



# Outside the standard test: The development of a model-based assessment and corresponding rubric

Lana M. Minshev<sup>2</sup> · Janice L. Anderson<sup>1</sup> · Kerry A. Bartlett<sup>3</sup>

Received: 31 October 2019 / Accepted: 13 February 2022 / Published online: 23 March 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Models and modeling are central to both scientific literacy and practices as demonstrated by the Next Generation Science Standards. Through a design-based research framework, we developed a model-based assessment (MBA) and associated rubric as tools for teachers to understand and support students in their conceptualization of the flow of energy and matter within ecosystems. The MBA was piloted with four middle school students ( $n=4$ ) and implemented in two sixth grade student cohorts ( $n=89$  &  $n=98$ ). The MBA and rubrics went through several design iterations in order to best capture student understanding of complex systems. The design of the MBA allows students to express conceptual understanding while also capturing the transformation of their understanding as they are exposed to new information and experiences within the curricular content.

**Keywords** Middle school science · Model-based assessment · Rubrics · Design-based research

## Introduction

Modeling plays an important role in both scientific literacy and scientific practices as demonstrated by the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013). According to NGSS, students begin using models early in school and by middle school, students have progressed to the construction, use, and revision of conceptual models that enable students to communicate their understanding of disciplinary content as well as make predictions about abstract scientific phenomenon (van Joolingen et al., 2019). Most often, models are used to provide real world representations (Aduriz-Bravo, 2019). However, they also can be abstractions that allow scientists and learners to test theoretical models, and

---

✉ Janice L. Anderson  
anderjl@email.unc.edu

<sup>1</sup> School of Education, University of North Carolina at Chapel Hill, Peabody Hall 307A, CB#3500, Chapel Hill, NC 27599-3500, USA

<sup>2</sup> Human-Centered Design lab, Robert and Patricia Kern Institute for the Transformation of Medical Education, School of Pharmacy, Medical College of Wisconsin, Milwaukee, WI, USA

<sup>3</sup> School of Education, University of North Carolina at Chapel Hill, Peabody Hall 309, CB#3500, Chapel Hill, NC 27599-3500, USA

from these, gain a better understanding of complex systems (Constantinou et al., 2019). Further, models can be described as mental or expressed (Baek et al., 2011). Expressed models can include schematic representations, physical representations, mathematical representations, analogies or computer simulations. It is through students' expressed models that teachers and researchers can begin to understand students' mental models (Baek et al., 2011).

Ecosystems are thought to be complex systems and therefore present a cognitive challenge for K-12 students (Hmelo-Silver & Azevedo, 2006). Ecosystems require students to understand multiple inter-related systems and use knowledge from both the physical and life sciences (Bell-Basca et al., 2000; Hogan & Fisherkeller, 1996; Honward et al., 2010). While research studies have attempted to gain insight into student understanding of ecosystems, they have found it to be a difficult concept for students to master due to the complexity of the systems (Hogan, 2000; Hogan & Fisherkeller, 1996; Lin & Hu, 2003). The study of student understanding of ecological processes in ecosystems has typically focused on students acquiring conceptual structures about elements of an ecosystem and learning about linear cause-effect interactions such as food chains and predator/prey relationships (Hayes et al., 2017; Hogan, 2000; Manz, 2012). Despite the complexity of ecosystems' inter-related parts, studies have found that students need to be exposed to all aspects of ecosystems simultaneously (Barak et al., 1999). Ultimately, it is necessary to stress the inter-relatedness among biological systems rather than teach concepts in isolation which hinders student's ability to construct a complete picture of biological processes (Barak et al., 1999; Grotzer & Bell-Basca, 2003). The Next Generation Science Standards (NGSS) suggest that students should be able to develop a model to describe the cycling of matter and flow of energy in an ecosystem (MS-LS2-3; NGSS Lead States, 2013). This puts an emphasis on student understanding of multiple biological systems within an ecosystem rather than studying them in isolation.

Students typically have a reasonable understanding of simple aspects of ecological processes, such as energy moves from one organism to another (e.g., Barak et al., 1997, 1999; Opitz et al., 2015; Svandova, 2014). However, they struggle to explain where the energy goes once it reaches the end of a system like a food chain (Opitz et al., 2015; Svandova, 2014). Additionally, more intricate aspects, such as inter-relationships between processes are difficult for students to express (Grotzer & Basca, 2003). Other studies have found that students generally focus on visible aspects of ecosystems (i.e., plants, animals, and the sun) (Minshew et al., 2017, 2018; Hmelo-Silver et al., 2007, 2017; Zangori & Forbes, 2015). In addition, the visible components are represented as linear relationships, often depicted as separate pieces, and are never seen as part of a whole (Hmelo-Silver et al., 2007; Lin & Hu, 2003). However, some students do include non-visible components (i.e., oxygen, nitrogen, and bacteria) of an ecosystem on their post- maps and models; which demonstrates students are able to consider non-visible components and connect them to the visible components of an ecosystem (Minshew et al., 2017; Honward et al., 2010; Zangori & Forbes, 2015). By having students construct models it allows them to make their ideas visible, which enables them to begin to make meaning out of larger more complex systems (Hmelo-Silver et al., 2017).

Further, it is often assumed that students, particularly primary students, hold naïve explanatory models of ecological causal connections based upon evidence they obtain through observations and experiences (Lehrer & Schauble, 2012; Zangori et al., 2020). This is referred to as 'every day' reasoning and is personal because students situate themselves, their beliefs structures, and their emotional perspectives on their experiences (Zangori et al., 2020). Zangori et al. (2020) found that when students use their own models

to understand complex ecological relationships, their reasoning shifts beyond these naïve biological theories to begin to consider how and why phenomena occur. Students are able to increase the complexity and sophistication of their understanding and representation of ecological relationships (Lehrer & Schauble, 2012; Manz, 2012; Zangori & Forbes, 2015). For instance, students move beyond linear relationships to consider mutual causality cycles.

The Zangori et al (2020) study suggests that there may be an intermediate level of reasoning, called relational, which is situated between personal and causal. At the relational level, students are able to identify relationships between ecosystem components but not assign causality, only noting that a relationship existed (Zangori et al., 2020). This intermediate level may be the necessary stepping stone from personal to causal reasoning that students use when they infer a relationship, but do not yet have knowledge of the nature of the relationship.

Our work builds off of previous ecological modeling research (e.g., Hmelo-Silver et al., 2017; Honward et al., 2010; Zangori & Forbes, 2015) by examining students' model-based reasoning of ecological processes, such as the flow of energy and the cycling of matter. The developed model-based assessment (MBA) and associated three-tiered rubric are tools for teachers to understand and support students in conceptualizing and expressing their knowledge of how energy and matter flow in ecosystems. Our assessment was based upon the following research question: *Does a three-tiered rubric for a model-based assessment (MBA) capture student understanding of complex systems?* In the sections that follow, we will describe our iterative process of developing the MBA and corresponding rubric.

