Kaldaras and Wieman
International Journal of STEM Education (2023) 10:18
https://doi.org/10.1186/s40594-023-00409-8

International Journal of STEM Education

RESEARCH Open Access

Cognitive framework for blended mathematical sensemaking in science

Leonora Kaldaras[1,2*] and Carl Wieman[1]

Abstract

Background Blended mathematical sensemaking in science ("Math-Sci sensemaking") involves deep conceptual understanding of quantitative relationships describing scientific phenomena and has been studied in various disciplines. However, no unified characterization of blended Math-Sci sensemaking exists.

Results We developed a theoretical cognitive model for blended Math-Sci sensemaking grounded in prior work. The model contains three broad levels representing increasingly sophisticated ways of engaging in blended Math-Sci sensemaking: (1) developing qualitative relationships among relevant variables in mathematical equations describing a phenomenon ("qualitative level"); (2) developing mathematical relationships among these variables ("quantitative level"); and (3) explaining how the mathematical operations used in the formula relate to the phenomenon ("conceptual level"). Each level contains three sublevels. We used PhET simulations to design dynamic assessment scenarios in various disciplines to test the model. We used these assessments to interview undergraduate students with a wide range of Math skills. Interview analysis provided validity evidence for the categories and preliminary evidence for the ordering of the categories comprising the cognitive model. It also revealed that students tend to perform at the same level across different disciplinary contexts, suggesting that blended Math-Sci sensemaking is a distinct cognitive construct, independent of specific disciplinary context.

Conclusion This paper presents a first-ever published validated cognitive model describing proficiency in blended Math-Sci sensemaking which can guide instruction, curriculum, and assessment development.

Keywords Cognitive framework, Validity, Blended sensemaking, Math sensemaking, Science sensemaking

Introduction

Blended mathematical sensemaking in science ("Math-Sci sensemaking") is a special type of sensemaking that involves developing deep conceptual understanding of quantitative relationships and scientific meaning of equations describing a specific phenomenon (Kuo et al., 2013; Zhao et al., 2021). Blended Math-Sci sensemaking is an important component of expert understanding of science and expert mental models (Redish, 2017). While various aspects of the Math-Sci sensemaking have been described for specific disciplines (Bing & Redish, 2007; Hunter et al., 2021; Lythcott, 1990; Ralph & Lewis, 2018; Schuchardt, 2016; Schuchardt & Schunn, 2016; Tuminaro & Redish, 2007), there has been little work on formulating and testing a theory of mathematical sensemaking as a cognitive construct that applies across different scientific fields. This paper offers initial evidence that a unified blended Math-Sci framework is possible. Having a general framework for discussing, diagnosing, and supporting the development of blended Math-Sci sensemaking across disciplines will help improve instruction and assessment principles.

To design the unified framework for blended Math-Sci sensemaking, we build on previously published theoretical framework that outlines different ways students can

*Correspondence:
Leonora Kaldaras
kaldaras@stanford.edu
[1] Stanford Graduate School of Education, Stanford, CA, USA
[2] PhET Interactive Simulations Department of Physics, University of Colorado Boulder, Boulder, USA

engage in Math and Science sensemaking separately (Zhao & Schuchardt, 2021). We use the Math and Science dimensions defined in Zhao and Schuchardt (2021) to develop a new theoretical cognitive model ("framework") for blended Math-Science sensemaking to answer the first research question of the study: (RQ 1): *How can one characterize the different ways of engaging in blended Math-Sci sensemaking?* The framework outlines qualitatively different proficiency levels reflecting increasingly sophisticated ways of engaging in blended Math-Sci sensemaking. Then, we investigated whether this theoretical framework indeed represents the various ways in which students engage in such sensemaking. This is the second research question of the study (RQ2): *To what degree does the validity evidence support the theoretical framework for blended Math-Sci sensemaking?* To answer RQ 2 we specifically focused on evaluating whether the validity evidence supported the existence of the categories and the order of the categories of the theoretical cognitive framework developed as part of RQ 1.

To answer RQ 2, we probe the levels of the framework by leveraging the capabilities of PhET simulations. Specifically, one of the key features of the sensemaking process is its dynamic nature focused on continuously revising an explanation based on new evidence to figure something out (Oden & Russ, 2019). The dynamic nature of PhET simulations provides a unique and suitable environment for assessing blended Math-Sci sensemaking skills. This supports revisions of explanations by calling on blended understanding of the scientific concepts and the underlying mathematical relationships.

In the context of blended Math-Sci sensemaking, the relevant mathematical equations represent processes described by specific variables that reflect a certain natural phenomenon. Simulations, in turn, represent a scientific model of the physical behavior that reflects the natural phenomenon with certain variables that control that behavior. The simulation allows learners to explore how the behavior depends on different variables, both qualitatively and quantitatively, therefore providing a meaningful context for engaging in blended Math-Sci sensemaking. Simulations provide a simplified (but not too simplified) system for exploring the mathematical complexity of the phenomenon described in the simulation.

We designed an interview protocol aimed at probing the levels of the theoretical blended Math-Sci sensemaking framework in the context of PhET simulations spanning physics, chemistry, and energy conversion disciplinary contexts. The range of scientific contexts was chosen to explore the extent to which the blended Math-Sci sensemaking varied with context. We collected and analyzed interviews with 25 undergraduate science and non-science majors with a wide range of Math skills to test the validity of the theoretical framework. The interview analysis provided evidence of the validity of the proposed theoretical framework, including both the existence of the categories and preliminary evidence for the order of the categories.

Literature review

Blended sensemaking refers to the process of combining separate cognitive resources to generate a new, blended understanding (Fauconnier & Turner, 1998). In the context of blended Math-Sci sensemaking, the two cognitive resources are the Science and the Math sensemaking, respectively. We define the blended Math-Sci sensemaking process as the process of integration of both Math and Science cognitive resources to make sense of phenomena as opposed to using only one of the cognitive resources (e.g., either Science or Math sensemaking). In this regard, blended Math-Sci sensemaking refers to a distinct cognitive construct that incorporates the elements of Math and Science sensemaking but exhibits unique features that result from the blending process. This is similar to the way Fauconnier and Turner (1998) describe the blending structure resulting from two mental spaces, in this case Math and Science sensemaking dimensions.

Evidence from prior studies suggests that the ability to engage in blended Math-Sci sensemaking reflects higher level, expert-like understanding (Redish, 2017), and has been shown to help students in solving complex quantitative problems in science (Schuchardt & Schunn, 2016). The framework for blended Math-Sci sensemaking presented here focuses on defining what proficiency in blended Math-Sci sensemaking looks like at various levels of sophistication. As shown in the description of the levels below, the essence of blended Math-Sci sensemaking lies in the student's ability to demonstrate the blending of the Math and Science dimensions when making sense of a phenomenon, as opposed to engaging in each dimension separately.

Defining what proficiency looks like at various levels of sophistication is important, because it provides understanding of how students develop competence in a cognitive construct (National Research Council [NRC], 2000). This understanding is essential for designing effective assessment, curriculum, and instructional strategies for supporting student learning (NRC, 2000, 2012a) and outlining what proficiency looks like and how it develops over time (NRC, 2000). Cognitive frameworks represent learning as a developmental process (NRC, 2012a, b; Smith et al., 2006) and provide a "road map" for the pathway that students can follow to achieve this

understanding (Duschl et al., 2007; NRC, 2012a, b; Smith et al., 2006).

To support students in blended Math-Sci sensemaking, it is important to understand what proficiency looks like at different levels of sophistication. The different levels of sophistication in this context refer to the ability to integrate the two dimensions in a more cognitively complex manner at the higher proficiency levels. At present, there has not been a coherent framework for characterizing proficiency in blended Math-Sci sensemaking. However, there has been considerable work characterizing different ways students can engage in Math and Science sensemaking separately (Zhao & Schuchardt, 2021) as well as characterizing blended sensemaking from different educational perspectives that are not describing different levels of proficiency (Bain et al., 2019; Gifford & Finkelstein, 2020). Gifford and Finkelstein (2020) developed a cognitive framework for mathematical sensemaking in physics which describes the process of sensemaking and relates it to basic cognition. Bain et al. (2019) study high- and low-quality Math sensemaking in the context of chemical kinetics, but they do not examine different levels of proficiency in blended Math-Sci sensemaking or how this extends across disciplines, which is the focus of this study.

Zhao and Schuchardt (2021) have provided a major advance in presenting a framework that captures sensemaking opportunities for mathematical equations in science grounded in the review of relevant literature. The framework presents the sensemaking along two separate dimensions: Science sensemaking and Math sensemaking. The categories within the dimensions are ordered theoretically to represent increasingly sophisticated levels of sensemaking. The framework presented by Zhao and Schuchardt is theoretical and has been used to characterize sensemaking opportunities provided by instructors (Zhao et al., 2021). However, Zhao and Schuchardt framework has not been validated in terms of characterizing the types of blended Math-Sci sensemaking that students can demonstrate in practice. Moreover, while the framework presented by Zhao and Schuchardt can be used for characterizing both Math and Science sensemaking and identify opportunities for blended sensemaking during instruction, the framework does not offer explicit guidance for supporting *blended* Math-Sci sensemaking at different levels of sophistication, which is the focus of the current work. This work builds on the work of Zhao and Schuchardt (2021) and develops the two separate cognitive dimensions of Science and Math into a unified cognitive dimension of blended Math-Sci sensemaking. The current study focuses on defining and distinguishing different proficiency levels of blended Math-Sci sensemaking for assessing and scaffolding instruction.

Theoretical framework

Developing cognitive model for defining proficiency

A cognitive model (also called a model of cognition) describes how students represent knowledge and develop proficiency in a domain (National Research Council [NRC], 2001). Proficiency refers to describing what mastery looks like in a domain. The understanding of how proficiency develops is essential for designing effective instructional and assessment strategies. Cognition models allow for empirical testing and valid interpretation of assessment results, aligning curriculum, instruction, and assessment with the purpose of helping students achieve higher proficiency in a given concept (NRC, 2001).

Blended Math-Sci sensemaking is a cognitive construct that has been studied in various fields of science. Zhao and Schuhardt (2021) provided categories divided into two dimensions (see Fig. 1): Science sensemaking and Mathematics sensemaking (Zhao & Schuhardt, 2021). Their Science sensemaking dimension includes four categories organized in the order of increasing sophistication of understanding: scientific label ("Sci Label"), scientific description ("Sci Description"), scientific pattern ("Sci Pattern") and scientific mechanism ("Sci Mechanism"). The Math sensemaking dimension includes five categories in order of increasing sophistication: "Math Procedure", "Math Rule", "MathStructure", "Math Relation" and "Math Concept". For example, logically, engaging in "Sci Mechanism" type of sensemaking requires first being able to identify specific properties and the corresponding variables relevant to characterizing a given phenomenon ("Sci Description"). Once the variables have been identified, it is possible to engage in identifying specific patterns among the relevant variables ("Sci -Pattern") in increasing order of sophistication. Finally, once the relevant patterns have been identified, it is possible to engage in developing a causal mechanistic account of the phenomenon ("Sci Mechanism"). Similar logic applies to the Mathematics sensemaking dimension. Zhao & Schuchardt note the need to empirically test these levels of sophistication for both dimensions. The current work extends the work of Zhao and Schuchardt and their categories of sophistication to develop and empirically validate a cognitive model that combines the two dimensions to achieve *blended* Math-Sci sensemaking.

