendix D). Using these protocols provided a structure for the researcher to collect the data; however, the data collection could have been enhanced if, instead of just using audio recording and field notes to document the cognitive lab, the researcher would have recorded the on-screen interactions between the student and the cluster along with the students' gestures and body language. This would have provided useful data regarding how students changed their answers throughout the interaction, what parts of the screen the students were paying attention to, and more in depth understanding of their explanations as is provided by gestures. Collecting this type of data would have provided information regarding the cognitive moves students were making as they were interacting with the assessment, the range of ways students communicate their understanding, and the extent to which the large-scale assessment captures that understanding.

Additionally, this research did not focus on the clusters holistically; rather, each item was analyzed individually. Further research is warranted to determine if there are advantages to using a holistic approach versus a disaggregated approach (Deverel-Rico & Furtak, 2021). The quality of the phenomenon was not considered in this study, but we know that the quality of the phenomenon and a task's ability to elicit multidimensional thinking from students relies heavily on the extent to which the phenomenon is problematized and presented in a way that elicits uncertainty from the students (Achieve, 2018).

Implications

The implications for this research center around the ways in which large-scale assessments are designed and implemented. There are three areas where large-scale assessment can learn from this research: (a) Design processes; (b) Products; and (c) Interpretation.

Large-scale Assessment Design Processes

A principled design approach, like the one used in Michigan, offers opportunities to design assessment tasks that elicit what students know and can do in science (DeBarger et al., 2016; Harris et al., 2019). However, these design processes have proven to be most effective in the design of classroom formative and summative assessments for NGSS. While large-scale assessment designers can borrow from these processes, there seem to be some unique features of large-scale assessment for NGSS that need to be considered. First, at any level of assessment, capturing and interpreting students' reasoning and sensemaking is difficult (Alonzo & Ke, 2016, Herman et al., 2007; Pellegrino, 2014). Add to that the constraints of large-scale assessment, including testing time, limited item type availability, and overwhelming number of students participating, designing an assessment for the NGSS three-dimensional standards becomes quite a challenge.

However, the findings from this research may provide more insight to inform design processes. First, one of the major findings supports an in-depth look at the tensions of embedded dimensions and dimensional density. The unpacking process used for the development of this assessment, derived from Harris et al. (2019), did not include consideration of what counts as evidence of dimensions when instances of embedded dimensions occur. One recommendation is to provide a scaffolded set of questions that would encourage cluster writers to carefully consider the pairings of dimensions chosen for each item and implications of the evidence provided by the item with respect to embedded dimensions. Finally, policy-capturing conversations (Aiman-Smith et al., 2002) must occur to determine how much of each dimension needs to be addressed in an item for it to serve as proxy for the whole dimension to account for grade-level determinations.

Large-scale Assessment Products

Clusters were thought to be the optimal way to assess the NGSS (NRC, 2014; National Academies of Sciences, 2017; SAIC, 2015). However, this research revealed that as the sophistication of core ideas and practices grow across the grade-levels, the format of the items within clusters provide less evidence of students' understanding of the ideas set forth in the standards. The practices require for students to "do" something, while force choice clusters only provide the opportunity for students to "choose" something. These two actions are not equal. Additionally, the cluster format requires a great amount of scripting in order for the stimulus and items to be accessible for all students. The degree to which scripting occurs impacts the cognitive complexity of the cluster and thereby impacts the opportunities for students to "do science" (Achieve, 2019). One recommendation would be to change the item type availability for large-scale assessments like those used in Michigan. Examples of innovations in large-scale assessment can be found in the NAEP-TEL and the PISA where students are required to engage in reasoning with science ideas as described in the Framework (Pellegrino, 2013). By designing item types that include simulations, animations, multimedia-based tasks, and open-ended response options, the large-scale assessment product can provide more opportunities to capture students' knowledge and abilities in science.

Large-scale Assessment Interpretation

It has been clearly stated that assessments must be designed and implemented only when a clear purpose for the assessment has been set forth (NCME/APA/AERA, 2014). When considering the implications for interpretation of large-scale assessment data, we must consider the extent to which the assessment design and resulting product provide an opportunity for those analyzing the data to draw conclusions about what students know and can do in science and what

that means for science programmatic implementation. Essentially, the assessment provides an argument that should be supported by the resulting data. Therefore, can an assessment, like the one presented in this research, provide the evidence necessary to support claims about what students know and can do in relation to the MSS/NGSS? One recommendation is to decrease the dependency on large-scale assessment data by increasing efforts to develop and use NGSS-aligned assessments throughout the assessment system. This recommendation is aligned with the BOTA report (NRC, 2013)

Conclusion

This research set out to determine the extent to which large-scale assessment clusters aligned with the NGSS standards in Michigan by collecting cognitive lab data from students interacting with the large-scale assessment. The findings indicate that clusters can elicit the three dimensions; however, the extent to which the clusters elicit dimensions is dependent on the pairings of dimensions, the available large-scale assessment item types, and the scaffolding designed into the cluster. The implications for this work include changes to design processes to include careful pairings of dimensions, and ensuring students have the opportunity to interact with constructed response item types to show their abilities. Future work stemming from this research can include scaling up of cognitive labs to determine if the findings in this study are generalizable across student populations and states, in-depth studies regarding which large-scale assessment item types are better suited to elicit evidence of the dimensions, and studies regarding the design processes for large-scale assessment.

APPENDICES

APPENDIX A: CLUSTER WRITING WORKSHOP AGENDA

Day	Activity	Description	Learning Goal for ICW teams
Monday	Training	Michigan Assessment Update	Understand the history and process of the State of Michigan Science Assessment and the implementation timeline for the new science assessment.
		New Michigan K-12 Science	Understand that the new Michigan K-12 Science Standards are the performance expectations from the NGSS formatted in the Nov. 2015 adoption document.
		Standards Structure	Review the three-dimensional nature of the performance expectations.
		Evidence Centered Design	Understand the rationale of using ECD to inform the overarching process for developing claims for the assessment.
		Topic Bundles	Understand the structure of the performance expectations being used for the state assessment.
		Clusters	Understand the structure of assessment tasks they will be designing.
		Item Pool	Utilize existing assessment items as models for three-dimensional questions.
		Unpacking	Begin a domain analysis of the topic bundle to understand the Disciplinary Core Ideas that must be assessed.
	Writing	Phenomenon Brainstorm	Understand the characteristics of an anchoring phenomenon for use on large-scale assessments and determine several options for phenomena that can be explained by the content of their assigned topic bundle.
		Stimulus Draft	Understand the relationship between phenomenon and stimulus and the characteristics

			of a good stimulus. Create a stimulus to drive the cluster.
Tuesday	Training	Bias and Sensitivity Training	Understand the equity issues associated with large-scale assessment and determine ways of developing items that are fair for the targeted population of students.
		Item Type Training	Understand how to utilize the various item types in order to elicit evidence from students in the task design process.
	Writing	Unpacking	Continue a domain analysis of the topic bundle to understand all of the elements that must be assessed.
		Cluster Outline	Begin development of the cluster including the story line, the stimulus, and item types designed to assess the topic bundle.
		Item Templates	Understand how to use to the item templates to present item types and make alignment of the item explicit while stating the evidence that the task elicits to support a claim about what students know and can do in science.
Wednesday	Writing	Cluster writing	Continue development of the cluster including the story line, the stimulus, and item types designed to assess the topic bundle.
	Peer Review	View and provide feedback on one cluster developed in same domain	Exchange feedback with another ICW team to gain perspective on the extent to which the cluster assessed the intended topic bundle.
Thursday	Writing	Incorporate feedback from peer review into cluster	Utilize feedback from peer review to enhance and further develop cluster.
	Content Review	View and provide feedback on all clusters in domain	Exchange feedback with other ICW teams to gain perspective on the extent to which the cluster assessed the intended topic

			bundle. Make policy-capturing decisions to influence future cluster development and item specifications.
Friday	Revise	Make revisions to cluster	Utilize feedback from content review to enhance and complete development of cluster.
	Submit	Submit final draft of all stimuli and items to MDE	Complete a cluster for use on Michigan Science Assessment.

APPENDIX B: UNPACKING DOCUMENT TEMPLATES

Unpacking the Disciplinary Core Idea	
1. Select the Disciplinary Core Idea	
2. What are the main ideas that are present in the grade band endpoints?	
3. What are the main ideas that are present in each element? What additional ideas are critical for the learner to understand? <mark>Element = the bullets in the foundation boxes</mark>	
4. What is the intended meaning of each element of the core idea? • Is there one idea or several separate ideas in the statement? • What terminology is explicitly used in the core idea?	

Unpacking the Core Idea	
5. Define Boundary condition • What peripheral ideas or terms are **not** essential for understanding the core idea?	
6. Describe Prior-Knowledge • What other knowledge and skills (both from **this** topic and from **other** topics) do students need in order to achieve an understanding of this core idea?	
7. Describe Student Challenges • Are there any commonly-held ideas that differ in important ways from the scientifically accepted understanding? • What methods can be used to determine students' current understandings?	
8. Brainstorm Phenomena • What phenomena would provide an **e**xample of this disciplinary core idea?	

Unpacking of Science and Engineering Practices

1. Describe the Science and Engineering Practice. What are the essential components of this practice? What possible intersections might there be with other practices? Components = Bullets in Foundation Boxes	**Science and Engineering Practice:** **Components of the SEP:** **Intersections with other Practices:**
2. List the knowledge and skills needed by students in order to successfully perform the practice. What knowledge and skills do students need to use in order to show that they can perform the practice?	**Knowledge and Skills for Performing the Practice:**
3. Identify the evidence that you would expect to see for each component of the practice. What is a high level of performance that you would expect to see for each component? What are the different levels of performance for each component?	**Evidence for each Component of the Practice:**

Unpacking of Crosscutting Concept

2. Describe the Crosscutting concept. What are the essential components of this crosscutting concept? What explanatory value does this crosscutting concept have? (i.e. how might this help a student/teacher explain a phenomenon?) ==Components = Bullets in Foundation Boxes==	**Crosscutting concept:** **Components of the CCC:** **Intersections with other Crosscutting concepts:**
3. Identify intersections with science and engineering practices and disciplinary core ideas Which SEPs provide meaningful connections with this crosscutting concept? What are some concepts and/or contexts in life, earth, and physical science that would provide good opportunities for students to explore this crosscutting concept?	**Interactions with SEPs and DCIs:**
4. Identify the evidence that you would expect to see for each component of the crosscutting concept. What is a high level of performance that you would expect to see for each component? What are the different levels of performance for each component? How might a student's understanding of this crosscutting concept grow over time?	**Evidence for the crosscutting concept:**

	SEP			DCI							CC		
	Analyzing & Interpreting Data	Develop and use scientific mode	Designing solution	PS1.A: Structure and properties of matter		PS1.B Chemical reactions		ET S1.B Developing possible solutions	ET S1.C Optimizing the design solution		Patterns	Energy and matter	
	Analyze and interpret data to determine similarities and differences in findings	Develop a model to describe unobservable mechanisms	Undertake a design project, engaging in the design cycle, to construct and/or implement a solution that meets specific design criteria and constraints	Each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it	Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants	The total number of each type of atom is conserved, and thus the mass does not change	Some chemical reactions release energy, others store energy	A solution needs to be tested, and then modified on the basis of the test results, in order to improve it	Although one design may not perform the best across all tests, identifying the characteristics of the design that performed the best in each test can provide useful information for the redesign process—that is, some of the characteristics may be	The iterative process of testing the most promising solutions and modifying what is proposed on the basis of the test results leads to greater refinement and ultimately to an optimal solution	Macroscopic patterns are related to the nature of microscopic and atomic-level structure	Matter is conserved because atoms are conserved in physical and chemical processes	The transfer of energy can be tracked as energy flows through a designed or natural system
1A	X			x	x								
2A	x					x						X	
3A		x		x	x							X	
Alt3A		x		x	x							X	
4B	x										X		
5B	x	x				x							x
6B	x	X						X					
7B	x	x							X				

M-Step™

Michigan Student Test of Educational Progress

2017 Science Pilot – Cognitive Lab

Grade 5

Form 1

Lights Out Cluster

Facilitator Name: _____

Student Name: _____

Student Grade: _____

School: _____

Date/Time: _____

Cog Lab Overview

The science pilot directions and a screen shot of each item that the student will be working on are included in this booklet. The screen shots of the items will show you what the student sees on the computer screen. As the items are administered on the computer, you will need to document the student's answers. While the student works through the items, you will document your observations in this booklet.

At the beginning of the cognitive lab, you will be provided with a script of directions to read aloud to the student and features to point out to the student. There are no accommodations for this pilot assessment, therefore, text-to-speech, masking, color contrasting, and color choice are not available.

Remember:

- Sit behind or next to the student, but not in their personal space.

- Do not interrupt the student to ask why he or she is responding in a particular way. Only prompt if they are not describing what they are thinking.

- **DO NOT** tell the student if he/she is right or wrong.

Description of Cognitive Interview Process

Step 1. The student begins by reading the Stimulus aloud. Ask the two questions provided and record the student's comments.

Step 2. For the Item, ask the student to read the question aloud, remind the student to tell what he/she is thinking. If the student is not talking, you may want to ask the question provided. Mark the student responses in the spaces provided.

Step 3. The student moves onto the next task. *[Repeat Steps 1 and 2]*

Step 4. At the end of the science items, there are a few items that are survey questions. The student will respond to those questions and you will have some follow-up questions. Record the student responses in the spaces provided.

Directions for Administration –
Grade 5 Form 3 ES: Derecho

* *

*Before beginning this test with the student, log in and be sure the "**Display Settings Test**" screen is active on the computer.*

Then, using the following script, read aloud anything that is printed in red and preceded by [SAY].

[SAY] Thank you for helping me today. My name is _____. As your teacher probably told you, I am here because I am interested in learning about what students think about different kinds of test questions. I have found that the best way to find out what people are thinking about is to have them think out loud while they work on the questions. This will help us design better test questions.

As you answer these questions, I would like you to say out loud everything that you normally think to yourself while solving the questions. Just act as if you are alone in the room speaking to yourself. After you finish each question, I will ask you a few more questions about how you solved it.

Even though this is called a test, your test answers will not be graded, and this test will not affect your grade in your class. Don't worry about getting the questions right or wrong. We are really interested in figuring out how we can make the test questions better for students.

Do you have any questions before we begin?

Pause to answer questions. Proceed when there are no more questions.

Think Aloud Practice

[SAY] To give you some practice with thinking out loud, I'd like you to think out loud while solving a math problem. I'll do one first.

I'm going to add 56 + 79.

[Interviewer talks out loud about what they're thinking while solving the problem (e.g., First, I add 6 plus 9 and I get fifteen, so I'm putting a 5 down here and I'm going to carry the 1. Then, I'm going to add 1 plus 5 plus 7 and I get 13. So the answer is 135.]

[SAY] Now you try one. 68 + 26

After the participant finishes, give her/him feedback about talking aloud if needed.

Answer any questions the student may have.

*Then, make sure the student is now viewing the computer "**Display Setting Test**" screen, and continue this script, while the student follows along through the screens on the computer.*

Note that some screens should NOT be read aloud, which is indicated in this script.

Display Settings Test (DO NOT *READ SCREEN*)

On the next several screens you will see information for the assessment. Remember, even thought it's called a test, you will not be graded. Some of these screens will not be important for you and the work we are doing today. Others will help you understand the online tasks.

Test Security (DO NOT *READ SCREEN*)

Although most of the information on this screen does not apply to our task today, one thing is important. You should not talk about the questions you will see today to anyone.

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow at the bottom of your screen.

Introduction

Read these directions carefully before beginning. To look at these directions again, select the **Help** button and choose the **Test Directions** tab.

Now, select the **Next** arrow at the bottom of your screen.

Answering Questions

- This test has passages and pictures that you will read and use to answer different types of questions.

- Carefully read each passage and look at each picture before answering the questions.

- Be sure the **Pointer** tool is selected and choose your answer. Some questions may require more than one answer.
- The **Pointer** is already selected for you when you select the **Next** arrow and go to the next question.
- To change an answer, use the **Pointer** tool to choose a different answer.
- For questions that ask you to type your response or show your answer, use the keyboard or the online tools for that question to provide your response.
- For some questions you will be required to drag words, numbers or objects from one place on the screen and drop them at another location in order to show your answer.
- Some questions will have a **Click to Respond** button. Selecting this button will open a larger version of what is seen in the small window. This is where you will enter your response.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Navigation

After you have answered a question, select the **Next** arrow at the bottom of the screen to go to the next question.

Sometimes you will need to use a **scroll bar** to see the entire part of a passage or question. Scroll up or down using the scroll bar on the right side of a passage or question.

Use the **Back** arrow to go back to a question you have already seen or answered. You can also go back to any picture or passage in the set using the **Back** arrow.

To move quickly to any question on the test, select the **Down** arrow next to the question number and select the question you'd like to see.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Tools

There are several tools to help you with the questions.

- **Pointer:** Use the pointer tool to choose an answer.
- **Cross-Off:** Use the cross-off tool to mark answers that you believe are NOT correct.
- **Highlighter:** Use the highlighter tool to highlight important information.
- **Magnifier:** Use the magnifier tool to enlarge the information on the screen.
- **Line Guide:** Use the line guide tool to help you track a single line of text as you are reading.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

- **Sticky Notes:** Use the **Sticky Note** tool to type notes on the screen.
- **Help:** Use the **Help** button to find more information, such as instructions on how to use the tools.
- **Eraser:** Use the **Eraser** button to reset your answer and start the question over.
- **Flag:** Use the **Flag** button, located at the bottom of the page, to mark a question you want to review at a later time. Later, the **Flag** will serve as a reminder to return to that question.

Do you have any questions before we move on?

[SAY] Use the **Next** arrow to skip through the **Helpful Testing Hints** and the **Completing the Test** screens.

(NOTE to Proctor: Ignore the "Helpful Testing Hints" and "Completing the Test" screens, and move on to the "Begin the Test" screen.)

[SAY] We are going to begin the test now. Remember to select the **Pause** button if you need to pause your test for any reason.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Select the green **Begin the Test** button at the bottom of the screen.

Please start by reading aloud the passage on the left of the screen.

Stimulus Part 1 (below)

Ask the student to read the passage aloud.

Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)

Stimulus Part 1 & Item 1

Gr 5 Science

Question 1

Lights Out! (Part 1)

Students are learning about eyes in science class. During their class discussion, a power outage occurs and the lights go out in the classroom. While the teacher looks for a flashlight, one student exclaims, "I can't see anything!"

The teacher turns on a flashlight and points it across the classroom to a plant on a table. The teacher says, "This makes me wonder how we are able to see the plant."

More Text Below

Which statement **best** describes how the students are able to see the plant?

(a) Once the plant produces its own light, the students can observe the plant.

(b) Once the plant absorbs all the light from the flashlight, the students can observe the plant.

(c) The light from the flashlight is reflected toward the students' eyes and then back to the plant.

(d) The light from the flashlight is reflected off of the plant and then enters the students' eyes.

N/A	PS4.B	CC 2: Cause and Effect
	An object can be seen when light reflected from its surface enters the eyes	Cause and effect relationships are routinely identified.

· · · · · 5 · · · · ·

Item 1 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Stimulus Part 1 & Item 1

Gr 5 Science
Question 1

Lights Out! (Part 1)

Students are learning about eyes in science class. During their class discussion, a power outage occurs and the lights go out in the classroom. While the teacher looks for a flashlight, one student exclaims, "I can't see anything!"

The teacher turns on a flashlight and points it across the classroom to a plant on a table. The teacher says, "This makes me wonder how we are able to see the plant."

More Text Below

Which statement **best** describes how the students are able to see the plant?

(a) Once the plant produces its own light, the students can observe the plant.

(b) Once the plant absorbs all the light from the flashlight, the students can observe the plant.

(c) The light from the flashlight is reflected toward the students' eyes and then back to the plant.

(d) The light from the flashlight is reflected off of the plant and then enters the students' eyes.

N/A	PS4.B	CC 2: Cause and Effect
	An object can be seen when light reflected from its surface enters the eyes	Cause and effect relationships are routinely identified.