## Theoretical framework

### Models and model-based reasoning

One of the dimensions of the Next Generation Science Standards (NGSS) is student engagement in science and engineering practices (NGSS Lead States, 2013). Science and engineering practices are conceptualized as the disciplinary norms of science that lead students towards their construction of knowledge around a core concept, their evaluation of these ideas, and the communication of their findings to a broader community (e.g. their fellow students and teachers; Dushl et al., 2007). The development of models and model-based reasoning is one component of science and engineering practices that allows for the elaboration of student ideas and explanations that can potentially lead to more complex understandings of a disciplinary core concept (Aduriz-Bravo, 2019; Constantinou et al., 2019; Crawford & Cullin, 2004; Hokayem & Schwarz, 2014; Lehrer & Schauble, 2007).

Models and Model-based reasoning have a range of types and uses within the domain of science. These models can be represented by physical representations, illustrations or model diagrams (Krajcik & Merritt, 2012) and can range from traditional static models such as anatomical structures, solar system models or animal cell models to interactive computer simulations that demonstrate the relationship between the Earth, Moon and Sun or concepts such as how ion channels work within the animal cell membrane. In addition, models can also be utilized as both quantitative and qualitative measures. For example, as a quantitative measure, the Hardy–Weinberg Equilibrium Principle can be used as a model to describe genetic frequencies of populations over a period of time; while qualitatively, students can describe the phenotypic differences observed within the same populations. By creating opportunities for students to engage with models and model-based reasoning,

students learn to construct, compare, revise, evaluate and validate the models that they have created, a skill that has been identified as an important scientific practice that contributes to student learning (Lead States, 2013; Nicolaou & Constantinou, 2014; Schwarz et al., 2009).

Scientific models also function as epistemological components that are used to represent specific phenomenon (Giere, 1999, 2004; Hughes, 1997; Passmore & Stewart, 2002). They are the intermediate structures that exist between students' abilities to describe a scientific phenomenon and how it operates (Berland & Reiser, 2009; Braaten & Windschitl, 2011); models also function as mechanisms of interpretation of identified phenomenon (National Research Council, 2012; Nicolaou & Constantinou, 2014; Schwarz et al., 2009). Building on this perspective, Schwarz and colleagues (2009) argue that students need to construct their own models rather than work from models created for them by their teachers, textbooks or others with scientific expertise. This enables students to demonstrate their own understanding of the phenomenon. Schwarz and associates (2009) also demonstrated that when given appropriate scaffolding, students are able to develop models that exhibit a sophistication around a phenomenon and utilize the model to make predictions about other events.

One of the difficulties that arise in the use of models is the need to engage with the phenomenon from a systems-thinking perspective. This includes the consideration of boundaries of the system, components of the system, interactions of the system with other complex systems and emergent properties that result from behaviors of the system (Harrison & Treagust, 2000; Hmelo-Silver & Azevedo, 2006; Passmore et al., 2014). The models that are developed by the students become the tools that allow them to understand the complexities of the system(s) and the changes that may arise (Yoon et al., 2015). This type of systems-thinking requires understanding the causal links that exist within the system and the dynamic nature that influences how students are able to predict outcomes within these complex models (Hmelo-Silver et al., 2007; Orgill et al., 2019). Within this study, we define scientific models as a set of representations, rules and reasoning structures that allow students to develop explanations about the cycling of energy and matter within an ecosystem; teachers then evaluate the sophistication of student understanding of this complex scientific phenomenon.

In order to capture student conceptual understanding of the flow of energy and the cycling of matter in an ecosystem, a MBA was developed. Using the MBA is consistent with previous research (e.g. Hogan, 2000; Lin & Hu, 2003; Zangori & Forbes, 2015), that utilizes model building to demonstrate student conceptual understanding concerning the flow of energy and the cycling of matter in ecosystems. These studies were able to effectively capture patterns and trends expressed at pre- and post-time points by having students construct models to demonstrate their knowledge. Student constructed models is one way to capture student understanding of complex systems which have multiple levels of organization (Hmelo-Silver & Azevedo, 2006).

### **Knowledge in pieces—resources**

Student learning from the MBA was approached from a constructivist perspective which suggests that students construct or build their understanding of phenomena by engaging with their environment both socially and alone. Additionally, the constructivist approach describes learning as occurring in phases and recognizes that expert understanding does not develop rapidly but is built up over time (Smith et al., 1994). Constructivism also emphasizes the

important role of prior knowledge as an important facet of learning (Smith et al., 1994). From this perspective individuals develop their knowledge in fragmented pieces that some call phenomenological primitives (diSessa, 1988, 2002), resources (Hammer et al., 2005), or simply ideas (Clark & Linn, 2003). For the purposes of this paper, we utilize the term “Resources,” in alignment with Hammer and colleagues’ (Hammer & Elby, 2003; Hammer et al., 2005) conception, to denote these pieces of knowledge. It is through continued exposure and experience that individuals are able to piece together a more complete understanding of phenomena through their constructed model.

The importance of a Resources perspective (Hammer & Elby, 2003; Hammer et al., 2005) is that it recognizes that students have prior knowledge of scientific concepts, albeit naive knowledge. A Resources approach to student learning was conceptualized as a means to explain how students acquire new knowledge and reconcile it with their existing understanding. It suggests that novices start to develop an expert understanding through knowledge reorganization. Students’ knowledge fragments are a resource for learning and are not extinguished when students learn scientific concepts (diSessa, 2002). The development of resources in this fashion is described as the “reweaving” (diSessa, 2014, p.98) of knowledge. In this way, the activated resources are interconnected with each other and new information in various ways and student conceptual understanding is transformed. A resources approach to student conceptual understanding asserts, “even a fully-compiled conception is assumed to be built from finer-grained knowledge elements that have become tightly linked” (Hammer et al., 2005, p. 7). Resource activation is context dependent, meaning that certain scenarios may activate different resources, sometimes accurately and sometimes not (Hammer & Elby, 2003; Hammer et al., 2005). A variety of experiences regarding a single concept enable students to activate and connect their various resources together, thus refining their understanding. One way that this can be achieved is through Model-based Learning and assessments.

## Methodological frameworks

### Design based research (DBR)

The methodological approach for this study was Design-Based Research (DBR), an iterative process of data collection and analysis that simultaneously informs the design of educational innovation and develops theory (Cobb et al., 2003; Collins et al., 2004; Easterday et al., 2014; McKenney & Reeves, 2018). DBR as a research approach offers flexibility of implementation and design to fully understand the context in which the study is conducted (Cobb et al., 2003). The Easterday and colleagues (2014) model (Fig. 1) that was utilized in this study has six distinct and iterative phases: Focus, Understand, Define, Conceive, Build, and Test. While each phase produces important research and is used iteratively to refine and advance both educational innovations and theory, our study reports two successive test phases in which researchers, along with a participating classroom teacher, implemented and learned about the developed MBA.