Developing theoretical cognitive model for blended Math-Sci sensemaking

We used the Science and Mathematics sensemaking categories described by Zhao and Schuchard and blended them together to design new categories that

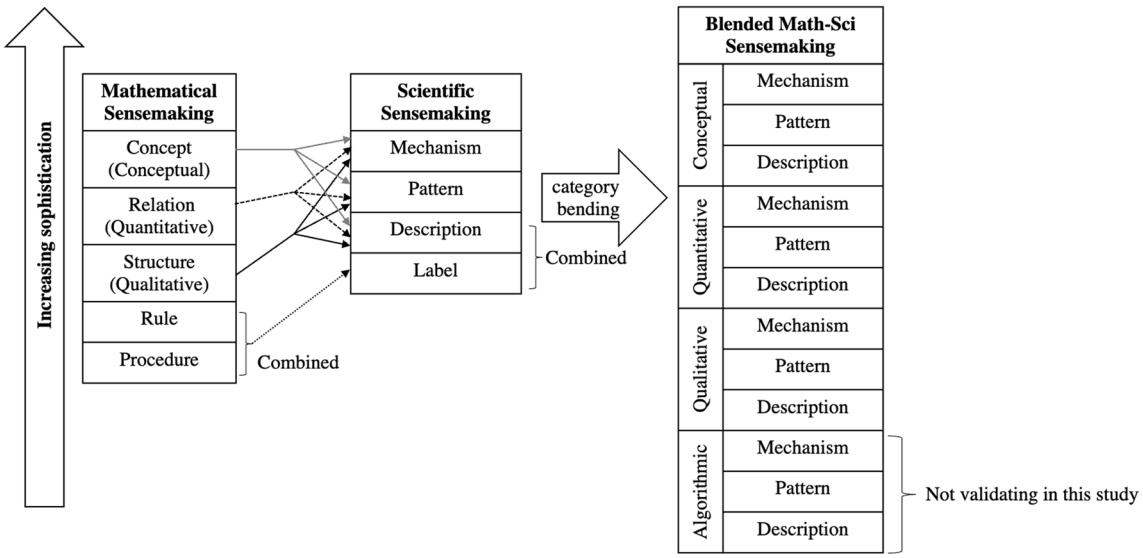

Fig. 1 Science and Mathematics categories of Zhao and Schuchardt (2021) blended into new categories making up the blended Math-Sci framework

each combine a Mathematics and a Science dimension to reflect the blended nature of the cognitive model. The blending process is illustrated in Fig. 1 and discussed in detail below. Since the focus of the cognitive framework is quantitative understanding of scientific phenomena, the levels of framework aim to describe increasingly sophisticated ways of providing quantitative accounts of phenomenon in question. Following the developmental approach premise, the lowest level (level 1) of the framework reflects limited ability to develop exact quantitative relationship focusing instead on providing qualitative accounts of phenomenon in question that will serve as steppingstones for developing the exact quantitative relationships with different degree of sophistication at levels 2 and 3. Furthermore, levels 2 and 3 of the framework reflect the ability to both develop quantitative relationships and demonstrate quantitative understanding of the previously known mathematical relationships at different levels of sophistication. Therefore, this framework can be used to guide instruction and assessment in various contexts ranging from supporting learners in developing new mathematical relationships and attaining deeper understanding of the known mathematical relationships describing various scientific phenomena.

Before blending the dimensions, we combined the two lowest categories of Zhao and Schuchardt Mathematics sensemaking dimension ("Math Procedure" and "Math Rule") into one Math sensemaking category, because they represent closely related types of Math sensemaking (student knowledge of Math procedures and Math rules respectively). We called the resulting category "Algorithmic", because it represents very basic Math sensemaking skills that only allow for the most rudimentary type of blended sensemaking. While this type of sensemaking is an important prerequisite to engaging in more complex blended sensemaking and would be relevant to students in lower grade levels, in the current study, we are interested in looking at more advanced types of blended Math-Sci sensemaking focusing on how it would happen in real-life settings when a mathematical explanation of the observed phenomenon is sought. The focus of the current study is validating the levels of the blended Math-Sci sensemaking framework that reflect the types of blended sensemaking occurring in a scenario when the mathematical relationship is not provided to the learner but is the end goal of the sensemaking activity. Since both "Math Procedure" and "Math Rule" would not be meaningful for this context, we will not be providing the validity evidence for these categories in the current study. However, we believe that it is important to include these categories into the theoretical framework presented in this paper, because they are important prerequisites for developing higher level blended sensemaking ability. These categories could be validated in the future studies.

Each blended category was developed by combining the four Math sensemaking dimension categories described by Zhao and Schuchardt ("Algorithmic", "Structure", "Relation" and "Concept") with three Science sensemaking dimension categories ("Description", "Pattern", "Mechanism"). We included their "Science Label" category as part of our "Description" blended category,

because both categories focus on naming and describing the variables or system properties.

The resulting framework has each of the four Math sensemaking categories ("Structure", "Relation", "Concept") subdivided into three Science sensemaking categories ("Description", "Pattern", "Mechanism"). The order of sophistication in the blended Math-Sci sensemaking framework followed that suggested by Zhao and Schuchardt. As shown in Fig. 1, the lowest Math sensemaking category termed "Algorithmic" was combined with each of the three Science sensemaking categories to yield the lowest level (level 0) of the blended framework. This level will be theoretically described but not validated in the current study due to the reasons described above. Similarly, the next lowest Math sensemaking category termed "Structure" was combined with each of the three Science sensemaking categories to yield three blended Math-Sci sensemaking categories shown as the lowest ("qualitative") sublevel of the blended framework. This is the lowest sublevel that will be validated in this study. Similar logic applied to blending each of the three Science sensemaking categories with the two higher level Math sensemaking categories including "Relation" and "Concept". The twelve resulting categories are divided into four broad levels (0-3) with respect to the Math sensemaking dimension which we label as "algorithmic", "qualitative", "quantitative", and "conceptual". These reflect different levels of proficiency in quantitatively describing phenomena. Levels 1-3 of the framework contain three sub-levels reflecting the Science sensemaking dimension, as shown in Table 1. The theoretical "algorithmic" level is labeled as level 0 of the framework in Table 1.

The resulting framework consists of new categories that are adapted from the categories proposed by Zhao and Schuchardt and follow similar ordering but reflect proficiency in *blended* Math-Sci sensemaking. The process of developing the blended categories focused on identifying the aspects of student thinking relevant for characterizing scientific phenomena mathematically at various levels of sophistication. The blended categories were developed following discussions with educational and subject matter experts. The resulting framework was reviewed by educational and subject-matter experts. The experts who reviewed the framework had both subject matter knowledge (chemistry, physics, biology) and educational expertise and all of them held PhDs in the respective fields. In addition, all the experts had over 5 years of teaching experience in the respective fields and had professional interest in pedagogy. As a result of the review no major changes were made to the description of the first two levels of the framework shown in Table 1. Furthermore, the clarifications were made on the description of the highest level of the framework, level 3. Specifically, the description of the lowest sublevel ("Description") was revised to emphasize student ability to notice both unobserved or not directly obvious variables and constants (the original iteration only focused on the variables). Furthermore, the description of the highest sublevel ("Mechanism") was revised to clarify what it means to provide causal explanation of the equation structure, that is how the equation (the variables and the mathematical operations among the variables) describe the causal mechanism of the scientific phenomenon.

Theoretical cognitive model for blended Math-Sci sensemaking

The detailed description of the blended categories is shown in Table 1. The lowest, "algorithmic" (level 0) category reflects recognizing mathematical relationships in a provided formula. The higher categories apply to contexts where students are developing the mathematical relationship from observations or recall the formula and recognize that it applies to describe a given phenomenon mathematically. This work concentrates on the latter context, so we did not examine the "algorithmic" level in our student validation interviews.

The lowest level that will be validated in the current study, "qualitative", reflects the ability to identify *qualitative* aspects that are important for characterizing the phenomenon mathematically. While students can engage in sensemaking of various aspects of the phenomenon in question, their sensemaking is limited to qualitative conclusions. The term "qualitative" in this context refers to providing descriptive accounts grounded in language (more vs. less, small vs. large) as opposed to "quantitative", which refers to providing numerical, measurable accounts grounded in numbers ("increase by 5 vs. decrease by 10"). As noted above, this level reflects blended Math-Sci sensemaking that serves as a stepping-stone to developing exact quantitative accounts of phenomena at the next level.

The intermediate level, "quantitative", reflects the ability to develop a quantitative description of the phenomenon and the sublevels mirror those at the qualitative level but reflect the ability to go beyond qualitative accounts. The blended sensemaking at this level builds on the previous, "qualitative" level. At the lowest ("Description") sublevel of the "quantitative" level students recognize numerical values of the relevant variables (as opposed to the lowest sublevel of the "qualitative" level, where they simply identify the relevant variables). Furthermore, at the intermediate ("Pattern") sublevel of the "quantitative" level students use the numerical values to identify specific patterns among the numerical values of the variables. Finally, at the highest ("Mechanism") sublevel of the "quantitative" level students translate these quantitative

Table 1 Theoretical blended Math-Sci sensemaking framework

0	Students recognize mathematical relationships in a provided formula	
1 Qualitative	Description	Students use observations to identify which measurable quantities (variables) contribute to the phenomenon *Example: force and mass make a difference in the speed of a car*
	Pattern	Students recognize patterns among the variables identified using observations and can explain *qualitatively* how the change in one variable affects other variables, and how these changes relate to the scientific phenomenon in question *Example: the smaller car speeds up more than the big car when the same force is exerted on both*
	Mechanism	Students demonstrate *qualitative* understanding of the underlying causal mechanism (cause-effect relationships) behind the phenomenon based on the observations but can't define the exact mathematical relationship *Example: it is easier to move lighter objects than heavy objects, so exerting the same force on a lighter car as on a heavy car will cause the lighter car to speed up faster*
2 Quantitative	Description	Students recognize that the variables identified using the observations provide measures of scientific characteristics and can explain *quantitatively* how the change in one variable affects other variables (but not recognizing the quantitative patterns yet), and how this change relates to the phenomenon. Students are not yet able to express the phenomenon as an equation *Example: recognizing that when variable A has value of X, variable B has value of Y*
	Pattern	Students *recognize quantitative patterns* among variables and explain *quantitatively* how the change in one parameter affects other parameters, and how these changes relate to the phenomenon in question. Students not yet able to relate the observed patterns to the operations in a mathematical equation and can't develop the exact and accurate mathematical relationship yet *Example: recognizing mathematical relationships such is direct linear and inverse linear among others*
	Mechanism	Students can explain *quantitatively* (express relationship as an equation)*(express relationship as an equation)* how the change in one variable affects other variables based on the quantitative patterns derived from observations. Students include the relevant variables that are not obvious or directly observable. Students are not yet able to explain conceptually why each variable should be in the equation beyond noting that the specific numerical values of variables and observed quantities match with this equation. Students cannot explain how the mathematical operations used in the equation relate to the phenomenon, and why a certain mathematical operation was used. Students can provide qualitative causal account for the phenomenon *Example: In $F_{net} = ma$, multiplication makes sense because when applied force on the mass of 50 kg increases from 10 to 20 N, acceleration increases by 2. That only makes sense for a multiplication operation*
3 Conceptual	Description	Students can describe the observed phenomenon in terms of an equation, and they can explain why all variables or constants (including unobservable or not directly obvious ones) should be included in the equation. Students are not yet able to explain how the mathematical operations used in the formula relate to the phenomenon *Example: In $F = ma$, the F is always less than applied force by a specific number, so there must be another variable subtracted from $F_{applied}$ to make the equation work. The variable involves the properties of the surface. So, the equation should be modified: $F_{applied}$-(variable)$= ma$*
	Pattern	Students can describe the observed phenomenon in terms of an equation, and they can explain why all variables or constants (including unobservable or not directly obvious ones) should be included in the equation. Students are not yet able to provide a causal explanation of the equation structure *Example: In $F_{net} = ma$, multiplication makes sense because as applied force on the same mass increases, acceleration increases linearly, which suggests multiplication*
	Mechanism	Students can describe the observed phenomenon in terms of an equation, and they can explain why all variables or constants (including unobservable or not directly obvious ones) should be included in the equation. Students can fully explain how the mathematical operations used in the equation relate to the phenomenon in questions and therefore provide causal explanation of the equation structure, that is how the equation (the variables and the mathematical operations among the variables) is describing the causal mechanism of the scientific phenomenon *Example: Since greater acceleration is caused by applying a larger net force to a given mass, this shows a positive linear relationship between a and F_{net}, which implies multiplication between m and a in the equation, or $F_{net} = ma$*