· · · · · 6 · · · · ·

127

Please indicate the student's response to the item by checking the box:

☐ *a* ☐ *b* ☐ *c* ☐ *d*

While working on this question, the student's commentary showed evidence that the question elicited his/her thinking around:

☐ *light is reflected from a surface enters the eye*

☐ *the light being reflected causes the ability to see*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy* ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

Item 2 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)
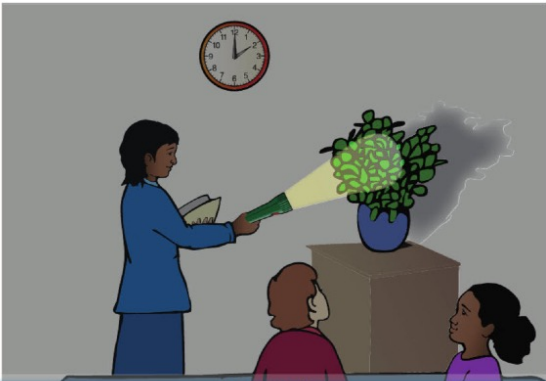
_____

Item 2

N/A	LS1.A: Structure and Function	CCC 4: Systems and System Models
	Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.	A system can be described in terms of its components and their interactions.
	LS1.D: Information Processing	
	Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain. Animals are able to use their perceptions and memories to guide their actions.	

· · · · · 8 · · · · ·

129

Please indicate the student's response to the item by checking the box:

☐ a ☐ b ☐ c ☐ d

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ nervous system allows light to be processed in the brain

☐ sense receptors are part of the nervous system

☐ the nervous system has different components that interact

☐ nervous system structures help animals / people survive

☐ other _____

_____

☐ none of the above

While working on the item, the student:

☐ appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.

☐ appeared to be engaged with the item.

☐ appeared to be able to read and understand what the item was asking.

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ Easy ☐ Hard

[Say] Why was it easy or hard?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

SECURE – DO NOT COPY

Item 3 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Item 3

SEP 2: Developing and Using Models Develop a model to describe phenomena.	PS4.B: Electromagnetic Radiation An object can be seen when light reflected from its surface enters the eyes. LS1.D: Information Processing Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain. Animals are able to use their perceptions and memories to guide their actions.	CCC 4: Systems and System Models A system can be described in terms of its components and their interactions.

· · · · · 10 · · · · ·

131

Please indicate the student's response to the item by drawing lines that connect each option to the box in which it was placed in the enlarged image *(below)*:

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ how the model represents the phenomenon

☐ the nervous system has different components that interact

☐ nervous system structures help animals / people survive

☐ light is reflected from a surface enters the eye

☐ the light being reflected causes the ability to see

☐ other _____

_____

☐ none of the above

While working on the item, the student:

☐ appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.

☐ appeared to be engaged with the item.

☐ appeared to be able to read and understand what the item was asking.

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ Easy   ☐ Hard

[Say] Why was it easy or hard?

_____

_____

_____

_____

_____

_____

_____

_____

_____

Item 3
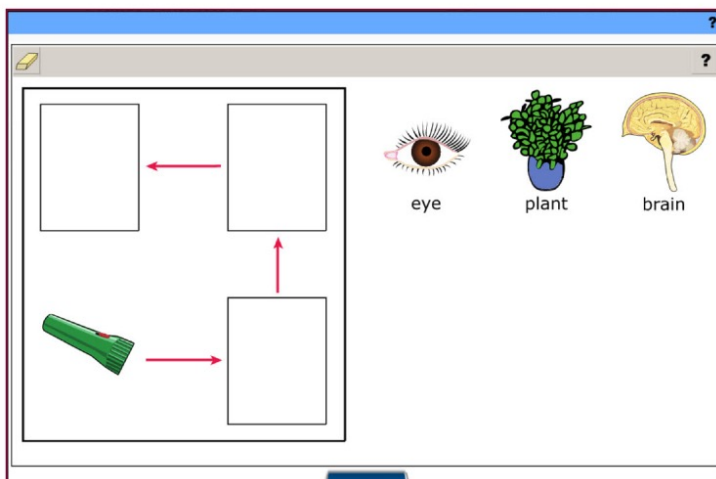
(Indicate the student's response with lines that connect each option to the box in which it was placed.)

←

Please provide any additional information/ observations that you made about the student's interaction with the item and technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Stimulus 2 (below)

Ask the student to read the passage aloud.

Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)
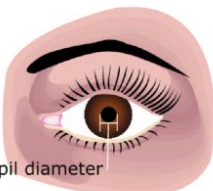
Stimulus 2, Item 4

Gr 5 Science
Question 4

Lights Out! (Part 2)
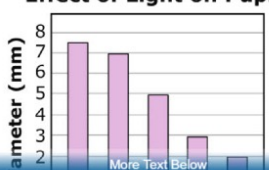
Human eyes have specialized structures to help people see. One of these structures is the pupil. The pupil is the dark center opening in the middle of the eye.

pupil diameter

The size of a pupil can be measured by its diameter. A scientist shines different amounts of light on a pupil and measures the diameter of the pupil. The graph shows the data collected by the scientist.

Effect of Light on Pupil

Diameter (mm): 8, 7, 6, 5, 4, 3, 2

More Text Below

This question has two parts.

Part A: After a while, the lights come back on in the classroom. How will the students' pupils **most likely** change?

(a) The students' pupil diameters will increase.

(b) The students' pupil diameters will decrease.

Part B: Which statement **best** explains the change in pupil diameter described in Part A?

(a) Pupil diameter increases when there is low light.

(b) Pupil diameter increases when there is bright light.

(c) Pupil diameter decreases when there is low light.

(d) Pupil diameter decreases when there is bright light.

SEP 7 - Engaging in Argument from Evidence	LS1.A: Structure and Function	CC 2: Cause and Effect
Construct an argument with evidence, data, and/or a model	Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.	Cause and effect relationships are routinely identified.

SECURE – DO NOT COPY   · · · · · 12 · · · · ·

133

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Item 4 (bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Part A: *Please indicate the student's response to the item by checking the box:*

☐ *a*   ☐ *b*

Part B: *Please indicate the student's response to the item by checking the box:*

☐ *a*   ☐ *b*   ☐ *c*   ☐ *d*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *evidence from the stimulus supports the claim*

☐ *the function of the pupil is to regulate the amount of light into the eye*

☐ *changes in light intensity inversely affect pupil size*

☐ *other* _____

_____

☐ *none of the above*

*While working on the item, the student:*

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____
_____
_____
_____

_____   _____

_____   _____

_____   _____

_____

Item 5 (bottom)

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

_____

_____

_____

_____

_____

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Item 5

SEP 7 - Engaging in Argument from Evidence	LS1.A: Structure and Function	CC 2: Cause and Effect
Construct an argument with evidence, data, and/or a model.	Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.	Cause and effect relationships are routinely identified.

· · · · · 14 · · · · ·

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

☐ *the intensity of the light causes the pupils to change*

☐ *the function of the pupil is to regulate the amount of light let into the eye*

☐ *the evidence and reasoning statements must support the claim*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

*Please indicate the student's response to the item by drawing a line from each chosen response to the box in which it was placed on the enlarged image **(below)**:*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

Item 5

| ? |
| |
| ? |

Claim	The pupil diameter changes with different amounts of light.
Evidence	
Reasoning	

Evidence Statements

The diameter was largest in the lowest light and smallest in the brightest light.

The diameter was largest in the brightest light and smallest in the lowest light.

The diameter of the pupil increased as the light increased.

The diameter of the pupil decreased as the light increased.

Reasoning Statements

When there is less light, the pupil gets bigger to let in more light.

When the pupil is smaller, it lets in more light so a person can see better in less light.

When there is bright light, the pupil lets in more light so a person can see better.

(Indicate the student's response by drawing a line from each chosen response to the box in which it was placed.)

←

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

(Transition Message in Test that comes up before Survey Questions)

[Say] Congratulations! You have reached the end of the Michigan Science Standards Pilot. Now, take a short survey.

[Say] Select the **Begin Section** button to answer the survey questions.

(NOTE to Proctor: Please indicate the student's response to each survey question with an X on the screen shots on the next 2 pages.)

Question 6 (bottom)

[Say] Can you tell me what you have learned about how people can see things in class? How was this group of questions different or the same from what you learned in class?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

Question 7 (bottom)

[Say] Can you tell me what kinds of real life examples your teacher has used in class to teach about science and solve problems?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Q 6

Gr 5 Science
Question 6

Before I took this test, I had learned _____ in my science class about how people can see things.

ⓐ a lot

ⓑ some or a little

ⓒ nothing

(Indicate the student's response with an X on this image.)

Q 7

Gr 5 Science
Question 7

My science teacher uses examples of things that happen in real life and the world around us to teach about science and to solve problems in my science class.

ⓐ a lot or all the time

ⓑ sometimes

ⓒ never

(Indicate the student's response with an X on this image.)

138

Question 8

[Say] Can you tell me what was different or the same about reading passages and using data in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____

Question 9

[Say] Can you tell me what was different or the same about writing to explain your answer in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____

Q 8

Gr 5 Science
Question 8

In my science class, the tests my teacher gives us include things we have to read and data we need to look at and interpret to answer the questions on the test.

(a) every test in science class

(b) for some tests in science class

(c) not on any tests in science class

(Indicate the student's response with an X on this image.)

Q 9

Gr 5 Science
Question 9

In my science class, the tests have questions that I have to write my answer and also include an explanation for my answer.

(a) every test in science class

(b) for some tests in science class

(c) not on any tests in science class

(Indicate the student's response with an X on this image.)

· · · · · 18 · · · · ·

139

What to do at the end of the Cognitive Lab

[Say]: Thank you very much for your time and feedback about the science test. Your comments will help us make the test better for students in Michigan.

*The student should **NOT** select the **End Test** button. You can navigate back to student responses if you need to review or capture information you may have missed. After all information has been captured, select the **End Test** button and exit the Test Development Environment.*

SECURE – DO NOT COPY

2017 Science Pilot – Cognitive Lab

Grade 8

Form 2

Hand Warmers Cluster

Facilitator Name: _____

Student Name: _____

Student Grade: _____

School: _____

Date/Time: _____

Cog Lab Overview

The science pilot directions and a screen shot of each item that the student will be working on are included in this booklet. The screen shots of the items will show you what the student sees on the computer screen. As the items are administered on the computer, you will need to document the student's answers. While the student works through the items, you will document your observations in this booklet.

At the beginning of the cognitive lab, you will be provided with a script of directions to read aloud to the student and features to point out to the student. There are no accommodations for this pilot assessment, therefore, text-to-speech, masking, color contrasting, and color choice are not available.

Remember:

- Sit behind or next to the student, but not in their personal space.

- Do not interrupt the student to ask why he or she is responding in a particular way. Only prompt if they are not describing what they are thinking.

- **DO NOT** tell the student if he/she is right or wrong.

Description of Cognitive Interview Process

Step 1. The student begins by reading the Stimulus aloud. Ask the two questions provided and record the student's comments.

Step 2. For the Item, ask the student to read the question aloud, remind the student to tell what he/she is thinking. If the student is not talking, you may want to ask the question provided. Mark the student responses in the spaces provided.

Step 3. The student moves onto the next task. *[Repeat Steps 1 and 2]*

Step 4. At the end of the science items, there are a few items that are survey questions. The student will respond to those questions and you will have some follow-up questions. Record the student responses in the spaces provided.

Directions for Administration –
Grade 8 Form 2 PS: Hand Warmers

· ·

*Before beginning this test with the student, log in and be sure the "**Display Settings Test**" screen is active on the computer.*

*Then, using the following script, read aloud anything that is printed in red and preceded by **[SAY]**.*

[SAY] Thank you for helping me today. My name is _____. As your teacher probably told you, I am here because I am interested in learning about what students think about different kinds of test questions. I have found that the best way to find out what people are thinking about is to have them think out loud while they work on the questions. This will help us design better test questions.

As you answer these questions, I would like you to say out loud everything that you normally think to yourself while solving the questions. Just act as if you are alone in the room speaking to yourself. After you finish each question, I will ask you a few more questions about how you solved it.

Even though this is called a test, your test answers will not be graded, and this test will not affect your grade in your class. Don't worry about getting the questions right or wrong. We are really interested in figuring out how we can make the test questions better for students.

Do you have any questions before we begin?

Pause to answer questions. Proceed when there are no more questions.

Think Aloud Practice

[SAY] To give you some practice with thinking out loud, I'd like you to think out loud while solving a math problem. I'll do one first.

I'm going to add 56 + 79.

[Interviewer talks out loud about what they're thinking while solving the problem (e.g., First, I add 6 plus 9 and I get fifteen, so I'm putting a 5 down here and I'm going to carry the 1. Then, I'm going to add 1 plus 5 plus 7 and I get 13. So the answer is 135.]

[SAY] Now you try one. 68 + 26

After the participant finishes, give her/him feedback about talking aloud if needed.

Answer any questions the student may have.

*Then, make sure the student is now viewing the computer "**Display Setting Test**" screen, and continue this script, while the student follows along through the screens on the computer.*

Note that some screens should NOT be read aloud, which is indicated in this script.

Display Settings Test *(DO NOT READ SCREEN)*

On the next several screens you will see information for the assessment. Remember, even thought it's called a test, you will not be graded. Some of these screens will not be important for you and the work we are doing today. Others will help you understand the online tasks.

Test Security *(DO NOT READ SCREEN)*

Although most of the information on this screen does not apply to our task today, one thing is important. You should not talk about the questions you will see today to anyone.

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow at the bottom of your screen.

Introduction

Read these directions carefully before beginning. To look at these directions again, select the **Help** button and choose the **Test Directions** tab.

Now, select the **Next** arrow at the bottom of your screen.

Answering Questions

- This test has passages and pictures that you will read and use to answer different types of questions.
- Carefully read each passage and look at each picture before answering the questions.

· · · · · 3 · · · · ·   **SECURE – DO NOT COPY**

- Be sure the **Pointer** tool is selected and choose your answer. Some questions may require more than one answer.
- The **Pointer** is already selected for you when you select the **Next** arrow and go to the next question.
- To change an answer, use the **Pointer** tool to choose a different answer.
- For questions that ask you to type your response or show your answer, use the keyboard or the online tools for that question to provide your response.
- For some questions you will be required to drag words, numbers or objects from one place on the screen and drop them at another location in order to show your answer.
- Some questions will have a **Click to Respond** button. Selecting this button will open a larger version of what is seen in the small window. This is where you will enter your response.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Navigation

After you have answered a question, select the **Next** arrow at the bottom of the screen to go to the next question.

Sometimes you will need to use a **scroll bar** to see the entire part of a passage or question. Scroll up or down using the scroll bar on the right side of a passage or question.

Use the **Back** arrow to go back to a question you have already seen or answered. You can also go back to any picture or passage in the set using the **Back** arrow.

To move quickly to any question on the test, select the **Down** arrow next to the question number and select the question you'd like to see.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Tools

There are several tools to help you with the questions.

- **Pointer:** Use the pointer tool to choose an answer.
- **Cross-Off:** Use the cross-off tool to mark answers that you believe are NOT correct.
- **Highlighter:** Use the highlighter tool to highlight important information.
- **Magnifier:** Use the magnifier tool to enlarge the information on the screen.
- **Line Guide:** Use the line guide tool to help you track a single line of text as you are reading.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

- **Sticky Notes:** Use the **Sticky Note** tool to type notes on the screen.
- **Help:** Use the **Help** button to find more information, such as instructions on how to use the tools.
- **Eraser:** Use the **Eraser** button to reset your answer and start the question over.
- **Periodic Table:** You may find the **Periodic Table of Elements** helpful in answering test questions. You can view more detail about an element by clicking on the element's box. The box will open in a new window and can be moved around to different parts of the screen by selecting the box and dragging it to a new location.
- **Flag:** Use the **Flag** button, located at the bottom of the page, to mark a question you want to review at a later time. Later, the **Flag** will serve as a reminder to return to that question.

Do you have any questions before we move on?

[SAY] Use the **Next** arrow to skip through the **Helpful Testing Hints** and the **Completing the Test** screens.

(NOTE to Proctor: Ignore the "Helpful Testing Hints" and "Completing the Test" screens, and move on to the "Begin the Test" screen.)

[SAY] We are going to begin the test now. Remember to select the **Pause** button if you need to pause your test for any reason.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Select the green **Begin the Test** button at the bottom of the screen.

Please start by reading aloud the passage on the left of the screen.

Stimulus Part 1 (below)

Ask the student to read the passage aloud.

Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

Stimulus Part 1 & Item 1

Gr 8 Science
Question 1

How do Hand Warmers Work? (Part 1)

Two students are outside in the cold, waiting for a bus. One of the students has a package of hand warmers and offers to share them with the other student. The student opens the package and they each put a hand warmer bag in one of their gloves.

Hand Warmer ← hand warmer package

Hand Warmer ← hand warmer bag

After a few minutes, the students notice that the hand warmer bags start to feel warm. The students want to know how hand warmer bags get warm. They decide to ask their science teacher if they can test the materials inside the hand warmer bags. After reading the ingredients on the hand warmer package, the students decide to focus on iron because it is the most common ingredient.

The students designed the following procedure.
1. Open a new hand warmer package.
2. Cut open the hand warmer bag. *More Text Below*

This question has two parts.

Part A: Use the data table to complete the following statement.

The students can tell that a chemical reaction involving iron _____ because a new substance _____ form overnight.

Part B: Choose one set of properties that **best** supports the completed statement in Part A.

ⓐ density and color

ⓑ color and volume

ⓒ volume and texture

ⓓ texture and mass

ⓔ mass and density

SEP 4: Analyzing and Interpreting Data	PS1.B: Chemical Reactions	CCC 1: Patterns
Analyze and interpret data to determine similarities and differences in findings.	Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.	Macroscopic patterns are related to the nature of microscopic and atomic-level structure

· · · · · **5** · · · · ·   **SECURE – DO NOT COPY**

_____

_____

_____

_____

_____

_____

If the student is not talking, you may want to ask the following question:

— Why did you answer that way?

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Item 1 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

_____

_____

_____

_____

_____

Stimulus Part 1 & Item 1

Gr 8 Science

Question 1

How do Hand Warmers Work? (Part 1)

Two students are outside in the cold, waiting for a bus. One of the students has a package of hand warmers and offers to share them with the other student. The student opens the package and they each put a hand warmer bag in one of their gloves.

Hand Warmer ← hand warmer package

Hand Warmer ← hand warmer bag

After a few minutes, the students notice that the hand warmer bags start to feel warm. The students want to know how hand warmer bags get warm. They decide to ask their science teacher if they can test the materials inside the hand warmer bags. After reading the ingredients on the hand warmer package, the students decide to focus on iron because it is the most common ingredient.

The students designed the following procedure.
1. Open a new hand warmer package.
2. Cut open the hand war... More Text Below

This question has two parts.

Part A: Use the data table to complete the following statement.

The students can tell that a chemical reaction involving iron _____ ▾ because a new substance _____ ▾ form overnight.

Part B: Choose one set of properties that **best** supports the completed statement in Part A.

ⓐ density and color

ⓑ color and volume

ⓒ volume and texture

ⓓ texture and mass

ⓔ mass and density

SEP 4: Analyzing and Interpreting Data Analyze and interpret data to determine similarities and differences in findings.	PS1.B: Chemical Reactions Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.	CCC 1: Patterns Macroscopic patterns are related to the nature of microscopic and atomic-level structure

· · · · · 6 · · · · ·

146

_____

_____

_____

_____

_____

_____

_____

Part A: Please indicate the student's responses to the item by checking the boxes:

☐ occurred   ☐ did not occur

☐ did   ☐ did not

Part B: Please indicate the student's response to the item by checking the box:

☐ a   ☐ b   ☐ c   ☐ d   ☐ e

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ chemical reactions can result in substances with new properties

☐ changes in specific properties indicate a chemical reaction has occurred

☐ the patterns in the data provide evidence of a chemical reaction

☐ other _____

_____

☐ none of the above

While working on the item, the student:

☐ appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.

☐ appeared to be engaged with the item.

☐ appeared to be able to read and understand what the item was asking.

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ Easy   ☐ Hard

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

SECURE – DO NOT COPY

Item 2 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)
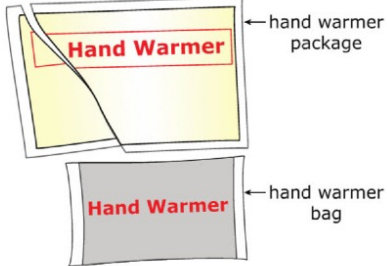
_____

_____

_____

_____

_____

_____

Item 2

Gr 8 Science
Question 2

Hand Warmer package

Hand Warmer ← hand warmer bag

After a few minutes, the students notice that the hand warmer bags start to feel warm. The students want to know how hand warmer bags get warm. They decide to ask their science teacher if they can test the materials inside the hand warmer bags. After reading the ingredients on the hand warmer package, the students decide to focus on iron because it is the most common ingredient.