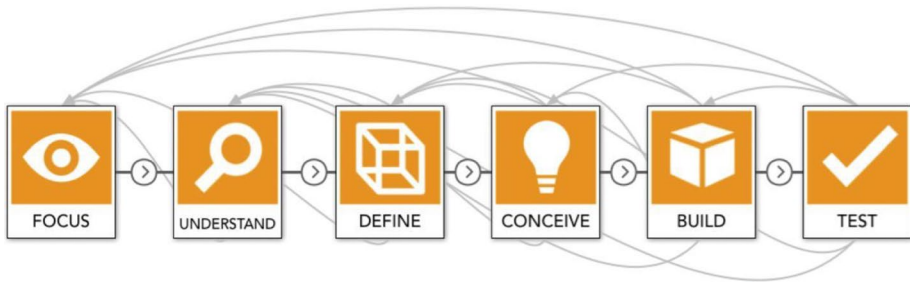


Fig. 1 Easterday model

## Contexts, methods, and data analysis

### Context

This study was conducted over three iterations of the design cycle—a pilot and two implementations in a grade 6 classroom. To capture students' changing understanding of the flow of energy, the cycling of matter, and the process of decomposition a unique assessment was needed. Following the Resources perspective, MBA was created to allow students to construct and represent their understanding in an open and flexible manner. This assessment was based upon the consensus model (Fig. 2) developed for the curriculum.

### Curriculum context

This study utilized the (Bio-Sphere Compost) unit within a life sciences curriculum entitled (Bio-Sphere Compost), created by researchers at the (Bio-Sphere Compost) and modified by the research team at the (Bio-Sphere Compost). The (Bio-Sphere Compost) curriculum is an eight weeklong, project-based, technology supported science curriculum for middle school students. Students build, collect data on, and modify a compost bio-reactor in order to develop compost that decomposes quickly. Students utilize computer simulations to run tests on virtual compost piles and collect secondary inquiry information via an online reference tool. The curriculum consists of hands-on activities that provide students with experiences surrounding decomposition, waste, ecosystems, the cycling of matter, and the flow of energy.

In order to develop an understanding of the concepts and relationships relevant to energy and matter that were central to the unit, a consensus model was developed by the research team. The consensus model was developed in a three-phase process that work to validate the model. First, the research team, including two science experts, developed a list of concepts that were central to the areas of the flow of energy and cycling of matter that would be expected to fall within the scope of the unit of study. The identified concept areas were also compared to the NGSS Middle School Disciplinary Core Ideas. Second, a member of the team drafted an initial model that was presented to all team members for review and feedback. In the final step, changes were made to the model based upon the team feedback (See Fig. 2). The consensus model was then used to develop the initial MBA and the

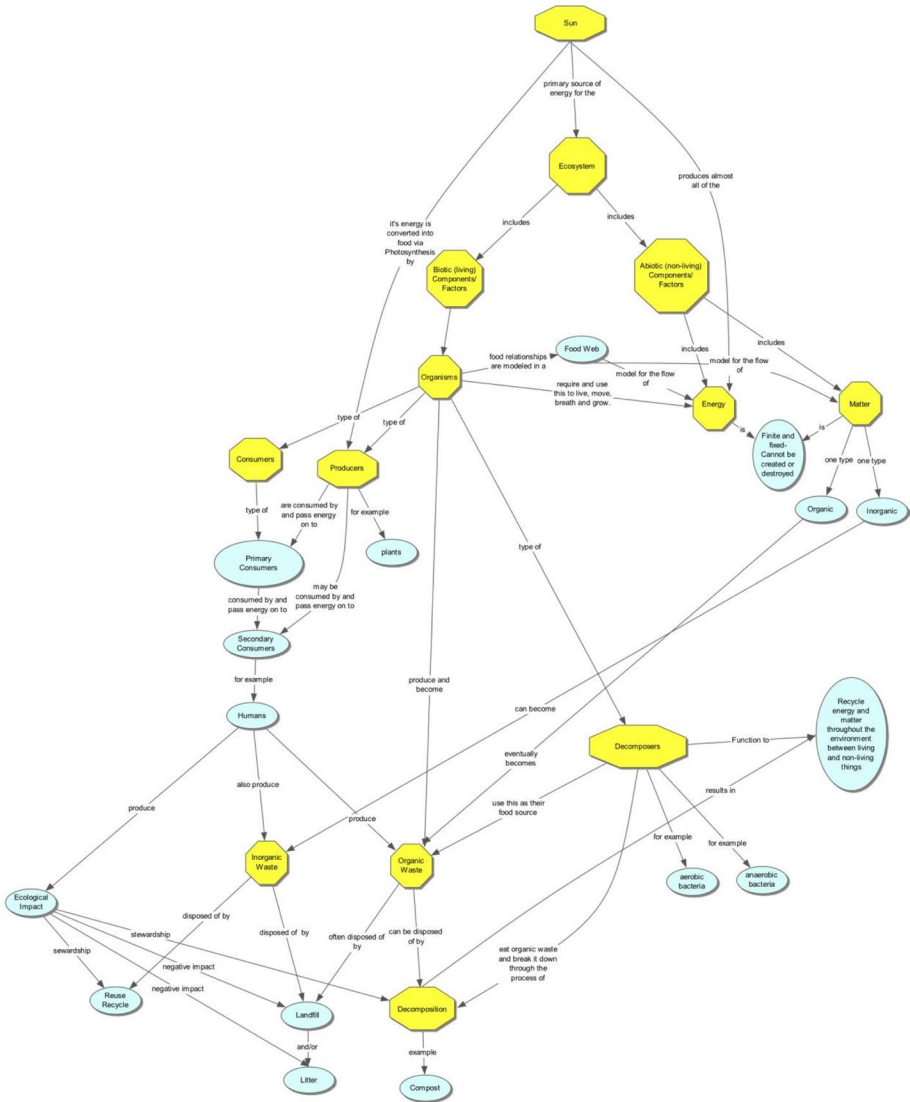


Fig. 2 Consensus model

eleven core ideas in an effort to elicit evidence of student understanding of key concepts and relationships that are relevant to energy and matter.

### Study context

The pilot study and both iterations (Iteration 1 and Iteration 2) were conducted in a small, rural, middle school in the southeastern United States. The middle school had a science, technology, engineering and mathematics (STEM) focus and was implementing a 1:1 laptop initiative. The school was identified as a Title I school, with nearly 50% of the student



**Table 1** Student demographics

White	African American	Hispanic	Asian	Bi-racial	Native American	Free or reduced lunch
33.2%	43%	15.5%	1.9%	6.0%	0.4%	98.7%

population receiving free or reduced lunch prices. The two student cohorts ( $n=89$  &  $n=98$ ) were from demographically diverse backgrounds (see Table 1). All students had the same science teacher and experienced the (name Bio-Sphere Compost) curriculum during the regular science course. Only student data from those who consented to participate in the study were used in the analysis. All analyses reported were conducted using the combined data set that included student MBA scores from Iterations 1 and 2. Additionally, students with incomplete data were removed from the combined data set ( $n=177$ ).