Examples provided in the table assume students are working towards developing a mathematical relationship describing the scientific phenomenon in question

patterns among the variables into the appropriate mathematical relationship that describes the scientific phenomenon in question. The justification they provide for the mathematical relationship is grounded in the numerical values of the variables as opposed to higher level conceptual justification. An example of this would be developing a mathematical relationship describing how acceleration depends on the applied force ($F_{net} = ma$) and using numerical values of each variable to support the choice of variables and the operation among the variables in the equation. Alternatively, a student might recall the equation and recognize that it applies in a given context. Irrespective of whether the mathematical relationship is developed or recalled, at this level students justify it using numerical values of the relevant variables.

Finally, the highest level, "conceptual", indicates a causal understanding of quantitative relationships describing the scientific phenomenon. At the lowest ("Description") sublevel students not only identify additional unobservable variables needed to characterize the phenomenon mathematically, but also justify why they are scientifically important to include in the mathematical relationship. Furthermore, at the intermediate ("Pattern") sublevel students define quantitative patterns among the relevant variables and directly translate those patterns into the mathematical operations in the relationship describing the phenomenon. Students justify the mathematical operations by directly connecting the observed quantitative patterns to the mathematical operations in the equation. Alternatively, students might recall the mathematical relationship and recognize that it applied for describing a given phenomenon mathematically. In this case students justify the applicability of the mathematical relationship by directly relating mathematical operations in the equation to the observed quantitative patterns. An example of this would be making a connection between linear patterns among the variables and the multiplication operation that would quantitatively describe the relationship between these variables (see sample response for this sublevel in Table 1). This is in contrast to the lower, "quantitative" level where students use numerical values of the relevant variables to justify the derived mathematical relationship describing the phenomenon. However, at this sublevel they are not yet providing causal explanation of the equation structure, that is how the equation (the variables and the mathematical operations among the variables) is describing the causal mechanism of the scientific phenomenon. This ability is indicative of the highest ("Mechanism") sublevel. Specifically, at the highest sublevel students recognize that the mathematical relationship provides a causal quantitative account of the phenomenon. An example of this would be not just recognizing that a linear relationship between the variables reflects multiplication operation in the corresponding equation but recognizing that the equation reflects the scientifically sound causal relationship, such as recognizing that applying a greater force to a given mass causes greater acceleration, which is consistent with the linear relationship between force and acceleration reflected in the mathematical relationship for Newton's Second Law. This contrasts with the lower, "quantitative" level where students justify the mathematical operations using the numerical values of the relevant variables, but provide causal account of the phenomenon at the qualitative level without relating the equation structure to the underlying scientific mechanism.

This hierarchy of levels makes sense in terms of the level of mathematical abstraction required for the different reasoning processes. In this study, we provide initial empirical evidence for the ranking based on the interviews presented below. We see that students routinely demonstrate blended sensemaking at the levels below their highest demonstrated level of proficiency, but they never demonstrate a high proficiency level while failing to achieve the criteria for lower proficiency levels. However, given the limitation of the sample size and the assessment scenarios, additional evidence is needed to confirm the level hierarchy proposed here.

Validating the theoretical cognitive model for blended Math-Sci sensemaking

Validating a cognitive model starts with developing a theoretical model reflecting what different levels of proficiency look like in a domain. This model is shown in Table 1. We empirically tested the model using assessment interview scenarios from three different subject domains (physics, chemistry and energy conversion). These interviews probed how well student thinking fit within the levels and sublevels of the framework shown in Table 1. The student responses were the data used to test the validity of the model. This method is following the Standards for Educational and Psychological Testing as appropriate for test validity evidence (American Educational Research Association [AERA], 2018) and has been previously used to validate cognition models such as learning progressions (Kaldaras, 2020; Kaldaras et al., 2021a, b).

Response process-based validity is obtained by evaluating the correspondence between responses to assessments measuring the construct for a population of students and the various cognitive model levels. If there is sufficient evidence of correspondence between the variation in student responses and the theoretical model levels, one can conclude that a given cognitive model exhibits response process-based validity (AERA, 2018).

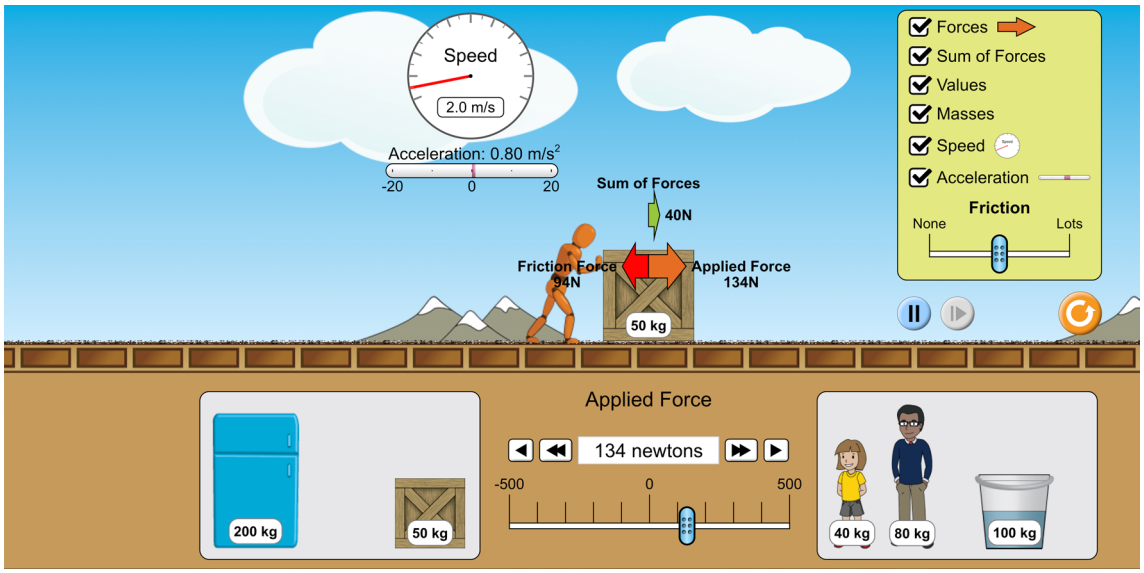

Fig. 2 Snapshot of the Acceleration simulation

We used a sample of students with a wide range of Math and Science proficiency to adequately test the model.

Methods

PhET simulations allow learners to explore how the behavior depends on different variables, both qualitatively and quantitatively, therefore providing a meaningful context for engaging in blended Math-Sci sensemaking. This study uses PhET simulations as an assessment context to test the cognitive model for blended Math-Science sensemaking shown in Table 1.The framework is designed to be widely applicable for guiding curriculum and assessment design.

To test the theoretical cognitive model for blended Math-Sci sensemaking, we first selected PhET simulations spanning three subject areas that would be suitable to provide assessment scenarios. Then we developed an interview protocol for each scenario that would probe each level and sublevel of the model. That was followed by interviewing undergraduate students spanning a range of majors and Math proficiency on all three scenarios, then coding their responses and comparing with the levels of the theoretical model. We discuss each step in more detail below.

Choosing disciplinary contexts and PhET simulations

Most studies on mathematical, scientific, and blended sensemaking have been conducted in the fields of physics, chemistry and biology (Zhao & Schuchardt, 2021). We initially planned to use those three disciplines as a context for the current study. The next step was to choose an appropriate phenomenon in each of those fields. The main criteria were: (1) the simplicity of the mathematical relationship describing the phenomenon; (2) the observational simplicity of the phenomenon; and (3) the wide applicability of the scientific idea underlying the phenomenon. As described below, we could not find suitable assessment scenarios for biology that met all the defined criteria, so we chose a different context for the third scenario.

In terms of mathematical simplicity, the criteria were phenomena that were described by a simple mathematical relationship (e.g., direct, or inverse linear multiplicative relationships). This allowed a substantial fraction, though far from all, of the interviewees (and presumably our target population) to express the mathematical relationship based on their interactions with the PhET simulation describing the phenomenon. We also chose phenomena described by a mathematical relationship with only one type of mathematical operation to reduce the mathematical complexity. Furthermore, we chose PhET simulations that model phenomena that most students are familiar with from everyday life and/or their coursework. Finally, we chose phenomena that are based on a widely applicable science ideas, so that the assessment might offer a useful learning opportunity to the student volunteers.

Following these criteria, we chose a PhET simulation modeling acceleration on an object as a function of applied force[1] (Newton's Second Law) for physics.

[1] https://phet.colorado.edu/sims/html/forces-and-motion-basics/latest/forces-and-motion-basics_en.html

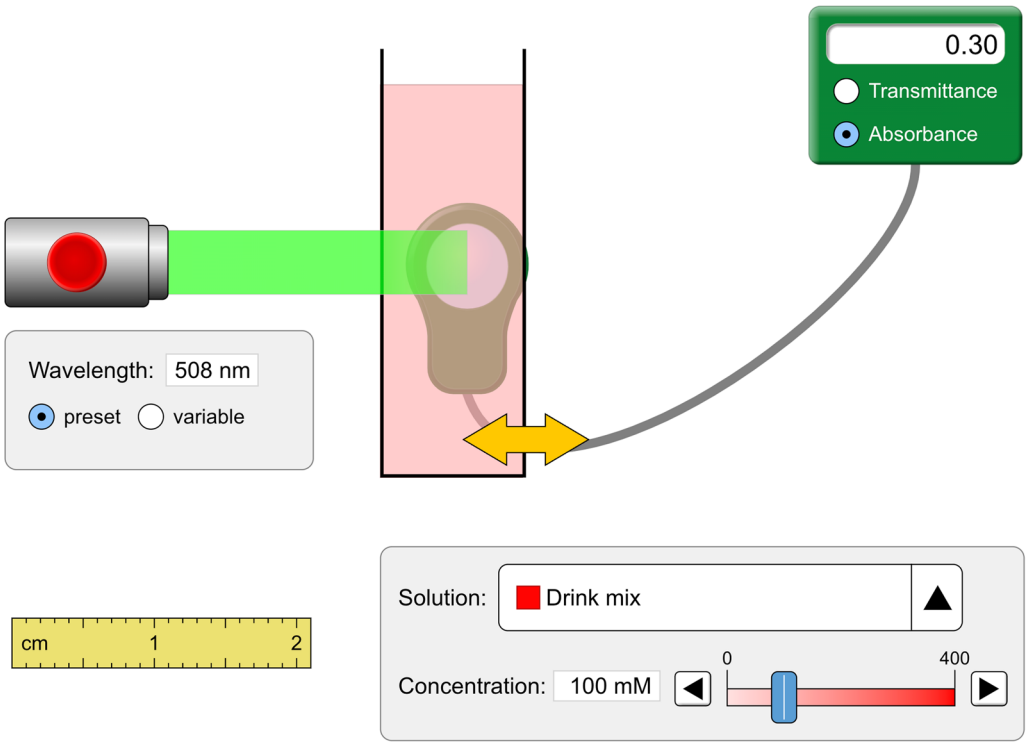

Fig. 3 Snapshot of Beer's Law simulation

The phenomenon is described by the formula $F_{\text{net}} = ma$, where "F_{net}" is a net force exerted on an object (calculated by subtracting applied force from the force of friction), "m" is mass of an object and "a" is the acceleration of the object. The formula involves a simple linear relationship and describes a familiar scenario.

