



## Research Memorandum

ETS RM-21-06

# Conceptualization and Development of a Performance Task for Assessing and Building Elementary Preservice Teachers' Ability to Facilitate Argumentation-Focused Discussions in Science: The Mystery Powder Task

Jamie N. Mikeska  
Heather Howell  
Joseph Ciofalo  
Adam Devitt  
Elizabeth Orlandi  
Kenneth King  
Michelle Lipari  
Glenn Simonelli

December 2021



# ETS Research Memorandum Series

---

## EIGNOR EXECUTIVE EDITOR

Laura Hamilton  
*Associate Vice President*

## ASSOCIATE EDITORS

Beata Beigman Klebanov  
*Senior Research Scientist*

Brent Bridgeman  
*Distinguished Presidential Appointee*

Heather Buzick  
*Senior Research Scientist*

Tim Davey  
*Research Director*

John Davis  
*Research Scientist*

Marna Golub-Smith  
*Consultant*

Priya Kannan  
*Research Scientist*

Sooyeon Kim  
*Principal Psychometrician*

Jamie Mikeska  
*Senior Research Scientist*

Gautam Puhan  
*Psychometric Director*

Jonathan Schmidgall  
*Research Scientist*

Jesse Sparks  
*Research Scientist*

Michael Walker  
*Distinguished Presidential Appointee*

Klaus Zechner  
*Senior Research Scientist*

## PRODUCTION EDITORS

Kim Fryer  
*Manager, Editing Services*

Ayleen Gontz  
*Senior Editor*

---

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

**Conceptualization and Development of a Performance Task for Assessing and Building  
Elementary Preservice Teachers' Ability to Facilitate Argumentation-Focused  
Discussions in Science: The Mystery Powder Task**

Jamie N. Mikeska

ETS, Princeton, New Jersey, United States

Heather Howell

ETS, Princeton, New Jersey, United States

Joseph Ciofalo

ETS, Princeton, New Jersey, United States

Adam Devitt

California State University, Stanislaus, College of Education

Elizabeth Orlandi

Consultant

Kenneth King

Aurora University, Department of Elementary Education

Michelle Lipari

Consultant

Glenn Simonelli

Consultant

December 2021

Corresponding author: J. N. Mikeska, E-mail: [jmikeska@ets.org](mailto:jmikeska@ets.org)

Suggested citation: Mikeska, J. N., Howell, H., Ciofalo, J., Devitt, A., Orlandi, E., King, K., Lipari, M., & Simonelli, G. (2021). *Conceptualization and development of a performance task for assessing and building elementary preservice teachers' ability to facilitate argumentation-focused discussions in science: The Mystery Powder task* (Research Memorandum No. RM-21-06). ETS.

Find other ETS-published reports by searching the ETS ReSEARCHER  
database at <https://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit  
<https://ets.org/research/contact/>

**Action Editor:** Laura Hamilton

**Reviewer:** Dante Cisterna-Alburquerque

Copyright © 2021 by Educational Testing Service. All rights reserved.

ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS).

All other trademarks are the property of their respective owners.

## **Abstract**

In this research memorandum, one of a series of eight such reports, we describe the development process by which we produced a series of performance tasks designed for preservice elementary teachers for formative assessment use in the context of teacher education programs. Each performance task provides an opportunity for preservice elementary teachers to practice facilitating an argumentation-focused discussion targeting a student learning goal in elementary mathematics or science. One unique aspect of this work is that the discussions take place within an online simulated classroom environment that consists of five upper elementary student avatars. This report documents the development process at three levels. First, we define the overarching teaching competency that each task targets—the ability to facilitate argumentation-focused discussions—by describing the general approach and processes used to develop the full set of eight tasks and the key components embedded within each task. Next, we describe the academic content addressed in the subset of four science tasks and how the content conceptualization supports the use of the tasks individually or as a set. We then discuss the specific task that is the focus of this research memorandum, outlining how it was designed to capture evidence of the targeted teaching competency.

*Keywords:* performance task, elementary education, simulated classrooms, virtual reality, discussion, argumentation, preservice teachers, teacher education, science, matter and its interactions

## **Acknowledgments**

This study was supported by a grant from the National Science Foundation (Award No. 1621344). The opinions expressed herein are those of the authors and not the funding agency.

The development of these performance tasks, especially the interactor training materials, was enhanced by the feedback from the talented group of Mursion interactors who worked on this project and who served as the human-in-the-loop during these simulated discussions. In addition, we are grateful for the advice and critical review of these tasks from our advisory board members, assessment developers, and research colleagues. Finally, we are appreciative of the teacher educators and preservice teachers who provided substantive feedback on how to improve these tasks for future use in teacher education.

## Preface

This research memorandum is one of eight reports in which we describe the development process by which we produced a series of performance tasks designed for preservice elementary teachers for formative assessment use in the context of teacher education programs. The following table provides an overview of the eight performance tasks.

### Descriptions of the Eight Performance Tasks

| Task name                   | Task description  |
|-----------------------------|---|
| <b>Mathematics</b>          |   |
| Ordering Fractions          | The teacher leads a discussion of three student-generated strategies for ordering a set of given fractions from least to greatest.  |
| Fractions Between           | The teacher leads a discussion with the students about an unconventional student-generated method for generating fractions between two given fractions. The discussion is focused on the strengths and weaknesses of the strategy, and its applicability to other situations.         |
| Birdseed                    | This discussion is grounded in students' work on a story problem in which they have used fraction multiplication. Prior to the discussion, the students individually critiqued one another's work, making the critique aspect of argumentation more clearly available to the teacher. |
| Eight Divided By One-Fourth | This discussion focuses on students' work to generate meaningful understandings and representations of division by a fraction.  |
| <b>Science</b>              |   |
| Mystery Powder              | This discussion focuses on reaching group consensus around the identity of an unknown powder based on its properties and what is known about a set of common powders. In addition to identifying the mystery powder, students discuss which properties are most useful and why.       |
| Conservation of Matter      | In this task, the teacher supports the students in discussing whether the amount of matter is conserved during a physical change, in this case the mixing of ingredients to produce lemonade.   |
| Modeling Matter             | This task focuses on critiquing and revising visual models for explaining what happens after a drop of red food coloring is dropped into a cup of water.  |
| Changing Matter             | This discussion builds on students' prior work mixing together different combinations of substances and forming claims about whether each combination produced a new substance, with an emphasis on using evidence to support those claims.   |

Each report is dedicated to a singular task and provides a full description and corresponding appendix text for that particular task. All of the reports include a description of the general development process that applies to the full set of tasks. Additional materials to support the use of the performance tasks, such as interactor training and scoring documentation, are not included in these reports but are archived and publicly available through the Qualitative Data Repository housed at Syracuse University (<https://data.qdr.syr.edu/dataverse/go-discuss>).



The first section of this report details the development of the performance tasks, including a description of the construct, the task type, and the process used to develop each of the eight tasks. In the second section, we discuss the content focus of the set of science tasks and of the Mystery Powder task in particular. In the final section, we describe the resulting set of materials that make up the stable components of the task itself and use examples from the Mystery Powder task to illustrate what these components look like and how they function together in the performance task.

### **Section 1: Development of the Performance Tasks**

In this section of the report, we share our conceptualization of the teaching practice of facilitating argumentation-focused discussions, describe what a simulated teaching performance task is, and explain how our use of the performance task maps onto the conceptualization of the teaching practice. We finish by outlining the process steps that we used to develop the tasks.

#### **Construct Definition: Facilitating Argumentation-Focused Discussions**

Our construct of interest is the teaching practice of facilitating discussions that engage students in argumentation, or what we refer to as “facilitating argumentation-focused discussions.” We focused on this teaching practice for a number of reasons. First, facilitating argumentation-focused discussions is an ambitious teaching practice that is critically important for teachers to learn how to do well in order to support student conceptual learning within content areas (Kazemi & Stipek, 2009; Russell et al., 2017; Stylianides et al., 2016; Walshaw & Anthony, 2008). Second, this practice is hard to learn how to do it well, and many teachers—even experienced teachers—tend to have had little opportunity to learn how to do it well (Barkai et al., 2002; Reid & Zack, 2009). Finally, the focus on argumentation was purposeful. Although teachers may facilitate many kinds of discussions with K–12 students, both the *Common Core State Standards* and *Next Generation Science Standards* (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; National Research Council, 2013) identify argumentation as one of the key mathematical and scientific practices that K–12 students need to master.