The students designed the following procedure.
1. Open a new hand warmer package.
2. Cut open the hand warmer bag.
3. Separate the materials by using a magnet to attract the iron.
4. Place the iron on a dish.
5. Make initial observations and calculations to record properties of the iron.
6. Leave the iron in the dish overnight.
7. Record final observations and calculations the next day.

Iron from Hand Warmer Bag

More Text Above   More Text Below

Complete the statements below.

The final mass of the material on the plate the next day is
_____ ▾ the initial mass of the material.

This could happen if
_____ ▾ the environment.

| SEP 4: Analyzing and Interpreting Data

Analyze and interpret data to determine similarities and differences in findings. | PS1.B: Chemical Reactions

Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.

The total number of each type of atom is conserved, and thus the mass does not change. | CCC 5: Energy and Matter

Matter is conserved because atoms are conserved in physical and chemical processes. |

· · · · · 8 · · · · ·

148

_____

_____

_____

_____

_____

_____

Please indicate the student's responses to the item by checking the boxes:

☐ more than   ☐ less than   ☐ equal to

☐ iron atoms are escaping into

☐ iron atoms are combining with matter in

☐ iron atoms are being produced and released into

☐ iron atoms are being exchanged in equal amounts with

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ chemical reactions can result in substances with new properties

☐ changes in specific properties indicate a chemical reaction has occurred

☐ atoms are conserved in a chemical process

☐ other _____

_____

☐ none of the above

While working on the item, the student:

☐ appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.

☐ appeared to be engaged with the item.

☐ appeared to be able to read and understand what the item was asking.

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ Easy   ☐ Hard

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

149

Item 3 (below, and bottom of next page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Item 3, Part A

Gr 8 Science
Question 3
Page 1 of 2

How do Hand Warmers Work? (Part 1)

Two students are outside in the cold, waiting for a bus. One of the students has a package of hand warmers and offers to share them with the other student. The student opens the package and they each put a hand warmer bag in one of their gloves.

Hand Warmer ← hand warmer package

Hand Warmer ← hand warmer bag

After a few minutes, the students notice that the hand warmer bags start to feel warm. The students want to know how hand warmer bags get warm. They decide to ask their science teacher if they can test the materials inside the hand warmer bags. After reading the ingredients on the hand warmer package, the students decide to focus on iron because it is the most common ingredient.

The students designed the following procedure.
1. Open a new hand warmer package.
2. Cut open the hand war... More Text Below

This question has two parts.

The teacher tells the students that iron often reacts with oxygen in the air. The students decide to study the reaction of iron and oxygen to determine whether this reaction is occurring in the hand warmer bags.

Part A: The students begin to develop a model to represent this reaction.

Complete the model by dragging the appropriate number of atoms from the key into the model.

Iron Atoms Before | Oxygen Atoms Before | After | Key: Fe Iron atom, O Oxygen atom

Click To Respond

SEP 2. Developing and Using Models	PS1.B: Chemical Reactions	CCC 5: Energy and Matter
• Develop a model to predict and/or describe phenomena.) • Develop a model to describe unobservable mechanisms	Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants. • The total number of each type of atom is conserved, and thus the mass does not change.	Matter is conserved because atoms are conserved in physical and chemical processes.

· · · · · 10 · · · · ·

150

Part A: *Please indicate the student's response to the item by drawing his/her response on the enlarged image (below).*

Part B (1): *Please record (as close as possible) the student's written response to the first prompt:*

_____

_____

_____
_____
_____
_____
_____
_____
_____
_____
_____

Item 3, Part A

(Indicate the student's response to Part A by drawing.)

Item 3, Part B

Part B (2): *Please record (as close as possible) the student's written response to the second prompt:*

_____

_____

_____

_____

_____

_____

_____

_____

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *chemical reactions can result in substances with new properties*

☐ *the model represents an unobservable mechanism*

☐ *atoms are conserved in a chemical process*

☐ *other* _____

_____

Item 3, Part B

Gr 8 Science
Question 3
Page 2 of 2

How do Hand Warmers Work? (Part 1)

Two students are outside in the cold, waiting for a bus. One of the students has a package of hand warmers and offers to share them with the other student. The student opens the package and they each put a hand warmer bag in one of their gloves.

Hand Warmer ← hand warmer package

Hand Warmer ← hand warmer bag

After a few minutes, the students notice that the hand warmer bags start to feel warm. The students want to know how hand warmer bags get warm. They decide to ask their science teacher if they can test the materials inside the hand warmer bags. After reading the ingredients on the hand warmer package, the students decide to focus on iron because it is the most common ingredient.

The students designed the following procedure.
1. Open a new hand warmer package.
2. Cut open the hand warmer bag. **More Text Below**

This question has two parts.

The teacher tells the students that iron often reacts with oxygen in the air. The students decide to study the reaction of iron and oxygen to determine whether this reaction is occurring in the hand warmer bags.

Part B: Use the model in Part A to help explain how some of the properties of the iron changed overnight.

Describe how one of the properties changed.

0/500

Describe how the model explains this property change.

0/500

SEP 2. Developing and Using Models	PS1.B: Chemical Reactions	CCC 5: Energy and Matter
• Develop a model to predict and/or describe phenomena.) • Develop a model to describe unobservable mechanisms	Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants. • The total number of each type of atom is conserved, and thus the mass does not change.	Matter is conserved because atoms are conserved in physical and chemical processes.

∙ ∙ ∙ ∙ ∙ 12 ∙ ∙ ∙ ∙ ∙

□ *none of the above*

While working on the item, the student:

□ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

□ *appeared to be engaged with the item.*

□ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

□ *Easy*   □ *Hard*

[Say] Why was it easy or hard?

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction with the item and technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

SECURE – DO NOT COPY

153

Stimulus Part 2 (below)

Ask the student to read the passage aloud. Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Item 4 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Stimulus 2, Item 4

Gr 8 Science — Question 4

How do Hand Warmers Work? (Part 2)

After observing what happens with the iron in the hand warmers, the students decide to work on improving the hand warmers to make them get warm faster. The students observe that the hand warmer bag is made of fabric containing tiny holes.

Hand Warmer

They think that if the iron has to react with oxygen to produce heat, maybe only a small amount of oxygen is getting through the fabric. To test this, the students design an investigation to record the change in temperature when the hand warmer ingredients are left in the hand warmer bag and when they are taken out of the hand warmer bag. The students set up two systems.

Investigation Plan

- Open two hand warmer packages. Place one hand warmer bag on a plate labeled System 1.
- Cut open the other hand warmer bag and pour the materials from inside the bag onto the plate labeled System 2.
- Place a thermometer under the hand warmer bag in System 1 and a thermometer inside the materials in System 2.
- Record the temperature of both systems every five minutes until one of

More Text Below

Compare the two sets of data in the graph. Select **two** similarities and **two** differences between System 1 and System 2.

Similarities
The temperature increases in both sy...
The temperature decreases in both sy...
The temperature remains constant in ... systems.

Click To Respond

| SEP 4: Analyzing and Interpreting Data | N/A | CCC 5: Energy and Matter |
| Analyze and interpret data to determine similarities and differences in findings | | The transfer of energy can be tracked as energy flows through a designed or natural system. |

· · · · · 14 · · · · ·

154

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please indicate the student's response to the item by circling the similarities and differences selected in the enlarge table **(below)**.

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ the data in the graph indicate a change in temperature of the two systems

☐ energy in the form of heat (temperature) is changing in the systems

☐ other _____

_____

☐ none of the above

While working on the item, the student:

☐ appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.

☐ appeared to be engaged with the item.

☐ appeared to be able to read and understand what the item was asking.

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ Easy   ☐ Hard

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Item 4

Similarities	Differences
The temperature increases in both systems. The temperature decreases in both systems. The temperature remains constant in both systems.	The temperature of System 1 increases more quickly than the temperature of System 2. The temperature of System 2 increases more quickly than the temperature of System 1. System 1 reaches a greater maximum temperature than System 2 reaches. System 2 reaches a greater maximum temperature than System 1 reaches.

OK

(Indicate the student's response circling the similarities and differences selected.)

←

155

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

Item 5

Item 5 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Gr 8 Science
Question 5

How do Hand Warmers Work? (Part 2)

After observing what happens with the iron in the hand warmers, the students decide to work on improving the hand warmers to make them get warm faster. The students observe that the hand warmer bag is made of fabric containing tiny holes.

Hand Warmer

They think that if the iron has to react with oxygen to produce heat, maybe only a small amount of oxygen is getting through the fabric. To test this, the students design an investigation to record the change in temperature when the hand warmer ingredients are left in the hand warmer bag and when they are taken out of the hand warmer bag. The students set up two systems.

Investigation Plan

- Open two hand warmer packages. Place one hand warmer bag on a plate labeled System 1.
- Cut open the other hand warmer bag and pour the materials from inside the bag onto the plate labeled System 2.
- Place a thermometer under the hand warmer bag in System 1 and a thermometer inside the materials in System 2.
- Record the temperature of both systems every five minutes until one of

More Text Below

Identify and explain the temperature pattern in the graph. Select and move the **best** claim statement, evidence statement, and reasoning statement that can be used to explain the pattern.

Claim	
Evidence	
Reasoning	

Claim Statements:
Energy is transferred from each system to the thermometers.
Energy is transferred from the thermometers to each system.
Energy is not transferred in either of the systems.

Evidence Statements:
The thermometers measured a higher temperature at 50 minutes than at 0 minutes.
The thermometers measured a lower temperature at 50 minutes than at 0 minutes.
The temperature did not change between 0 minutes and 50 minutes in either system.

Reasoning Statements:
The iron absorbed energy from the oxygen during the chemical reaction.
The hand warmer bag absorbed energy from the thermometer.
Energy was released when the iron reacted with the oxygen in the air.
Energy was released when the hand warmer package was opened.

Click To Respond

SEP 4: Analyzing and Interpreting Data	PS1.B: Chemical Reactions	CCC 5: Energy and Matter
Analyze and interpret data to determine similarities and differences in findings.	Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.	The transfer of energy can be tracked as energy flows through a designed or natural system.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please indicate the student's response to the item by drawing a line from each chosen response to the box in which it was placed on the enlarged image (below):

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the data in the graph indicate a change in temperature of the two systems*

☐ *energy in the form of heat (temperature) is changing in the systems*

☐ *chemical reactions can result in substances with new properties*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

Item 5

Claim	
Evidence	
Reasoning	

Claim Statements:
Energy is transferred from each system to the thermometers.
Energy is transferred from the thermometers to each system.
Energy is not transferred in either of the systems.

Evidence Statements:
The thermometers measured a higher temperature at 50 minutes than at 0 minutes.
The thermometers measured a lower temperature at 50 minutes than at 0 minutes.
The temperature did not change between 0 minutes and 50 minutes in either system.

Reasoning Statements:
The iron absorbed energy from the oxygen during the chemical reaction.
The hand warmer bag absorbed energy from the thermometer.
Energy was released when the iron reacted with the oxygen in the air.
Energy was released when the hand warmer package was opened.

(Indicate the student's response by drawing a line from each chosen response to the box in which it was placed.)

←

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

Item 6

Gr 8 Science
Question 6   [Item ID] [?]

How do Hand Warmers Work? (Part 2)

After observing what happens with the iron in the hand warmers, the students decide to work on improving the hand warmers to make them get warm faster. The students observe that the hand warmer bag is made of fabric containing tiny holes.

Hand Warmer

They think that if the iron has to react with oxygen to produce heat, maybe only a small amount of oxygen is getting through the fabric. To test this, the students design an investigation to record the change in temperature when the hand warmer ingredients are left in the hand warmer bag and when they are taken out of the hand warmer bag. The students set up two systems.

Investigation Plan

- Open two hand warmer packages. Place one hand warmer bag on a plate labeled System 1.
- Cut open the other hand warmer bag and pour the materials from inside the bag onto the plate labeled System 2.
- Place a thermometer under the hand warmer bag in System 1 and a thermometer inside the materials in System 2.
- Record the temperature of both systems every five minutes until one of
  More Text Below

The students want to decrease the amount of time it takes for the hand warmer to get warm. One student suggests changing the design of the hand warmer bag by adding more small holes to the fabric of the bag.

Which evaluation of the design **best** explains why making more small holes in the fabric of the bag **would** or **would not** work?

(a) This design change would not work because the additional holes in the bag would allow oxygen from the air to move into the bag and cool the iron.

(b) This design change would not work because the only way to increase the temperature is to generate more energy by adding iron.

(c) This design change would work because when the iron in the hand warmer is exposed to more oxygen in the air, the temperature will increase at a faster rate.

(d) This design change would work because when pieces of iron are able to move out of the additional holes, they will be in contact with more oxygen, which will quickly increase the temperature.

SEP 6. Constructing Explanations and Designing Solutions	ETS1.C: Optimizing the Design Solution	CCC 1. Patterns
Undertake a design project, engaging in the design cycle, to construct and/or implement a solution that meets specific design criteria and constraints.	Although one design may not perform the best across all tests, identifying the characteristics of the design that performed the best in each test can provide useful information for the redesign process - that is, some of the characteristics may be incorporated into the new design.	• Different patterns may be observed at each of the scales at which a system is studied and can provide evidence for causality in explanations of phenomena. Energy and Matter • The transfer of energy can be tracked as energy flows through a designed or natural system.

· · · · · 18 · · · · ·

158

_____
_____
_____
_____
_____

Item 6 (bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Please indicate the student's response to the item by checking the box:

☐ *a*　　☐ *b*　　☐ *c*　　☐ *d*

_____
_____
_____
_____

_____
_____
_____
_____
_____
_____
_____

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the criteria of getting the hand warmer faster needs to be considered in the design*

☐ *the patterns observed in the investigation plan would hold true for the re-design*

☐ *energy would flow in the designed system in a similar way to the investigation*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*　　☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____
_____

　　SECURE – DO NOT COPY

159

Item 7

Gr 8 Science

Question 7

How do Hand Warmers Work? (Part 2)

After observing what happens with the iron in the hand warmers, the students decide to work on improving the hand warmers to make them get warm faster. The students observe that the hand warmer bag is made of fabric containing tiny holes.

Hand Warmer

They think that if the iron has to react with oxygen to produce heat, maybe only a small amount of oxygen is getting through the fabric. To test this, the students design an investigation to record the change in temperature when the hand warmer ingredients are left in the hand warmer bag and when they are taken out of the hand warmer bag. The students set up two systems.

Investigation Plan

- Open two hand warmer packages. Place one hand warmer bag on a plate labeled System 1.
- Cut open the other hand warmer bag and pour the materials from inside the bag onto the plate labeled System 2.
- Place a thermometer under the hand warmer bag in System 1 and a thermometer inside the materials in System 2.
- Record the temperature of both systems every five minutes until one of

More Text Below

Which statement **best** describes a trade-off if a new hand warmer bag is designed to get warm faster than the original design?

(a) The hand warmer bag would be more expensive than the original design.

(b) The hand warmer bag would contain more iron and be bigger than the original design.

(c) The hand warmer bag would not get as warm as the original design.

(d) The hand warmer bag would cool down faster than the original design.

SEP 6. Constructing Explanations and Designing Solutions	ETS1.B: Developing Possible Solutions	CCC 5: Energy and Matter
Undertake a design project, engaging in the design cycle, to construct and/or implement a solution that meets specific design criteria and constraints.	A solution needs to be tested, and then modified on the basis of the test results, in order to improve it. ETS1.C: Optimizing the Design Solution Although one design may not perform the best across all tests, identifying the characteristics of the design that performed the best in each test can provide useful information for the redesign process - that is, some of the characteristics may be incorporated into the new design	The transfer of energy can be tracked as energy flows through a designed or natural system.

_____

_____

_____

_____

_____

Item 7 (bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— Why did you answer that way?

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please indicate the student's response to the item by checking the box:

☐ *a*   ☐ *b*   ☐ *c*   ☐ *d*

_____

_____

_____

_____

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the trade-offs of getting the hand warmer faster needs to be considered in the design*

☐ *energy would flow in the designed system in a similar way to the investigation*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

161

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

(Transition Message in Test that comes up before Survey Questions)

[Say] Congratulations! You have reached the end of the Michigan Science Standards Pilot. Now, take a short survey.

[Say] Select the **Begin Section** button to answer the survey questions.

(NOTE to Proctor: Please indicate the student's response to each survey question with an X on the screen shots below and on the next pages.)

Question 8 (bottom)

[Say] Can you tell me what you have learned about how handwarmers work? How was this group of questions different or the same from what you learned in class?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Q 8

Gr 8 Science
Question 8

Before I took this test, I had learned _____ in my science class about how handwarmers worked.

(a) a lot

(b) some or a little

(c) nothing

(Indicate the student's response with an X on this image.)

Question 9 (bottom of next page)

[Say] Can you tell me what kinds of real life examples your teacher has used in class to teach about science and solve problems?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____

Question 10 (bottom)

[Say] Can you tell me what was different or the same about reading passages and using data in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Q 9

Gr 8 Science
Question 9

My science teacher uses examples of things that happen in real life and the world around us to teach about science and to solve problems in my science class.

(a) a lot or all the time

(b) sometimes

(c) never

(Indicate the student's response with an X on this image.)

Q 10

Gr 8 Science
Question 10

In my science class, the tests my teacher gives us include things we have to read and data we need to look at and interpret to answer the questions on the test.

(a) every test in science class

(b) for some tests in science class

(c) not on any tests in science class

(Indicate the student's response with an X on this image.)

163

Question 11 (bottom)

[Say] Can you tell me what was different or the same about writing to explain your answer in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

What to do at the end of the Cognitive Lab

[Say]: Thank you very much for your time and feedback about the science test. Your comments will help us make the test better for students in Michigan.

*The student should **NOT** select the **End Test** button. You can navigate back to student responses if you need to review or capture information you may have missed. After all information has been captured, select the **End Test** button and exit the Test Development Environment.*

Q 11

In my science class, the tests have questions that I have to write my answer and also include an explanation for my answer.

(a) every test in science class

(b) for some tests in science class

(c) not on any tests in science class

(Indicate the student's response with an X on this image.)

· · · · · 24 · · · · ·

Grade 5

Item 1

Which statement **best** describes how the students are able to see the plant?

(a) Once the plant produces its own light, the students can observe the plant.

(b) Once the plant absorbs all the light from the flashlight, the students can observe the plant.

(c) The light from the flashlight is reflected toward the students' eyes and then back to the plant.

(d) The light from the flashlight is reflected off of the plant and then enters the students' eyes.

Item	SEP	DCI	CCC
1	n/a	PS4.B: An object can be seen when light reflected from its surface enters the eyes.	Cause and Effect
Key: D			

Item 1 Coding Rules

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.
How the DCI is coded in this item: Student states that light must be present for the plant to be seen AND that light must reflect off the plant (Ref) AND that light must enter the eye after reflecting off the plant (Eye).   The language for "reflect" can include   • "directs,"   • "bounces off of,"   • "goes back,"   • etc   Code as 0 when   • If only the flashlight and seeing the plant is mentioned

● If the order or causal mechanism are incorrect Non-codable: ● If the student repeats the answer options or the prompt verbatim		
Code	**Definition**	**Example**
Ref	Reflection of light off the surface of an object. States that light must be present for the plant to be seen and the light must reflect off the plant	(510) ~~P: Student reads option A - no I don't think so~~ ~~P: Student reads option B and C~~ P: Student reads option D. Well I think that is right. Do I just click…. ~~R: Why do you think D is the right answer?~~ P: Because if you shine a light on something from a flashlight, It's going to reflect off the plant... Well the thing. And then you can see it.
Eye	Light enters the eyes for objects to be seen. States that light must enter the eye after reflecting off plant for plant to be seen	(53) P: I think they're able to see the plants now because the light is reflecting off of their eyes To the plant so they can see it. "~~Once the plant produces its own light the students can observe for the plant. Once the plant absorbs all the light from the flash light the students can observe the plant. The light from the flash light is reflected off the student eyes and then back to the plant. The light from the flash light is reflected off the plant and then enters the student eyes.~~" I'm going to say D. ~~R: D? Can you tell me why you answered that way?~~ P: The light reflects into their eyes and then they can see the plant.