Figure 2 shows a snapshot of the Acceleration simulation that students used during the interview. Students could use different objects to change the mass, they could apply force of different magnitude, change the magnitude of friction, and observe how the acceleration and speed change as a result of applied force on different masses. At the beginning of the interview all students were given time to interact with the simulation before responding to interview questions. Then, when they indicated they have explored the simulation enough to start the interview, they were asked to respond to the interview questions. Students were interacting with the simulation as they were responding to the interview questions. This was the case for all the three disciplinary scenarios.

For chemistry, we chose a PhET simulation modeling the relationship between concentration of a substance and the resulting absorbance at a given wavelength (Beer's Law[2]). The phenomenon is described by the formula A = c b e, where "A" is the absorption at a given wavelength, "c" is the concentration of a substance, "b" is the width of the substance's container and "e" is a molar absorption coefficient constant reflecting an internal property of a substance. Figure 3 shows a snapshot of the Beer's Law simulation that students were interacting with during the interview. Students could use different substances and change their concentration to investigate how absorbance and transmittance change as a result. They could also change the width of the container and the wavelength at which the signal was detected. Note that the molar absorption coefficient, which is a constant in the Beer's Law formula, is not a variable that students can change in the simulation. However, it can be noticed that at the same concentration and wavelength different substances absorb differently, which implies that there is some parameter unique to the substance that also affects the absorption.

We selected an interdisciplinary phenomenon for our third scenario, the conversion of energy and the efficiency of the conversion across different systems. The phenomenon is described by an efficiency formula which could be represented in one of the two ways: a) *Fraction of the*

[2] https://phet.colorado.edu/sims/html/beers-law-lab/latest/beers-law-lab_en.html

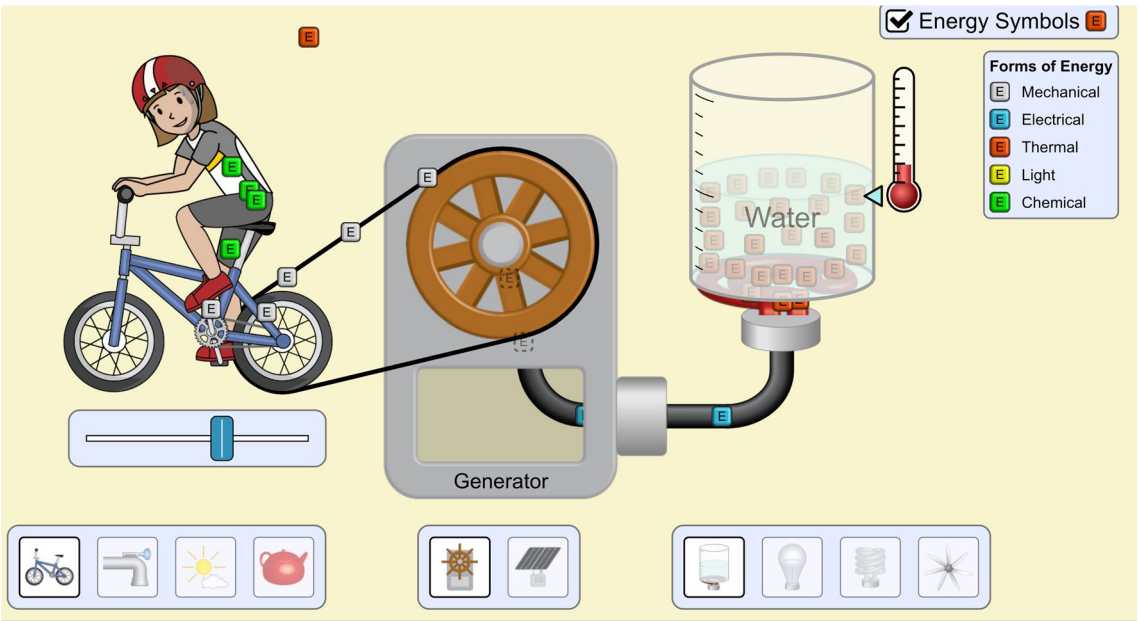

Fig. 4 Snapshot of Energy Conversion simulation

Energy Used = Useful Energy output/Energy input or b) *Useful Energy Output = Energy Input- Energy Lost to useless forms*. This phenomenon was chosen, because it represented simple mathematical relationships (inverse or subtraction) and involved a familiar and important idea that spans different disciplines. Finally, the PhET simulation showing energy conversion[3] allowed learners to investigate the energy efficiency of various systems. We label this context "energy conversion". Figure 4 shows a snapshot of the Energy Conversion simulation. Students could change components of the system (energy input, energy generator, energy output) and observe the amount of energy converted and lost at each step within a system. They could quantify the amount of energy by counting the number of energy boxes and their type (mechanical, electrical, thermal, light, chemical).

Developing interview protocol

We developed an interview protocol that was used for all three assessment scenarios (physics, chemistry, energy conversion). The interview protocol is provided in the Appendix (Additional file 1).

The interview questions focused on asking students to use the PhET simulation to explore the phenomenon and then characterize the behavior mathematically. Then, a set of questions probed the mastery of the lowest ("qualitative") level of the framework in Table 1 by asking students to identify the relevant variables, note qualitative patterns among the variables, and qualitatively explain causal relationships between the variables.

Next, student thinking at the intermediate ("quantitative") level was probed by asking students to determine the numerical values of the relevant variables and the quantitative patterns among the variables. This was followed by asking them to develop mathematical relationships (express a mathematical relationship or equation) among the variables and justify that quantitative relationship.

At this point, if students were struggling to provide a mathematical relationship based on their interaction with and observations of the simulation, they were provided with data that was collected from the simulation (or relevant to the simulation, as the case with energy conversion). This data was presented to them in a table which reflected how numerical values of the relevant variables change with respect to each other. For example, for the Acceleration simulation, students were provided with three data tables: one showing how acceleration changes as different forces are applied to the same mass; a second one showing how acceleration changes as the same force is applied to different masses; and a third table showing the acceleration for the combination of different objects with the same resulting mass (to demonstrate that it is the resulting mass that matters, and not the combination of objects). As students were studying the data provided to them, they were also allowed to go back

[3] https://phet.colorado.edu/sims/html/energy-forms-and-changes/latest/energy-forms-and-changes_en.html

Table 2 Participants' level of preparation

Student	Physics	Chemistry	SAT Math Score	Major
1	Algebra-based	HS Chemistry	Not available	Undecided
2	Algebra-based	General Chem. I	540	X-Ray Technician
3	Calculus-based	HS Chemistry	750	Aerospace Engineer
4	Calculus-based	General Chem. II	760	Integrated Physiology
5	Calculus-based	HS Chemistry	720	Mechanical Engineer
6	Calculus-based	General Chem. II	680	Mechanical Engineer
7	Calculus-based	AP/IB Chemistry	Not available	Mechanical Engineer
8	Algebra-based	HS Chemistry	740	Mechanical Engineer
9	Calculus-based	HS Chemistry	Not available	Mechanical Engineer
10	Algebra-based	HS Chemistry	Not taken	Civil Engineer
11	Calculus-based	Organic Chem. II	790	Biology
12	Honors physics	HS Chemistry	560	Undecided
13	Calculus-based	General Chem. I	760	Human Biology
14	Calculus-based	General Chem. I	730	Symbolic Systems
15	Calculus-based	HS Chemistry	760	Computer Science/Astronomy
16	Calculus-based	AP Chemistry	690	Environmental System Engineering
17	Calculus-based	HS Chemistry	670	Environmental systems engineering
18	Calculus-based	Honors Chem	660	Undecided
19	Calculus-based	Organic Chem. II	760	Symbolic systems/pre-med
20	Calculus-based	HS Chemistry	Not taken	Undecided/plan to do Physics
21	Algebra-based	HS Chemistry	710	Elementary Education
22	HS Physics	General Chem. I	490	Elementary Education
23	Algebra-based	HS Chemistry	560	Elementary Education
24	None	HS Chemistry	480	Elementary Education
25	None	HS Chemistry	580	Elementary Education

and forth between the data and the simulation to see if that helps them figure out the mathematical equation.

If they were still unable to give a suitable equation, they were presented with a list of possible equations and asked to use the simulation and the data to see if they can figure out which of these equations properly described the phenomena. The list of possible formulas was purposefully made very long with similar combinations of variables to minimize the possibility of guessing the correct formula simply based on which variables it contained. The data tables with lists of possible formulas for all the assessment scenarios are provided in the Appendix (Additional file 1).

Finally, those students who mastered levels 1 and 2, by giving a correct equation based on either the PhET simulation alone, or a combination of the data and the list of possible formulas provided to them were assessed as to their mastery of level 3, conceptual. The students were asked to justify the mathematical relationships among the variables in the equation, and explicitly relate the mathematical operations to the observations in the simulation. This probed whether students could accurately translate observation patterns to specific mathematical operations. In addition, students were also asked to provide a causal explanation for the equation structure that they proposed, focusing on probing whether students understand cause-effect relationships reflected in the equation.

If at any point during the interview students recalled an equation from memory (such as $F_{net} = ma$) and recognized that it applies in a given situation, they were asked to provide justification for why they thought the equation applied based on all the information they have. They were asked probing question consistent with each of the levels of the framework to help gauge the degree of sophistication of their justification as related to the framework. For example, for level 2 ("quantitative"), they were asked to use the quantitative patterns they have identified to justify the recalled mathematical relationship and its applicability in a given context.