### Study limitations

We recognize the limitations of this study, with its small population and rural context. In addition, the MBA and corresponding rubrics need to be tested and validated in other contexts and with a broader population. However, the iterative design and development of the MBA and three-tier rubric, as described below, suggest the strength of the assessment.

## Methods

In the section that follows, we will describe the methods for each of the three phases of the study: Pilot Phase, Iteration 1 and Iteration 2. Figure 3 below provides an overview of the process of each iteration of the development of the MBA and the corresponding three-tier rubric.

### Pilot phase

The MBA was initially designed with a general prompt that instructed students to create their own model of how they understand and interpret the concepts through the world around them. The directions prompted students to start by thinking about where energy comes from and where it goes in an environment. Students were not provided with any direction on how to construct a model. The initial MBA was piloted with a small number (four total, two girls and two boys) of grade seven students from our participating middle school. Students did not experience the (name Bio-Sphere Compost) curriculum; however, these students had previously participated in a after school program with the researchers. Selected students completed the MBA in order to help refine it prior to in-class implementation with our grade six cohorts. Students worked individually and were given as much time as needed to complete the MBA. The instructions were read aloud to each student prior to beginning the MBA, and students were allowed to ask clarifying questions throughout. After a student completed their MBA a member of the research team



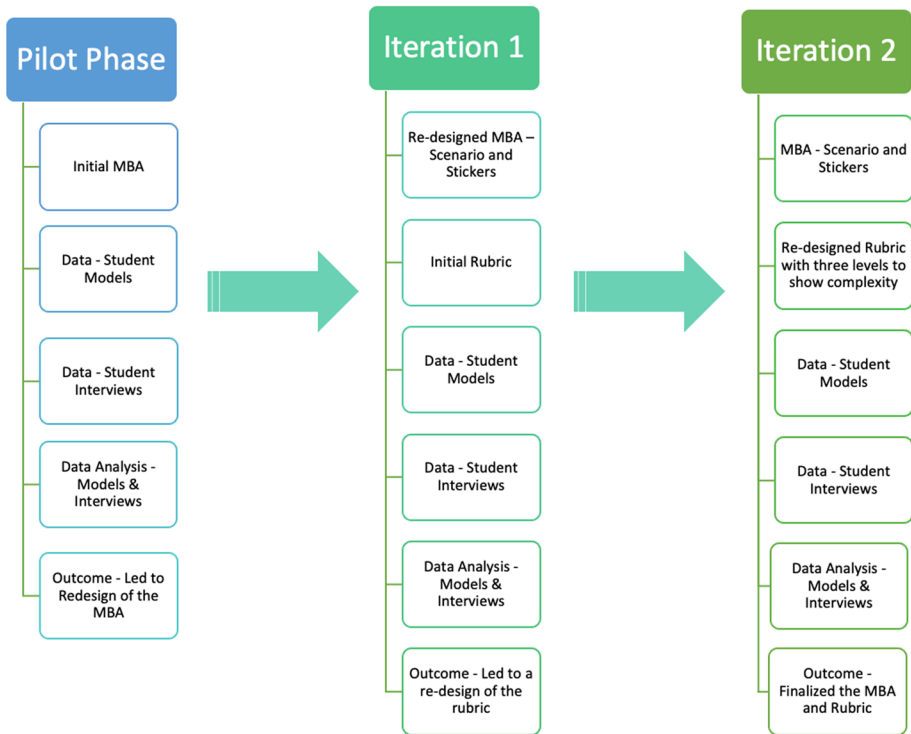


Fig. 3 Overview of study methods and design

conducted an informal interview with each student. Individual students were probed about their experience and were asked to provide feedback about the MBA. Field notes documented student responses and captured student thoughts and feedback about the MBA.

### Iteration 1

Building upon the insights gained from the pilot implementation and the corresponding interviews, the MBA was re-designed. In this re-design, a scenario to set the context for the model was created and a set of stickers developed to be provided to the students. This addressed issues that emerged in the pilot implementation concerning the model context and the students’ drawing capabilities. An initial coding scheme was developed for the MBA that focused on how student MBAs would be assessed for understanding of the flow of energy, the cycling of matter, and the process of decomposition in an ecosystem.

The re-designed MBA was administered two times, pre- and post-implementation of the 8-week curricular unit by the cooperating science teacher during students’ regular science instructional time. Students were provided the MBA page with the revised environmental scenario and a sheet of 30 pre-printed removable stickers along with additional blank stickers to build their model representing their understanding of how energy flows, matter cycles, and the process of decomposition in an ecosystem. The blank stickers were purposefully provided, and aligned with the Resources perspective, in order to encourage

students to bring in elements from their own experiences that were meaningful to their individual sense-making.

Instructions and the scenario were provided to the students and the teacher read them out loud to the students prior to releasing them to construct their models. The same procedure was followed for each implementation of the MBA and students were given 20 min to construct their model. Student models were collected after each implementation, de-identified, and digitized. The MBA was a unique activity for students, as they did not have prior experience representing their knowledge through model building, in this capacity, during their grade six science class. A subset of students were interviewed about their models to gain insight into their thinking. This also allowed the researchers to triangulate data between student interviews, constructed models and rubric scores and provided a basis for the design changes that resulted.

A member of the research team was present for the first administration of the MBA to provide support to the teacher. Students generally understood the directions and what they were expected to do regarding the MBA. Feedback from the teacher and the low number of inquiries from students during model construction led to no additional changes to the design of the MBA. This allowed the research team to shift their focus to how to best analyze the MBAs for student understanding of relevant scientific concepts.

The eleven codes (see Table 2 for code descriptions) initially developed to assess the MBAs served as a first step in examining student learning. Code Presence and Code Count of the 11 codes were used as the mechanism for analysis. However, after the analysis was completed the research team felt that these eleven codes did not capture the complexity and nuance of student knowledge expressed in their MBAs. This was apparent through the intricate connections students made between the abiotic and biotic components of their MBAs, and the additional information students included via their written explanations. These elements of student MBAs were not captured by the 11 codes, thus necessitating a more sophisticated analysis process to capture student understanding and the growth of their knowledge. For example, a student could have increased the presence of Code 4, decomposers break down organic waste, in their post-MBA; yet, based upon the analysis process it could not be determined if a student fully understood how energy flows and matter cycles through an ecosystem. The Code Presence and Code Counts analyses did not provide enough detail to ascertain if students understood

**Table 2** 11 codes created from consensus model

Code	Description
1	Sun/light to plant/producer
2	Producers consumed by consumers
3	Primary to secondary consumers
4	Decomposers break down organic waste
5	Humans produce and/or become organic waste
6	Humans produce inorganic waste
7	Organisms produce and/or become organic waste
8	Decomposers recycle energy and matter through the environment
9	Compost related to decomposition
10	Biotic and Abiotic factors interacting
11	Sun produces nearly all of the energy

the interconnectedness of the complex system. From the analysis of this iteration, the rubric was re-designed using a systems-thinking approach to allow for the analysis of the complexity of the student thinking.