CCC: Cause and Effect
Overarching Rule for Cause and Effect: Student states a relationship between two occurrences where one occurrence leads to the other (or needs the other to occur). The language should include linking words such as "because," " and then" (but just having a linking word is not sufficient to get a code of "present" - the linking words have to link the occurrences). If there is a sequence of intermediate events that link the cause and effect, the student states some intermediate events.

Code	**Definition**	**Example**
Lnk	Includes a link between a cause and an effect (the light is needed for the plant to be seen)	(52) I think it's D because while she's pointing at the plant there's a flashlight pointing at the plant. And the students are able to see where the plant is because of the flashlight.

166

Seq	Includes code 1 and a sequence of intermediate events that link the cause and effect, the student states some intermediate events (the plant being seen because of reflection of the light off the plant and the reflected light entering the eye. Student must reference either the reflection of light off the plant or the light entering the eye).	(51) The light hits the plants and it directs to your eyes. So I think it would be D because the plants into the student's eye because of the flashlight's light that is given to the plant. It can direct to your eye

Item 2

Eyes collect information about the world in the form of light. Which statement **best** describes how this information is processed?

(a) Light is sensed by the brain and then transferred to the eyes.

(b) Light is processed in the eyes, allowing the object to be seen immediately.

(c) The eyes reflect light back to the object as the information about the object is processed.

(d) The eyes have structures that sense light, and then the information is sent to the brain to be processed.

Item	SEP	DCI	CCC
2	n/a	LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain. Animals are able to use their perceptions and memories to guide their actions.	Cause and Effect
Key: D			

Item 2 Coding Rules

DCI: LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain.

How the DCI is coded in this item: Student states the eyes are sense receptors that take in light information (Sns) AND that light information taken in by the eyes is processed in the brain (Brn).
The language for "sense" can include
- "feel,"
- "take in,"
- "notice"
- etc

Code as 0 when
- If only the eyes are mentioned
- If the order or causal mechanism are incorrect

Non-codable:

● If the student repeats the answer options or the prompt verbatim		

Code	Definition	Example
Sns	Sense receptors specialized for information.States that the eyes are sense receptors that take in light information	No examples in student responses
Brn	Information is processed by the animal's brain. States that the light information taken in by the eyes is processed in the brain	Ne examples in student responses

CCC: Cause and Effect	
Overarching Rule for Cause and Effect: Student states a relationship between two occurrences where one occurrence leads to the other (or needs the other to occur). The language should include linking words such as "because," " and then" (but just having a linking word is not sufficient to get a code of "present" - the linking words have to link the occurrences). If there is a sequence of intermediate events that link the cause and effect, the student states some intermediate events.	

Code	Definition	Example
Lnk	Includes a link between a cause and an effect (the eyes sense light and the brain processes the information or an incorrect link)	(510) P:I think it's A because... well no I think it's B **because the light affects the eyes kind of and then it seems like….. to your eyes in order for the object to be seen immediately.** So I think it's B
Seq	Includes code 1 and a sequence of intermediate events that link the cause and effect, the student states some intermediate events (the light information being sent from the eye to the brain to be processed or an incorrect sequence).	(56) P: I think it might be C. Wait not it is not C. I think it might be D actually. Actually I am going to change it to A. ~~Light is sensed by the brain and then transferred to the eyes.~~ I think the **brain might sense the light** and **then sends it. It produces a picture and then sends it to the eyes**

169

Item 3

Drag each picture into the model to complete the path that light and information follow to allow the plant to be seen.

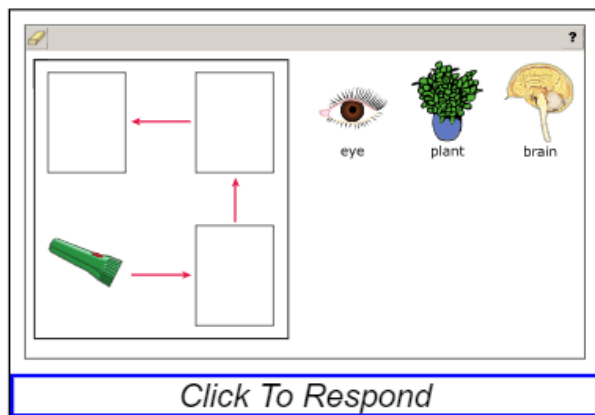| Item	SEP	DCI	CCC
3	Developing and Using Models	PS4.B: An object can be seen when light reflected from its surface enters the eyes. LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain. Animals are able to use their perceptions and memories to guide their actions.	Systems and System Models
Key: plant, eye, brain			

Item 3 Coding Rules

DCI: PS4.B: An object can be seen when light reflected from its surface enters the eyes.

How the DCI is coded in this item: Student states that light must be present for the plant to be seen AND that light must reflect off the plant (Ref) AND that light must enter the eye after reflecting off the plant (Eye).
The language for "reflect" can include
  ● "directs,"
  ● "bounces off of,"
  ● "goes back,"
  ● etc
Code as 0 when.
  ● If only the flashlight and seeing the plant is mentioned
  ● If the order or causal mechanism are incorrect
Non-codable:
  ● If the student repeats the answer options or the prompt verbatim

---

Code	Definition	Example
Ref	Reflection of light off the surface of an object. States that light must be present for the plant to be seen AND the light must reflect off the plant.	(56) P: First the flashlight goes to the plant and **the light bounces off the plant** into the eyes and then it goes up to the brain so it can process the information.
Eye	Light enters the eyes for objects to be seen. States that light must enter the eye after reflecting off plant for plant to be seen.	(59)P: Well to see the plant you have to have a plant. P: and then once the flashlight turns on, your eyes see it next and then to actually process what it is it goes... Like what's happening in your brain. Because you can't really see stuff when it's in your brain because you can't go through your whole body. R: Okay say more about that P: if you turn on a flashlight it's not going to go into your skin and like through your head into your brain. **It has to go through your eyes** because they're open and they're easier to get into. And then that tracks into your brain so that's why I would say like that it goes before the brain.

DCI: LS1.D: Different sense receptors are specialized for particular kinds of information, which may be then processed by the animal's brain.

How the DCI is coded in this item: Student states the eyes are sense receptors that take in light information (Sns) AND that light information taken in by the eyes is processed in the brain (Brn).
The language for "sense" can include
- "feel,"
- "take in,"
- "notice"
- etc

Code as 0 when.
- If only the eyes are mentioned
- If the order or causal mechanism are incorrect

Non-codable:
- If the student repeats the answer options or the prompt verbatim

Code	Definition	Example
Sns	Sense receptors specialized for information.States that the eyes are sense receptors that take in light information	No examples in student responses

Brn	Information is processed by the animal's brain. States that the light information taken in by the eyes is processed in the brain	(57) P: Because the teacher is trying to reflect the light off the plant and then it got into the students' eyes. **And then the brain now tries to process it so that it can be looked at in the brain** and then you can see

SEP: Modeling

Overarching rule for modeling: Student states connections/interactions between the different components of the model (all components are given). For these items, the "arrows" are what is counting as "modeling" because the arrows represent mechanisms by which ….So language for "arrows" could include "leads to" "causes" "and then" .... _____ **Item 3 - has 3 arrows that could be discussed** The description of what the arrow means does not have to be scientifically accurate (e.g., does not need to say "reflect") ● Flashlight to eye (or flipped for all) ● Flashlight to plant ● Flashlight to brain ● Eye to plant ● Eye to brain ● Plant to eye ● Plant to brain *** "And then you can see the plant" is coded as an arrow when the incorrect model is given **There are some students who included different aspects of modeling (e.g., using the idea that the light does not go through the skin as a way to rationalize model) - this is not included in the modeling code, but will be included in the notes section to potentially examine further

Code	Definition	Example
1	Includes what one arrow represents in the model	(58)P: it said that [**it's sense by the brain**] [**and then it goes to the eyes**] /[**and then you can see the plant**]. if it went to the plant and then the eyes and the brain…. it goes to your brain and your brain senses the light and then it goes to your eyes and then your eyes can see the plant. Brain eye plant
2	Includes what two arrows represents in the model	(54)P: So basically you see it with your eyes [**and then it goes to your brain**]/ [**and then you see the plant**]. I don't know. Is it the other way around? I don't know if it is the other way around between the eyes and the brain R: ok so what is make you question that

172

		P: In order..to like see...cause your brain allows you to see stuff. If your blind you basically can't see stuff. So then something is wrong with your brain and you can't see. ~~R: So you are saying that if you are blind there is something wrong with your brain?~~ P: Isn't there like some parts...cause your **eyeball is connected to your brain**. I think it is the other way around Eye, brain, plant *NOTE: Eyeball is connected to your brain is the same "arrow" as "you see it with your eyes and then it goes to your brain"*
3	Includes what three arrows represents in the model	(56) P: First the [**flashlight goes to the plant]** and the[**light bounces off the plant into the eyes]** and **[then it goes up to the brain so it can process the information].**

CCC: Systems and System Modeling
If an item is designed to be aligned to the SEP of modeling and the CCC of Systems and systems models, there is only one code which is an SEP/CCC code.

Item 4

This question has two parts.

Part A

After a while, the lights come back on in the classroom. How will the students' pupils **most likely** change?

(a) The students' pupil diameters will increase.

(b) The students' pupil diameters will decrease.

Part B

Which statement **best** explains the change in pupil diameter described in Part A?

(a) Pupil diameter increases when there is low light.

(b) Pupil diameter increases when there is bright light.

(c) Pupil diameter decreases when there is low light.

(d) Pupil diameter decreases when there is bright light.

Item	SEP	DCI	CCC
4 A&B	Engaging in Argument from Evidence	LS1.A: Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.	Cause and Effect
Key: B;D			

Item 4 Coding Rules

DCI: LS1.A: Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.
How the DCI is coded in this item: Student states that the pupil regulates the amount of light entering the eye as a function to promote, survival,. In this item, this is only seen with some phrases that indicate the function of the pupil is to regulate light due to the body's response system. ● For example: "eyes hurt when the lights come on" ● "The muscles in the eyes make the change…" ● "Pupil needs to open to process light" ● "The pupil's diameter doesn't have to open" Non-codable:

● If the student repeats the answer options or the prompt verbatim	

Code	Definition	Example
Stf	States that the pupil regulates the amount of light entering the eye as a function to promote growth, survival, behavior, and reproduction.	(55)P: I think the students' diameter will decrease because as more light comes in the less the pupil needs to open to process light. P: D because the pupil's diameter...**when there is bright light the pupil's diameter doesn't <u>have to</u> open as much**. So it doesn't open as much

CCC: Cause and Effect
Overarching Rule for Cause and Effect: Student states a relationship between two occurrences where one occurrence leads to the other (or needs the other to occur). The language should include linking words such as "because," " and then" (but just having a linking word is not sufficient to get a code of "present" - the linking words have to link the occurrences). If there is a sequence of intermediate events that link the cause and effect, the student states some intermediate events.

Code	Definition	Example
Lnk	Includes a link between a cause and an effect	(56) P: I think the students' **diameter will decrease because as more light comes in the less the pupil needs to open to process light.** P: D because the pupil's diameter...when there is bright light the pupil's diameter doesn't have to open as much. So it doesn't open as much
Seq	Includes code 1 and a sequence of intermediate events that link the cause and effect, the student states some intermediate events	Seq is not required by the item. Just the link between cause and effect is required.

SEP: Argument from Evidence
Overarching Rule for Argument from Evidence: Students must indicate that they are using (1) evidence given in the item, (2) evidence from prior knowledge or (3) information from other sources within the item cluster. The evidence does not need to be correct. In addition, for reasoning, students must indicate that they are explaining connections between the evidence and the claim. The reasoning does not need to be scientifically accurate, but must be clear that they

are attempting to make a connection		
Code	Definition	Example
Ev	Evidence from item cluster: Students are drawing on the given data in the stimulus or ideas from prior items in the cluster or prior experiences/knowledge to support their claim/answer the question.	(54)P: Isn't it decrease because basically first it (the graph) is all the way up and then it (the graph) goes down. And here (on the x-axis) it says increasing. So basically it is decreasing by the different lights it is showing. So if it is bright light, then it will be this high (indicating height of the bar on graph) but if they do different shades of light basically this light is decreasing. P: Is it D? (referring to Part B answer option)
Rsn	Reasoning: students explain how the evidence they stated or chose supports the claim they stated or chose. (the reasoning must go beyond stating that a relationship to the evidence exists but must attempt to explain the relationship (the "why").	(56) P: I think the students' diameter will decrease because as more light comes in the less the pupil needs to open to process light. P: D because the pupil's diameter...**when there is bright light the pupil's diameter doesn't have to open as much. So it doesn't open as much** (59) P: increases when there is bright light. oh wait. Cuz I said... Cuz it increases when there is low light. Oh yeah it does increase when there's bright light. It decreases because it wouldn't get smaller when it's dark because it just gives it less room to see. ~~Decreases when there's bright light~~...~~It increases when there's low light~~. So not that one...D. because on A it says it ~~increases when there's low light~~. **But it says it explains the change in part A and I said it would decrease when the lights turn back on and since I said it wouldn't be reflecting on part A.** *NOTE from Alicia: Students cannot just say that they are connecting the two parts of the question, they have to explain the connection using science ideas. Even if they are not correct.*

Item 5

Based on all of the information, complete the scientific explanation about the effect of light on pupil diameter.

Drag into the chart **two** evidence statements and **one** reasoning statement that support the claim for the effect of light on pupil diameter.

Claim	The pupil diameter changes with different amounts of light.
Evidence	
Reasoning	

Evidence Statements

The diameter was largest in the lowest light and smallest in the brightest light.

The diameter was largest in the brightest light and smallest in the lowest light.

The diameter of the pupil increased as the light increased.

The diameter of the pupil decreased as the light increased.

Reasoning Statements

When there is less light, the pupil gets bigger to let in more light.

When the pupil is smaller, it lets in more light so a person can see better in less light.

When there is bright light, the pupil lets in more light so a person can see better.

Click To Respond

Item 5 Coding Rules

DCI: LS1.A: Plants and animals have both internal and external structures that serve various functions in growth, survival, behavior, and reproduction.
How the DCI is coded in this item: Student states that the pupil regulates the amount of light entering the eye as a function to promote growth, survival, behavior, and reproduction. In this item, this is only seen with some phrases that indicate the function of the pupil is to regulate light due to the body's response system. • For example: "eyes hurt when the lights come on" • "The muscles in the eyes make the change…" • "Pupil needs to open to process light" • "The pupil's diameter doesn't have to open" Non-codable: • If the student repeats the answer options or the prompt verbatim

Code	Definition	Example
Stf	States that the pupil regulates the amount of light entering the eye as a function to promote growth, survival, behavior, and reproduction.	(58)P: ~~the diameter of the pupil increases when the light decreases~~ this is **so the light doesn't all go into your eyes cuz it's bad for your eyes like when you look at the Sun . So it is not so bright on your eyes.** It decrease as more light came in. Because if it got bigger all of the light would come in. And then for the second one, ~~The diameter was light largest in the lowest light and smallest in the brightest light~~. It's pretty much the same thing. If it's bigger it will let more light in because if it's dark and it was really small then you wouldn't be able to see really because it's really dark And you need more so it gets bigger. And then for the reasoning ~~when there is less light the pupil gets bigger to let in more light~~. It just gets bigger to let in more light.

CCC: Cause and Effect
Overarching Rule for Cause and Effect: Student states a relationship between two occurrences where one occurrence leads to the other (or needs the other to occur). The language should include linking words such as "because," " and then" (but just having a linking word is not sufficient to get a code of "present" - the linking words have to link the occurrences). If there is a sequence of intermediate events that link the cause and effect, the student states some intermediate events.

Code	Definition	Example
Lnk	Includes a link between a cause and an effect (t	(510)P: I answered this one because it says **when there is less light it helps you see better.** It says it changes.
Seq	Includes code 1 and a sequence of intermediate events that link the cause and effect, the student states some intermediate events	Seq is not required by the item. Just the link between cause and effect is required.

SEP: Argument from Evidence
Overarching Rule for Argument from Evidence: Students must indicate that they are using (1) evidence given in the item, (2) evidence from prior knowledge or (3) information from other sources within the item cluster. The evidence does not need to be correct. In addition, for

reasoning, students must indicate that they are making connections between the evidence and the claim. The reasoning does not need to be scientifically accurate, but must be clear that they are attempting to make a connection

Code	Definition	Example
Ev	Evidence from item cluster: Students are drawing on the given data in the stimulus or ideas from prior items or prior experiences/knowledge to support their claim/answer the question.	(59)P: ~~The pupil changes with different amounts of light. It was largest in the lowest of light and smallest in the brightest of light.~~ Yeah that could be one. ~~Largest in the brightest light~~ No it's not the second one. ~~The pupil increased as the light increased.~~ No because it increases with the light decreases. ~~The pupil decreased as the light increased.~~ It got smaller as the light got...**I just used from the last question.** It does the opposite of what the lights doing. Like when it's low light it will increase and then in the bright light it will decrease. P: ~~When there is less light the pupil gets bigger to let more light in.~~ Yeah that one. ~~when the pupil is smaller it lets more light~~…. I don't think it's the second one because it really doesn't go with my evidence. ~~When there's bright light the pupil lets in more light so a person can see better~~. No When there's bright light it makes it darker so there's not too much light. So I'm going to pick the first one.
Rsn	Reasoning: students explain how the evidence they stated or chose supports the claim they stated or chose. (the reasoning must go beyond stating that a relationship to the evidence exists but must attempt to explain the relationship (the "why")	(56) P: Because ~~the diameter was largest in the lowest light and smallest in the brightest light.~~ Like I said when it is brighter the pupil closes. It is smaller so it doesn't have to take in that much light. And the diameter of the pupil decreases as the light increases and when there is less light the pupil gets bigger to let in more light. Yeah, **the pupil gets bigger to let in more light so you can see when it is dark.**

179

Grade 8

Item 1

This question has two parts.

Part A

Use the data table to complete the following statement.

The students can tell that a chemical reaction involving iron
[▾] because a new substance [▾] form
overnight.

Part B

Choose one set of properties that **best** supports the completed
statement in Part A.

(a) density and color

(b) color and volume

(c) volume and texture

(d) texture and mass

(e) mass and density

Item	SEP	DCI	CCC
1 A&B	Analyzing and interpreting data	PS1.B.1: Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants. PS1.A: Each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it.	Patterns
Key: Part A: Occurred/Did; Part B: A - density and color			

Item 1 Coding Rules

DCI: PS1.B.1: Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these new substances have different properties from those of the reactants.

How the DCI is coded in this item: Student explains why density and/or color are the properties they can use to determine a chemical reaction has occurred (Chp). Non-codable: If the student repeats the answer options or the prompt verbatim		
Code	Definition	Example
Chp	Explain why density and/or color are the properties used to determine a chemical reaction has occurred.	83) P: color and volume. **It did gain some volume and mass. The density was a lot lower (1)** so I will have to go with the density and color P: I chose **density and color because the color identifies that it did change (2), the substance did change overnight. That is what the density tells me too. (Chp)**

DCI: PS1.A: Each pure substance has characteristic physical and chemical properties (for any bulk quantity under given conditions) that can be used to identify it.		
How the DCI is coded in this item: Student explains why density and/or color are the properties they can use to determine a chemical reaction has occurred (Chp). Non-codable: If the student repeats the answer options or the prompt verbatim		
Code	Definition	Example
Chp	Explain why density and/or color are the properties used to determine a chemical reaction has occurred.	(83) P: color and volume. **It did gain some volume and mass. The density was a lot lower (1)** so I will have to go with the density and color P: I chose **density and color because the color identifies that it did change (2), the substance did change overnight. That is what the density tells me too. (Chp)**

SEP: Analyzing and Interpreting Data
Overarching rule for analyzing and interpreting data: Student states patterns and relationships in the data and describe why they are meaningful to the investigation question. For these items, language indicating patterns or relationships could include…   - Quantitative or Qualitative description of change presented in data   *Just indicating a "change" happened is not enough for 1*   And for describing why the data is meaningful could include…   - Identifies relationships: Students analyze the data to identify patterns (i.e., similarities and differences), including the changes   - Interpret the data about the properties of each substance before and after the interaction   - Students use data to determine whether a chemical reaction has occurred

	● Students support their interpretation of the data by describing that the change in properties of substances is related to the rearrangement of atoms in the reactants and products in a chemical reaction

Code	Definition	Example
1	Patterns and Relationships: Identifies patterns and relationships that exist in the data	(82) P: **Well this one got bigger and the mass smaller or the density got smaller and the mass actually got bigger and the color change too. (1)** Involving iron.
2	Includes 1 and describes why those patterns are meaningful to the investigation question	(85) I think that because the **color changed from grey to red, the mass went up by 11 grams. Or 9 grams I mean.(1)** The volume went up and so did density so it must be something different. P: Probably mass and density because I know like **when they compare elements like gold and stuff they look at the mass and density.(2)**

CCC: Patterns
If an item is designed to be aligned to the SEP of analyzing and interpreting data and the CCC of Patterns, there is only one code which is an SEP/CCC code.