Interviews

Interviews were conducted via zoom using standard zoom recording features. Each interview lasted between 40 and 60 min during which students were given time to interact with the simulation and answer interview questions provided in the Appendix (Additional file 1). Each

Table 3 Interview coding rubric for Beer's Law simulation

1	Description	Students identify concentration and width of the container as variables that affect absorbance and transmittance
	Pattern	Students identify that for specific wavelength the larger the concentration the larger the absorbance, and the smaller the transmittance
	Mechanism	Students recognize that the concentration of substance is the main causal factor behind the changing absorbance and transmittance but can't define the exact mathematical relationship for Beer's law
2	Description	Students quantitatively describe how the change in the concentration and the container width affect absorbance and transmittance but don't recognize quantitative patterns yet *Example: when I use concentration X for substance A, the absorbance changes to Y*
	Pattern	Students recognize that the relationship between concentration/container width and absorbance is positive linear, and between concentration/container width and transmittance is not linear (may say logarithmic or inverse). Students are not yet able to relate the observed patterns to the operations in a mathematical equation and can't develop exact mathematical relationship for Beer's law
	Mechanism	Students can explain *quantitatively* (express the relationship as an equation) for how the change in concentration and container width affects absorbance. The formula derived: $A =$ concentration • width of vial• molar absorption coefficient (MAC). Students can't fully explain why MAC should be included in the equation and can't justify multiplication operations beyond the fact that numerical values of the variables otherwise don't agree. Students recognize that the cause for changing absorbance is concentration of the substance *Note: MAC is an unobserved variable because it is not reflected in the PhET simulation and can only be inferred by noticing that absorbance at a given concentration and wavelength is different across various substances. MAC is provided to students in data tables*
3	Description	Students can express the relationship as an equation for absorbance ($A =$ concentration • width of vial • molar absorption coefficient (MAC)) and explain that MAC relates to specific properties of a given substance, and therefore should be included in the equation. Students can't explain why multiplication is their operation of choice beyond the fact that the numerical values of the variables otherwise don't agree
	Pattern	Students can develop the equation for absorbance and explain how the patterns among variables in the formula relate to observations. Specifically, students recognize that concentration and container width have a positive linear relationship to absorbance, which suggests multiplication operation. They also recognize that concentration and container width relate to absorbance through the factor of MAC, which also suggests multiplication operation. Students are not yet able to provide a causal explanation of the equation structure
	Mechanism	Students recognize that the cause for the change in absorbance is primarily the change in concentration (all other factors such as cuvette width and MAC being related to concentration) and can relate all the variables and operations in the equation to the observations of the phenomenon

student completed three interviews, each focusing on one of the three subject areas and PhET simulations, respectively.

Participants

Participants were first- and second-year undergraduate students recruited from a large public university and a private university in Western US. Participants were chosen from the list of volunteers to represent a sample of students with varying levels of Math and Science preparation. The participants were recruited by sending an email to the list of volunteers introducing the interview opportunity and asking volunteers to sign up. A total of 26 students were interviewed. One student was dropped from the interview analysis, because they did not finish all three interviews. The relevant information on the participants' level of preparation is shown in Table 2. Specifically, Table 2 shows the highest level of Physics and Chemistry taken by each student as well as their Math SAT score and major. All participants were compensated for each interview with $20 gift cards.

Interview analysis

Rubric development

Interview analysis was conducted using rubrics designed for each of the three PhET simulations. Each rubric was aligned to the framework shown in Table 1 and described in detail what student responses should contain at each sublevel for each specific simulation. An example of Beer's Law rubric is shown in Table 3. The Newton's Law and Energy Conversion rubrics are provided in the Appendix (Additional file 1). Note that all the rubrics are very specific and contain samples of student responses that would be consistent with a given sublevel of the framework for each disciplinary scenario. The rubrics were reviewed by the same disciplinary and educational experts that reviewed the framework shown in Table 1 to ensure alignment with the framework. No significant changes were made to the rubrics upon the review.

Interview coding

All interviews were analyzed by the first author following the respective rubrics. The coder listened to each interview and recorded the time stamps for the instances where they believed the interviewee demonstrated the blended Math-Science sensemaking consistent with specific sublevel of the framework. To ensure that three

interviews from the same students were coded randomly and the level assignments from the interviews of the same student were not known, the coder paid special attention not to code more than one interview from the same student within a short time frame.

Inter-rater reliability

Inter-rater reliability (IRR) was obtained by having a science education researcher unfamiliar with the project code two interviews in each discipline using the respective rubric. Before coding for establishing the IRR, the researcher was trained in using the framework shown in Table 1 and the rubrics for each disciplinary scenario. The IRR was established by having the coder listen to six interviews (two from each disciplinary scenario) and record the time stamps for the instances where they believed the interviewee demonstrated the blended Math-Science sensemaking consistent with specific sublevel of the framework. The time stamps were recorded and compared to those assigned by the first author during discussion. When discussing rating for each interview, the researchers compared final level assignment and evidence for all other sub-levels of the sensemaking framework detected in the interview. This approach ensured that the same information from student responses is taken as evidence for all level assignments.

For Beer's Law and Energy Conversion simulations, the IRR was 100% on the first try. That is, the timestamps assigned by both coders were the same for both interviews. For Acceleration simulation, the IRR was 100% for one of the interviews on the first try. For the other interview, one of the coders assigned level 2 "Mechanism" and the other coder assigned level 3 "Description" (difference 1 sub-level). Upon discussion both conders concluded that the interviewee did in fact demonstrate level 3 "Description" blended Math-Science sensemaking.

Results

Validity evidence supporting the existence of the categories of the theoretical framework

Below are sample responses for each sub-level of the framework for all three disciplinary scenarios. Tables 4, 5 and 6 show sample responses for the sub-levels of "Description", "Pattern" and "Mechanism" respectively at each of the broad levels ("qualitative", "quantitative" and "conceptual"). We were able to identify evidence of blended Math-Science sensemaking corresponding to every level and sublevel of the framework.

Table 4 shows sample responses for each of the three disciplinary scenarios for the "Description" sublevel at each of the broad levels of the framework. The "Description" sublevel at the lowest "qualitative" level reflects student ability to identify the observable variables relevant for characterizing the phenomenon mathematically. For *Acceleration* simulation, the variables are acceleration, mass, and force. For *Beer's Law* simulation, the variables are concentration, container width and absorbance. For *Energy Conversion* simulation, the variables are energy input and energy output. Student responses shown in Table 4 for this sublevel are examples of proficiency in blended Math-Science sensemaking focusing on student ability to identify the variables relevant to characterizing the phenomenon mathematically, but not provide any quantitative account of these variables yet. For example, for *Acceleration* simulation sample response states "In the sim you can **change the mass and the force and see how that affects acceleration**". The bolded phrase reflects blended Math-Science sensemaking. Math sensemaking is reflected in student recognizing that changing the values of mass and force changes acceleration. It's a qualitative evaluation at this level, so numerical values or quantitative patterns of these changes are not identified yet. Science sensemaking is reflected in recognizing the relevant scientific variables describing the phenomenon observed in the simulation: mass, force, and acceleration. The two types of sensemaking cannot be meaningfully separated, because one cannot occur without the other: as students interacts with the simulation and change the numerical values for mass and force, they notice that this also changes the numerical value for acceleration, which they otherwise cannot control. Therefore, mass and applied force as well as acceleration are relevant for characterizing the observed phenomenon mathematically. Similar assessment holds for the other two disciplinary scenarios and for all the sublevels.

The "Description" sublevel at the next level, "quantitative", reflects the student's ability to notice the values of the variables corresponding to a particular situation, but not noticing the patterns of behavior corresponding to changes in any of those variables. For *Acceleration* simulation, this involves noticing there are specific numerical values for acceleration, mass, and force. For example, for *Acceleration* simulation sample response states "**If you apply a force of 500 N and you have a mass of 50, then the net force is 500, you see that the acceleration is 10**". The bolded phrase reflects blended Math-Science sensemaking, because it contains elements of Math and Science sensemaking, but similarly to the above example they cannot be meaningfully separated. Specifically, in the sample response the student is recognizing the numerical values of the variables (Math sensemaking) that are relevant for providing quantitative account of the observed phenomenon (Science sensemaking). Similarly, for *Beer's Law* simulation, this means noticing the

Table 4 Sample responses for sub-level "Description" for all levels of the framework

Level	Acceleration	Beer's law	Energy conversion
1 Qualitative	"In the sim you can **change the mass and the force, and see how that affects acceleration**"	"I feel like the **variables that are really affecting absorbance are the concentration, the type of solution and the length**"	Student: **may be the starting amount of energy**. **Then more and more energy gets put in, and then more energy ends up at the end heating the water** Interviewer: any other energy? Student: **May be the final energy**? The energy that is released after the function has been done
2 Quantitative	"**If you apply a force of 500 N and you have a mass of 50, then the net force is 500 (***no friction***), you see that the acceleration is 10**"	"For absorbance, this one is a decimal number much smaller (*than transmittance*), the closer you are (*to the light source*), the smaller the number. **As you come up here** (*increase the distance to the light source*) **it comes from 0.16 to 1.3** (*absorbance*)"	"Some of the energy is going away as thermal energy in the very beginning, but what energy IS going into the system into the generator is the same amount that's coming out. **So, if three things of mechanical energy go in, three things of mechanical energy come out**"
3 Conceptual	"**The friction force was 84, in order to counter it, you would need, like 84 newtons of applied force**, and then the weight of this mass is 50 kg. **I solved on paper what the acceleration should be (***using F = ma***), and it should be 1.68 m/s^2, and that's what the sim is showing**"	Student: **I guess there is something between the wavelength and the solution type, some constant, like a certain variable that is specific to the solution that determines what wavelength gets through**. Going back to the data, you would have to divide the absorbance and the length by the molar absorption coefficient to get the concentration Interviewer: why do you need to divide? Student: **just by manipulating the numbers, if I divided it every time it would give me the correct number**	Student: **electrical energy equals 1/6 thermal plus 5/6 of the mechanical** Interviewer: why you derived it in that form? Student: because the input outlet has to be equal to the output Interviewer: if you generalize from this specific set-up to across set-ups, how would you change your equation? Student: **I think every energy will have a different way of showing stuff. The lightbulb would generate more heat, then, let's say, the fan. The ratio of the thermal energy to the ratio of the light energy would be a little different**

*Bold text indicates key evidence in student responses for the specific sub-level

**Italics text indicates clarification comments from the authors

Table 5 Sample responses for sub-level "Pattern" for all levels of the framework

Level	Acceleration	Beer's law	Energy conversion
1 Qualitative	"**As you apply more force and the mass stays the same, that changes acceleration and makes it faster**"	"Transmittance and absorbance have sort of an opposite relationship**, as absorbance goes up transmittance goes down with increasing concentration**"	"**A lot more energy is being put into it than coming out**"
2 Quantitative	"The more mass that you get or the greater acceleration that you are going at, the greater the force will be. So, the force is **directly** related to the mass and the acceleration"	"I think the transmittance is **non-linear** with the concentration, and the absorbance is **linear** with the concentration when the wavelength is held at like a standard wavelength and the volume doesn't change (*container width*)."**"If it is 1 cm (*the vial width*), the maximum absorbance goes up by 0.5 every 100 mM**"	"It's not a 100% mechanical when it goes out because there is going to be the heat generated. **Every six electrical energy units that goes in, there is going to be one unit that comes out as a thermal** (*from the sim*)"
3 Conceptual	"As the mass increases, it will require more force. **If you do like F = a/m** *(as opposed to F = ma)***, that means that as the mass increases, it will take less force to move the object**, which doesn't make sense, I think. As the mass increase, the object should be harder to move"	"As the concentration is higher, the absorbance rate gets higher as well, so that's proportional, on top. And then, opposite of the radius, as the radius gets…oh…. (*explores the sim*)…I was wrong in that. As the radius gets bigger, the absorbance will go up as well, **so it probably will be concentration times radius at a certain wavelength will equal to the rate of absorbance**"	Student: **energy output equals some sort of conversion rate, depending on what you are using, like a normal light bulb vs you are using like heating the water. The conversion rate represents the efficiency of the transfer of energy times, like, the input of the energy** Interviewer: and you said that it is efficiency times the energy output. Why did you decide that it should be multiplication? Student: **the efficiency was not a whole number, but a small number, like a percentage, so ranging from like a 1% to like a 100%. Multiplied just takes the maximum amount of energy that could be in an ideal system, and then shrinks that to like what it actually is**