## Iteration 2

In the second iteration, the research team embarked on re-designing and developing the rubric and analysis process that was able to capture the complexity of knowledge displayed in student MBAs. The initial rubric was sufficient to capture the components students displayed in their MBAs but lacked the depth necessary to capture their systems thinking, specifically the relationships and connections students depicted in their MBAs. The research team was inspired by Forbes and colleagues' (2015) work with 3rd grade students' model-driven explanation construction and Zangori and associates' (2017) work on students' model-based reasoning about carbon cycling and climate change. These studies developed and utilized multiple rubrics to capture the many facets of students' models. The current study utilized specific aspects of Forbes and colleagues (2015) and Zangori and associates (2017) work to develop a three-tier rubric to capture the complexity and range of students' model-based reasoning to capture students' systems thinking.

Prior studies (e.g., Forbes et al., 2015; Zangori et al., 2017) examined student models based upon a three-tier rubric that included *Components*, *Sequences*, and *Explanatory Processes* students used to express their understanding of scientific concepts. The *Components* reflect the individual pieces of knowledge that students possess. These are the elements, both visible and non-visible, that students include in their models (Forbes et al., 2015). The *Sequences* reflect students' organization and reorganization of knowledge from pre- to post MBA. *Sequences* captures the relationships between system processes and subprocesses, in addition to the mechanisms that occur within a system (Forbes et al., 2015). Finally, the *Explanatory Processes* provide a glimpse into the deeper understanding students have regarding the concepts of the flow of energy and the cycling of matter in ecosystems through the extended information students provide on their MBAs. *Explanatory processes* are the mechanisms that explain the process of sequences (Forbes et al., 2015), which provides additional insight into student understanding. Modifications were made to each of the three tiers in the rubric and they were adapted to fit the science content areas of energy flow and matter cycling in ecosystems, as well as the process of decomposition.

Two research team members developed the three-tier rubric iteratively; at each stage of development, the research team discussed the rubrics and modifications were made to each tier. Throughout the development process the rubrics were applied to a small sample of models from Iteration 1 to test the three rubrics ability to capture students' model-based reasoning. Additional detail on the development of each of the three tiers of the rubrics follows.

The *Components* rubric, the first tier, serves as a means to determine what elements of an ecosystem students included in their model. The eleven codes developed from the consensus model and the stickers provided to the students served as the initial list of *Components*. As student MBAs were analyzed, additional components were added based upon student's MBAs. Table 3 displays the *Components* rubric, how the components were scored, and examples of what constitutes an individual component. Most stickers that were provided were classified as an individual component (i.e., oxygen, the sun,

**Table 3** Description of the components, sequences, and explanatory processes rubrics

Level	Description	Example*
<i>Components</i>		
0	No components represented in the model	Blank model
1	Any 1–3 components	Organisms (any organism sticker), sun (or light), oxygen, carbon dioxide, soil, waste (poop), waste (trash can), human, inorganic matter, organic matter, water, dead remains, heat, nutrients
2	Any 4–6 components	
3	Any 7 or more components	
<i>Sequences</i>		
0	No sequences are considered	No direct connections can be identified
1	1 or more sequences are represented that describe a simple relationship between 2 components (sequences are in isolation; only demonstrates one relationship)	Sun → producer Producer → consumer Consumer → consumer Human → waste (poop) Consumer → waste (poop) Human → oxygen Human → carbon dioxide
2	1–2 unique sequences with 3 components connected (representing different/unique sequences of relationships e.g. Sun → producer → consumer AND/OR Producer → consumer → consumer) Students may also have Level 1 sequences represented *** Arrows have to be pointing in the same direction for it to be considered a sequence. As long as they are all pointing in the same direction it is good	Sun → producer → consumer Producer → consumer → consumer Producer → consumer → decomposer Waste → decomposer → soil Carbon dioxide → producer → oxygen Carbon dioxide → Producer → consumer Sun → producer → decomposer Sun → producer → human Producer → consumer → human Oxygen → bacteria → waste

**Table 3** (continued)

Level	Description	Example*
3	<p>3 or more unique sequences with 3 or more components connected (representing different/unique sequences of relationships e.g. Sun → producer → consumer AND Producer → consumer → decomposer AND Producer → consumer → decomposer AND Students may also have Level 1 sequences represented)</p> <p>OR</p> <p>1 sequence with 4 or more components connected (no other sequences of 3 or more; may have 2-component sequences)</p>	<p>3 of the identified sequences in level 2</p> <p>OR</p> <p>Sun → producer → consumer → consumer            Sun → producer → consumer → decomposer            Producer → consumer → consumer → decomposer            Producer → consumer → consumer → consumer            Human → organic matter → decomposer → soil → producer            Producer → oxygen → human → carbon dioxide</p>
4	<p>1 sequence with 4 or more components connected AND</p> <p>1 or more sequences with 3 or more components connected (representing two types of relationships in a single chain)</p> <p>The sequence of 3 or more does not have to represent new unique sequence of relationships</p> <p>*A student automatically is a Level 4 if they have 7 or more components connected in a single sequence. Because you can logically separate the sequence into two parts a 4 component and 3 component sequence</p> <p>Students may also have sequences representative of the previous Levels present</p>	<p>Any of the above examples</p> <p>7 or more component example:</p> <p>sun → producer → consumer → consumer → waste (poop) → decomposer → soil → producer            Sun → producer → consumer → consumer → decomposer → waste (poop) → nutrients → soil            Sun → producer → consumer → consumer → human → dead remains → decomposer → soil            Sun → producer → consumer → consumer → human → waste (poop) → decomposer → soil            Sun → producer → consumer → consumer → organic matter → decomposer → soil</p>
0	<p><i>Explanatory Processes &amp; Generalizations</i>—pertaining to the flow of energy or cycling of matter</p>	<p>No identifiable verbiage about the model</p>

Table 3 (continued)

Level	Description	Example*
1	<p>A model contains only one of the following instances in the model</p> <p>Explicit naming of processes (energy transfers)</p> <p>Generalizations/large ideas about <i>the flow of energy or cycling of matter</i></p> <p>Parallel of structures—two structures (sequences) that demonstrate that the student can generalize the information across different contexts</p>	<p>Explicit naming:</p> <p>Student identifies that the connection between two components indicates that energy or matter is being transferred</p> <p>Generalizations:</p> <p>Common student examples:</p> <p>Student identifies that things are made of matter. Note that the sun provides energy</p> <p>“<i>Plants store energy from the sun which turns the energy into chemical energy.</i>”</p> <p>“<i>When the waste is decomposed it turns into nutrients.</i>”</p> <p>“<i>The energy travels from producer to consumer and to decomposers.</i>”</p> <p>“<i>All animals in my model produce waste.</i>”</p> <p>“<i>All living things in my model produce waste.</i>”</p> <p>“<i>Gives it off</i>”—oxygen connected to producer, carbon dioxide connected to owl</p> <p>“<i>Made by everything.</i>” “<i>What this thing makes.</i>” And “<i>Everything uses this.</i>” Then the student identifies items such as waste, organic matter, inorganic matter, etc. as being connected to these brief statements count as generalization</p> <p>NON-Examples of Generalizations –</p> <p>“<i>Eaten by _____.</i>”</p> <p>“<i>It is in the air.</i>”</p> <p>“<i>Connects to almost everything.</i>”</p>