To define the construct that we were aiming to measure—preservice elementary teachers’ ability to facilitate argumentation-focused discussions—we began by reviewing the empirical and practitioner literature as well as the current student standards in mathematics and science to identify the core aspects of this teaching practice. Building on this review, we identified five dimensions of high-quality, argumentation-focused discussions: (a) attending to students’ ideas, (b) developing a coherent and connected storyline, (c) encouraging student-to-student interactions, (d) developing students’ conceptual understanding, and (e) engaging students in argumentation. Table 1 provides details about the specific focus of each dimension.

**Table 1. Dimensions of a Scoring Rubric to Evaluate Preservice Teachers’ Ability to Facilitate Argumentation-Focused Discussions**

| Dimension                                     | Description: Degree to which the teacher. . .   |
|---|---|
| Attending to students’ ideas                  | . . . is being responsive to students, with a focus on making sure the discussion is grounded in the ideas the students bring with them   |
| Developing a coherent and connected storyline | . . . is able to shape a coherent discussion, with a focus on building and connecting ideas toward an instructional goal  |
| Encouraging student-to-student interactions   | . . . organizes the classroom community and the social interactions so students respond directly to one another’s thinking  |
| Developing students’ conceptual understanding | . . . makes productive decisions about how to address particular ideas, especially students’ misunderstandings  |
| Engaging students in argumentation            | . . . emphasizes disciplinary argumentation (e.g., consideration of opposing claims; facilitates critique and rebuttals; encourages students to draw upon evidence and reasoning) |

*Note.* Adapted from “Using Performance Tasks Within Simulated Environments to Assess Teachers’ Ability to Engage in Coordinated, Accumulated, and Dynamic (CAD) Competencies” by J. N. Mikeska, H. Howell, and C. Straub, 2019, *International Journal of Testing*, 19(2), p. 138 (<https://doi.org/10.1080/15305058.2018.1551223>). Copyright 2019 by Taylor & Francis.

The first dimension, attending to students’ ideas, focuses on the extent to which teachers are responsive to students’ ideas in equitable ways, ensuring that the discussion is grounded in students’ ideas and that all students are engaged in meaningful aspects of the discussion. The second dimension, developing a coherent and connected storyline, targets the

degree to which the teacher can shape a coherent discussion by building and connecting ideas toward a learning goal. The third dimension, encouraging student-to-student interactions, pays attention to how teachers facilitate the discussion so students are the ones responsible for interacting directly with each other and engaging with one another's ideas. The fourth dimension, developing students' conceptual understanding, targets the extent to which the teacher and students are involved in evaluating the accuracy and validity of key ideas and how well the teacher productively addresses students' misunderstandings. The fifth dimension, engaging students in argumentation, emphasizes the degree to which students are invited to and engage in argument construction and critique during the discussion.

### Simulated Teaching Performance Tasks

The overall goal of our research was to develop a set of simulation-based performance tasks that could be used to assess and build preservice elementary teachers' ability to facilitate argumentation-focused discussions. We conducted this work in the context of an innovative, mixed reality platform (see Figure 1)—an upper elementary simulated classroom composed of five student avatars.

**Figure 1. Image of an Upper Elementary Simulated Classroom**



Credit: Image courtesy of Mursion

The student avatars are controlled on the back end by a human in the loop, called an *interactor*, who is trained to respond as each of the five student avatars during the discussion. The preservice teacher does not see the interactor but instead views the student avatars on a television or computer screen and can interact with the student avatars in real time during the discussion. We hypothesized that the simulated classroom could serve as a practice-based space for preservice teachers to hone their skill in this teaching competency. Each performance task was designed to be deployed within the upper elementary classroom environment.

The teaching competency of facilitating argumentation-focused discussions is one that involves complex interactions between a teacher and students around specified content. It requires a practice space that provides opportunities for extended interactions to unfold over time, as a teacher's ability to engage in this practice is observable only across these patterns of interactions (Mikeska et al., 2019). In earlier writing, we describe this competency as one that is "coordinated," "accumulated," and "dynamic" (Mikeska et al., 2019, pp. 132–133). By *coordinated*, we mean that the teacher is required to manage multiple, sometimes competing, considerations simultaneously—for example, trying to balance the goal of engaging students in argumentation with addressing students' erroneous conceptual understanding. *Accumulated* refers to the nature of the evidence that needs to be captured, as the teaching competency is observed over time across the patterns of interactions and not by examining individual, disparate interactions. By *dynamic*, we mean that this teaching competency is observed as teachers respond to the constantly changing nature of various task conditions. Each one of these aspects has implications for task design.

First, to ensure that we were adequately measuring this teaching competency, we had to ensure that our task design afforded teachers the opportunity to manage various considerations at the same time. Second, we had to ensure that the tasks provided substantial opportunities to capture evidence at various time points. For example, the tasks needed to provide us opportunity to observe how teachers prompt (or fail to prompt) direct student dialogue and the ways that students begin to engage in specific behaviors more (or less) frequently based on this teacher prompting over time. Finally, we had to create variable task

situations so that the teacher would be required to respond to the changing nature of the situation over time, for example, creating dynamic student profiles where students can “learn” based on their interactions with other students and the teacher, as described in the final section of this report. In the next section, we explain our process for developing each performance task, which includes both the preservice teacher-facing task materials and the interactor-facing task materials.

### **Overview of the Task Design Process**

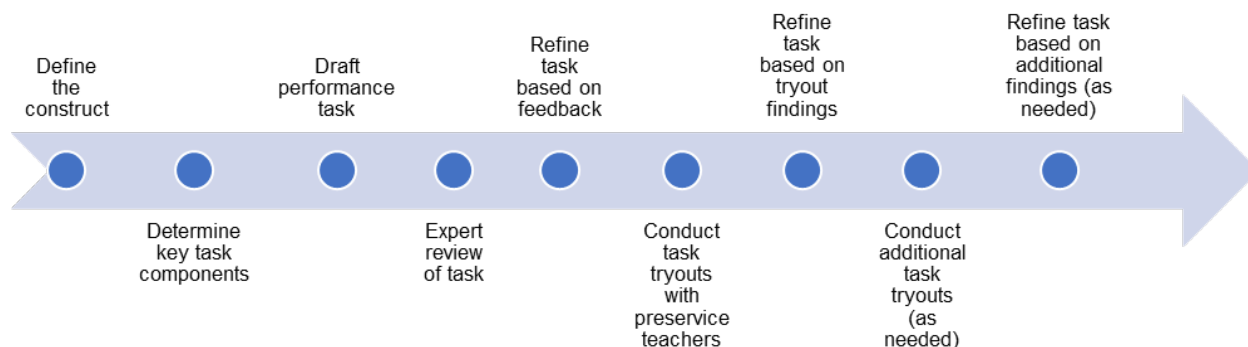
Because the overall goal of using these tasks was to be able to make valid inferences about preservice teachers’ ability to facilitate discussions that engage students in the practice of argumentation, we drew upon the process of evidence-centered design (Mislevy et al., 2002) to develop our evidence model. We then used this evidence model to inform the overall design of each performance task. Our first step was to use our construct definition to develop an evidence model to articulate the observable behaviors that could serve as evidence of preservice teachers’ ability to engage successfully in each dimension. For example, for the first dimension—attending to students’ ideas—we identified three indicators of that dimension of this overall teaching practice, including the preservice teachers’ abilities to incorporate ideas from the students’ written prework into the discussion, to elicit substantive ideas from all students, and to make use of students’ ideas to move the lesson forward in regard to the discussion’s specified student learning goal. We further elaborated each one at three levels of proficiency—beginning novice, developing novice, and well-prepared novice—to describe the observable behaviors one would gather evidence about to inform assessment of that indicator. For example, for the previously discussed dimension, attending to students’ ideas, under the second indicator, eliciting substantive ideas, the observable behaviors specified (Figure 2) indicate that elicitation of substantive ideas from students is related both to the teacher’s sustained efforts to elicit such contributions and to the teacher’s success in eliciting such contributions from all students. Substantive ideas are defined as those that go beyond yes/no statements or restatements of the work the student completed before the discussion.