*Bold text indicates key evidence in student responses for the specific sub-level

**Italics text indicates clarification comments from the authors

Table 6 Sample responses for sub-level "Mechanism" for all levels of the framework

Level	Acceleration	Beer's law	Energy conversion
1 Qualitative	"**Acceleration would stop when you stop pushing, and you could see speed decrease**"	Interviewer: What is the main cause for what you observe? Student: **the amount of particles dissolved and the types of substance**	"**I think everything has thermal in it, it is not all transferred into the kind of energy you actually want. There is going to be some thermal to it**"
2 Quantitative	Student: Let's see, in table A *(data table)* we did 250 divided by 50, we get five. **So, it's applied force minus friction over mass gives the acceleration** Interviewer: Why did you state it in this form? Student: **I guess because acceleration is the furthest right on the table** *(the data table)***, and usually in the equation what you try to figure out is on the right of the equal sign** Interviewer: what is the cause for all of these outcomes? Student: **I think the applied force is the most important variable, it's the only thing that is really changing**	Student: I was able to figure out that the **absorbance is the molar absorption coefficient times the concentration times the container width** Interviewer: and why did you choose to express it that way? Student: mostly because with a compound, the molar coefficient will be a constant, and **because the math worked out for the molar coefficient times the concentration time container width to equal the absorption** Interviewer: what do you think is the main causal factor behind changing absorbance and transmittance? Student: **I'd say the concentration of the solution is the main changer**	Student: I am assuming, the **LED, since it has more energy output as light, the incandescent light bulb is going to have more thermal energy than the LED** Interviewer: what is the percent efficiency now (*bottom right table*)? Student: It's now a 125% Interviewer: Does that supports your model? Student: yes, it still supports the model, it's just different Interviewer: why does it support the model? Student: **because this percentage, in this case it's like 125%, the model took into account that there is a constant. So, like, it's a constant because it is 125% for this one, and 125% for this one** (*going down the column*), **and so on and so forth. It took into account that there is a constant, it just the constant can change depending on the situation**
3 Conceptual	Student: **I think it's $F = ma$ because the more mass that you get or the greater acceleration that you are going at, the greater the force will be. So, the force is directly related to the mass and the acceleration** Interviewer: Do you think force is the outcome? Student: Actually no, force is not the outcome. **The cause mechanism is a force which is causing the acceleration to go in a negative or positive direction, increase or decrease in magnitude**	"M**aybe the absorbance has to do with multiplying by whatever the concentration is because multiplying by zero (when the concentration is zero) will give you zero absorbance**. Because no matter what all the other variables are, when concentration is zero absorbance is always zero. If it was addition or subtraction of the thickness of the container or the wavelength, it wouldn't matter what those are, if they get multiplied by concentration which is zero, you would still get zero. **If a concentration is zero, it's just water; adding anything to water makes it less clear; if anything is less clear, it makes it harder for the light to go through**"	"**In every equation there is going to be a different ratio of thermal energy in the right side of the equation.** Example- the light bulb (LED with solar panel). This one is very efficient. So, 1/8th of the total energy would be thermal, and the 7/8th that's produced will be light"

*Bold text indicates key evidence in student responses for the specific sub-level

**Italics text indicates clarification comments from the authors

specific numerical values for concentration and container width, which result on changes for the values of the absorbance. For *Energy Conversion* simulation, this involves noticing the numerical values for the variables of energy input and output. At this level, students can connect their qualitative observations with the specific numerical values of the variables from the simulation or data provided to them (like with *Beer's Law* and *Energy Conversion* examples), but they do not notice any quantitative relationships between the variables (compare to the "Pattern" sublevel of quantitative level shown in Table 5, where students can recognize specific numerical patterns from the data or the simulation).

Finally, the "Description" sublevel at the highest level, "conceptual" reflects the student's ability to both identify and justify the need for the variables that are not directly observed or those that are not directly obvious. Students at this level can justify their mathematical relationship by stating that it is supported by the patterns among the numerical values of the relevant variables.

For *Acceleration* simulation, this involves specifically recognizing that F_{net} is calculated by subtracting the applied force from friction force and justifying the mathematical relationship ($F_{net} = ma$) using numerical values of the variables to show that the formula works. For example, for *Acceleration* simulation sample response states "**The friction force was 84, to counter it, you would need, like 84 newtons of applied force**, and then, the weight of this mass is 50 kg. **I solved on paper what the acceleration should be (*using* $F_{net} = ma$), and it should be 1.68 m/s^2, and that's what the sim is showing**". The bolded phrase reflects blended Math-Science sensemaking, because it contains elements of Math and Science sensemaking, but they cannot be meaningfully separated, because one cannot occur without the other. Specifically, the student is explicitly connecting their observations of how variables in the simulation affect each other to generate the phenomenon e.g. acceleration (Science sensemaking) to the numerical values of the associated variables required to characterize the phenomenon mathematically (Math sensemaking). In addition, in this specific example the student recalled $F_{net} = ma$ from memory and recognized that it applied.

Blended sensemaking at the "conceptual" level with *Beer's Law* simulation involves recognizing that there is an additional variable that needs to be accounted for apart from concentration and container width to find the specific absorption properties of a substance. This variable is molar absorption coefficient (MAC), and it is not part of the PhET simulation on Beer's Law. Students can infer the information about this variable by noticing that different substances at a given wavelength and the same concentration absorb differently. Most students were provided MAC as part of the data on Beer's Law simulation (see Appendix (Additional file 1) for data tables provided to students), and they could use the information on MAC to help them develop the exact mathematical relationship. This is reflected is the sample quote, where the student first recognizes that there should be an additional variable, then confirms it using the provided data: "**I guess there is something between the wavelength and the solution type, some constant, like a certain variable that is specific to the solution that determines what wavelength gets through**. Going back to the data, you would have to divide the absorbance and the length by the molar absorption coefficient to get the concentration". The student is explicitly connecting their observations of the scientific phenomenon in the sim (noticing something specific to the solution that determines what wavelength gets through) to the numerical values of the associated variables. The student justifies the derived Math relationship using the numerical values of the variables (see the rest of the example).

Finally, *Energy Conversion* simulation blended sensemaking at the "conceptual" level involves recognizing the variable of "energy lost as thermal" in any system during the process of energy conversion. This thermal energy is not used for the purposes of the system (e.g., generating electricity). The PhET simulation shows thermal energy loss at every step of the process in the form of energy units leaving the system and not being used, but it is hard to notice this lost energy. This represents an unobserved variable for this phenomenon. The mathematical relationship can be derived either in the form of a) *Fraction of the Energy Used = Useful Energy output/Energy input* or b) *Useful Energy Output = Energy Input- Energy Lost to Useless Forms*. The mathematical relationship is justified by using the numerical values of the relevant variables to show that they make sense for the specific form of the mathematical relationship. In the example quote, the student is recognizing that the thermal energy should be part of the equation and derives the equation for specific case as "**Electrical energy equals 1/6 thermal plus 5/6 of the mechanical"**. This is an acceptable variation of the equation in part b where the useful energy is the mechanical energy and energy input is electrical energy in the student's example. The student derived it for a specific case observed in the simulation and recognized that the type and amount of useful energy will be different for different systems (see the rest of the example). Like in above examples, Math and Science sensemaking cannot be separated: the student uses observations of the scientific phenomenon (e.g.. energy transfer and conversion through the system) to quantify the amounts and types of energies at each stage of the process to develop the mathematical relationship using the simulation.

Table 5 shows sample responses for each of the three disciplinary scenarios for the "Pattern" sublevel at each of the broad levels of the framework. The "Pattern" sublevel at the lowest level, "qualitative", reflects student ability to identify the qualitative patterns among the observable variables relevant for characterizing the phenomenon. At this level students cannot translate the identified qualitative patterns into the exact mathematical relationship. As shown in sample student responses, Math and Science sensemaking cannot be separated: as students change the numerical values for various variables, they not only notice the changes in the numerical values of other variables (which is consistent with lower, "Description" sublevel) but also provide a qualitative evaluation of this change using words such as "more", "less", "higher", "lower" etc. as opposed to specific numerical and quantitative patterns, which happens at the next level. This qualitative evaluation (Math sensemaking) occurs as students are making sense of how the variables relate to each other and the phenomenon (Science sensemaking).

The next level, "quantitative", involves noticing the exact quantitative patterns among the observable variables. The quantitative patterns include recognizing direct and inverse relationships, or verbally describing a specific pattern (e.g., as variable A increases by X units, variable B increases by Y units). At this level, students cannot translate the identified quantitative patterns into the exact mathematical relationship. Math and Science sensemaking dimensions cannot occur separately: students identify quantitative patterns relevant for mathematical description of the phenomenon (Math sensemaking) as they are making sense of how the variables relate to each other and the overarching phenomenon (Science sensemaking).

Finally, at the highest level, "conceptual", students can translate the quantitative patterns they have noticed into the exact mathematical operations and develop a quantitative relationship for the phenomenon. They can fully explain the choice of the mathematical operations (and argue against choosing alternative mathematical operations using observations as evidence) and relate them to specific observations of the phenomenon. For example, sensemaking at the "conceptual" level for the *Acceleration* simulation would involve relating observations (larger mass requires more force to move) to the multiplication operation in the formula. For the *Beer's law* simulation, sensemaking would involve relating the observations (increasing container width and concentration leads to increased absorbance) to the multiplication operation in the formula for absorption. Finally, for the *Energy Conversion* simulation, sensemaking at the conceptual level involves relating observations (useful energy is always a fraction of energy input) to the division (or multiplication operation if the conversion rate is known) in the formula. Students at this level cannot provide a causal explanation for the equation structure (see Table 6 "conceptual" level sample responses for comparison). As can be seen from the examples, Math and Science sensemaking dimensions are blended: students relate qualitative or quantitative observations to the mathematical operations in the equation (Math sensemaking) that is used to quantitatively describe the scientific meaning of the phenomenon (Science sensemaking).

Table 6 shows sample responses for each of the three disciplinary scenarios for the "Mechanism" sublevel at each of the broad levels of the framework. The "Mechanism" sublevel at the lowest level, "qualitative", reflects student's ability to identify qualitative causal relationships among the relevant variables. For the *Acceleration* simulation, this involves recognizing that applied force causes acceleration. For the *Beer's Law* simulation, this involves recognizing that concentration is the main causal factor behind changing absorbance. For the *Energy Conversion* simulation, this involves recognizing that the reason all energy input cannot be converted into useful energy is because there is always thermal energy loss in the system. However, at this level students cannot develop the exact mathematical relationship describing the phenomenon. The Math and Science dimensions are blended, because this qualitative evaluation (Math sensemaking) occurs as students are working towards establishing causal relationship among the relevant variables that describe the phenomenon (Science sensemaking). For example, for *Acceleration* simulation stating that **"Acceleration would stop when you stop pushing"** reflects figuring out that acceleration is caused by applied force, because if there is no applied force, there is no acceleration, which is the essence of the scientific meaning of the phenomenon. Similar logic applies to all examples in Table 6 for this level.