Item 2

Complete the statements below.

The final mass of the material on the dish the next day is
[▾] the initial mass of the material.

This could happen if
[▾] the
environment.

Item	SEP	DCI	CCC
2	Analyzing and interpreting data	PS1.B.2: The total number of each type of atom is conserved, and thus the mass does not change.	Energy and matter
Key: more than/iron atoms are combining with atoms in			

182

Item 2 Coding Rules

DCI: PS1.B.2: The total number of each type of atom is conserved, and thus the mass does not change.

How the DCI is coded in this item: Student explains why the Law of Conservation supports their chosen response (C).
The language for explaining can include
- "It has to be the same mass before and after"
- "The mass of the air added to the iron"

Non-codable:
- If the student repeats the answer options or the prompt verbatim

Code	Definition	Example
C	Explain why the Law of Conservation supports their chosen response	

SEP: Analyzing and Interpreting Data

Overarching rule for analyzing and interpreting data: Student states patterns and relationships in the data and describe why they are meaningful to the investigation question.
For these items, language indicating patterns or relationships could include…
- Quantitative or Qualitative description of change presented in data
- *Just indicating a "change" happened is not enough for 1*

And for describing why the data is meaningful could include…
- Identifies relationships: Students analyze the data to identify patterns (i.e., similarities and differences), including the changes
- Interpret the data about the properties of each substance before and after the interaction
- Students use data to justify their response
- Students support their interpretation of the data by describing that the change in properties of substances is related to the rearrangement of atoms in the reactants and products in a chemical reaction

Code	Definition	Example
1	Patterns and Relationships: Identifies patterns and relationships that exist in the data	(85) P: ~~The final mass of the material the next day is~~ more ~~than the initial mass of the material this could happen if iron atoms are escaping into the environment iron atoms are combining with matter in the environment iron atoms are~~

		being produced and released into the environment iron atoms are being exchanged in equal amounts with the environment. I don't know what an iron atom is. It's nothing equal because **it's getting bigger,** escaping would mean it would get lighter so I guess it combines.
2	Includes 1 and describes why those patterns are meaningful to the investigation question	(83) P: **Here is more** than the initial mass. This could happen if atoms are combining with matter inside the iron itself. Or there...equal amounts...released into. iron atoms are combining with matter inside the environment. **Due to more of the mass and volume of the material that is left over on the final day it tells me that there is more stuff inside of there but it is less dense.** Or it could be exchanged with equal amounts but with less of a density within the entire object. So it might also be that one. But I could really choose on either or. I am going to have to go with the second answer because I changed my mind quite a bit just thinking about it.

CCC: Patterns

If an item is designed to be aligned to the SEP of analyzing and interpreting data and the CCC of Patterns, there is only one code which is an SEP/CCC code.

Item 3

This question has two parts.

The students decide to develop a model to explain why the iron appeared to change overnight. To complete the model the students need to include:

- iron atoms from the hand warmer
- oxygen atoms from the air
- the final substance they observed

Part A

Complete the student model by dragging the appropriate number of atoms from the key into the model.
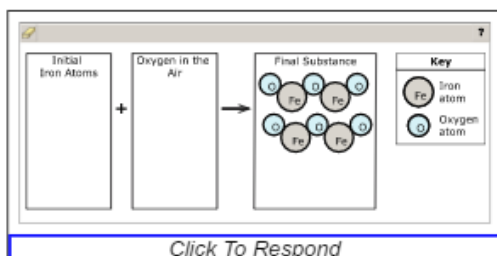
Part B

Select one limitation of the model shown in Part A.

(a) The model does not show conservation of matter.

(b) The model does not show the color change of the final substance.

(c) The model does not show how the atoms are organized in the final substance.

Item	SEP	DCI	CCC
3 A&B	Developing and using models	PS1.B.2: The total number of each type of atom is conserved, and thus the mass does not change.	Energy and matter
Key: Part A: 4&6; Part B: B			

Item 3 Coding Rules

> **DCI: PS1.B.2:** The total number of each type of atom is conserved, and thus the mass does not change.

How the DCI is coded in this item: The student must reference to the number of atoms in the final substance Non-codable: If the student repeats the answer options or the prompt verbatim		

Code	Definition	Example
Atc	Student references the final substance when determining the number of atoms	(81) Alright, so **I am going to put 4 iron atoms since there's one, two, three, four. Then I'll put six oxygen atoms which will make it up the yes the final substance.** That is what I am thinking. ~~The model does not show how the atoms are organized in the final substance.~~ I'm going to say B ~~Does not show the color change of the final substance~~ because it doesn't it only shows like the atomic make up I guess.

SEP: Modeling	

Overarching rule for modeling: The limitations portion of the question (Part B) is the focus of the modeling SEP in this item.

Limitations:
Student must say more than the limitation option they picked. They must explain what is missing in the model that would cause the limitation to be valid or explain why they chose the limitation.

For example:
Student must say more than "the color changed" but must also explain what color change occurred and that it is not shown in the model.
Or
Acknowledging that the actual substance is red but the model shows only grey and blue

Code	Definition	Example
Lim	Student must explain why they chose the particular limitation or articulate what is missing in the model.	(82) I'm putting oxygen atoms into the oxygen in the air box... there's like six of them And then same with the iron. So there's four and then six. (Atc) I'm going to go with B because **there's not really a color. Or a way to show the color change in the model (Lim)**

CCC: Energy and Matter		
Overarching rule for Energy and Matter: Describes how mass and/or energy are conserved in a particular system by including relevant features of the system that demonstrate conservation.		
Code	Definition	Example
EM	Conservation	No examples in Cognitive Lab data

Item 4

Item	SEP	DCI	CCC
4	Analyzing and interpreting data	PS1.B.3: Some chemical reactions release energy, others store energy.	Patterns
Key: Similarities 1&2; Differences 2&4			

Item 4 Coding Rules

DCI: PS1.B.3: Some chemical reactions release energy, others store energy.		
How the DCI is coded in this item: Student explains how they know that the reaction is releasing energy in the form of heat (E). The language for explaining can include ● "It is releasing heat because the temperature is increasing" Non-codable: ● If the student repeats the answer options or the prompt verbatim		
Code	Definition	Example
E	Explain how they know that the reaction is releasing energy in the form of heat	(83) The differences for system two and system one. ~~The temperature of system 2 increased more quickly than system 1~~ so system 1 is the original stock handwarmer that they get out of the bag and let it sit there in the dish but since there is smaller holes to let oxygen in slowly instead of just hitting the gas pedal and pouring all of the fuel into the machine. ~~System two reaches a greater maximum temperature than system one reaches.~~ **So system 2 there is a lot more fuel being burned at one time so it allows the**

188

		temperature to rise a lot more than system one. **System one slowly burns that fuel.**

SEP: Analyzing and Interpreting Data

Overarching rule for analyzing and interpreting data: Student states patterns and relationships in the data and describe why they are meaningful to the investigation question. For these items, language indicating patterns or relationships could include…
- Quantitative or Qualitative description of change presented in data
- *Just indicating a "change" happened is not enough for 1*

And for describing why the data is meaningful could include…
- Identifies relationships: Students analyze the data to identify patterns (i.e., similarities and differences), including the changes
- Interpret the data about the properties of each substance before and after the interaction
- Students use data to justify their response
- Students support their interpretation of the data by describing that the change in properties of substances is related to the rearrangement of atoms in the reactants and products in a chemical reaction

Code	Definition	Example
1	Patterns and Relationships: Identifies patterns and relationships that exist in the data	(86) I would say for the two differences ~~system 1 reaches a greater maximum temperature.~~ system 1 I think this is system 1. Oh no I messed up I should have checked I didn't see that. So ~~system 2 reach has a greater maximum temperature in system 1~~. And I would also say ~~the temperature system 2 increases more quickly than the temperature system 1~~ **because system 1 was constant at 110 degrees Fahrenheit.**
2	Includes 1 and describes why those patterns are meaningful to the investigation question	(83) System 1 is completely enclosed in its original packaging. Temperature. Ok so system 2. I am guessing...yeah...drops a lot faster because system 1...it's slowly letting that oxygen in so it will have a lot more run time compared to system 2 burning all of its fuel and dropping. Two similarities and two differences. ~~System 2 has a greater maximum temperature than system one~~ **because it burns all of its fuel more at one time increasing the temperature of the actual model itself. It does not remain constant.** ~~It decreases in both systems eventually~~

189

| | | over time. I didn't know that this graph showed that until I really looked at it. |
| | | The differences for system two and system one. ~~The temperature of system 2 increased more quickly than system 1~~ so system 1 is the original stock handwarmer that they get out of the bag and let it sit there in the dish but since there is smaller holes to let oxygen in slowly instead of just hitting the gas pedal and pouring all of the fuel into the machine. |

CCC: Patterns

If an item is designed to be aligned to the SEP of analyzing and interpreting data and the CCC of Patterns, there is only one code which is an SEP/CCC code.

Item 5

Identify and explain the temperature pattern in the graph.
Select and move the **best** claim statement, evidence
statement, and reasoning statement that can be used to
explain the pattern.

Claim	
Evidence	
Reasoning	

Claim Statements:
Energy is transferred from each system to the thermometers.
Energy is transferred from the thermometers to each system.
Energy is not transferred in either of the systems.

Evidence Statements:
The temperature was higher at 50 minutes than at 0 minutes.
The temperature was lower at 50 minutes than at 0 minutes.
The temperature did not change between 0 minutes and 50 minutes in either system.

Reasoning Statements:
The iron absorbed energy from the oxygen during the chemical reaction.
The hand warmer bag absorbed energy from the thermometer.
Energy was released when the iron reacted with the oxygen in the air.
Energy was released when the hand warmer package was opened.

OK

Item	SEP	DCI	CCC
5	Constructing Explanations	PS1.B.3: Some chemical reactions release energy, others store energy.	Energy and matter
Key: C: Energy is transferred from each system to the thermometers. (1) E: The temperature was higher at 50 minutes than at 0 minutes. (1) R: Energy was released when iron reacted with the oxygen in the air.(3)			

Item 5 Coding Rules

DCI: PS1.B.3: Some chemical reactions release energy, others store energy.
How the DCI/CCC is coded in this item: Students have to identify that energy is released from the system The language for this can include: "The system got hotter" "The temperature went up or was higher"

"Heat was released"
- Reasoning and the DCI evidence may be the same for some cases

Non-codable:
- If the student repeats the answer options or the prompt verbatim

Code	Definition	Example
Re	Students identify that energy is released form the system in the form of heat	(83) I am guessing so...after 50 minutes on both. **Well system one is a lot higher than system 2. (Re)** Is this for system 1 or system 2? I am guessing it is just in general.

SEP: Constructing Explanations

Overarching Rule for Constructing Explanations: Students must indicate that they are using (1) evidence given in the item, (2) evidence from prior knowledge or (3) information from other sources within the item cluster. The evidence does not need to be correct. In addition, for reasoning, students must indicate that they are making connections between the evidence and the claim. The reasoning does not need to be scientifically accurate, but must be clear that they are attempting to make a connection NOTE: Same as Argument from Evidence in Grade 5 Item 5

Code	Definition	Example
Ev	Evidence from item cluster: Students are drawing on the given data in the stimulus or ideas from prior items or prior experiences/knowledge to support their claim/answer the question.	(85) Evidence statements. ~~Temperature was higher at 50 minutes than at 0 minutes.~~ In which one? **There's two bags. Actually, 50 minutes it is like 90 degrees. 0 minutes it is like 70 degrees (Ev)**. So yeah, that's not true. That is not true either.
Rsn	Reasoning: students explain how the evidence they stated or chose supports the claim they stated or chose. (the reasoning must go beyond stating that a relationship to the evidence exists but must attempt to explain the relationship (the "why")	(83) the energy was released when the hand warmer package was opened because the **oxygen gets to it as you open the package which allows it to kind of heat up and make that chemical reaction (Re/Rsn)** *NOTE from Alicia: Students cannot just say that they are connecting the two parts of the question, they have to explain the connection using science ideas. Even if they are not correct.*

Item 6

The students want to redesign the hand warmer bag in such a way that the hand warmer would take less time to get warm but would also remain a safe product. The new design includes more small holes in the fabric of the hand warmer bag.

Which statement **best** evaluates if the new design **would** or **would not** work?

(a) The new design would not work because the additional holes in the bag would allow oxygen from the air to move into the bag and cool the iron.

(b) The new design would not work because the only way to increase the temperature is to generate more energy by adding iron.

(c) The new design would work because when the iron in the hand warmer is exposed to more oxygen in the air, the temperature will increase at a faster rate.

(d) The new design would work because when pieces of iron are able to move out of the additional holes, they will be in contact with more oxygen, which will quickly increase the temperature.

Item	SEP	DCI	CCC
6	Designing Solutions	ETS1.B: A solution needs to be tested, and then modified on the basis of the test results, in order to improve it.	Energy and matter
Key: C			

Item 6 Coding Rules

DCI: ETS1.B: A solution needs to be tested, and then modified on the basis of the test results, in order to improve it.
How the DCI/SEP is coded in this item: Students need to discuss how their response supports the criteria, the handwarmer needs to get warmer faster. Code as 0 when. ● If the student repeats the answer options verbatim or the prompt verbatim, this is considered non-codable portions of the transcript and are not considered here.

Code	Definition	Example
Sol	Students need to discuss how their response supports the criteria, the handwarmer needs to get warmer faster.	No examples in the Cognitive Lab Data

SEP: Designing Solutions: Undertake a design project, engaging in the design cycle, to construct and/or implement a solution that meets specific design criteria and constraints.

If an item is designed to be aligned to the DCI ETS1.B and the SEP of Designing Solutions, there is only one code which is a DCI/SEP code.

CCC: Energy and Matter: The transfer of energy can be tracked as energy flows through a designed or natural system.

Overarching Coding Rule: Student states a path that energy takes from one component of a system to another indicating the changes in forms of energy at various points in the system. The language should include words such as "heat," "temperature increase or decrease"

Code	Definition	Example
Trn	Student discusses the heat transfer in the system.	(83) **It would get really hot really fast but it wouldn't be as consistent it would slowly decline over the 80 minute mark** they had marked compared to the system 1 which is the original hand warmer that the company has designed in which it gradually goes up slowly and stays somewhat consistent throughout that time.

Item 7

Which statement **best** describes a trade-off if a new hand warmer bag is designed to get warm faster than the original design?

(a) The hand warmer bag would be more expensive than the original design.

(b) The hand warmer bag would contain more iron and be bigger than the original design.

(c) The hand warmer bag would not get as warm as the original design.

(d) The hand warmer bag would cool down faster than the original design.

Item	SEP	DCI	CCC
7	Designing Solutions	ETS1.C.1: Although one design may not perform the best across all tests, identifying the characteristics of the design that performed the best in each test can provide useful information for the redesign process - that is, some of the characteristics may be incorporated into the new design.	n/a
Key: D			

Item 7 Coding Rules

DCI: ETS1.C.1: Although one design may not perform the best across all tests, identifying the characteristics of the design that performed the best in each test can provide useful information for the redesign process - that is, some of the characteristics may be incorporated into the new design.
How the DCI/SEP is coded in this item: Students discuss how the redesign of the handwarmer impacts the trade-offs they considered. Non-codable: ● If the student repeats the answer options or the prompt verbatim

Code	Definition	Example

ReD	Students discuss how the redesign of the handwarmer impacts the trade-offs they considered.	No examples in the Cognitive Lab Data

SEP: Designing Solutions: Undertake a design project, engaging in the design cycle, to construct and/or implement a solution that meets specific design criteria and constraints.

If an item is designed to be aligned to the DCI ETS1.B and the SEP of Designing Solutions, there is only one code which is a DCI/SEP code.

APPENDIX F: CONSENT FORMS

Research Participant Information and Consent Form

EXPLANATION OF THE RESEARCH

Study Title: Examining Content Validity for a Three-Dimensional State Science Assessments

Your child is being asked to participate in a research study of how students interact with the Michigan M-STEP Science Assessment.

The researcher will be meeting with your child to conduct an online assessment study session called a Cognitive Lab. Your child will have the chance to see one item cluster (a passage and group of items) on a computer. Your child will answer each of the questions, but these questions will not be scored or graded in any way and no decisions or judgments will be made about your knowledge or skills. While your child is answering each question, the researcher will be audio recording the conversation and making notes about your experience with the questions on the computer. In addition, your child will be asked some questions so that you can provide important feedback about the test questions and what he/she/they liked and didn't like. This will take approximately 30 minutes of his/her/their time. I hope that your child will enjoy giving opinions and sharing ideas with me about the test. What your child thinks about these online/computer sample test questions will help provide important information for the researcher, which will be used to better understand the interactions between students and the Michigan M-STEP Science Assessment.

YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW:

Your child does not have to participate in this study. It is up to him/her/them. Your child can say no now, or he/she/they can even change your mind later. No one will be upset with him/her/them if he/she/they decide not to be in this study.

Your child's grades and relationship with his/her/their school, teachers, and classmates, will not be affected if he/she/they choose to not participate in the study or if he/she/they choose to stop participating at any point. If he/she/they choose to not participate, he/she/they can stop at any time.

COSTS AND COMPENSATION FOR BEING IN THE STUDY

Being in this study will bring your child no harm. There are no direct benefits to your child for participating in this study. It will hopefully help us learn more about the things your child thinks about while taking the M-STEP Science Assessment.

CONTACT INFORMATION FOR QUESTIONS AND CONCERNS

If you or your child have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury, please contact the researcher Tamara (Heck) Smolek at (517) 706-9130, or smolekt@michigan.gov.

If you or your child 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.

DOCUMENTATION OF INFORMED CONSENT

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

_____
Student's Name

_____   _____
Parent/Guardian's Signature   Date

Participant Assent Form

I am from Michigan State University and I am asking you to be in a research study. We do research studies to learn more about how the world works and why people act the way they do. In this study, we want to learn about how students interact with the Michigan M-STEP Science Assessment questions.

What we are asking you to do:
We would like to ask you to take a M-STEP science item cluster and talk out loud as you answer the questions. This will take about 30 minutes. You can skip any question if it makes you uncomfortable.

Do I have to be in this study?
You do not have to participate in this study. It is up to you. You can say no now, or you can even change your mind later. No one will be upset with you if you decide not to be in this study.

Your grades and your relationship with your school, teachers and classmates will not be affected if you choose to not participate in the study or if you choose to stop participating at any point. If you do not participate, you can stop at any time.

Will being in this study hurt or help me in any way?
Being in this study will bring you no harm. There are no direct benefits to you for participating in this study. It will hopefully help us learn more about the things you think about while taking the M-STEP Science Assessment.

What will you do with information about me?
We will be very careful to keep your answers to the assessment questions private. Before and after the study we will keep all information we collect about you locked up and password protected.

If you want to stop doing the study, contact Tamara (Heck) Smolek at 517-706-9130 or smolekt@michigan.gov. If you choose to stop before we are finished, any answers you already gave will be destroyed. There is no penalty for stopping.

If you have questions about the study, contact:
Tamara (Heck) Smolek
517-706-9130
smolekt@michigan.gov

If you have questions about your rights in the study, contact:
Human Research Protection Program Institutional Review Board
Michigan State University
Phone number: 517-355-2180
Email address: irb@ora.msu.edu

Agreement:
By signing this form, I agree to be in the research study described above.
Name: _____
Signature: _____ **Date:** _____
You will receive an electronic copy of this form.

Science Task Screener

Introduction

The purpose of the Next Generation Science Standards (NGSS) Task Screener is **1)** to determine whether classroom assessment tasks are high quality, designed to elicit evidence of three-dimensional performances, and designed to support the purpose for which they will be used, and **2)** to provide a group of reviewers with a common set of features to ground conversations about what it "looks like" for students to demonstrate the kinds of performances expected by three-dimensional standards. This Screener builds off the criteria in Category III of the EQuIP Rubric for Science by more clearly specifying features for the assessment tasks embedded in lessons and units.