**Figure 2. Example of Observable Behaviors for Indicator 1b: Elicits Substantive Contributions**

| Indicator  | Level 1<br>Beginning Novice   | Level 2<br>Developing Novice   | Level 3<br>Well-Prepared Novice  |
|--|---|--|--|
| <b>1b.<br/>Elicits Substantive<br/>Contributions</b> | <p>The teacher does not probe students for substantive contributions or does so only once or twice.</p> <p>OR</p> <p>The teacher interacts with only one student from each group.</p> | <p>The teacher probes students for substantive contributions intermittently during the lesson.</p> <p>AND</p> <p>The teacher does not elicit a substantive contribution from at least one student.</p> | <p>The teacher probes students for substantive contributions consistently throughout the lesson.</p> <p>AND</p> <p>The teacher succeeds in eliciting one or more substantive contribution from <b>every</b> student.</p> |

Our project’s advisory board, made up of teacher educators, content specialists, and researchers in mathematics and science teacher education, conducted an expert review of these dimensions and indicators. Their goal was, first, to ensure that they were adequately aligned to the construct and previous literature in mathematics and science teacher education and, second, to provide feedback on whether our characterization of high-quality discussions in the context of disciplinary argumentation adequately addressed the ways in which this teaching practice is used, valued, and characterized within each of the disciplines (elementary mathematics and science). The advisory board also identified and offered suggestions for any aspects of our construct definition that were missing, misrepresented, or not sufficiently addressed. Finally, they considered whether the progressions seemed logical, comprehensive, and scoreable and captured the most important observable teacher behaviors for each dimension and indicator.

In the design of the performance tasks themselves, we used a design-based research approach (Design-Based Research Collective, 2003) in developing and refining the task materials<sup>1</sup> at multiple stages and leveraging various expertise from teacher educators, researchers, preservice and in-service teachers, content experts, and assessment developers (Figure 3). After defining the construct of interest, including the specific dimensions and indicators of this teaching practice, the next step in our task design process was to determine the key task components that would provide opportunities for the preservice teachers to engage in these dimensions of this teaching practice and support us in capturing adequate evidence across all five dimensions.

**Figure 3. Progression of the Design-Based Research Process**

The task materials include two types of components: the preservice teacher-facing and the interactor training materials. The preservice teacher-facing materials include a written document that provides information to the preservice teacher about the simulated discussion’s student learning goal, where this discussion fits into a larger instructional sequence, and what the instructional activities are that the student avatars engaged in prior to this discussion. This document also shows the preservice teacher written work samples that the student avatars generated prior to the discussion, which provides the preservice teachers with insight into the students’ sense-making about the specific mathematics problem or science investigation that is the focus of the discussion. In addition, we developed materials to train the interactor. These training materials are designed to help the interactor learn about the student avatars’ initial ideas and understandings related to the mathematics problem or science investigation that is the focus for the discussion. These materials also support the interactor in learning about the circumstances under which the student avatars can arrive at new understandings based on ideas and arguments that the preservice teacher or the other students make during the discussion. This level of training is also critical in helping to support the standardization of opportunity across preservice teachers so that experiences in the simulated classroom are comparable in the level and nature of challenge each preservice teacher encounters (Howell & Mikeska, 2021). We describe specific components for these task materials, including how we designed them to capture adequate evidence of the five dimensions of this teaching competency, in the final section of this manuscript.

The next step in our task development process included an expert review of the performance task materials to ensure that the task components (both preservice teacher and interactor facing) worked synergistically to gather observable evidence of the preservice teachers' ability to facilitate argumentation-focused discussions. Reviewers included our advisory board members and assessment development experts, who reviewed the preservice teacher-facing materials to ensure that (a) the task provided opportunities for the preservice teachers to engage the student avatars in the practice of mathematical or scientific argumentation; (b) the written student work samples captured a range of typical responses for upper elementary students regarding the specific mathematics problem or science investigation that was the focus of the discussion; and (c) the educative features in the task would be useful to support preservice elementary teachers in learning how to facilitate high-quality, argumentation-focused discussions centered around these student learning goals. Reviewers also considered whether each of the preservice teacher-facing task components—such as the student learning goal, specific instructional scenario, task description, and student profiles—were clear, appropriate, and sufficient for the intended audience. In terms of the interactor training materials, reviewers focused on ensuring that we identified reasonable responses for the interactor to use as the discussion unfolds in the simulated classroom and that the responses did not limit or misrepresent the preservice teachers' ability to engage in this teaching practice. Our research team then revised these task materials based on the experts' feedback. These revisions included a variety of different changes across these tasks, such as more clearly articulating the discussion's student learning goal, modifying the written student responses to better align with grade-level expectations, refining the teaching tips to provide more robust educative supports for the preservice teachers, and updating the lesson overview and background sections to ensure that the preservice teachers understood where they were being dropped into a larger instructional sequence.

Once we had developed task materials that we hypothesized would allow us to make valid inferences about preservice teachers' ability to facilitate argumentation-focused discussions, we then engaged in a set of tryouts for each performance task. For each tryout, we recruited five to 10 preservice teachers to pilot the task with us within the simulated classroom.



Prior to each tryout, our research team trained interactors on how to enact the student avatars' responses in alignment with the student thinking profiles developed for each task. During the interactor training, we systematically gathered additional information to inform revisions of those materials, assigning a team member as a dedicated observer for each section of training to document where the interactors needed additional support.

For the tryouts, each participating preservice teacher reviewed the preservice teacher-facing task document to prepare for their simulated discussion and then facilitated a discussion for up to 20 minutes with the five student avatars in the simulated classroom. Our research team video recorded these discussions and later scored each one based on the scoring rubric we had developed from the five dimensions of our construct and the progression levels for each dimension. We also gathered self-reported data from each preservice teacher via a task survey and semistructured interview to learn about their perceptions of the task authenticity, interactions with the student avatars, their discussion performance, and the usefulness of the simulated teaching experience integrated within mathematics and science elementary method courses. Our research team analyzed these data sources to identify patterns in the preservice teachers' perceptions of these task materials and their discussion performances and then used the tryout findings to refine the task materials and our scoring rubric further.

These revisions, like those that took place after expert review, included attention to the clarity of wording throughout the preservice teacher-facing materials, which involved simplifying wording and presentation, refining the teaching tips to call attention to points that had been misunderstood, and in a few cases, revising the core content of the task to better fit the 20-minute time limit. We also revised interactor training materials to provide more support in areas that we had observed to be difficult and to refine language where we had observed it to be confusing to one or more interactors during training. For four of the eight tasks, the resulting revisions were substantial enough to warrant a second round of tryouts using a similar process of data collection, analysis, and task refinement.

Once we finalized the task materials and scoring rubric based on the tryout findings, we then used them in the research project's main study within multiple sections of elementary mathematics and science courses at three different universities in the United States.

## **Section 2: Content Focus of the Performance Tasks**

Each set of performance tasks in mathematics or science is grounded in a single high-leverage content area, which, as described in Martin-Raugh et al. (2016), is operationalized following the model of Ball and Forzani's (2011) high-leverage practice framework to include content of the student curriculum that is foundational, spans multiple grade levels, and makes up a significant component of the student curriculum and in which students often struggle absent strong instruction. In other words, it is the content that is most consequential for students to learn well and, therefore, most important for teachers to teach skillfully. In science, the content area of focus is matter and its interactions; in mathematics the content area of focus is fractions and operations with fractions.

### **Content Focus of the Science Tasks**

In 2018, Martin-Raugh et al. identified matter and its interactions as one of the high-leverage content topics within the elementary science curriculum via a systematic analysis of the science performance expectations in the Next Generation State Standards (National Research Council, 2013) and a survey of elementary teachers and faculty members. Research literature has also widely acknowledged that matter and its interactions is a difficult content area for teachers to learn how to teach; however, there exists a broad empirical research base on which to model common student understandings and misunderstandings (Smith et al., 2006; Stevens et al., 2010).

One goal of this focus on high-leverage content was to create a coherent and connected set of performance tasks that would fit together across the time span of a semester, make sense in sequence, and include core content that teacher educators likely would have made a focus of instruction in their work with preservice elementary teachers. Within the set of science performance tasks, the Mystery Powder performance task, unlike the others, was designed to be used as a pre and post measure at the beginning and end of an elementary methods course and therefore needed to be conceptualized such that it would be reasonable for preservice teachers to engage in before and after the other three tasks. Identifying materials based on their properties is a topic that teachers often return to at different points of the curriculum and in which students can engage with different levels of sophistication across multiple grades,

making it a good fit to this purpose. The other three performance tasks—Conservation of Matter, Modeling Matter, and Changing Matter—were designed to be presented in order across the semester, as this was our envisioned use case, while still standing alone if used individually outside of this sequence.