\*List of examples are not exhaustive

and organic matter) and earned a student a single point per sticker placed in their model. However, all organism stickers and any student drawings of organisms (i.e., all plants, animals, and decomposers) were classified together as a single component “organism”. Findings in previous studies conducted by the research team (e.g., Minshew et al., 2017, 2018) suggested students had a strong understanding of the concept of organisms and the relationships that exist between organisms in an ecosystem. Therefore, in order to explore and capture the other elements that students used in their MBAs; organisms were classified as a single component. For example, if a student used the grass, rabbit, and mushroom stickers in their model these counted as a single component, “organism”.

The *Sequences* rubric, the second tier, captured the relationships between components that students represented in their depiction of how energy and matter move through an ecosystem (see Table 3). The *Sequences* rubric was developed by examining existing student MBAs and identifying patterns and connections. Specifically, the number of connections between components and the unique sequences as these distinct elements were important to capture. For example, a student might connect a plant to a rabbit (e.g., plant → rabbit); this demonstrates that the student had an understanding that a relationship existed between the two components. The more continuous connections a student represented in their model the higher their score on the rubric. In addition, it was determined that *Sequences* must be reasonably aligned with scientific accuracy and a comprehensible inferable relationship. Hence, the directionality of the arrows used in students’ constructed models were essential in determining student understanding of energy flow and matter cycling in an ecosystem.

The third and final tier of the rubric captured students’ *Explanatory Processes* that were depicted in their models (see Table 3). These were explanations or specific naming of processes related to the flow of energy and the cycling of matter in ecosystems. These explanations could be simple identification of a process such as “energy transfers” or could be more in-depth explanations such as “Plants store energy from the sun which turns the energy into chemical energy.” In addition to providing explanations, when a student created parallel structures in their model these too were characterized as a student’s ability to generalize about a scientific concept. Being able to generalize a concept is important for student scientific understanding because it indicates that the learner can expand their understanding beyond a single example (Forbes et al., 2015). Similar to the *Sequences* rubric, the more explanations students were able to provide the more points they received.

## Data analysis

### Pilot phase

Data analysis of the student interviews were conducted by the research team in order to identify the major concerns students had while constructing their model. Since the informal interviews were not audio recorded the field notes were analyzed by the research team to gain an understanding of student perceptions regarding the MBA and the task posed to the students. The major ideas expressed by the students were discussed among the research team and potential solutions to student issues were crafted.



## Iteration 1

The research team used the consensus model to identify eleven concepts that students could reasonably be expected to display on the MBA in response to the scenario, and these concepts became the initial eleven codes used to analyze the MBAs (see Table 2 for list and description of codes). Each model was initially evaluated along two dimensions: Code Presence and Code Count. Code Presence considered dichotomously whether or not each code was represented on a given MBA. Expanding upon this, Code Count considered the prevalence of each type of code, including code replications within an individual model. Analysis determined that Code Count did not provide significant information beyond Code Presence data and, therefore was not analyzed further.

Two members of the research team scored the student MBAs and any discrepancies were discussed and resolved. Average scores of the Code Presence on pre- and post-MBA were calculated to examine trends over time for each code. The Code Presence scores were aggregated into on Total Code Presence for each MBA. Paired t-tests were conducted for the pre- and post-MBAs for the Total Code Presence scores. The data were analyzed using R (Mac version 3.5.1).

## Iteration 2

The re-designed rubric consisting of the three tiers -Components, Sequences, and Explanatory Processes- that students displayed in their MBAs were initially applied to six student MBAs (approximately 3% of the MBAs). The research team discussed how well the rubrics captured students' understanding and adjustments were made to each rubric ensuring distinction between each score among the three rubrics. From here three team members scored a small sample of 25 MBAs (approximately 14% of the data set) to determine the inter-rater agreement for each rubric. The raters discussed discrepancies between initial scores and talked about the nuances of the rubric to ensure succinct application of the rubrics to each student model. Inter-rater agreement was calculated at 93% overall for all three rubrics. Inter-rater reliability, using Randolph's (2005) free-marginal multi-rater kappa, was calculated at 0.75 overall-all for all three rubrics which is accepted as strong agreement. The remaining student MBAs were randomly divided among the three raters and the MBAs were scored independently.

The first step in the statistical analysis consisted of descriptive statistics for visual patterns among the three rubric scores. Then mean scores were calculated for each student for pre- and post-MBAs for each of the three rubrics. An overall average was calculated for each rubric at each time point and a paired test was conducted to determine differences between pre- and post-MBA scores.

In the sections that follow, we will discuss the findings from the pilot, iteration 1 and iteration 2, in greater detail and will review how the data that was collected contributed to the development of the MBA and the associated rubrics.

## Findings

### Designing the model-based assessment—pilot phase findings

From the pilot test, we learned that the students needed more direction on how to construct a model, simply instructing students to model their understanding was not sufficient for our specific context. All four participants paused while working on the MBA and asked a clarifying question, such as “*what do you mean by model?*” The students also expressed that they were unsure of which organisms to include in their model with one student asking, “*I can draw anything?*”, suggesting that students again needed more guidance. These support the notion that students often have difficulty identifying which elements of a system are critical to represent in order to scientifically reason about how and why a system works, thus requiring scaffolds to support them as their understanding grows (Zangori et al., 2017). Finally, students disclosed that they did not want to draw the organisms, two students indicated that they were “*bad drawers*” and preferred having options of things to include in their model.