The "Mechanism" sublevel at the next level, "quantitative", reflects the student's ability to develop an exact quantitative relationship, justify that relationship using numerical values of the relevant variables, and recognize the qualitative causal mechanism behind the phenomenon. At this level students cannot provide a causal explanation for the equation structure and cannot justify the choice of the mathematical operation by directly relating their choice to patterns in their observations. That is what distinguishes it from the highest level, "conceptual". For the *Acceleration* simulation, this involves developing the mathematical relationship for Newton's Law, justifying the relationship using data (either from the simulation, or the data provided to the students), and recognizing that applied force causes acceleration. For the *Beer's Law* simulation, this involves developing a mathematical relationship

Table 7 Sample responses from one student for Acceleration simulation for all framework levels

Level	Description	Pattern	Mechanism
1 Qualitative	I apply a certain force, and after I increase mass, that's kind of subtracting away from acceleration	With greater mass in order to keep the same acceleration I would need greater force	With greater force comes more acceleration
2 Quantitative	*Note: student didn't demonstrate this level but defined quantitative patterns (see next cell) right away*	Force and mass will be directly proportional to one another	*Note: Student didn't struggle to explain why division makes sense in the formula, was assigned level 3 mechanism (see the cell below)*
3 Conceptual	With a lot of friction, even with a lot of force we can see that the acceleration is not as great. We are at about 8 m/s^2, but as we increase friction, it's only 6 m/s^2	Acceleration is directly proportional to force, but mass is inversely proportional, so I can form $a = F/m$	Application of a force on an object creates acceleration, and that is a certain value, and can vary with friction. What causes acceleration is the interaction between how much force is applied onto the object because it contributes positively or negatively to the overall acceleration. Currently it is $a = F_{net}/m$, so maybe I can break down F_{net} into applied force and frictional force to yield $a = (F_{applied}\text{-}F_{net})/m$. Acceleration is just a rate of speed increases, and with higher friction acceleration will have a lower rate of increase

for Beer's Law, justifying the relationship using data (either from the sim, or the data provided to the students), and recognizing that change in concentration of the substance causes a change in the absorbance. Finally, for the *Energy Conversion* simulation, this involves developing the mathematical relationship for efficiency in the form of *Fraction of Energy Used = Useful Energy output/Energy input* and justifying the equation using the data provided. Students at this level can qualitatively recognize that there is always energy lost as thermal in the process of energy conversion, but they struggle to relate it explicitly to the equation. In the example shown in Table 6 the student recognized that the reason the LED lightbulb is more efficient than incandescent is because the incandescent lightbulb loses more thermal energy. However, when given wrong data that shows a system with over 100% efficiency (see the bottom right data table for *Energy Conversion* simulation in the Appendix (Additional file 1)) the student applied the formula derived earlier (*Efficiency = Useful Energy output/Energy input*) and stated that 125% efficiency is acceptable since it was a constant that holds across that system. This example demonstrates that while the student has qualitative understanding of the causal mechanism (recognizes that there is thermal energy lost from the system) and can derive the mathematical relationship for the phenomenon, they cannot relate the mathematical relationship meaning to the causal mechanism. This is the distinguishing feature between this level and the highest level, "conceptual". The Math and Science dimensions are blended, because the students develop a quantitative relationship (Math sensemaking) that reflects how the relevant variables describe the overarching scientific phenomenon (Science sensemaking).

Finally, the "Mechanism" sub-level at the highest level, "conceptual", reflects the ability to develop the exact quantitative relationship, justify the relationship by explicitly relating the choice of the mathematical operation to the observations, and provide causal explanation for the equation structure. For the *Acceleration* simulation, this involves developing the mathematical relationship for Newton's Law, justifying the choice of multiplication operation by directly relating observations (the force is directly related to mass and acceleration) to the choice of multiplication as the operation in the equation, and recognizing that the equation structure supports the premise that applied force causes acceleration. For the *Beer's Law* simulation this involves developing the mathematical relationship for Beer's Law, justifying the choice of multiplication operation by directly relating observations (the absorbance is directly related to concentration and container width) to the choice of multiplication as the operation in the equation, and recognizing that the equation structure supports the premise that concentration causes absorption. For *Energy Conversion*, this level would involve deriving the relationship for efficiency (either in the form of a) *Fraction of the Energy Used = Useful Energy output/Energy input* or b) *Useful Energy Output = Energy Input- Energy Lost to useless forms*) and relating the observations of energy lost to useless forms (such as thermal) to the formula derived across various systems (beyond specific cases).The Math and Science dimensions are blended, because students directly relate the equation structure to the observations of the phenomenon.

Validity evidence supporting the order of the categories in the theoretical framework

Below are sample responses from one student for each sublevel of the framework for Newton's Second Law disciplinary scenario (*Acceleration* simulation). Table 7 shows sample responses for the sublevels of "Description", "Pattern" and "Mechanism", respectively, at each of the broad levels ("qualitative", "quantitative" and "conceptual") for the same student. These sample responses illustrate that the same student engaged in blended Math-Sci sensemaking at all the sublevels of the framework. The only two levels the student did not demonstrate were level 2 ("qualitative") "Description" and level 2 "Mechanism". Level 2 "Description" was not demonstrated, because the student started defining the quantitative patterns right away (level 2 "Pattern") without reiterating the data from the simulation first as would be indicative of this sublevel (for example, "when I use mass X, acceleration changes to Y" type of sensemaking). However, it is reasonable to assume that this student could demonstrate level 2 "Description" type of sensemaking, and they simply skipped over it during the interview. This is because it is unlikely that the student would demonstrate level 2 ("quantitative") "Pattern" type of sensemaking (e.g."Force and mass will be directly proportional to one another" as shown in Table 7) if they were not able to reiterate the data from the simulation. Therefore, it is reasonable to assume that the student simply skipped over this step in the blended Math-Sci sensemaking process during the interview. Furthermore, level 2 "Mechanism" was not demonstrated, because the student justified the operations in the formula by directly relating to the observed quantitative patterns, which is consistent with the higher level 3 ("conceptual") "Mechanism" instead of using the numerical values, which is the indicator of the missing, level 2 ("quantitative") "Mechanism". This response pattern was common for students who were assigned the highest level of the framework: they did not always justify

Table 8 Final level assignment for each simulation

Student	Acceleration	Beer's law	Energy conversion	SAT Math
1	Level 3 Pattern	Level 3 Mechanism*	Level 3 Mechanism	Not available
2	*Level 1 Mechanism*	*Level 1 Mechanism**	*Level 2 Description*	540
3	**Level 3 Mechanism**	**Level 2 Pattern***	**Level 3 Mechanism**	750
4	***Level 2 Description***	***Level 2 Pattern***	***Level 2 Pattern***	760
5	Level 2 Pattern	Level 2 Pattern	Level 2 Pattern	720
6	***Level 3: Pattern***	***Level 3: Pattern***	***Level 3: Mechanism***	680
7	***Level 2: Mechanism***	***Level 2: Mechanism***	***Level 2: Pattern***	Not available
8	Level 2: Mechanism	Level 2: Mechanism	Level 2: Mechanism	740
9	**Level 3: Mechanism**	**Level 3: Mechanism**	**Level 2: Mechanism**	Not available
10	***Level 3: Pattern***	***Level 3: Mechanism***	***Level 3: Pattern***	Not taken
11	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	790
12	***Level 3: Description***	***Level 3: Description***	***Level 3: Mechanism***	560
13	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	760
14	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	730
15	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	760
16	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	690
17	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	670
18	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	660
19	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	760
20	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	Not taken
21	Level 3: Mechanism	Level 3: Mechanism	Level 3: Mechanism	710
22	Level 1: Mechanism	Level 1: Mechanism	Level 1: Mechanism	490
23	***Level 1: Pattern***	***Level 1: Mechanism***	***Level 1: Mechanism***	560
24	Level 1: Pattern	Level 1: Pattern	Level 1: Pattern	480
25	**Level 2: Mechanism**	**Level 1: Pattern**	**Level 2: Mechanism**	580

*Students were not provided molar absorption coefficient data

their formula using numerical values of the relevant variables obtained from the simulation or from the sample data but would directly relate the identified quantitative patterns to the mathematical operations in the formula, which is consistent with level 3 ("conceptual"). Therefore, for students who are engaging in level 3 ("conceptual") blended sensemaking level 2 ("quantitative") "Mechanism" sublevel would sometimes not be observed. However, it is also reasonable to assume that they could demonstrate this sublevel of sensemaking which was skipped during the interview. This is because, as consistent with the developmental approach reflected in the framework, justifying the mathematical relationship using numerical values of the relevant variables (level 2 "Mechanism") is cognitively easier than justifying the relationship by directly relating mathematical operations to the observations (level 3 "Mechanism"). Therefore, it is reasonable to assume that the student can easily justify the mathematical relationship using numerical values but chose to use a more cognitively demanding justification path consistent with level 3 "Mechanism", because they were able to engage in blended Math-Sci sensemaking at that level.

The student whose responses are shown in Table 7 demonstrated the same pattern of engaging in all the sublevels of the framework for the other two disciplinary scenarios. These data provide evidence that a student who is assigned the highest level of the framework can engage in blended Math-Sci sensemaking at all the levels below that highest level. Similarly, students who were assigned any other level of the framework were also able to demonstrate engagement in blended Math-Sci sensemaking at all levels below the assigned sublevel, but not above. This pattern was observed for all the three disciplinary scenarios. These findings indicate that the theoretical ordering of the framework categories shown in Table 1 is supported by the response patterns irrespective of the disciplinary context. These results provide preliminary response process-based validity evidence that the theoretical ordering of the categories in the framework proposed in this study is plausible (AERA, 2018; Kaldaras et al., 2021a). However, given the small interview sample

and few assessment scenarios, further work is needed to confirm the theoretical ordering of the categories.

Validity evidence supporting blended Math-Sci sensemaking as a distinct cognitive construct

Table 8 shows final level assignment for all interviewed students for all three disciplinary scenarios. The general trend (15 students out of 25) was that students were assigned the same level and sublevel across the three disciplinary contexts. The other ten students exhibited different degree of variability in level assignment across the three disciplinary scenarios. Specifically, six students out of 25 were assigned the same broad level ("qualitative", "quantitative", "conceptual") across all three disciplinary contexts (physics, chemistry, and energy conversion), but the sublevel assignment ("Description", "Pattern", "Mechanism") varied within that level. The level assignments for these students are shown in bold italic in Table 8.

Furthermore, the levels assignment for one of the students (student 2) differed by 1 sublevel only for one of the simulations. Specifically, student 2 was assigned level 1 "Mechanism" on *Acceleration* and *Beer's Law* simulations but scored one sublevel higher at level 2 "Description" on the *Energy Conversion* simulation. The level assignments for this student are shown in Italic in Table 8.

Only three students out of 25 were assigned different broad levels across the three disciplinary scenarios. The levels assignments for these students are shown in bold in Table 8. Each had a unique feature. Student 3 was assigned level 3 "Mechanism" on *Acceleration* and *Energy Conversion* simulation but scored four sublevels below at Level 2 "Pattern" on the *Beer's Law* simulation. Notably, this student was an early interview and unlike the subsequent 22 of the others was not provided with the molar absorption coefficient on *Beer's Law* simulation. We believe this affected that student's ability to bring together the quantitative observations made while interacting with the simulation to develop the exact mathematical relationship. However, notice that students 1 and 2 were also not provided the molar absorption coefficient. This, however, did not affect the ability to demonstrate level 3 "Mechanism" for student 1. Regarding student 2, it is likely that it was the student's low level of blended sensemaking proficiency that affected their ability to transition to higher levels of the framework as opposed to the lack of information on molar absorption coefficient. Student 9 was assigned level 3 "Mechanism" on the *Acceleration* and *Beer's Law* simulations but scored three sublevels below at on *Energy Conversion* simulation. This student expressed a strong incoming pre-conception about the energy conversion process, which interfered with their sensemaking in the energy context (the student was sure that an incandescent lightbulb uses thermal energy rather than electricity to produce light.) Finally, student 25 was assigned level 2 "Mechanism" on the *Acceleration* and *Energy Conversion* simulations but scored four sublevels lower at level 1 "Pattern" on the *Beer's Law* simulation. This student was unfamiliar with the subject matter and seemed more confused than any of the other students as to what the simulation was showing. That appeared to affect their sensemaking during the exploration of this simulation.