The Conservation of Matter task focuses on determining whether matter is conserved when combining sugar, water, and lemon juice to make lemonade and deciding what observations to use as evidence to make that determination. The Modeling Matter task is organized around three different drawn models showing what happens when a drop of food coloring is added to a container of water and involves the students in coming to a consensus about the features of an improved model that would better describe and explain this phenomenon. The Changing Matter task is grounded in the question of what evidence indicates that a new substance has been formed and involves the students in considering whether a new substance was formed when mixing different substances together (baking soda and pepper; white vinegar and baking soda; and white vinegar and milk).

This ordering of ideas meets two prespecified criteria. First, this sequence allows the student avatars to appear to advance through a typical science instructional sequence by moving from a focus on considering the structure and properties of matter during physical changes to then considering specific properties of matter and how they can be used as evidence to differentiate between physical and chemical changes. Second, the specific science of each task does not depend directly on the science of the prior task, minimizing instances where a preservice teacher might expect a particular student avatar to remember the exact content of the prior task. The Conservation of Matter, Modeling Matter, and Changing Matter tasks are described in detail in other reports in this series. In the following section we describe in more detail the Mystery Powder task, which is the focus of this report.

### **The Mystery Powder Science Task**

The Mystery Powder performance task is focused on identifying an unknown mystery powder based on its properties and coming to consensus about which properties are most useful to identify the unknown powder. Prior to the discussion, the students already collected data on various properties (e.g., color, texture, reaction with vinegar) of six known powders

(e.g., salt, sugar, baking soda) and one unknown mystery powder, which is one of the six known powders. The preservice teacher is provided a packet of materials (the preservice teacher-facing materials) prior to facilitating the discussion in the simulated classroom following a template used across the full set of tasks. The packet describes the work that five students have already completed on the science investigation, including showing written work from each student or group of students. For this particular task, one of the students, Carlos, has worked alone, and the other four have worked in pairs: Mina with Will and Jayla with Emily. Carlos systematically compared the properties of the known powders with the properties of the unknown mystery powder to correctly identify the mystery powder, although he recognized that weight<sup>2</sup> was not useful in determining the identity of the mystery powder and did not use that property to make his determination about the mystery powder's identity. Mina and Will focused on too few properties, only comparing a few of the mystery powder's properties to those of the known powders. In doing so, they failed to identify the mystery powder correctly or demonstrate an in-depth understanding of the properties of various powders and how they can be used to identify an unknown powder. Jayla and Emily, the last group, used all the properties to make comparisons between the mystery powder and the six known powders and found the one known powder that matched exactly with the mystery powder's properties. Although they did correctly identify the mystery powder using this approach, their approach failed to attend to the properties that were most useful and included one property (weight) that was not useful in determining the mystery powder's identity. Each approach offers useful aspects for consideration, although specific decisions the students made are variably useful for identifying the unknown mystery powder. This intentional variation across the three groups provides opportunities for the preservice teacher to guide the students in critiquing and comparing the approaches, to elicit justifications for their claims about the mystery powder's identity, and to guide them toward consensus about both the mystery powder's identity and their conclusions around which properties are most useful in identifying the unknown mystery powder.

### Section 3: The Generalized Task Design

As referenced previously, one of the outcomes of the design-based research process described in the first section was the development of a stable set of task components to be used across all eight (four science and four mathematics) tasks and designed to support a consistent experience for preservice teachers. The resulting template can be used to support future development by providing a structure for newly developed tasks with different content and is described here in the context of the Mystery Powder task.

Each task is made up of two types of components: the preservice teacher-facing materials and the interactor training materials. Table 2 lists the task components of the preservice teacher-facing materials, which includes three documents for each task. The Introduction to the Simulated Classroom and the Warm-Up Task are separate handouts that are used in common across all eight performance tasks and provide an overview of how the simulation works and a brief familiarization exercise to get the preservice teacher started before they lead the discussion. The main document is the performance task itself (see the appendix for the full text of the Mystery Powder performance task), which is designed to help the preservice teacher plan for and lead the discussion. Task components included in the preservice teacher-facing task document include the sections Introduction to the Task, Lesson Overview, Student Responses, Making Sense of the Student Work, Shared Workspace Pages, Features of High-Quality Discussions Focused on Argumentation, and Video Examples of High-Quality Discussions Focused on Argumentation. Teaching tips appear throughout the document rather than as a separate section.

**Table 2. Components of the Preservice Teacher–Facing Task Materials**

| Component  | Purpose   |
|--|---|
| Introduction to the Simulated Classroom (separate handout)             | This stand-alone handout acquaints the preservice teacher with the basic functionality of the simulated classroom as well as introduces them to each of the five students via short bios. It also includes links to short videos in which the students introduce themselves.  |
| The Warm-Up Task: Taking the Students' Lunch Orders (separate handout) | The warm-up task, which takes about five minutes, is a scripted task in which the preservice teacher takes the students' lunch orders. It is intended to allow the preservice teacher to become accustomed with the simulated environment before starting the discussion.   |
| Introduction to the Task   | This task component orients the preservice teacher to the task. It includes a clear statement of the student learning goal and what the preservice teachers should aim to do during the discussion.   |
| Lesson Overview  | This task component situates the 20-minute discussion within the larger lesson and instructional sequence, describing students' background knowledge as well as what transpired in the class before the discussion began.   |
| Student Responses  | This task component provides each student group's written work, which was generated prior to the discussion.  |
| Making Sense of the Student Work                                       | This task component complements the student responses and provides explanatory text to help the preservice teacher understand the students' written work. The explanatory text identifies salient features of the students' ideas that might inform the discussion.   |
| Shared Workspace Pages   | This task component includes copies of the written student work and any other relevant reference material (e.g., class data table). It can be printed out for use during the discussion.  |
| Features of High-Quality Discussions Focused on Argumentation          | This task component is a short list of the key features of high-quality discussions as we have defined them and includes a set of questions about each feature. The preservice teacher can use the questions before or after the discussion to support them in considering how well their discussion will or did meet the task's specified student learning goal. |

| Component   | Purpose   |
|---|---|
| Video Examples of High-Quality Discussions Focused on Argumentation | This task component provides links to publicly available examples of classroom discussions that illustrate some of the features of high-quality argumentation-focused discussions. The preservice teacher can use the examples to better understand these features and how to incorporate them into their discussion. |
| Teaching Tips   | This task component is embedded throughout the preservice teacher-facing materials and includes teaching tip bubbles that call attention to important ideas about how the discussion might be planned and enacted.  |

**Table 3. Components of the Interactor Training Materials**

| Component   | Purpose   |
|---|---|
| Non-Task-Specific Training                            | The non-task-specific training materials cover the discussion construct, direct interactors in how to be responsive to teacher prompts to engage in student-to-student interaction, and include the “testing the waters” guidelines. This component also includes independent study of the warm-up activity materials and culminates with an interactive practice session between the interactor and a trainer. |
| Task-Specific Lessons 1 & 2                           | For each task, Lesson 1 is an overview of the task and Lesson 2 is an overview of the student profiles for that task, including independent video-guided study of what each student thinks initially as well as how their thinking may shift over time.   |
| Task-Specific Lesson 3: The Student Profile Check Out | Lesson 3 is a face-to-face session in which the trainer leads the interactor through a standardized set of questions to ensure adequate mastery of the student profiles for the task.   |
| Task-Specific Lesson 4: The Observational Workshop    | For Lesson 4, the interactor meets with two trainers, one of whom plays the part of a teacher and enacts four separate practice discussions while the second trainer provides targeted feedback on the interactor’s performance. The four teacher profiles are carefully constructed to represent the breadth of discussion approaches the interactor is likely to encounter.                                   |
| Task-Specific Lesson 5: The Final Check Out           | Lesson 5 is also a face-to-face session with a trainer who enacts two more teacher profiles. Recordings of the session are uploaded and scored by the trainer for adequate fidelity to interactor training guidelines.  |

Table 3 lists the components of the interactor training materials, including a series of lessons that combine self-study modules with planned interactive practice with a content expert or trainer in order to help the interactor master the delivery of the task in the simulated classroom.