The directions for using the Task Screener assume an understanding of *A Framework for K–12 Science Education* and the NGSS, including how the NGSS are different from past standards as outlined in Appendix A of the NGSS and the Innovations of the NGSS. The Task Screener focuses on determining whether what is new and different about three-dimensional expectations are accurately represented in the tasks being evaluated. For more information about how the Task Screener was developed and fits into the EQuIP suite of tools, please see these Frequently Asked Questions.

Task Screener Overview

The Task Screener is organized around four criteria:

A. Tasks are driven by **high-quality scenarios that focus on phenomena or problems.**

B. Tasks require **sense-making using the three dimensions.**

C. Tasks are **fair and equitable.**

D. Tasks support their **intended targets and purpose.**

Each criterion includes:

1. A set of indicators to help reviewers determine whether the criterion is met.

2. A set of response forms for gathering and analyzing evidence, providing suggestions for improvement, and rating the task.

To use the Task Screener effectively, users should use the indicators and response forms to collect specific and detailed evidence from the task under review. Then, users should consider the body of evidence to determine how well each criterion is addressed within the task.

While it is possible for the Screener to be applied by an individual, the real power of the Task Screener lies in the meaningful conversations it can drive among a team of reviewers as part of a collaborative process. Just as when using other resources in the EQuIP suite of tools, collaborative teams of users should:

1. Individually record criterion-based evidence using the provided response forms;

2. Individually make suggestions for improvement; and then

3. Collaboratively discuss findings with team members before checking one of the boxes under the "Evidence of Quality?" section included at the end of the screening process. As part of these discussions, reviewers should address any differences in how they interpreted the criteria and indicator language, as well as the evidence they found, to support a common understanding of the task, the expectations outlined in the screener, and how well the task met those expectations. A rating of "Adequate" means that the task meets the criterion. If the collaborative feedback is being used to improve the task or make decisions about how it should be used, use a blank set of response sheets to capture the consensus feedback.

1

Science Task Screener

Using the Task Screener. Use this tool to evaluate tasks designed for three-dimensional standards. For each criterion, record your evidence for the presence or absence of the associated indicators. After you have decided to what degree the indicators are present within the task, revisit the purpose of your task and decide whether the evidence supports using it.

Before you begin: Complete the task as a student would. Then, consider any support materials provided to teachers or students, such as contextual information about the task and answer keys/scoring guidance.

A. Tasks are driven by high-quality scenarios that are grounded in phenomena or problems.	B. Tasks require sense-making using the three dimensions.
i. Making sense of a phenomenon or addressing a problem is necessary to accomplish the task. ii. The task scenario—grounded in the phenomena and problems being addressed—is sufficient, engaging, relevant, and accessible to a wide range of students.	i. Completing the task requires students to use reasoning to sense-make about phenomena or problems. ii. The task requires students to demonstrate grade-appropriate: a. SEP element(s) b. CCC element(s) c. DCI element(s) iii. The task requires students to integrate multiple dimensions in service of sense-making and problem-solving. iv. The task requires students to make their thinking visible.

C. Tasks are fair and equitable.	D. Tasks support their intended targets and purpose.
i. The task provides ways for students to make connections of meaningful local, global, or universal relevance. ii. The task includes multiple modes for students to respond to the task. iii. The task is accessible, appropriate, and cognitively demanding for all learners, including students who are English learners or are working below or above grade level. iv. The task cultivates or explicitly builds upon students' interest in and confidence with science and engineering. v. The task focuses on performances for which students' learning experiences have prepared them (opportunity to learn considerations). vi. The task uses information that is scientifically accurate.	i. The task assesses what it is intended to assess, and supports the purpose for which it is intended. ii. The task elicits student artifacts that provide evidence of how well students can use the targeted dimensions together to make sense of phenomena and design solutions to problems. iii. Supporting materials include clear answer keys, rubrics, and/or scoring guidelines that are connected to the targeted three-dimensional standards and provide the necessary and sufficient guidance for interpreting student responses relative to all three dimensions and the target as a whole. iv. The task's prompts and directions provide sufficient guidance for the teacher to administer it effectively and for the students to complete it successfully while maintaining high levels of students' analytical thinking as appropriate.

Criterion A.
Tasks are driven by high-quality scenarios that are grounded in phenomena or problems.

Tasks designed for the NGSS include clear and compelling evidence that:	What was in the task, where was it, and why is this evidence?				
i. Making sense of a phenomenon or addressing a problem is necessary to accomplish the task.	1) Is a phenomenon and/or problem present? 2) Is information from the scenario necessary to respond successfully to the task?				
ii. The task scenario is engaging, relevant, and accessible to a wide range of students.	**Features of engaging, relevant, and accessible tasks (Check the appropriate box, then describe rationale with evidence)**				

Features of scenarios	Yes	Somewhat	No	Rationale
Scenario presents real-world observations				
Scenarios are based around at least one specific instance, not a topic or generally observed occurrence (e.g., observations related to a specific hurricane rather than "hurricanes" in general)				
Scenarios are presented as puzzling/intriguing				
Scenarios create a "need to know"¹				
Scenarios are explainable using grade-appropriate SEPs, CCCs, DCIs				

3

Features of scenarios	Yes	Somewhat	No	Rationale
Scenarios effectively use at least 2 modalities (e.g., images, diagrams, video, simulations, textual descriptions)				
If data are used, scenarios present real/well-crafted data				
The local, global, or universal relevance of the scenario is made clear to students[2]				
Scenarios are comprehensible to a wide range of students at grade-level				
Scenarios use as many words as needed, no more				
Scenarios are sufficiently rich to drive the task				

Across all indicators, there is _____ evidence of quality of this criterion (choose one).

No

Inadequate

Adequate

Extensive

1. When considering whether the scenario creates a need to know for students, consider whether the scenario makes the uncertainty associated with explaining a phenomenon or solving a problem central, in ways that are likely to 1) connect with students' own experiences or knowledge, and 2) connect to disciplinary core ideas (regardless of whether those ideas are explicitly named or required by the task).

2. Consider whether an authentic stakeholder group is interested in the outcome of the scenario, and/or whether students are given enough information to answer the question "why should I care?".

4

Suggestions for improvement of the task for Criterion A:

5

Tasks designed for the NGSS include clear and compelling evidence that:	What was in the task, where was it, and why is this evidence?	
i. Completing the task requires students to use reasoning to sense-make about phenomena or problems.	*Consider in what ways the task requires students to use reasoning to engage in sense-making and/or problem solving.*	
ii. The task requires students to demonstrate grade-appropriate: • SEP element(s) • CCC element(s) • DCI element(s)	Evidence of **SEPs** (which element [s], and how does the task require students to demonstrate this element in use?)	
	Evidence of **CCCs** (which element [s], and how does the task require students to demonstrate this element in use?)	
	Evidence of **DCIs** (which element [s], and how does the task require students to demonstrate this element in use?)	
iii. The task requires students to integrate multiple dimensions in service of sense-making and/or problem-solving.	*Consider in what ways the task requires students to use multiple dimensions together to sense-make and/or problem-solve.*	
iv. The task requires students to make their thinking visible.	*Consider in what ways the task explicitly prompts students to make their thinking visible. Look for evidence of how the task surfaces current understanding, abilities, gaps, and problematic ideas.*	

6

Across all indicators, there is _____ evidence of quality of this criterion (choose one).

No

Inadequate

Adequate

Extensive

Suggestions for improvement of the task for Criterion B:

NEXT GENERATION SCIENCE STANDARDS TASK SCREENER VERSION 1.0

Criterion C.
Tasks are fair and equitable.

Tasks designed for the NGSS include clear and compelling evidence of the following:	What was in the task, where was it, and why is this evidence?				
i. The task provides ways for students to make connections of local, global, or universal relevance.	*Consider specific features of the task that enable students to make local, global, or universal connections to the phenomenon/problem and task at hand. Note: This criterion emphasizes ways for students to find meaning in the task; this does not mean "interest." Consider whether the task is a meaningful, valuable endeavor that has real-world relevance—that some stakeholder group locally, globally, or universally would be invested in.*				
ii. The task includes multiple modes for students to respond to the task.	*Describe what modes (written, oral, video, simulation, direct observation, peer discussion, etc.) are expected/ possible for student responses.*				
iii. The task is accessible, appropriate, and cognitively demanding for all learners, including students who are English learners or are working below or above grade level.	*Consider how the task supports all learners, including:*				
		Yes	Somewhat	No	Rationale
	Task includes appropriate scaffolds				
	Tasks are coherent from a student perspective				
	Tasks respect and advantage students' cultural and linguistic backgrounds				

3- For more information about culturally and linguistically responsive classroom assessments, please see this resource.

Tasks designed for the NGSS include clear and compelling evidence of the following:	What was in the task, where was it, and why is this evidence?				
iii. *(continued)*		Yes	Somewhat	No	Rationale
	Tasks provide both low- and high-achieving students with an opportunity to show what they know				
	Tasks use accessible language				
iv. The task cultivates students' interest in and confidence with science and engineering.	*Consider how the task cultivates students interest in and confidence with science and engineering, including opportunities for students to reflect their own ideas as a meaningful part of the task; make decisions about how to approach a task; engage in peer/self-reflection; and engage with tasks that matter to students.*				
v. The task focuses on performances for which students' learning experiences have prepared them (opportunity to learn considerations).	*Consider the ways in which provided information about students' prior learning (e.g., instructional materials, storylines, assumed instructional experiences) enables or prevents students' engagement with the task and educator interpretation of student responses.*				

9

Tasks designed for the NGSS include clear and compelling evidence of the following:	What was in the task, where was it, and why is this evidence?
vi. The task presents information that is scientifically accurate.	*Describe evidence of scientific inaccuracies explicitly or implicitly promoted by the task.*

Across all indicators, there is _____ evidence of quality of this criterion (choose one).

No

Inadequate

Adequate

Extensive

Suggestions for improvement of the task for Criterion C:

208

Criterion D.
Tasks support their intended targets and purpose.

Before you begin:

1. Describe what is being assessed. Include any targets provided, such as dimensions, elements, or PEs. :

2. What is the purpose of the assessment? (check all that apply)

Formative (including peer and self-reflection)

Summative

Determining whether students learned what they just experienced

Determining whether students can apply what they have learned to a similar but new context

Determining whether students can generalize their learning to a different context

Other (please specify) _____

Tasks designed for the NGSS include clear and compelling evidence that:	What was in the task, where was it, and why is this evidence?
i. The task assesses what it is intended to assess and supports the purpose for which it is intended.	*Consider in what ways:* *1) The assessment target is necessary to respond to the task.* *2) Any ideas, practices, or experiences **not** targeted by the assessment are necessary to respond to the task. Consider the impact this has on students' ability to complete the task and interpretation of student responses.* *3) The student responses elicited support the purpose of the task (e.g., if a task is intended to help teachers determine if students understand the distinction between cause and correlation, does the task support this inference?).*

11

NEXT GENERATION SCIENCE STANDARDS TASK SCREENER VERSION 1.0

Tasks designed for the NGSS include clear and compelling evidence that:	**What was in the task, where was it, and why is this evidence?**
ii. The task elicits artifacts from students as direct, observable evidence of how well students can use the targeted dimensions together to make sense of phenomena and design solutions to problems.	*Consider what student artifacts are produced and how these provide students the opportunity to make visible their **1)** sense-making processes, **2)** thinking across all three dimensions, and **3)** ability to use multiple dimensions together[note: these artifacts should connect back to the evidence described for Criterion B].*
iii. Supporting materials include clear answer keys, rubrics, and/or scoring guidelines that are connected to the three-dimensional target. They provide the necessary and sufficient guidance for interpreting student responses relative to the purpose of the assessment, all targeted dimensions, and the three-dimensional target.	*Consider how well the materials support teachers and students in making sense of student responses and planning for follow up (grading, instructional moves), consistent with the purpose of and targets for the assessment. Consider in what ways rubrics include:* *1) Guidance for interpreting student thinking using an integrated approach, considering all three dimensions together as well as calling out specific supports for individual dimensions, if appropriate:* *2) Support for interpreting a range of student responses, including those that might reflect partial scientific understanding or mask/misrepresent students' actual science understanding (e.g., because of language barriers, lack of prompting or disconnect between the intent and student interpretation of the task, variety in communication approaches):* *3) Ways to connect student responses to prior experiences and future planned instruction by teachers and participation by students:*

210

Tasks designed for the NGSS include clear and compelling evidence that:	What was in the task, where was it, and why is this evidence?
iv. The task's prompts and directions provide sufficient guidance for the teacher to administer it effectively and for the students to complete it successfully while maintaining high levels of students' analytical thinking as appropriate.	*Consider any confusing prompts or directions, and evidence for too much or too little scaffolding/supports for students (relative to the target of the assessment—e.g., a task is intended to elicit student understanding of a DCI, but their response is so heavily scripted that it prevents students from actually showing their ability to apply the DCI).*

Across all indicators, there is _____ evidence of quality of this criterion.

No

Inadequate

Adequate

Extensive

Suggestions for improvement of the task for Criterion D:

Consider the task purpose and the evidence you gathered for each criterion. Carefully consider the purpose and intended use of the task, your evidence, reasoning, and ratings to make a summary recommendation about using this task. While general guidance is provided below, it is important to remember that the intended use of the task plays a big role in determining whether the task is worth students' and teachers' time.

Final recommendation

Use this task (all criteria had at least an "adequate" rating)

Modify and use this task

Do not use this task

14

APPENDIX H: TASK ANNOTATION PROJECT IN SCIENCE (TAPS) ANALYSIS

Grade 5 Item 1: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking.	The information in the scenario is not necessary to answer the question. The stated CCC is not measured in the item and very little reasoning is required. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 2: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate.	The information in the scenario is not necessary to answer the question. The DCI is not used in service of sensemaking. The stated CCC is not measured in the item and very little reasoning is required. The item did not require sensemaking because the response is very close to the DCI and could be rote. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 3: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking.	The information in the scenario is not necessary to answer the question. The SEP is not measured. The stated CCC is not measured. The item requires a visualization of the DCI but does not assess the SEP. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 4: TAPS Findings

Strengths	Improvement Opportunities

The information in the scenario is necessary to answer the item. A substantial portion of the SEP is required to answer the question.	The SEP is different from the identified SEP. It is measured below grade-level and is not used in service of sensemaking because students are expected to read the graph but do not have to apply any ideas from it. The stated DCI is not measured. The stated CCC is not measured. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 5: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP is required to answer the question. A substantial portion of the DCI is required to answer the question and is grade appropriate. The DCI is used in service of sensemaking. The students must connect the data and their understanding that light is needed to see. Multiple dimensions are used together. The item measures what is intended.	The SEP is measured below grade-level. The CCC is not measured.

Grade 8 Item 1: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required. Overall, the item does assess what it is intended to assess.	The stated CCC is not measured.

Grade 5 Item 2: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required, however not at grade level. Overall, the item does assess what it is intended to assess.	The SEP is not engaged at grade level. The stated CCC is not measured.

Grade 5 Item 3: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required. Overall, the item does assess what it is intended to assess.	The application of the DCI is at a low level. The stated CCC is not measured.

Grade 5 Item 4: TAPS Findings

Strengths	Improvement Opportunities

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. The SEP and CCC are both engaged in this item. Multiple dimensions are used together and sensemaking or problem solving is required.	The SEP and CCC are not engaged at the appropriate grade level. The DCI is not measured. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 5: TAPS Findings

Strengths	Improvement Opportunities
The information in the scenario is necessary to answer the item. A substantial portion of the SEP and DCI is required to answer the question and is used in service of sensemaking. Multiple dimensions are used together and sensemaking or problem solving is required.	The stated SEP is not measured with the item, rather the reviewers suggested Analyzing and Interpreting Data was being assessed. They argued that selecting options for a CER is not engaging in the cited SEP. The stated CCC is not measured. Overall, the item does assess what it is intended to assess.

Grade 5 Item 6: TAPS Findings

Strengths	Improvement Opportunities
A substantial portion of the DCI is required to answer the question and is used in service of sensemaking.	The information in the scenario is not necessary to answer the item. The stated SEP is not measured because students do not have to evaluate the design to answer the question. The stated CCC is not measured. Multiple dimensions are not used together and sensemaking or problem solving is not required. Overall, the item does not assess what it is intended to assess.

Grade 5 Item 7: TAPS Findings
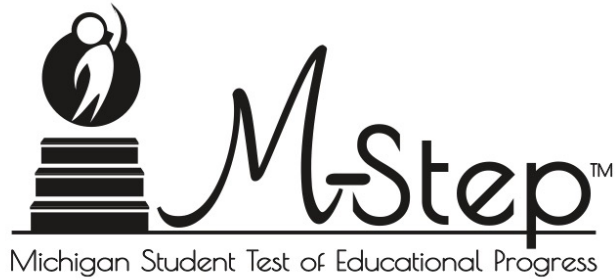
Strengths	Improvement Opportunities

A substantial portion of the DCI is required to answer the question and is used in service of sensemaking. An additional alignment to PS1.B.3 is warranted.	The information in the scenario is not necessary to answer the item. The stated SEP is not measured. The stated CCC is not measured. Multiple dimensions are not used together and sensemaking or problem solving is not required. Overall, the item does not assess what it is intended to assess.

M-Step™

Michigan Student Test of Educational Progress

2017
Science Pilot –
Cognitive Lab

Grade 11

Form 3

Atmosphere Cluster

Facilitator Name: _____

Student Name: _____

Student Grade: _____

School: _____

Date/Time: _____

Cog Lab Overview

The science pilot directions and a screen shot of each item that the student will be working on are included in this booklet. The screen shots of the items will show you what the student sees on the computer screen. As the items are administered on the computer, you will need to document the student's answers. While the student works through the items, you will document your observations in this booklet.

At the beginning of the cognitive lab, you will be provided with a script of directions to read aloud to the student and features to point out to the student. There are no accommodations for this pilot assessment, therefore, text-to-speech, masking, color contrasting, and color choice are not available.

Remember:

- Sit behind or next to the student, but not in their personal space.

- Do not interrupt the student to ask why he or she is responding in a particular way. Only prompt if they are not describing what they are thinking.

- **DO NOT** tell the student if he/she is right or wrong.

Description of Cognitive Interview Process

Step 1. The student begins by reading the Stimulus aloud. Ask the two questions provided and record the student's comments.

Step 2. For the Item, ask the student to read the question aloud, remind the student to tell what he/she is thinking. If the student is not talking, you may want to ask the question provided. Mark the student responses in the spaces provided.

Step 3. The student moves onto the next task. *[Repeat Steps 1 and 2]*

Step 4. At the end of the science items, there are a few items that are survey questions. The student will respond to those questions and you will have some follow-up questions. Record the student responses in the spaces provided.

· · · · · 2 · · · · ·

219

Directions for Administration – Grade 11 Form 3 ES: Atmosphere

*Before beginning this test with the student, log in and be sure the "**Display Settings Test**" screen is active on the computer.*

*Then, using the following script, read aloud anything that is printed in red and preceded by **[SAY]**.*

[SAY] Thank you for helping me today. My name is _____. As your teacher probably told you, I am here because I am interested in learning about what students think about different kinds of test questions. I have found that the best way to find out what people are thinking about is to have them think out loud while they work on the questions. This will help us design better test questions.

As you answer these questions, I would like you to say out loud everything that you normally think to yourself while solving the questions. Just act as if you are alone in the room speaking to yourself. After you finish each question, I will ask you a few more questions about how you solved it.

Even though this is called a test, your test answers will not be graded, and this test will not affect your grade in your class. Don't worry about getting the questions right or wrong. We are really interested in figuring out how we can make the test questions better for students.

Do you have any questions before we begin?

> *Pause to answer questions. Proceed when there are no more questions.*

Think Aloud Practice

[SAY] To give you some practice with thinking out loud, I'd like you to think out loud while solving a math problem. I'll do one first.

I'm going to add 56 + 79.

> *[Interviewer talks out loud about what they're thinking while solving the problem (e.g., First, I add 6 plus 9 and I get fifteen, so I'm putting a 5 down here and I'm going to carry the 1. Then, I'm going to add 1 plus 5 plus 7 and I get 13. So the answer is 135.]*

[SAY] Now you try one. 68 + 26

After the participant finishes, give her/him feedback about talking aloud if needed.