### Refining the pilot assessment

The consensus model, which displayed the relationships between relevant science concepts of the curricular unit, helped inform the re-design of the pilot MBA as it helped to create a scenario that would elicit student understanding of key concepts and relationships in the MBA. The re-design of the MBA also incorporated student feedback from the pilot study. First, the prompt was changed to reflect a specific scenario instead of a general statement about a nondescript environment. In the revised prompt, the students were asked, “Imagine you are in a park that is surrounded by a field, forest, and

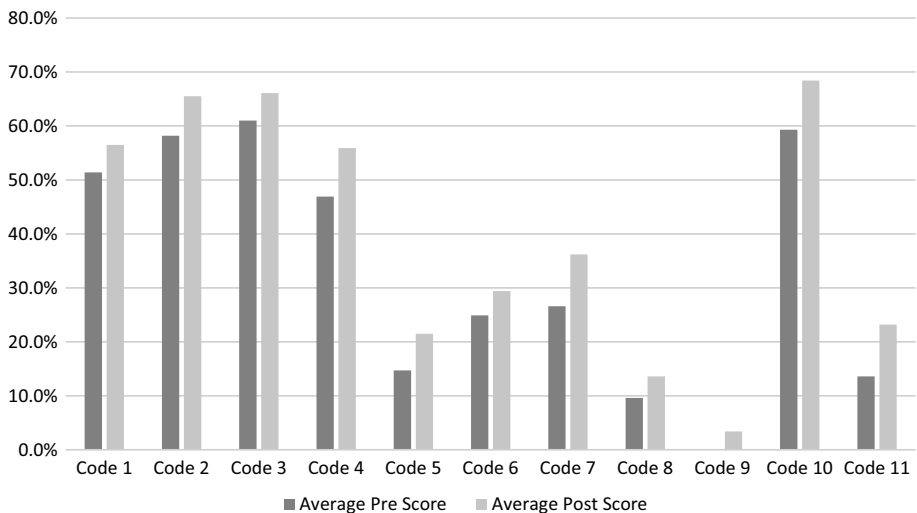
**Table 4** Re-design of model-based assessment prompt

Iteration 1: initial design	Re-design
<p>Prompt provided to students: Energy is all around us. During this unit you will learn about energy and how it travels through the environment. But First, create your own model of how you think energy flows through the world around you. You can use drawings, words You can start by thinking about where energy comes from and where it goes from there</p>	<p>New prompt: Imagine... You are in a park that is a part of the a large field. You are on a picnic with your family. The field is bordered by a forest on one side, a parking lot and a pond on the other side. May people and animals spend time in this place and sometimes leave stuff behind Explain by drawing, using pictures, and words, how you think <i>energy and matter</i> move through the field and park ecosystem</p> <ul style="list-style-type: none"> <li>• Think: Start by thinking about where energy and matter come from and where they go from there</li> <li>• Arrows: Use arrows to show the direction that you think the energy flows</li> <li>• Stickers: You can use any, all, or none of the stickers provided to show your thinking. You can also draw things that you do not have stickers for. Use whatever works for you</li> </ul>

a pond”, this provided students with a specific context in which to create their MBA. Next, additional support was provided to students regarding how to construct a model. The instructions included a brief statement that instructed students to “Explain by drawing, using pictures and words, how you think energy and matter move through the field, forest, and park ecosystem.” Students were also prompted to think about where energy and matter come from, where energy and matter go, and to use arrows to show the direction they think energy flows through the ecosystem (See Table 4 for design changes). Finally, students were provided thirty removable organism stickers to use in order to support the construction of their models. Stickers included both biotic and abiotic components of an environment. The re-designed MBA was reviewed by the research team and the participating teacher. Consensus regarding the different supports provided to the students to construct a model was achieved.

### Iteration 1—first classroom implementation

Utilizing the design changes that emerged from the findings of the Pilot, the revised MBA was implemented in the first classroom study. The initial rubric used to analyze the MBAs was created using the curriculum’s consensus model and resulted in 11 codes (see Table 2 for description of the codes) that represented eleven key concepts presented in the curriculum. The student models were scored using Code Presence. The analysis of Code Presence allowed us to investigate trends, specific relationships, and scientific concepts that were displayed in student MBAs. For instance, the pre-MBA data revealed that while roughly half of the students understood that producers were consumed by consumers (Code 2), no students included compost as part of decomposition (Code 9) and only two students indicated the sun was the origin of nearly all energy in an ecosystem (Code 11). The average Code Presence scores for the combined data set ( $n=177$ ), increased from pre to post-MBA for each code (Fig. 4). Likewise, mean Total Code Presence increased from pre- ( $m=3.7$ ,



**Fig. 4** Code presence. This figure shows the percent of student models which included representations of concepts relating to each of the 11 codes

$sd=2.5$ ) to post-MBA ( $m=4.4$ ,  $sd=2.5$ ) demonstrating that students were representing more of the targeted concepts in their post-MBAs than their pre-MBAs. Paired t-test revealed this increase was statistically significant,  $t(176)=4.1$ ,  $p\text{-value} < 0.00$ .

### Refining the iteration 1 assessment

Evidence of gains based on the Code Presence scores was encouraging, however, this approach to analyzing student MBAs was limited. While this method did allow us some insight into student understanding by capturing whether they represented the eleven key concepts within their MBAs, it did not provide information about how the information was connected to the larger ecological system. Thus, the initial 11 codes only captured the components students included in their models and no insight into the synthesis of system components or their implementation (Orgill et al., 2019). This prompted our research team to begin to explore methods of assessing the MBAs which would capture not only the key concepts but also student representations of the relationships and connections among the components present in their MBAs.

### Iteration 2—second classroom implementation

The re-designed three-tier rubric, consisting of Components, Sequences, and Explanatory Processes, was applied to all MBAs generated during the two classroom implementations ( $n=177$ ) of the (name Bio-Sphere Compost) curricular unit. Student scores were highest on the *Components* rubric ( $m=2.53$ ; maximum score=3), with many students scoring the maximum score of 3 on the pre-MBA. Student scores were second highest on the *Sequences* rubric ( $m=2.11$ ; maximum score=4), and lowest on the *Explanatory Processes* rubric ( $m=0.60$ ; maximum score=4). Overall, the mean post-scores were higher than the mean pre-scores and all three rubrics showed significant increases from the pre to post (see Table 5). The individual rubric scores were combined into a single Total Rubric Score for each pre-MBA ( $m=5.2$ ,  $sd=2.2$ ) and post-MBA ( $m=6.3$ ,  $sd=2.2$ ). A paired t-test of the Total Rubric Score revealed that students scored significantly higher on the post-MBA compared to the pre-MBA,  $t(177)=5.3$ ,  $p\text{-value} < 0.000$ .

In addition to being sensitive to changes over time regarding the overall averages, the rubric scores revealed individual growth that may otherwise be missed through visual inspection of the MBA alone. Table 6 shows three examples of student pre- and post-MBAs. Each student had growth from pre- to post-MBA for the Sequences and Explanatory Processes rubrics. As noted in the statistical analysis, the students had high Components scores on both MBAs with students B and C obtaining the maximum scores on their

**Table 5** Mean pre and post scores for each rubric

Rubric	Average pre-score	Average post-score
Component	2.53 (0.7)	2.7 (0.6)
Sequences	2.11 (1.6)	2.6 (1.4)
Explanatory process	0.60 (1.0)	1.0 (1.3)
Total rubric score	5.2 (2.2)	6.3 (2.2)

Standard deviation is included in parentheses

**Table 6** Advanced rubric pre and post-MBAs and scores

Pre-Score	Pre-MBA	Post - Score	Post-MBA
<p><i>Student A</i></p> <p>C* = 2 S = 3 EP = 1 Total = 6</p>		<p>C = 3 S = 4 EP = 4 Total = 11</p>	
<p><i>B</i></p> <p>C = 3 S = 1 EP = 0 Total = 4</p>		<p>C = 3 S = 4 EP = 2 Total = 9</p>	
<p><i>C</i></p> <p>C = 3 S = 2 EP = 0 Total = 5</p>		<p>C = 3 S = 4 EP = 2 Total = 9</p>	

\*C components (maximum score = 3), S sequences (maximum score = 4), EP explanatory processes (maximum score = 4)

pre-MBA. Thus, demonstrating the students were able to effectively identify abiotic and biotic elements of an ecosystem.