The content of the preservice teacher and interactor components are deeply intertwined. For example, a core part of the preservice teacher-facing materials is the presentation of student work that the student avatars have completed in advance of the simulated discussion (component: Student Responses). Every task includes this component, although the number of student groups varies. For each task, then, this necessitates a parallel component of the interactor training (component: Task-Specific Lessons 1 & 2) in which the interactor learns the student avatars' initial ideas and the work they have done prior to the discussion as well as their dynamic content profiles that dictate how their understandings would change over time in response to the teacher's (or other student avatars') statements or questions. For example, for the Mystery Powder task, the written student work explains clearly what Mina and Will did to generate their claim about the mystery powder's identity and which specific properties they think are most useful to identifying the unknown mystery powder. The accompanying interactor training specifies that they have a basic understanding that the mystery powder must match at least some properties of a known powder, although their understanding of what the mystery powder is and how to identify the mystery powder using various properties could use some revisions (Figure 4).

All components were designed, to the greatest extent possible, to be uniform in ways that are adaptation-friendly, allowing for the insertion of new content as needed to create new tasks. We next discuss some of the critical design considerations that informed our design of the task components using specifics from the Mystery Powder task to illustrate how some of these considerations are taken up and addressed in this performance task.



**Figure 4. Linked Components of Materials Include Student Work (for the Preservice Teacher) and Instructional Videos (for the Interactor)**

Mina & Will

The mystery powder is flour. We identified it by looking at some of the important properties. The flour & mystery powder are both white & smooth, and both weigh 24 grams. These things are the same so we know it is flour.

The most helpful properties are the ones you can see and measure.

The powders needed to look the same & they did. We weighed the powder to be sure and it weighed the same as the flour so we know it is the same thing.

### Student Profile A Mina and Will

ETS | MURSION

### Overview of Mina and Will's Profile

|   |  |
|---|--|
| <p><b>Key Understandings:</b></p> <ul style="list-style-type: none"> <li>• They understand that among the samples provided, there are a variety of different powders with different properties.</li> <li>• They recognize that some powders can be differentiated by their appearance.</li> </ul> | <p><b>Incomplete Conceptions:</b></p> <ul style="list-style-type: none"> <li>• They use incomplete evidence. Mina and Will do not use enough evidence to differentiate among the powders.</li> <li>• They misunderstand other evidence. "The flour and the mystery powder weigh the same."</li> <li>• They do not apply evidence based on how these powders interact with vinegar or water.</li> </ul> |
|---|--|

ETS | MURSION

### Design Consideration: Knowing Where to Start

The stand-alone Introduction to the Simulated Classroom as well as the Introduction to the Task and Lesson Overview components of the preservice teacher-facing materials are collectively intended to support the preservice teacher in knowing how to begin the discussion. In early tryouts, we realized that one of the logistical elements of the simulation we needed to manage was launching the preservice teacher straight into the discussion, as each teacher has only 20 minutes of simulation time and needs to use it for the intended interactive work of facilitating a discussion. A preservice teacher who spends time doing something else might well use up the full 20 minutes without engaging in the intended content discussion. For example, a natural starting point for preservice teachers encountering new students is to review prior knowledge, but reviewing what they already know takes time away from addressing the student learning goal in the task during the discussion. The Introduction to the Simulated Classroom, Introduction to the Task, and Lesson Overview components acquaint the preservice

teacher with the students' prior knowledge and describe exactly what has come before the discussion so that the preservice teacher has a clear sense of where they are to begin the discussion.

### **Design Consideration: Understanding the Task Purpose**

One area in which we found it necessary to build in substantial support across task components was helping the preservice teachers understand that they should be encouraging students to interact with one another directly. Many novice teachers struggle to engage students in this way. In contrast, a frequent and less productive pattern of engagement is known as the initiate—respond—evaluate (IRE) response pattern (Cazden, 1988) in which the teacher interacts with individual students in turn, intervening at each step. One goal of these performance tasks is to support preservice teachers in learning to avoid this pattern. However, if the preservice teachers interpret the instruction to “facilitate a discussion” as asking them to engage in IRE, they may not realize that they are not attempting to meet the intended goal. Paired with a technology environment in which a preservice teacher may not realize the students can speak to one another directly, there is some risk of misdirection on the preservice teacher's part. That misdirection would have represented a source of measurement error for us as it would be difficult to distinguish performances in which the preservice teacher was unable to elicit student interaction from those in which the preservice teacher did not understand that student interaction was possible.

We sought to counter this challenge in several ways across components of both the preservice teacher-facing and interactor training materials. First, we clarified the discussion goal across all tasks to make it clear that student-to-student interaction was possible and desirable. For example, Figure 5 shows text from the Introduction to the Task component stating, “You can encourage the students to talk to one another, ask one another questions, and respond to one another's ideas.” Along with this instruction is a teaching tip bubble that cautions the preservice teacher to allow wait time for students to respond.

### Figure 5. Text From the Introduction to the Task Component Supporting Student-to-Student Dialogue

**TEACHING TIP:**

The student avatars are not computer programmed, so it may take them a little time to think and respond.

You will have up to 20 minutes to lead this discussion in a simulated classroom environment made up of five upper-elementary student avatars. The students will be able to hear and see you, and they will respond in real time just like students in a real classroom. You can encourage the students to talk to one another, ask one another questions, and respond to one another's ideas.

Depending on how the discussion unfolds, you may or may not reach a satisfying conclusion by the end of the session time, and it is fine if you do not. If you do not reach a satisfying conclusion, just wrap up the discussion and indicate that you will pick the discussion up during the next class.

Then, on the interactor side, we built in two deliberate instances of student-to-student dialogue intended to help make sure the preservice teacher is aware that direct student interaction is possible (both of these instances are addressed as part of non-task-specific training as they are common across all eight tasks). First, during the warm-up task, one student jumps in and speaks directly to another student. Second, at some point during the first few minutes of the discussion, the interactor is instructed to engage in what we call “testing the waters,” by having one student jump in and engage in a brief back-and-forth dialogue with another. In general, the interactor will not have the students engage in this way without prompting, as the preservice teacher is supposed to be learning how to elicit such interaction. But for testing the waters, the interactor makes an exception. This dialogue serves two purposes: First, it is an additional reinforcement to the preservice teacher that students can speak directly with one another, and second, it gives the interactor valuable information about the preservice teacher's initial stance toward how student centered they would like the discussion to be. If the preservice teacher tries to quiet the students or asks them to raise hands, these are signs that the preservice teacher may be discouraging direct student-to-student interaction. However, if the preservice teacher encourages or praises the students or tries to build on the interaction, these are signs that the preservice teacher may be encouraging it.

**Design Consideration: Support in Unpacking Student Thinking**

Each task is designed on the premise that students have already worked on a given science investigation in advance of being called together to discuss their work. The students' written work is provided ahead of time (task component: Student Responses) so that the preservice teacher can review and plan the discussion based on it. In addition, we provide information to the preservice teacher to help them make sense of the written student work (task component: Making Sense of the Student Work). This text specifies, for example, whether the claim the students have given is correct, partially correct, or incorrect and what they might have been thinking about and calls attention to important things the preservice teacher might notice or pay attention to in planning the discussion. For example, for Mina and Will's work discussed previously, this text states that their claim is incorrect and the approach they used to identify the mystery powder is also not supported by the empirical data collected (Figure 6). Further, it directs the teacher to notice that Mina and Will have a basic understanding that the mystery powder should match at least some properties of a known powder, but that they focused on too few properties and not always ones that were most useful for this purpose.

**Design Consideration: Static and Dynamic Student Profiles**

As mentioned previously, interactor training includes both static profiles for students' personalities and initial content ideas as well as dynamic profiles reflecting their likely patterns of change. One characteristic of these tasks is that the students will contribute most of the key ideas if the preservice teacher is facilitating the discussion in a productive way. This means that interactors need training in both when to introduce those ideas and how the individual student avatars should respond to those ideas or sets of ideas, whether presented by the preservice teacher or by other students in response to the preservice teacher's prompting.

**Figure 6. Making Sense of the Student Work Component for Mina and Will**

**Making Sense of Mina and Will's Work**

**Is the claim correct?**

- They are not correct that the mystery powder is flour.
- Their approach is not correct.