Answer any questions the student may have.

*Then, make sure the student is now viewing the computer "**Display Setting Test**" screen, and continue this script, while the student follows along through the screens on the computer.*

Note that some screens should NOT be read aloud, which is indicated in this script.

Display Settings Test *(**DO NOT** READ SCREEN)*

On the next several screens you will see information for the assessment. Remember, even thought it's called a test, you will not be graded. Some of these screens will not be important for you and the work we are doing today. Others will help you understand the online tasks.

Test Security *(**DO NOT** READ SCREEN)*

Although most of the information on this screen does not apply to our task today, one thing is important. You should not talk about the questions you will see today to anyone.

> *Pause to answer questions. Proceed when there are no more questions.*

[SAY] Now, select the **Next** arrow at the bottom of your screen.

Introduction

Read these directions carefully before beginning. To look at these directions again, select the **Help** button and choose the **Test Directions** tab.

Now, select the **Next** arrow at the bottom of your screen.

Answering Questions

- This test has passages and pictures that you will read and use to answer different types of questions.
- Carefully read each passage and look at each picture before answering the questions.

- Be sure the **Pointer** tool is selected and choose your answer. Some questions may require more than one answer.

- The **Pointer** is already selected for you when you select the **Next** arrow and go to the next question.

- To change an answer, use the **Pointer** tool to choose a different answer.

- For questions that ask you to type your response or show your answer, use the keyboard or the online tools for that question to provide your response.

- For some questions you will be required to drag words, numbers or objects from one place on the screen and drop them at another location in order to show your answer.

- Some questions will have a **Click to Respond** button. Selecting this button will open a larger version of what is seen in the small window. This is where you will enter your response.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Navigation

After you have answered a question, select the **Next** arrow at the bottom of the screen to go to the next question.

Sometimes you will need to use a **scroll bar** to see the entire part of a passage or question. Scroll up or down using the scroll bar on the right side of a passage or question.

Use the **Back** arrow to go back to a question you have already seen or answered. You can also go back to any picture or passage in the set using the **Back** arrow.

To move quickly to any question on the test, select the **Down** arrow next to the question number and select the question you'd like to see.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

Tools

There are several tools to help you with the questions.

- **Pointer:** Use the pointer tool to choose an answer.

- **Cross-Off:** Use the cross-off tool to mark answers that you believe are NOT correct.

- **Highlighter:** Use the highlighter tool to highlight important information.

- **Magnifier:** Use the magnifier tool to enlarge the information on the screen.

- **Line Guide:** Use the line guide tool to help you track a single line of text as you are reading.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Now, select the **Next** arrow to go to the next page.

- **Sticky Notes:** Use the **Sticky Note** tool to type notes on the screen.

- **Help:** Use the **Help** button to find more information, such as instructions on how to use the tools.

- **Eraser:** Use the **Eraser** button to reset your answer and start the question over.

- **Periodic Table:** You may find the **Periodic Table of Elements** helpful in answering test questions. You can view more detail about an element by clicking on the element's box. The box will open in a new window and can be moved around to different parts of the screen by selecting the box and dragging it to a new location.

- **Flag:** Use the **Flag** button, located at the bottom of the page, to mark a question you want to review at a later time. Later, the **Flag** will serve as a reminder to return to that question.

Do you have any questions before we move on?

[SAY] Use the **Next** arrow to skip through the **Helpful Testing Hints** and the **Completing the Test** screens.

(NOTE to Proctor: Ignore the "Helpful Testing Hints" and "Completing the Test" screens, and move on to the "Begin the Test" screen.)

[SAY] We are going to begin the test now. Remember to select the **Pause** button if you need to pause your test for any reason.

Do you have any questions before we move on?

Pause to answer questions. Proceed when there are no more questions.

[SAY] Select the green **Begin the Test** button at the bottom of the screen.

Please start by reading aloud the passage on the left of the screen.

Stimulus Part 1 (below)

Ask the student to read the passage aloud.

Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)

_____

_____

_____

Stimulus Part 1 & Item 1

Gr 11 Science
Question 1

Atmospheric Changes over Time (Part 1)

The gases that make up Earth's atmosphere have changed over time. Scientists measure the levels of carbon dioxide (CO_2) in Earth's atmosphere. The model shown represents a simplified version of the carbon cycle. Some of the locations where carbon is stored are identified and processes that move carbon from one location to another are indicated.

Carbon Cycle Model

Atmosphere 730 GtC

119 GtC/year cellular respiration
120 GtC/year photosynthesis
90 GtC/year diffusion
6.3 GtC/year fossil fuel combustion and industrial processes
88 GtC/year diffusion

plants and soils 2,000 GtC
ocean 38,000 GtC

GtC = gigatons of carbon

Enlarge

Scientists can use carbon cycle models to help make predictions about the amounts of carbon in different locations. Use this carbon cycle model to identify **all** the processes that would decrease CO_2 in the atmosphere if the rate of that process were to increase.

(a) photosynthesis

(b) cellular respiration

(c) fossil fuel combustion

(d) diffusion into the ocean

(e) diffusion into the atmosphere

SEP 2. Developing and Using Models	ESS2.A: Earth Materials and Systems	CCC 7. Stability and Change
Develop a model based on evidence to illustrate the relationships between systems or between components of a system.	Earth's systems, being dynamic and interacting, cause feedback effects that can increase or decrease the original changes.	Much of science deals with constructing explanations of how things change and how they remain stable. Feedback (negative or positive) can stabilize or destabilize a system.

_____

_____

_____

_____

_____

_____

_____

_____

Item 1 (below, and bottom of next page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— Why did you answer that way?

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

Stimulus Part 1 & Item 1

Gr 11 Science

Question 1

Atmospheric Changes over Time (Part 1)

The gases that make up Earth's atmosphere have changed over time. Scientists measure the levels of carbon dioxide (CO_2) in Earth's atmosphere. The model shown represents a simplified version of the carbon cycle. Some of the locations where carbon is stored are identified and processes that move carbon from one location to another are indicated.

Carbon Cycle Model

Atmosphere 730 GtC

119 GtC/year cellular respiration

120 GtC/year photosynthesis

90 GtC/year diffusion

6.3 GtC/year fossil fuel combustion and industrial processes

88 GtC/year diffusion

plants and soils 2,000 GtC

ocean 38,000 GtC

GtC = gigatons of carbon

Enlarge

Scientists can use carbon cycle models to help make predictions about the amounts of carbon in different locations. Use this carbon cycle model to identify **all** the processes that would decrease CO_2 in the atmosphere if the rate of that process were to increase.

(a) photosynthesis

(b) cellular respiration

(c) fossil fuel combustion

(d) diffusion into the ocean

(e) diffusion into the atmosphere

SEP 2. Developing and Using Models	ESS2.A: Earth Materials and Systems	CCC 7. Stability and Change
Develop a model based on evidence to illustrate the relationships between systems or between components of a system.	Earth's systems, being dynamic and interacting, cause feedback effects that can increase or decrease the original changes.	Much of science deals with constructing explanations of how things change and how they remain stable. Feedback (negative or positive) can stabilize or destabilize a system.

· · · · · 6 · · · · ·

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please indicate the student's response to the item by checking the box:

☐ *a*   ☐ *b*   ☐ *c*   ☐ *d*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the interactions between Earth's systems affect the CO2 in the atmosphere*

☐ *the model represents relationships in Earth's systems*

☐ *changes in CO2 could destabilize Earth's systems*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Item 2 (below, and bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

Item 2

Gr 11 Science
Question 2

Atmospheric Changes over Time (Part 1)

The gases that make up Earth's atmosphere have changed over time. Scientists measure the levels of carbon dioxide (CO_2) in Earth's atmosphere. The model shown represents a simplified version of the carbon cycle. Some of the locations where carbon is stored are identified and processes that move carbon from one location to another are indicated.

Carbon Cycle Model

Atmosphere 730 GtC

119 GtC/year cellular respiration

120 GtC/year photosynthesis

90 GtC/year diffusion

6.3 GtC/year fossil fuel combustion and industrial processes

88 GtC/year diffusion

plants and soils 2,000 GtC

ocean 38,000 GtC

GtC = gigatons of carbon

🔍 Enlarge

The graph shows data about atmospheric CO_2 concentrations collected in Mauna Loa, Hawaii, since 1960.

Average Atmospheric CO₂ at Mauna Loa

🔍 Enlarge

Use the data and the Carbon Cycle Model to explain the pattern of atmospheric CO_2 concentrations over time.

The data show that atmospheric CO_2 concentrations have [_____▾] over time. This pattern is **most likely** due to the process of [_____▾].

Item ID ❓

SEP 4. Analyzing and Interpreting Data	ESS2.D: Weather and Climate	CCC 7. Stability and Change
Analyze data using tools, technologies, and/or models (e.g., computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution.	Changes in the atmosphere due to human activity have increased carbon dioxide concentrations and thus affect climate.	• Much of science deals with constructing explanations of how things change and how they remain stable. (HS-ESS2-7) • Feedback (negative or positive) can stabilize or destabilize a system. (HS-ESS2-2)

_____
_____
_____
_____
_____

Please indicate the student's responses to the item by checking the boxes:

☐ *increased*   ☐ *decreased*
☐ *stayed the same*

☐ *photosynthesis*

☐ *cellular respiration*

☐ *fossil fuel combustion*

☐ *diffusion into the ocean*

☐ *diffusion into the atmosphere*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the slope in the data indicates the pattern of CO2 concentrations*

☐ *human activity causes changes in CO2 concentrations*

☐ *changes in CO2 could destabilize Earth's systems*

☐ *other _____*
   _____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

SECURE – DO NOT COPY

226

Item 3 (below, and bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Item 3

Gr 11 Science — Question 3

Atmospheric Changes over Time (Part 1)

The gases that make up Earth's atmosphere have changed over time. Scientists measure the levels of carbon dioxide (CO_2) in Earth's atmosphere. The model shown represents a simplified version of the carbon cycle. Some of the locations where carbon is stored are identified and processes that move carbon from one location to another are indicated.

Carbon Cycle Model

Atmosphere 730 GtC
119 GtC/year cellular respiration
120 GtC/year photosynthesis
90 GtC/year diffusion
6.3 GtC/year fossil fuel combustion and industrial processes
88 GtC/year diffusion
plants and soils 2,000 GtC
ocean 38,000 GtC
GtC = gigatons of carbon

Scientists think that Earth's early atmosphere was different from Earth's current atmosphere due to a large amount of volcanic activity. Scientists use analyses of gases emitted from recent volcanic activity to make inferences about which gases were present in Earth's early atmosphere.

Gases in Current Atmosphere		Gases Emitted from Recent Volcanic Activity	
Gases	Percentage of Atmosphere (%)	Gases	Percentage of Volcanic Emission (%)
N_2	78.0	H_2O	96.0
O_2	21.0	CO_2	2.0
Ar	1.0	SO_2	1.5
H_2O, CO_2	trace amounts	H_2S, HCl, NH_3, CH_4	trace amounts

Using fossil data, scientists estimate that the first land plants developed on Earth about 450 million years ago and changed Earth's atmosphere drastically.

Complete the statement below to describe the role of plants in changing Earth's atmosphere.

The amount of [_____] in the atmosphere [_____] over time because of the process of [_____].

SEP 4. Analyzing and Interpreting Data	ESS2.D: Weather and Climate	CCC 7. Stability and Change
Analyze data using tools, technologies, and/or models (e.g., computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution.	Gradual atmospheric changes were due to plants and other organisms that captured carbon dioxide and released oxygen.	• Much of science deals with constructing explanations of how things change and how they remain stable. • Feedback (negative or positive) can stabilize or destabilize a system.

· · · · · 10 · · · · ·

227

Please indicate the student's responses to the item by checking the boxes:

☐ *water* ☐ *oxygen* ☐ *carbon dioxide*

☐ *increased* ☐ *decreased*

☐ *cellular respiration* ☐ *photo synthesis*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *the data shows changes in the atmospheric gasses over time*

☐ *plants caused changes in the atmosphere*

☐ *changes in CO2 destabilized Earth's systems*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy* ☐ *Hard*

[Say] Why was it easy or hard?

_____
_____
_____
_____
_____
_____
_____
_____

Please provide any additional information/ observations that you made about the student's interaction with the item and technology while completing this item.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

SECURE – DO NOT COPY

228

Stimulus Part 2 (bottom of previous page)

Ask the student to read the passage aloud.

Then ask:

— **Did you understand the information?**

— **What new information did you learn?**

Record the student's comments and interaction to the stimulus and the technology features. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Stimulus Part 2, Item 4, Part A

Gr 11 Science

Question 4
Page 1 of 2

Atmospheric Changes over Time (Part 2)

Students are studying how a change in Earth's atmosphere also has an effect on other Earth systems such as the hydrosphere. They study maps showing the changes in Arctic sea ice between 2000 and 2012.

Changes in Arctic Sea Ice Boundaries
2000–2012

2000 sea ice

2012 sea ice

Enlarge

In addition to sea ice, the maps show land and liquid ocean water. To determine whether there is a difference between how solar radiation affects ocean water and how it affects land, they perform an investigation. The students used water to model the ocean and soil to model the land. They used equal volumes of each substance. They used a heat lamp to model the Sun. The heat lamp was turned on for 10 minutes and then turned off.

Investigation Setup

More Text Below

This question has two parts.

The graph shows the data collected during the investigation.

Temperature of Water and Soil Over Time

Enlarge

Part A: Based on the graph, which statement **best** explains the difference in the temperature pattern between the substances?

(a) Water absorbs and stores more energy than soil does.

(b) Soil absorbs and stores more energy than water does.

(c) Water absorbs more energy than soil but can quickly release the energy in a short period of time.

(d) Soil absorbs more energy than water but can quickly release the energy in a short period of time.

SEP 3: Planning and Carrying Out Investigations	ESS2.C: The Roles of Water in Earth's Surface Processes	CCC 5: Energy and Matter
Plan and conduct an investigation individually and collaboratively to produce data to serve as the basis for evidence, and in the design: decide on types, how much, and accuracy of data needed to produce reliable measurements and consider limitations on the precision of the data (e.g., number of trials, cost, risk, time), and refine the design accordingly.	The abundance of liquid water on Earth's surface and its unique combination of physical and chemical properties are central to the planet's dynamics. These properties include water's exceptional capacity to absorb, store, and release large amounts of energy, transmit sunlight, expand upon freezing, dissolve and transport materials, and lower the viscosities and melting points of rocks.	Energy drives the cycling of matter within and between systems.

229

Item 4 (bottom of previous page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Part A: *Please indicate the student's response to the item by checking the box:*

☐ a   ☐ b   ☐ c   ☐ d

Part B: *Please indicate the student's response to the item by checking the box:*

☐ a large   ☐ a small   ☐ no

☐ a large   ☐ a small   ☐ no

Stimulus Part 2, Item 4, Part B

Gr 11 Science
Question 4
Page 2 of 2

Atmospheric Changes over Time (Part 2)

Students are studying how a change in Earth's atmosphere also has an effect on other Earth systems such as the hydrosphere. They study maps showing the changes in Arctic sea ice between 2000 and 2012.

Changes in Arctic Sea Ice Boundaries 2000–2012

2000 sea ice   2012 sea ice

Enlarge

In addition to sea ice, the maps show land and liquid ocean water. To determine whether there is a difference between how solar radiation affects ocean water and how it affects land, they perform an investigation. The students used water to model the ocean and soil to model the land. They used equal volumes of each substance. They used a heat lamp to model the Sun. The heat lamp was turned on for 10 minutes and then turned off.

Investigation Setup
More Text Below

This question has two parts.

The graph shows the data collected during the investigation.

Temperature of Water and Soil Over Time

Enlarge

Part B: Complete the statement that **best** uses evidence from the investigation to support the explanation chosen in Part A.

The water had [▼] increase in temperature when the heat lamp was on. The water had [▼] decrease in temperature when the heat lamp was turned off.

· · · · · 13 · · · · ·

230

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

- ☐ *the data collected during the investigation is represented in the graph*

- ☐ *soil and water have different abilities to absorb and store energy*

- ☐ *energy cycles between Earth' systems*

- ☐ *other* _____

  _____

- ☐ *none of the above*

While working on the item, the student:

- ☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

- ☐ *appeared to be engaged with the item.*

- ☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

- ☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

Item 5 (bottom of next page)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), *and the technology features of this item. (Be specific about what technology features the student used.)*

_____

_____

_____

_____

_____

_____

_____

Item 5

Gr 11 Science — Question 5

More Text Above

2000 sea ice / 2012 sea ice — Enlarge

In addition to sea ice, the maps show land and liquid ocean water. To determine whether there is a difference between how solar radiation affects ocean water and how it affects land, they perform an investigation. The students used water to model the ocean and soil to model the land. They used equal volumes of each substance. They used a heat lamp to model the Sun. The heat lamp was turned on for 10 minutes and then turned off.

Investigation Setup

heat lamp, thermometers, water, soil — Enlarge

The students decided to modify the original investigation to answer a new question.

New Question: Does reflection of solar radiation have a role in the changing amount of sea ice?

Move one modification that would **best** help to answer the new question and move one reasoning that supports the modification into the chart.

Modification	Reasoning

Possible Modification for Investigation

Move the heat lamp farther away from the cup of water and the cup of soil.

Place a thin sheet of metal on top of a second cup of water, and put it under the heat lamp for ten minutes. Then compare the water temperature of the covered cup with that of the uncovered cup.

Place a sheet of dark color paper on top of a second cup of water, and put it under the heat lamp for ten minutes. Then compare the water temperature of the covered cup with that of the cup of soil.

Possible Reasoning for Selected Modification

Covering the cup of water represents snow and would reflect light from the heat lamp.

The position of the heat lamp would affect how much energy reaches substances in the cups before the energy is reflected by each of the substances.

The material used to cover the cup would reflect energy that would be absorbed by the substance in the other cup.

The heat lamp represents the energy from the Sun during winter when there is less solar radiation being absorbed and reflected by surfaces on Earth.

Click To Respond

SEP 3. Planning and Carrying Out Investigations	ESS2.C: The Roles of Water in Earth's Surface Processes	CCC 5: Energy and Matter
• Plan and conduct an investigation individually and collaboratively to produce data to serve as the basis for evidence, and in the design: decide on types, how much, and accuracy of data needed to produce reliable measurements and consider limitations on the precision of the data (e.g., number of trials, cost, risk, time), and refine the design accordingly.	• The abundance of liquid water on Earth's surface and its unique combination of physical and chemical properties are central to the planet's dynamics. These properties include water's exceptional capacity to absorb, store, and release large amounts of energy, transmit sunlight, expand upon freezing, dissolve and transport materials, and lower the viscosities and melting points of rocks.	Energy drives the cycling of matter within and between systems.

_____

_____

_____

_____

*Please indicate the student's response to the item by drawing a line that connects each selected option to the box in which it was placed in the enlarged table (**below**):*

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ *refining the design of the investigation by reflecting the light would best represent the system*

☐ *soil and water have different abilities to absorb and store energy*

☐ *energy cycles between Earth's systems*

☐ *other* _____

_____

☐ *none of the above*

While working on the item, the student:

☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

☐ *appeared to be engaged with the item.*

☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

Item 5

Modification	Reasoning

Possible Modification for Investigation	Possible Reasoning for Selected Modification
Move the heat lamp farther away from the cup of water and the cup of soil.	Covering the cup of water represents snow and would reflect light from the heat lamp.
Place a thin sheet of metal on top of a second cup of water, and put it under the heat lamp for ten minutes. Then compare the water temperature of the covered cup with that of the uncovered cup.	The position of the heat lamp would affect how much energy reaches substances in the cups before the energy is reflected by each of the substances.
Place a sheet of dark color paper on top of a second cup of water, and put it under the heat lamp for ten minutes. Then compare the water temperature of the covered cup with that of the cup of soil.	The material used to cover the cup would reflect energy that would be absorbed by the substance in the other cup.

The heat lamp represents the energy from the Sun during winter when there is less solar radiation being absorbed and reflected by surfaces on Earth. |

(Indicate the student's response by drawing a line from each selected option to the box in which it was placed.)