In general, the data indicate that students tend to be assigned the same level and sublevel of the framework irrespective of disciplinary context. Most fluctuations happen for within level assignment where students score in different sub-levels ("Description", "Pattern", "Mechanism") of the same broad level ("qualitative", "quantitative", "conceptual"). Finally, it is rare that students are assigned sublevels in different broad levels of the framework, but as noted, this was usually because of some unique difficulty with one of the contexts. In addition, the level assignments seem to be reasonably well related with SAT Math scores below 650 but does not distinguish well for scores above 650, as shown in Table 8.

Discussion

In this paper, we presented a theoretical framework for blended Math-Sci sensemaking grounded in prior research. The framework shown in Table 1 represents the first detailed categorization of the blended Math-Sci sensemaking process that has been validated by student response data. The levels of the framework were developed following the blending process of the selected theoretical categories for Math and Science sensemaking dimensions originally described by Zhao and Schuchardt (2021). The final theoretical framework for blended Math-Sci sensemaking shown in Table 1 was reviewed by educational and subject matter experts and represents a cognition model reflecting engaging in blended Math-Sci sensemaking process. The development of the framework helped answer the first RQ of our study: *How can one characterize the different ways of engaging in blended Math-Sci sensemaking?*

We gathered response process-based validity evidence for the theoretical framework by analyzing student responses from the interviews probing the levels of the theoretical framework shown in Table 1. The interviews were conducted in three disciplinary contexts, including physics, chemistry and energy conversion. The results of the interview analysis provided evidence for the existence of all the levels and sublevels of the framework shown in Table 1. Specifically, we were able to identify evidence for all the different types of blended Math-Sci sensemaking in student responses for each disciplinary context, as

illustrated in Tables 4–6. This finding provides evidence for the second RQ of out study: *To what degree does the validity evidence support the theoretical framework for blended Math-Sci sensemaking?*

Furthermore, we have observed that students who demonstrated reaching any level of the framework were able to demonstrate engagement in blended Math-Sci sensemaking at all levels below the demonstrated level, but not above (Table 7). This pattern was observed for all the three disciplinary scenarios. These findings suggest the theoretical ordering of the framework categories shown in Table 1, but further confirmation is needed.

This work has important implications for instruction on blended Math-Sci sensemaking skills. Specifically, we have demonstrated that blended sensemaking can be characterized by different increasingly sophisticated cognitive levels, which suggests that the framework shown in Table 1 can be used both as a diagnostic tool to accurately determine the level of individual students as related to blended Math-Sci sensemaking, and as a guide for what needs to be emphasized during instruction to help students attain higher blended sensemaking ability (NRC, 2001).

Furthermore, the level assignments seem to be reasonably well related with SAT Math scores below 650 but does not distinguish well for scores above 650, as shown in Table 8. This suggests that Math sensemaking is strongly integrated with the Science dimension in the context of blended Math-Sci sensemaking. Therefore, to effectively support learners in developing higher proficiency in blended Math-Sci sensemaking it is essential that the two dimensions are closely integrated during instruction and assessment. This strategy is similar to that put forth by the Framework for K-12 Science Education (NRC, 2012a, b) emphasizing the important of supporting students in developing the ability to integrate the three dimensions of science (disciplinary core ideas, scientific and engineering practices and crosscutting concepts) as opposed to supporting them in developing proficiency in each dimension separately. According to the Framework for K-12 Science Education it is the ability to demonstrate the integration of the three dimensions that is indicative of 3D learning, which reflects deep understanding of complex constructs. Similarly, we believe that it is the ability to demonstrate the integration of Math and Science sensemaking that is indicative of high proficiency in blended Math-Sci sensemaking. We believe that the parallel between 3D understanding and blended Math-Sci sensemaking is meaningful and accurate, because both represent examples of complex cognitive constructs (NRC, 2000) and therefore likely require similar strategies during the learning process.

Finally, the data analysis showed that the level of student sensemaking tends to be consistent across the various disciplinary contexts, as shown in Table 8. What little variation there is primarily occurs with sublevel assignment ("Description", "Pattern", "Mechanism") within a single broader level.

These findings suggest that blended Math-Sci sensemaking is a distinct cognitive construct irrespective of specific disciplinary context, which in turn has important implications for instruction. Specifically, it is likely that supporting students in developing Math-Sci sensemaking ability in one disciplinary context would help them apply such sensemaking in other subjects. This hypothesis should be further investigated.

The framework presented here will provide guidance for how to teach students to carry out blended Math-Sci sensemaking. Although it remains to be tested, it is likely that the levels of this framework serve as a learning progression for this type of sensemaking. Will students move efficiently from lower sensemaking levels to higher with appropriate learning experiences, and will they transfer this across different contexts? Exploring these questions will be the subject of future work. The capabilities of PhET simulations that facilitated this research will likely also be useful for teaching sensemaking.

Another area of future work is to extend this work to create an assessment instrument to easily and accurately diagnose individual student's level of blended Math-Sci sensemaking. We will specifically align individual assessment items to the sublevels of the framework to probe student blended sensemaking ability at each individual sublevel.

Finally, the lowest, "Algorithmic" level (level 0 in Table 1) of the framework should be validated in the appropriate context such as K-12 and across various scientific disciplines including chemistry, physics and biology.

The framework presented here reflects a natural path towards generating new knowledge by engaging in blended Math-Sci sensemaking with the purpose of characterizing an observed scientific phenomenon mathematically. Specifically, the levels of the framework shown in Table 1 represent an authentic progression of how scientists and engineers generate new knowledge in real-life settings. In particular, they start by observing the phenomenon of interest aiming at identifying relevant variables, qualitative and quantitative patterns of change among those variables, and ensuring that the resulting mathematical formula reflects a quantitative cause and effect relationship among the identified variables such that the formula explains the scientific phenomenon under study. These key steps are reflected in the levels of the framework shown in Table 1 and described in detail

at varying levels of cognitive sophistication. As such, the framework shown in Table 1 reflects the current vision for STEM education focused on supporting students in developing deep science understanding by helping them progress from novice to expert-like understanding (NRC, 2012a, b). Therefore, we believe that the framework will be widely applicable across STEM disciplines for guiding curriculum, instruction and assessment aimed at helping students build expertise in blended Math-Sci sensemaking.

Conclusion

In this work, we have developed and presented validity evidence for a cognitive framework for blended mathematical sensemaking in science. We hope that this framework can serve as a guide for curriculum, instruction, and assessment to help support students in developing higher proficiency in this important construct.

Abbreviations

RQ	Research question
Math-Sci sensemaking	Mathematical sensemaking in Science
MAC	Molar absorption coefficient
F_{net}	Net applied force
m	Mass
a	Acceleration
A	Absorbance
e	Molar absorption coefficient (as used in Beer's Law formula)
b	Width of the container (as used in Beer's Law formula)
c	Concentration of the substance (as used in Beer's Law formula)

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40594-023-00409-8.

Additional file 1. Appendix.

Acknowledgements

We would like to thank Dr. Jocelyn Nardo for helping conduct inter-rater reliability analysis and the PhET project for making this work possible by providing open-source simulations.

Author contributions

LK collected, analyzed, and interpreted the interview data. KW and LK developed and refined the framework for blended Math-Science sensemaking shown in Table 1. KW reviewed and supervised the interview data analysis. All authors read and approved the final manuscript.

Funding

This work is funded by the Yidan foundation.

Availability of data and materials

The data sets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors have no conflict of interest to disclose.

Received: 3 August 2022 Accepted: 17 February 2023
Published online: 07 March 2023

References

American Educational Research Association. (2018). *Standards for educational and psychological testing*. American Educational Research Association.

Bain, K., Rodriguez, J. M. G., Moon, A., & Towns, M. H. (2019). Mathematics in chemical kinetics: Which is the cart and which is the horse?. In *It's just math: Research on students' understanding of chemistry and mathematics* (pp. 25–46). American Chemical Society.

Bing, T. J., & Redish, E. F. (2007, January). The cognitive blending of mathematics and physics knowledge. In *AIP conference proceedings* (Vol. 883, No. 1, pp. 26–29). American Institute of Physics.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. (Eds.). (2007). *Taking Science to school: Learning and teaching Science in grades K-8*. National Academy Press.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*(2), 133–187.

Gifford, J. D., & Finkelstein, N. D. (2020). Categorical framework for mathematical sense making in physics. *Physical Review Physics Education Research, 16*(2), 020121.

Hunter, K. H., Rodriguez, J. M. G., & Becker, N. M. (2021). Making sense of sensemaking: Using the sensemaking epistemic game to investigate student discourse during a collaborative gas law activity. *Chemistry Education Research and Practice, 22*(2), 328–346.

Kaldaras, L. (2020). Developing and validating NGSS-aligned 3D learning progression for electrical interactions in the context of 9th grade physical science curriculum. Michigan State University.

Kaldaras, L., Akaeze, H., & Krajcik, J. (2021a). Developing and validating Next Generation Science Standards-aligned learning progression to track three-dimensional learning of electrical interactions in high school physical science. *Journal of Research in Science Teaching, 58*(4), 589–618.

Kaldaras, L., Akaeze, H., & Krajcik, J. (2021b). A methodology for determining and validating latent factor dimensionality of complex multi-factor science constructs measuring knowledge-in-use. *Educational Assessment, 26*(4), 241–263.

Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education, 97*(1), 32–57.

Lythcott, J. (1990). Problem solving and requisite knowledge of chemistry. *Journal of Chemical Education, 67*(3), 248.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (Expanded). National Academies Press.

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

National Research Council. (2012a). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Research Council. (2012b). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. National Academies Press.

Odden, T. O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education, 103*(1), 187–205.

Ralph, V. R., & Lewis, S. E. (2018). Chemistry topics posing incommensurate difficulty to students with low math aptitude scores. *Chemistry Education Research and Practice, 19*(3), 867–884.

Redish, E. F. (2017). Analyzing the competency of mathematical modelling in physics. In T. Greczyło & E. Dębowska (Eds.), *Key competences in physics teaching and learning, Springer Proceedings in Physics.* (Vol. 190). Springer.

Schuchardt, A. (2016). *Learning biology through connecting mathematics to scientific mechanisms: Student outcomes and teacher supports* (Doctoral dissertation, University of Pittsburgh).

Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education, 100*(2), 290–320.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). FOCUS ARTICLE: implications of research on children's learning for standards and

assessment: a proposed learning progression for matter and the atomic-molecular theory. *Measurement Interdisciplinary Research & Perspective, 4*(1–2), 1–98.
Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics Physics Education Research, 3*, 020101.
Zhao, F. F., Chau, L., & Schuchardt, A. (2021). Blended and more: Instructors organize sensemaking opportunities for mathematical equations in different ways when teaching the same scientific phenomenon. *International Journal of STEM Education, 8*, 1–18.
Zhao, F., & Schuchardt, A. (2021). Development of the Sci-math Sensemaking Framework: Categorizing sensemaking of mathematical equations in science. *International Journal of STEM Education, 8*(1), 1–18.

Publisher's Note