**Things to notice about Mina and Will's approach to identifying the mystery powder:**

- This approach does not use enough properties to identify the mystery powder.
- This approach uses the property of weight, which does not help to identify the unknown powder in this example.
- There are a number of things that make it clear that these students do not have a strong understanding of how to determine the identity of the unknown powder, including the following:
  - Their focus on too few properties
  - Their use of properties, such as color, that do not help to differentiate among the powders
  - Their failure to recognize that other powders had the same properties as the ones they relied on to make their selection
  - Their failure to use properties that can clearly demonstrate the identity of the mystery powder
  - Their failure to recognize that weight is not a property that can distinguish the powders from one another

**Things to notice about Mina and Will's explanation:**

- The explanation demonstrates a basic understanding that the mystery powder must match at least some properties of a known powder.
- The explanation does not demonstrate an in-depth understanding of the properties of the various powders.
- The explanation does not acknowledge that other powders (cornstarch, baking soda, and baking powder) have the same texture and color as the flour and the mystery powder.



In the Mystery Powder task, for example, one group, Mina and Will, proposed an incorrect claim and approach for identifying the mystery powder. The interactor training (Figure 7) notes specific ways in which their thinking would (and would not) shift, including noting that just seeing the other students' claims and approaches would be insufficient for them to change their thinking, and that seeing that their claim is wrong would convince them that they did something incorrect but would not help them understand why. The interactor training materials specify key ideas that would need to be addressed during the discussion to shift Mina and Will's understanding. First, they would need to understand why it is necessary to consider other properties, besides the ones they initially considered, to identify the mystery powder. Changes to their understanding about the mystery powder's identity could occur after they are shown empirical evidence that indicates how specific properties are similar or different between the known powders and the mystery powder, including their initial incorrect claim. Second, they would need to be convinced that weight is not a useful property to identify the

mystery powder, as the same powder could have a different weight depending on how much of the powder is being measured. Those ideas could come from classmates, the teacher, or a combination. It also clarifies what types of empirical evidence or arguments would be more or less helpful to convince Mina and Will about the limitations of their initial claim and evidence-based reasoning.

**Figure 7. Slides From the Interactor Training: Unlocking Mina and Will’s Thinking**

### Ways in which our thinking might evolve...

- We could be brought to understand that there are limitations to our approach, if someone could clearly explain why it is necessary to consider other properties.
  - We still say our approach “works,” it is just that we may have overlooked some things.
  - Challenging our use of the weight of the mystery powder – and how other powders can have the same weight - might help our thinking evolve.
- We could understand that there are other approaches that work, and even that they might be better in some cases.
  - We still like our approach best.
  - But we could be convinced that it might be better to use more properties to identify an unknown powder.
  - We might even be convinced that weight is not as important a property as we think.





MURSION
11

copyright © 2017 by educational testing service. all rights reserved. ets, the ets logo and measures the power of learning are registered trademarks of educational testing service (ets). all other trademarks are property of their respective owners. 8888

### In Summary...

Mina and Will understand that they need to use the properties of the known materials to identify the mystery powder.

| Incomplete conceptions that Mina and Will have  | Ways in which we might be “unlocked” by the teacher or other students   |
|---|---|
| Mina and Will do not understand that they did not use enough properties to correctly identify the mystery powder. The properties that you can directly see and measure are not enough in this case. | <i>The other students might tell them that, “The mystery powder is clear when you mix it with water, but the flour isn’t.”<br/>                     “The mystery powder bubbles when you add vinegar; flour doesn’t.”<br/>                     “Other powders are smooth and white, too, not just flour.”</i> |
| Mina and will do not understand that some pieces of evidence provide better support for making a decision.  | <i>“The bubbles that you see when you mix together the mystery powder and vinegar makes it really clear that the mystery powder is not flour.”</i>  |
| Mina and Will do not understand that weight is not a useful property to identify the mystery powder.  | <i>“If you measure half of the amount of mystery powder, it changes the weight, but doesn’t change the substance itself. Even if you have less or more it still will have all of the other properties and it will still be baking soda.”</i>  |



MURSION
12

copyright © 2017 by educational testing service. all rights reserved. ets, the ets logo and measures the power of learning are registered trademarks of educational testing service (ets). all other trademarks are property of their respective owners. 8888

### Conclusion

The preservice teacher-facing materials for the Mystery Powder task can be found in the appendix. Our goal in this project was to create a set of simulation-based performance tasks that can be used to support preservice teachers in learning how to facilitate argumentation-focused discussions in two content areas: mathematics and science. In that project work, we video recorded each preservice teacher's discussion session for each task and provided detailed written feedback as well as access to the video to both the preservice teacher and the course teacher educator. We hypothesized that that preservice teacher would be supported on multiple levels. First, there is an aspect of experiential learning, as the preservice teacher sees the student avatars engage in response to their prompts during the simulation. Second, the preservice teacher learns from the written feedback. Although we provided feedback to the preservice teachers, that feedback could also come from a teacher educator or coach, or the preservice teacher could be guided in self-reflection. Third, the performance tasks provide a type of formative assessment information to the teacher educator who can see, in looking across the videos or the feedback, patterns in class or the individual performance that allow the teacher educator to adjust instruction within the methods course.

Our design process was deliberately systematic and was intended to support productive adaptation of the task materials that resulted. Although our work took place in the context of preservice teacher learning and for use with the Mursion simulated classroom environment, the tasks could easily be used for professional development and adapted for use in other simulation environments using other technologies or nontechnological approaches. For example, a teacher educator or coach might use the materials for the basis of live role playing and adapt the interactor training materials to help preservice teachers play the role of students. The full set of project materials, including interactor training materials and guidelines for scoring the discussions, is archived in an online repository (<https://data.qdr.syr.edu/dataverse/go-discuss>) and is publicly available for use and adaptation.

## References

- Ball, D. L., & Forzani, F. M. (2011). Building a common core for learning to teach: And connecting professional learning to practice. *American Educator*, *35*(2), 17–38.
- Barkai, R., Tsamir, P., Tirosh, D., & Dreyfus, T. (2002). Proving or refuting arithmetic claims: The case of elementary school teachers. In A. D. Cockburn & E. Nardi (Eds.), *Proceedings of the Twenty-Sixth Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 57–64). University of East Anglia, School of Education and Professional Development.
- Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning*. Heinemann.
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8.  
<https://doi.org/10.3102/0013189X032001005>
- Howell, H., & Mikeska, J. N. (2021). Approximations of practice as a framework for understanding authenticity in simulations of teaching. *Journal of Research on Technology in Education*, *53*(1), 8–20. <https://doi.org/10.1080/15391523.2020.1809033>
- Kazemi, E., & Stipek, D. (2009). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *Journal of Education*, *189*(1–2), 123–137.  
<https://doi.org/10.1177/0022057409189001-209>
- Martin-Raugh, M., Mikeska, J. N., Steinberg, J., & Minsky, J. (2018). *Investigating science content knowledge and teaching practices for beginning elementary school teachers* (Research Memorandum No. RM-18-09). ETS.
- Martin-Raugh, M., Reese, C. M., Howell, H., Tannenbaum, R. J., Steinberg, J. H., & Xu, J. (2016). *Investigating the relevance and importance of high-leverage mathematical content knowledge areas for beginning elementary school teachers* (Research Memorandum No. RM-16-20). ETS.
- Mikeska, J. N., Howell, H., & Straub, C. (2019). Using performance tasks within simulated environments to assess teachers' ability to engage in coordinated, accumulated, and dynamic (CAD) competencies. *International Journal of Testing*, *19*(2), 128–147.  
<https://doi.org/10.1080/15305058.2018.1551223>



- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62. [https://doi.org/10.1207/S15366359MEA0101\\_02](https://doi.org/10.1207/S15366359MEA0101_02)
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*.
- National Research Council. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.
- Reid, D., & Zack, V. (2009). Aspects of teaching proving in upper elementary school. In D. A. Stylianou, M. L. Blanton, & E. J. Kruth (Eds.), *Teaching and learning proof across the grades: A K–16 perspective* (pp. 133–146). Routledge.
- Russell, S. J., Schifter, D., Bastable, T., & Kasman, R. (2017). *But why does it work?: Mathematical argument in the elementary classroom*. Heinemann.
- Smith, C., Wiser, M., Anderson, C., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic-molecular theory. *Measurement*, 14(1 & 2), 1–98. <https://doi.org/10.1080/15366367.2006.9678570>
- Stevens, S. Y., Delgado, C., & Krajcik, J. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, 47(6), 687–715. <https://doi.org/10.1002/tea.20324>
- Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutiérrez, G. C. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 315–351). Sense Publishers. [https://doi.org/10.1007/978-94-6300-561-6\\_9](https://doi.org/10.1007/978-94-6300-561-6_9)
- Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, 78(3), 516–551. <https://doi.org/10.3102/0034654308320292>