←

· · · · · 16 · · · · ·

233

Please provide any additional information/observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

Item 6

_____

_____

_____

Item 6 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

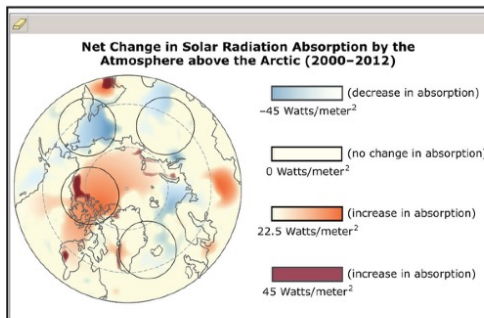| SEP 4. Analyzing and Interpreting Data	ESS2.D: Weather and Climate	CCC 7. Stability and Change
Analyze data using tools, technologies, and/or models (e.g., computational, mathematical) in order to make valid and reliable scientific claims or determine an optimal design solution.	The foundation for Earth's global climate systems is the electromagnetic radiation from the sun, as well as its reflection, absorption, storage, and redistribution among the atmosphere, ocean, and land systems, and this energy's re-radiation into space.	Much of science deals with constructing explanations of how things change and how they remain stable. Feedback (negative or positive) can stabilize or destabilize a system.

234

If the student is not talking, you may want to ask the following question:

— Why did you answer that way?

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____
_____
_____
_____
_____
_____
_____
_____

Please indicate the student's response by circling the selected option in the enlarge graphic (below):

Net Change in Solar Radiation Absorption by the Atmosphere above the Arctic (2000–2012)

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

- ☐ the oil alters the environment and the mosquitoes would have to adapt to survive

- ☐ the oil causes the mosquitoes to die, therefore not allowing the population to survive

- ☐ extinction occurs when a population can no longer survive

☐ other _____
_____

☐ *none of the above*

While working on the item, the student:

- ☐ *appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features.*

- ☐ *appeared to be engaged with the item.*

- ☐ *appeared to be able to read and understand what the item was asking.*

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ *Easy*   ☐ *Hard*

[Say] Why was it easy or hard?

Please record the student's response:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____
_____
_____
_____

_____

_____

_____

_____

_____

_____

_____

Item 7 (below)

Ask the student to read the item aloud.

Remind the student to tell what he/she is thinking with minimal prompting.

If the student is not talking, you may want to ask the following question:

— **Why did you answer that way?**

Record the student's comments and interaction to the answer(s), the item(s), and the technology features of this item. (Be specific about what technology features the student used.)

_____

_____

Item 7

Gr 11 Science
Question 7

Item ID ?

Atmospheric Changes over Time (Part 2)

Students are studying how a change in Earth's atmosphere also has an effect on other Earth systems such as the hydrosphere. They study maps showing the changes in Arctic sea ice between 2000 and 2012.

Changes in Arctic Sea Ice Boundaries 2000–2012

2000 sea ice   2012 sea ice

Enlarge

In addition to sea ice, the maps show land and liquid ocean water. To determine whether there is a difference between how solar radiation affects ocean water and how it affects land, they perform an investigation. The students used water to model the ocean and soil to model the land. They used equal volumes of each substance. They used a heat lamp to model the Sun. The heat lamp was turned on for 10 minutes and then turned off.

Investigation Setup
More Text Below

Read the excerpt below.

Enlarge

Based on the excerpt and your knowledge of atmospheric changes, describe the relationship between solar radiation absorption and the available amount of sea ice that affects the polar bears' access to prey in the Arctic.

Part A: Describe the relationship between solar radiation absorption and the amount of sea ice in the Arctic.

0/500

Part B: Describe how your answer in part A affects the polar bears' access to prey in the Arctic.

0/500

SEP 7: Engaging in Argument from Evidence	ESS2.D: Weather and Climate	CCC 7. Stability and Change
Construct an oral and written argument or counter-arguments based on data and evidence.	• The foundation for Earth's global climate systems is the electromagnetic radiation from the sun, as well as its reflection, absorption, storage, and redistribution among the atmosphere, ocean, and land systems, and this energy's re-radiation into space.	• Much of science deals with constructing explanations of how things change and how they remain stable. • Feedback (negative or positive) can stabilize or destabilize a system.

236

_____
_____
_____
_____
_____
_____

Please record (as close as possible) the student's written response to the first prompt (Species 1):

_____
_____
_____
_____
_____
_____
_____
_____
_____

Part A: _Please record (as close as possible) the student's written response to the first prompt:_

_____
_____
_____
_____
_____
_____
_____
_____

Part B: _Please record (as close as possible) the student's written response to the second prompt:_

_____
_____
_____
_____
_____
_____
_____
_____

While working on this item, the student's commentary showed evidence that the item elicited his/her thinking around:

☐ _the oil alters the environment and the mosquitoes would have to adapt to survive_

☐ _the oil causes the mosquitoes to die, therefore not allowing the population to survive_

☐ _extinction occurs when a population can no longer survive_

☐ _other_ _____
_____

☐ _none of the above_

While working on the item, the student:

☐ _appeared to rush in answering the item without understanding what was being asked and/or how to respond to the item using the technology-enhanced response features._

☐ _appeared to be engaged with the item._

☐ _appeared to be able to read and understand what the item was asking._

[Say] Was it easy or hard to use the information in the passage to answer the question?

☐ _Easy_   ☐ _Hard_

· · · · · 20 · · · · ·

[Say] Why was it easy or hard?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please provide any additional information/ observations that you made about the student's interaction and the technology while completing this item.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

(Transition Message in Test that comes up before Survey Questions)

[Say] Congratulations! You have reached the end of the Michigan Science Standards Pilot. Now, take a short survey.

[Say] Select the **Begin Section** button to answer the survey questions.

(NOTE to Proctor: Please indicate the student's response to each survey question with an X on the screen shots below and on the next pages.)

Question 8 (below)

[Say] Can you tell me what you have learned about the effects of carbon cycling in class? How was this group of questions different or the same from what you learned in class?

Please record the student's response:

_____

_____

_____

_____

_____

Q 8

Gr 11 Science
Question 8

Before I took this test, I had learned _____ in my science class about how Earth's atmosphere has changed over time.

(a) a lot

(b) some or a little

(c) nothing

(Indicate the student's response with an X on this image.)

· · · · · 21 · · · · ·

_____

_____

_____

_____

_____

_____

_____	_____

Question 9 (below)

[Say] Can you tell me what kinds of real life examples your teacher has used in class to teach about science and solve problems?

Please record the student's response:

_____

_____

_____

Question 10 (below)

[Say] Can you tell me what was different or the same about reading passages and using data in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____

_____

Q 9

Gr 11 Science
Question 9

My science teacher uses examples of things that happen in real life and the world around us to teach about science and to solve problems in my science class.

ⓐ a lot or all the time

ⓑ sometimes

ⓒ never

(Indicate the student's response with an X on this image.)

Q 10

Gr 11 Science
Question 10

In my science class, the tests my teacher gives us include material we need to read and data we need to look at and interpret in order to answer the questions on the test.

ⓐ every test in science class

ⓑ for some tests in science class

ⓒ not on any tests in science class

(Indicate the student's response with an X on this image.)

· · · · · 22 · · · · ·

239

_____

_____

_____

_____

_____

_____

_____

_____

Question 11 (below)

[Say] Can you tell me what was different or the same about writing to explain your answer in this group of questions than what your teacher has asked you to do in class?

Please record the student's response:

_____

_____

_____

_____

_____

_____

_____

_____

What to do at the end of the Cognitive Lab

[Say]: Thank you very much for your time and feedback about the science test. Your comments will help us make the test better for students in Michigan.

*The student should **NOT** select the **End Test** button. You can navigate back to student responses if you need to review or capture information you may have missed. After all information has been captured, select the **End Test** button and exit the Test Development Environment.*

Q 11

In my science class, the tests have questions that I have to write my answer and also include an explanation for my answer.

(a) every test in science class

(b) for some tests in science class

(c) not on any tests in science class

(Indicate the student's response with an X on this image.)

· · · · · 23 · · · · ·

REFERENCES

2014 MSTEP house bill 5314. (2014).

Abedi, J. (2002). Standardized Achievement Tests and English Language Learners: Psychometrics Issues. *Educational Assessment*, *8*(3), 231–257. https://doi.org/10.1207/S15326977EA0803_02

Achieve. (2018). Task screener.

Achieve. (2019). A framework to Evaluate Cognitive Complexity in Science Assessments. https://www.achieve.org/files/Science%20Cognitive%20Complexity%20Framework_Final_093019.pdf

Aiman-Smith, L., Scullen, S. E., & Barr, S. H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, *5*(4), 388–414.

Alonzo, A. C. (2013). *What can be learned from current large-scale assessment programs to inform assessment of the Next Generation Science Standards?* http://www.ets.org/Media/Research/pdf/alonzo.pdf

Alonzo, A. C., & Ke, L. (2016). Taking stock: Existing resources for assessing a new vision of science learning. *Measurement: Interdisciplinary Research and Perspectives*, *14*(4), 119–152. https://doi.org/10.1080/15366367.2016.1251279

American Educational Research Association (Ed.). (2011). *Report and recommendations for the reauthorization of the institute of education sciences*. American Educational Research Association.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, C. W., de los Santos, E. X., Bodbyl, S., Covitt, B. A., Edwards, K. D., Hancock, J. B., Lin, Q., Morrison Thomas, C., Penuel, W. R., & Welch, M. M. (2018). Designing educational systems to support enactment of the Next Generation Science Standards. *Journal of Research in Science Teaching*, *55*(7), 1026–1052. https://doi.org/10.1002/tea.21484

Appendix F. (n.d.). *Appendix F Science and Engineering Practices in the NGSS - FINAL 060513.pdf*

Appendix G. (n.d.). *Appendix G - Crosscutting Concepts FINAL edited 4.10.13 (1).pdf*

Beatty, P. C., & Willis, G. B. (2007a). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, *71*(2), 287–311. https://doi.org/10.1093/poq/nfm006

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students: *Journal of Research in Science Teaching*, *53*(7), 1082–1112. https://doi.org/10.1002/tea.21257

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in education. *Principles, Policy & Practice*, *5*(1), 7–74.

Blank, R. K., & Adams, E. (2018). *How well aligned are state assessments with state standards or Common Core/Next Generation Science Standards?* NORC.

Board on Science Education, Division of Behavioral and Social Sciences and Education, & National Academies of Sciences, Engineering, and Medicine. (2018). *Design, Selection, and Implementation of Instructional Materials for the Next Generation Science Standards: Proceedings of a Workshop* (H. G. Rhodes, Ed.). National Academies Press. https://doi.org/10.17226/25001

Brock, D. (2014). Developing assessments for the Next Generation Science Standards. *The Science Teacher*, *81*(8), 79.

Brown, D. E., & Sadler, T. D. (2018). Conceptual framing and instructional enactment of the Next Generation Science Standards: A synthesis of the contributions to the special issue. *Journal of Research in Science Teaching*, *55*(7), 1101–1108. https://doi.org/10.1002/tea.21509

Chesnutt, K., Jones, M. G., Hite, R., Cayton, E., Ennes, M., Corin, E. N., & Childers, G. (2018). Next generation crosscutting themes: Factors that contribute to students' understandings of size and scale. *Journal of Research in Science Teaching*, *55*(6), 876–900. https://doi.org/10.1002/tea.21443

Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences*, *6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Chudowsky, N., Glaser, R., & Pellegrino, J. W. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press.

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Practice partnerships: A strategy for leveraging research for educational improvement in school districts*. William T. Grant Foundation.

Conrad, F., Blair, J., & Tracy, E. (n.d.). *Verbal reports are data! A theoretical approach to cognitive interviews*. U.S. Bureau of Labor Statistics.

Corbin, J., & Strauss, A. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Sage.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Sage.

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2016). Building an assessment argument to design and use Next Generation Science Assessments in efficacy studies of curriculum interventions. *American Journal of Evaluation*, *37*(2), 174–192. https://doi.org/10.1177/1098214015581707

DeBoer, G. E., Lee, H. S., & Husic, F. (2008). Assessing integrated understanding of science. In *Coherent science education: Implications for curriculum, instruction, and policy* (pp. 153–182).

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Buttner, G., Fauth, B., … Hardy, I. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *American Educational Research Journal*, *52*(6), 1133–1159. https://doi.org/10.3102/0002831215596412

Developing Assessments for the Next Generation Science Standards. (2014). National Academies Press. http://www.nap.edu/catalog/18409

Duschl, R. A. (2012). The second dimension—Crosscutting concepts. *The Science Teacher*, *9*(2), 34–38.

Enrolled House Bill No. 5314, 97th Legislature, 2014 Reg. Sess. (Mi. 2014). http://www.legislature.mi.gov/documents/2013-2014/publicact/pdf/2014-PA-0196.pdf

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. The MIT Press.

Evagorou, M., Erduran, S., & Mäntylä, T. (2015a). The role of visual representations in scientific practices: From conceptual understanding and knowledge generation to "seeing" how science works. *International Journal of STEM Education*, *2*(1), 11. https://doi.org/10.1186/s40594-015-0024-x

Every Student Succeeds Act (ESSA). (2015). Public Law No. 114-95, S.1177, 114th Cong. (2015). https://www.congress. gov/114/plaws/publ95/PLAW-114publ95.pdf

Fick, S. J. (2018). What does three-dimensional teaching and learning look like? Examining the potential for crosscutting concepts to support the development of science knowledge. *Science Education*, *102*(1), 5–35. https://doi.org/10.1002/sce.21313

Fick, S. J., McAlister, A. M., Chiu, J. L., & McElhaney, K. W. (2021). Using students' conceptual models to represent understanding of crosscutting concepts in an NGSS-aligned curriculum unit about urban water runoff. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-021-09911-6

Furtak, E. M. (2017). Confronting dilemmas posed by three-dimensional classroom assessment: Introduction to a virtual issue of *Science Education*: *Science Education*, *101*(5), 854–867. https://doi.org/10.1002/sce.21283

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, *15*(3–4), 175–196. https://doi.org/10.1080/10627197.2010.530553

Furtak, E. M., Kang, H., Pellegrino, J., Harris, C., Krajcik, J., Morrison, D., Bell, P., Lakhani, H., Suárez, E., Buell, J., Nation, J., Henson, K., Fine, C., Tschida, P., Fay, L., Biddy, Q., Penuel, W., & Wingert, K. (2020). *Emergent Design Heuristics for Three-Dimensional Classroom Assessments that Promote Equity*. 9.

Gerard, L. F., Spitulnik, M., & Linn, M. C. (2010). Teacher use of evidence to customize inquiry science instruction. *Journal of Research in Science Teaching*, *47*(9), 1037–1063.

Gilbert, J. K. (2010). The role of visual representations in the learning and teaching of science: An introduction. *An Introduction*, *11*(1), 19.

Gorin, J. S., & Mislevy, R. J. (2013). Inherent measurement challenges in the next generation science standards for both formative and summative assessment. *Invitational Research Symposium on Science Assessment*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.5350&rep=rep1&type=pdf

Guide to evaluating CCSSO criteria test content. (n.d.). *Guide-to-Evaluating-CCSSO-Criteria-Test-Content-020316.pdf*

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice*, *38*(2), 53–67. https://doi.org/10.1111/emip.12253

Herman, J. L., Webb, N. M., & Zuniga, S. A. (2007). Measurement issues in the alignment of standards and assessments. *Applied Measurement in Education*, *20*(1), 101–126. https://doi.org/10.1080/08957340709336732

Johnstone, C., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. National Center on Educational Outcomes, University of Minnesota.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, *112*(3), 527.

Lissitz, R. W. (Ed.). (2009). *The concept of validity: Revisions, new directions and applications*. IAP.

Justi, R., & Gilbert, J. (2003). Models and modelling in chemical education. In J. K. Gilbert, O. Jong, R. Justi, D. F. Treagust, & J. H. Driel (Eds.), *Chemical education: Towards research-based practice* (Vol. 17, pp. 47–68). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47977-X_3

Justi, R. S., & Gilbert, J. K. (2002). Modelling, teachers' views on the nature of modelling, and implications for the education of modellers. *International Journal of Science Education*, *24*(4), 369–387. https://doi.org/10.1080/09500690110110142

Kind, P. M. (2013). Conceptualizing the science curriculum: 40 years of developing assessment frameworks in three large-scale assessments: *Science Education*, *97*(5), 671–694. https://doi.org/10.1002/sce.21070

Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? Understanding a framework for K-12 science education. *The Science Teacher*, *79*(3), 38.

Lee, O., Miller, E. C., & Januszyk, R. (2014). Next Generation Science Standards: All standards, all students. *Journal of Science Teacher Education*, *25*(2), 223–233. https://doi.org/10.1007/s10972-014-9379-y

Luecht, R. M. (2020). Generating performance-level descriptors under a principled assessment design paradigm: An example for assessments under the Next-Generation Science Standards. *Educational Measurement: Issues and Practice*, *39*(4), 105–115. https://doi.org/10.1111/emip.12356

Michigan K-12 Science Standards. (2015). Michigan Department of Education.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage.

National Academies of Sciences and Medicine. (2017). *Seeing students learn science: Integrating assessment and instruction in the classroom*. The National Academies Press. https://doi.org/10.17226/23548

National Research Council. (2006). *Systems for State Science Assessment*. Committee on Test Design for K-12 Science Achievement.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

National Research Council. (2014). *Developing assessments for the next generation science standards*. National Academies Press.

NGSS Lead States. (Ed.). (2013). *Next generation science standards: For states, by states*. National Academies Press.

No Child Left Behind Act of 2001. (2002). P.L. 107-110, 20 U.S.C. § 6319.

OECD. (2016). *PISA 2015 Assessment and analytical framework: Science, reading, mathematic and financial literacy*. OECD Publishing.

Passmore, C., Gouvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In M. R. Matthews (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1171–1202). Springer Netherlands. https://doi.org/10.1007/978-94-007-7654-8_36

Pellegrino, J. W., Chudowsky, N., Glaser, R., & National Research Council (U.S.). (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* National Academy Press.

Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 787–850). American Educational Research Association. https://doi.org/10.3102/978-0-935302-48-6_12

Penuel, W. R., Turner, M. L., Jacobs, J. K., Horne, K., & Sumner, T. (2019). Developing tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks. *Science Education, 103*(6), 1367–1395. https://doi.org/10.1002/sce.21544

Pruitt, S. L. (2014). The next generation science standards: The features and challenges. *Journal of Science Teacher Education, 25*(2), 145–156. https://doi.org/10.1007/s10972-014-9385-0

Quellmalz, E. S., Davenport, J. L., Timms, M. J., DeBoer, G. E., Jordan, K. A., Huang, C.-W., & Buckley, B. C. (2013). Next-generation environments for assessing and promoting complex science learning. *Journal of Educational Psychology, 105*(4), 1100–1114. https://doi.org/10.1037/a0032220

Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging students in the scientific practices of explanation and argumentation. *Science Scope, 35*(8), 6–11.

Research and Practice Collaboratory. (2016).

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema, 26*(1).

Rivet, A. E., Weiser, G., Lyu, X., Li, Y., & Rojas-Perilla, D. (2016). What are crosscutting concepts in science? Four metaphorical perspectives. In C. K. Looi, J. L. Polman, U. Cress, and P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS)* (Volume 2). International Society of the Learning Sciences.

Rulon, P. J. (1946). On the validity of educational tests. *Harvard Educational Review, 16*(4), 290–296.

SAIC. (n.d.). *Assessment framework.*

Schwarz, C., Passmore, C., & Reiser, B. J. (Eds.). (2016). *Helping students make sense of the world using next generation science and engineering practices.* National Science Teachers Association Press.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21–34. https://doi.org/10.1111/emip.12189

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, *45*(1), 83–117.

Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 20–37). Information Age Publishing.

Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *26*(1).

State Assessment Item Collaborative Framework. (2015). West Ed.

State of Michigan. (2016). 98th Legislature Regular Session of 2016 House Bill 4493, Sec. 1249 2a, i-ii.

Stiggins, R. (2017). *The perfect assessment system*. ASCD.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. *Journal of Research in Science Teaching*, *52*(5), 659–685. https://doi.org/10.1002/tea.21208

Willis, G. B. (2010). *Cognitive interviewing: A tool for improving questionnaire design*. Sage.

Wilson, M. R., & Bertenthal, M.W. (Eds.). *Board on Testing and Assessment, Center for Education*. The National Academies Press.

Yarroch, W. L. (1991). The implications of content versus item validity on science tests. *Journal of Research in Science Teaching*, *28*(7), 619–629.

Zucker, S. (2008). *Alignment in Educational Assessment*. 13.