Appendix: The Mystery Powder Preservice Teacher-Facing Task Materials

# ETS Research Study on Facilitating Student Discussions

## The Mystery Powder Discussion Task



Credit: Image courtesy of Mursion



## Table of Contents

|   |           |
|---|-----------|
| <b>Introduction to the Mystery Powder Discussion Task .....</b>   | <b>28</b> |
| What is the student learning goal for this discussion?  |           |
| What will you do?   |           |
| <b>Lesson Overview .....</b>  | <b>29</b> |
| A description of the scenario, teaching content, and information about the students’<br>background and what they have been working on |           |
| <b>Student Responses and Making Sense of the Student Work .....</b>   | <b>31</b> |
| Copies of the written student work, and notes on how you might make sense of it   |           |
| <b>Shared Workspace Pages .....</b>   | <b>37</b> |
| Copies of the pages that will be available on the shared workspace  |           |
| <b>Features of High-Quality Discussions Focused on Argumentation .....</b>  | <b>38</b> |
| A description of what it would look like if someone does well on this task  |           |
| <b>Video Examples of High-Quality Discussions Focused on Argumentation.....</b>   | <b>39</b> |
| Links to vignettes and video examples   |           |

**Note:** The materials provided in the following sections are designed to help you plan for your discussion and should help you understand what you are supposed to do. You will also find teaching tips embedded throughout the document. An additional document, “An Introduction to the Simulated Classroom and Student Avatars,” is also available for your use.

### **TEACHING TIP:**

The teaching tips are designed to enhance your understanding of the task and your performance. You are not required to use them; instead, they are here for you to use however you wish.

## Introduction to the Mystery Powder Discussion Task

### What is the student learning goal for this discussion?

#### Student Learning Goal

Students will construct an argument about the identity of a mystery powder based on its properties. They will come to a consensus about which properties are most useful in identifying the unknown powder.

### What will you do?

You will lead a discussion during which you will guide students to use their evidence and reasoning to come to a consensus about the identity of the mystery powder and the properties that were most useful in identifying it.

Your focus should be on engaging the students in discussion with one another and in the practice of argumentation. During the discussion, be sure to have students focus on considering multiple perspectives and coming to a shared consensus about the best explanation of the phenomena being explored.

#### TEACHING TIP:

Be sure to have students focus on constructing, defending, and critiquing their own and others' claims using scientific evidence and reasoning.

#### TEACHING TIP:

The student avatars may take a little time to think and respond. Provide wait time, just as you would in a classroom.

You will have up to 20 minutes to lead this discussion in a simulated classroom environment made up of five upper elementary student avatars. The students will be able to hear and see you and will respond in real time, just like students in a real classroom. You can encourage the students to talk to one another, ask one another questions, and respond to one another's ideas.

Depending on how the discussion unfolds, you may or may not reach a satisfying conclusion by the end of the session time, and it is fine if you do not. If you do not reach a satisfying

conclusion, just wrap up the discussion and indicate that you will pick the discussion up during the next class.

## Lesson Overview

**Student Learning Goal:** Students will construct an argument about the identity of a mystery powder based on its properties. They will come to a consensus about which properties are most useful in identifying the unknown powder.

**Background:** Students are in fifth grade. They have experience with, but may not be proficient in, the following:

- Making observations and taking accurate measurements of matter
- Collecting and recording data (observations and measurements) in a table
- Making claims and supporting those claims, using evidence and reasoning

**Note:** Not every student has the same level of understanding or ability with the content ideas and practices noted above, but these are the learning opportunities that all students in this classroom have previously experienced.

### Prior to starting the discussion:

- Students worked in three small groups to observe and measure the properties of six different powders (flour, cornstarch, baking soda, baking powder, sugar, and salt). Will and Mina worked together, Jayla and Emily worked together, and Carlos worked alone.
- The class created a class data table, which is shown on the next page. (The class data table will be available to you to write on and display on the tablet during your discussion with the students.)
- The students were then asked to conduct an investigation to identify one unknown sample (the mystery powder) that is identical to one of the known powders. All groups had the same mystery powder.
- Each group was asked to respond to the following prompt: "Refer to the data that you collected and explain how you used it to make your identification. Which properties were the most useful for identifying the unknown powder?"
- Students wrote their work on the shared workspace in preparation for sharing with the class. (Their written work is shown in the following sections and will be available to you to write on and display on the tablet during your discussion with the students.)

### TEACHING TIP:

Everything you do in planning for and leading this discussion should help the students make progress toward the student learning goal.

**TEACHING TIP:** Properties of materials are characteristics that help to identify a particular substance or differentiate among substances. Some properties are independent of the amount or quantity of the substance, while other properties change when the amount of the substance changes. You should encourage students to consider the full range of observations noted in the class data table and help build consensus about the properties that are most useful for verifying the mystery powder's identity.

**TEACHING TIP:** At the fifth-grade level, students likely will not differentiate between mass and weight and may use these terms interchangeably. In this task, we use the term weight.

### Class Data Table

| Powder        | Texture | Color | Weight   | Reaction with vinegar | Mix with water |
|---------------|---------|-------|----------|-----------------------|----------------|
| Flour         | Smooth  | White | 24 grams | No reaction           | Looks cloudy   |
| Cornstarch    | Smooth  | White | 20 grams | No reaction           | Looks cloudy   |
| Baking Soda   | Smooth  | White | 24 grams | Bubbles               | Looks clear    |
| Baking Powder | Smooth  | White | 24 grams | Bubbles               | Looks cloudy   |
| Sugar         | Rough   | White | 26 grams | No reaction           | Looks clear    |
| Salt          | Rough   | White | 22 grams | No reaction           | Looks clear    |



**Student Responses: Mina and Will's Work**

Mina & Will

The mystery powder is flour. We identified it by looking at some of the important properties. The flour & mystery powder are both white & smooth, and both weigh 24 grams. These things are the same so we know it is flour.

The most helpful properties are the ones you can see and measure.

The powders needed to look the same & they did. We weighed the powder to be sure, and it weighed the same as the flour so we know it is the same thing.

### Making Sense of Mina and Will's Work

#### Is the claim correct?

- They are not correct that the mystery powder is flour.
- Their approach is not correct.

#### Things to notice about Mina and Will's approach to identifying the mystery powder:

- This approach does not use enough properties to identify the mystery powder.
- This approach uses the property of weight, which does not help to identify the unknown powder in this example.
- There are a number of things that make it clear that these students do not have a strong understanding of how to determine the identity of the unknown powder, including the following:
  - Their focus on too few properties
  - Their use of properties such as color that do not help to differentiate among the powders
- They did not recognize that other powders had the same properties as the ones they relied on to make their selection:
  - Their failure to use properties that can clearly demonstrate the identity of the mystery powder
  - Their failure to recognize that weight is not a property that can distinguish the powders from one another

#### Things to notice about Mina and Will's explanation:

- The explanation demonstrates a basic understanding that the mystery powder must match at least some properties of a known powder.
- The explanation does not demonstrate an in-depth understanding of the properties of the various powders.
- The explanation does not acknowledge that other powders (cornstarch, baking soda, and baking powder) have the same texture and color as the flour and the mystery powder.



**Student Responses: Jayla and Emily's Work**

Name(s): Jayla and Emily

We think the mystery powder is baking soda because it matches all of the properties on the chart that baking soda has listed. Baking soda is white, smooth, bubbles when mixed with vinegar, looks clear when mixed with water, and weighs 24 grams. The mystery powder has all the same properties.

All of the properties are useful and needed to decide what the unknown powder is, we can't just use one or two. Some of the powders have some of the same properties. Like all the powders we looked at were white so we wouldn't be able to identify the mystery powder if we just looked at the color. We had to find out which powder had the most matching properties as the mystery powder.

**Making Sense of Jayla and Emily's Work****Is the claim correct?**

- They are correct that the mystery powder is baking soda.
- Their approach has some good components, but it is not fully correct.

**Things to notice about Jayla and Emily's approach to identifying the mystery powder:**

- They used all of the properties to identify the mystery powder.
- They do not appear to have had a strong understanding of how to determine the identity of the unknown powder, as evidenced by their use of all of the properties without attention to the ones that are most useful in this situation (for example, what happens when the powder is mixed with vinegar or with water).
- This approach uses the property of weight, although observations of weight do not help identify what the powder is; thus weight is not useful in determining the identity of the mystery powder.

**Things to notice about Jayla and Emily's explanation:**

- The explanation of the approach demonstrates the need to compare several properties of the powders.
- The explanation uses the property of weight, which is not helpful for determining the identity of the mystery powder.
- The explanation compares all properties of the powders without considering which properties are more useful.

**Student Responses: Carlos' Work**

Name : Carlos

My Identification:

I looked at the powder. It was smooth and white, so I knew it wasn't salt or sugar. Then I put vinegar on it and it bubbled up, so I knew it was baking soda or baking powder. So I mixed in some water and it was clear so I knew it was baking soda because that is the only thing it could be with all those things.

The properties that were most useful were texture, color, reaction with vinegar and reaction with water. Weight just tells you how much you have, so I didn't bother with it.

**TEACHING TIP:**

Carlos was the only student who recognized that weight was not an important property for identifying the mystery powder. He understood that the weight would change with the amount of powder, but the other properties would remain the same.

### Making Sense of Carlos' Work

#### Is the claim correct?

- He is correct that the mystery powder is baking soda.
- His approach is correct and well constructed.

#### Things to notice about Carlos' approach to identifying the mystery powder:

- He recorded some basic observations first and then used his observations to decide which tests to conduct.
- He was able to eliminate different possible candidates for the mystery powder as he worked, and this helped him to decide which test to conduct next.
- This approach works well for determining the identity of the mystery powder by comparing the properties of all the powders.
- There are a number of things that make it clear that he understands this approach well and has done a thorough job in his work, including realizing that weight is not useful and using enough properties to eliminate all other powders.
- He could have done fewer tests and still correctly identified the mystery powder.

#### Things to notice about Carlos' explanation:

- The explanation of the approach demonstrates the need to compare several properties of the powders.
- The explanation states that weight is not useful for determining the identity of the mystery powder.
- The explanation shows how Carlos used the results of preceding observations and tests to drive the direction of subsequent tests.

### Shared Workspace Pages

The following are images of the shared workspace pages that will be available on the tablet. You and the students will be able to access and interact with these pages during the discussion. The tools on the toolbar can be used to draw or write on the pages. Blank pages are also available for you to use during the discussion.

Class Data Table

| Powder        | Properties |       |          |                       |                |
|---------------|------------|-------|----------|-----------------------|----------------|
|               | Texture    | Color | Weight   | Reaction with Vinegar | Mix with Water |
| Flour         | Smooth     | White | 24 grams | No reaction           | Looks cloudy   |
| Cornstarch    | Smooth     | White | 20 grams | No reaction           | Looks cloudy   |
| Baking Soda   | Smooth     | White | 24 grams | Bubbles               | Looks clear    |
| Baking Powder | Smooth     | White | 24 grams | Bubbles               | Looks cloudy   |
| Sugar         | Rough      | White | 26 grams | No reaction           | Looks clear    |
| Salt          | Rough      | White | 22 grams | No reaction           | Looks clear    |

Mina and Will's Student Work

Mina & Will

The mystery powder is flour. We identified it by looking at some of the important properties. The flour & mystery powder are both white & smooth, and both weigh 24 grams. These things are the same so we know it is flour.

The most helpful properties are the ones you can see and measure. The powders needed to look the same & they did. We weighed the powder to be sure and it weighed the same as the flour so we know it is the same thing.

Jayla and Emily's Student Work

Name(s): Jayla and Emily

We think the mystery powder is baking soda because it matches all of the properties on the chart that baking soda has listed. Baking soda is white, smooth, bubbles when mixed with vinegar, looks clear when mixed with water, and weighs 24 grams. The mystery powder has all the properties.

All of the properties are useful and needed to decide what the unknown powder is, we can't just use one or two. Some of the powders have some of the same properties. Like all of the powders we looked at were white so we wouldn't be able to identify the mystery powder if we just looked at the color. We had to find out which powder had the most matching properties as the mystery powder.

Carlos' Student Work

Name: Carlos

My Identification:

I looked at the powder. It was smooth and white, so I knew it wasn't salt or sugar. Then I put vinegar on it and it bubbled up, so I knew it was baking soda or baking powder. So I mixed in some water and it was clear so I knew it was baking soda because that is the only thing it could be with all those things.

The properties that were most useful were texture, color, reaction with vinegar and reaction with water. Weight just tells you how much you have, so I didn't bother with it.



## Features of High-Quality Discussions Focused on Argumentation

The discussion task you have been asked to complete is complex, and there are multiple approaches that you might take. The following list is a series of reflection questions for you to consider as you plan to lead a productive **discussion focused on engaging students in the practice of argumentation**. These features identify the main characteristics of high-quality discussions focused on argumentation. You might expect a helpful observer, such as a coach, peer, or instructor, to provide feedback on these features when observing your teaching in order to help you reflect on and learn from the experience.

- 1. Attending to Students' Ideas:** Did I make sure every student's voice was heard and that all students' ideas were valued?
  - Did I give every student an opportunity to participate in meaningful ways?
  - Did I make sure to include all ideas that students shared in their previous work?
- 2. Facilitating a Coherent and Connected Discussion:** Did the discussion make sense and feel organized and purposeful to the students?
  - Did I help the students make connections and build toward a shared understanding?
  - Did I help the students make sense of the discussion so that they could summarize the main takeaways and know what was learned?
- 3. Encouraging Student-to-Student Interactions:** Did I succeed in getting students to engage in discussion with one another?
  - Did I encourage students to speak to one another directly?
  - Did I provide opportunities for students to pose questions to one another or comment on and critique one another's ideas?
- 4. Developing Students' Conceptual Understanding:** Did I support students in developing a correct content understanding during the discussion?
  - Did I represent science concepts correctly?
  - Did I give students opportunities to evaluate the correctness of content ideas so that they could learn how to be part of the process of critiquing those ideas?
  - Did I consider any science content errors students had during the discussion and support students in working together to address those areas of confusion?
- 5. Engaging Students in Argumentation:** Did the discussion allow students to engage in argumentation?
  - Did I focus the discussion on ideas that were worth debating?
  - Did I provide opportunities for the students to make claims or conjectures, support them with reasoning or evidence, and consider and critique their own and others' ideas?

**TEACHING TIP:** In order to engage students in the practice of scientific argumentation, your goal should be getting students to talk to one another and to critique and build on one another's claims, evidence, and reasoning.

### Video Examples of High-Quality Discussions Focused on Argumentation

Learning how to facilitate discussions focused on argumentation can be challenging. Observing examples of students and teachers engaged in these types of discussions can be helpful. The following video links will allow you to see what it looks like and sounds like when elementary and middle school students engage in productive argumentation in science classrooms. We have also provided you with some questions to think about as you view these video examples and prepare to lead a productive **discussion focused on engaging students in the practice of argumentation**.

| Video focus   | Things to notice  | Video link <sup>a</sup> |
|---|---|-------------------------|
| Promoting student-to-student interaction                            | <p>How did the students engage in discussion with one another?</p> <p>How did the teacher promote student interaction?</p>  | Student Interaction     |
| Promoting students' evaluation and critique of competing claims     | <p>How did the students question and critique each other's ideas?</p> <p>What activities and strategies did the teacher use to help students address competing ideas?</p>                 | Competing Claims        |
| Encouraging students' use and evaluation of the quality of evidence | <p>How did the students engage in discussion with one another?</p> <p>How did the teacher promote student interaction?</p>  | Quality Evidence        |
| Supporting students in making their reasoning clear                 | <p>How did the students use scientific ideas to connect their claim and evidence?</p> <p>What activities and strategies did the teacher use to help students explain their reasoning?</p> | Reasoning               |

<sup>a</sup>These video examples are part of the Argumentation Toolkit website at [www.argumentationtoolkit.org](http://www.argumentationtoolkit.org).

### Notes

- <sup>1</sup> The development of the simulated environment also included feedback from multiple stakeholders, including our advisory board, and a compilation of reviews and iterative refinements to the students' physical appearance as well as their voicing, background, and personality profiles. Although this development process happened concurrently, it is not described in this report.
- <sup>2</sup> Based on recommendations in the Next Generation Science Standards, we do not distinguish between weight and mass at the upper elementary grade band in our performance tasks.