The Sequences and Explanatory Processes rubrics are where growth by the students was more evident. Student organization of the components in their post-MBAs was more sophisticated than their pre-MBAs. For example, Student A, despite creating similar looking models had growth on the Sequences rubric obtaining the maximum score of four on their post-MBA. In their pre-MBA, Student A grouped the abiotic factors of carbon dioxide, oxygen, the sun, and light into one large group that connected to plants and had soil connecting to the grouped abiotic factors. Whereas in their post-MBA Student A showed the sun and light starting the sequence of components and no components connected back to the sun and light. On their pre-MBA, Student B was only able to make a few connections between components, with most of the stickers placed in groups (i.e., producers, consumers, decomposers) as a form of identification. However, Student B's post-MBA was more structured and included the explicit naming of processes that occur within an ecosystem, earning them the maximum score of four on the Sequences rubric and a two on the Explanatory Processes rubric. Student B continued to group the stickers appropriately in their post-MBA, as well as added arrows showing how the groups were connected in

the ecosystem. Student B also provided detailed explanations on their post-MBA such as “Plants store energy from the sun or light which turns the energy into chemical energy.” At a glance, student C created numerous connections on their pre-MBA, however, several of these connections were not plausible and therefore did not count towards their score. Whereas, for their post-MBA, Student C created three separate models in the space provided with numerous unique connections earning them the maximum score of four on the Sequences rubric. Finally, Student C included generalizations and had parallel structures represented in their post-MBA earning them a 2 on the Explanatory Processes rubric.

## Discussion of findings

In this study, we described the iterative development of a MBA, designed to elicit evidence of student understanding of key concepts and relationships relevant to a complex systems-thinking approach around the phenomenon of an ecosystem, specifically the flow of energy, the cycling of matter, and the process of decomposition. In addition, the study described two iterations of the rubric that was developed to examine student systems thinking from a Resources perspective. The iterative development of the MBA found the need for a structured and direct prompt to support students in expressing their understanding of the curriculum concepts. Students required context, guidance on how to construct a model, as well as support in the types of components potentially found in the prescribed ecosystem. The rubrics required refinement as it was observed that students engaged in systems thinking and the initial rubric did not capture the nuance of student understanding represented in the MBAs. The refined three-tiered rubric captured the Components, Sequences, and Explanatory Processes students included in the MBAs. The three-tiered rubric demonstrated that students had a firm understanding of the components in an ecosystem by obtaining high scores on their pre-MBA, and displayed the slight, yet statistically significant growth from pre- to post-MBA for the Sequences and Explanatory Processes.

The MBA provided information about deeper learning of complex systems thinking by requiring students to apply, construct, and validate their epistemological understandings of key concepts through the modeling of an ecosystem as an assessment (Passmore & Stewart, 2002). This builds on and extends the previous work of multiple researchers (e.g. Berland & Reiser, 2009; Braaten & Windschitl, 2011; Constantinou et al., 2019; Giere, 2004; Hughes, 1997; Passmore & Stewart, 2002) and applies it to the development of an alternative assessment. The information gathered in this assessment allowed both researchers and teachers to understand what concepts students were able to spontaneously incorporate and apply in their own systems thinking approaches. Analyzing this data allowed us to determine both salient and absent concepts and understandings, which informed the design of subsequent iterations of the MBA and rubrics. Further, the MBA showed the transition of student understanding from identifying the components and making simple-linear connections to students organizing the components to represent dynamic relationships and processes that exist in an ecosystem (Assaraf & Orion, 2005; Hmelo-Silver et al., 2007; Orgill et al., 2019).

While the initial code-based scoring scheme allowed for the measurement of student understanding through the presence or absence of key concepts, the three-tier rubric provided additional information on how student understanding was expressed through their MBA. The three-tier rubrics move beyond simply scoring what was included in student models (i.e., components) to capturing both what students included and how they arranged

their models to represent their understanding (i.e., sequencing & explanatory processes). This notion is supported by the (2019), Gogolin & Krüger (2018), and Nowak and colleagues (2013) who examined the roles of models for epistemological reasoning and models of scientific thinking. For example, Student A created models that appeared similar in the pre- and post-assessment; however, the connections and relationships depicted differed suggesting that their epistemological reasoning had developed. The three-tier rubric also provided insight into students' complex systems thinking, specifically, the relationships and connections students made between the different parts of an ecosystem, as well as student ability to describe how systems-level phenomena occur based on the interactions between the system's parts (e.g., energy released through the process of decomposition) (Orgill et al., 2019; Evagorou et al., 2009).

Students tended to include most, if not all, of the components we expected of them, indicating students understood what abiotic and biotic factors were relevant to ecosystems. Identification of components in an ecosystem is a necessary step in understanding the system as a whole (Forbes et al., 2015; Zangori & Forbes, 2015). The directed MBA scenario and the pre-printed stickers provided a scaffold for students as they constructed their model. The stickers correspond to the Resources perspective used to develop the curriculum. The stickers encouraged the activation of student prior knowledge for the pre-MBA and served as a support to help students represent and reconcile their understanding of new knowledge with their existing understanding. While the argument could be made that the high scores were influenced by the presence of the sticker sheet provided to students, students were encouraged to, and often did, draw their own abiotic and biotic factors. As depicted in our example students' MBAs, students often included additional animals, and other forms of organic and inorganic waste not depicted on the sticker sheet. This was reiterated in the student interviews about their MBAs.

Beyond merely capturing student understanding of what belongs in an ecosystem, the three-tiered rubric provided insight into the connections students made between components. The Sequences rubric provided parameters to assess student understanding of the relationships and connections they were making between components. It helped to identify meaningful and non-meaningful connections students made between the concepts using lines and arrows to show relationships. These constructed sequences display the relationships between the system subprocesses and show student understanding of the mechanisms of the system (Forbes et al., 2015; Zangori et al., 2017). The Sequences rubric was able to highlight this important conceptual understanding of the complexity of ecosystems (Zangori et al., 2017); it also helps in the development of systems-thinking approaches by students as described by Hmelo-Silver and Azevedo (2006) and utilized in this study to examine its impact on assessment of student understanding, using the MBA as an epistemic tool (Ritchey, 2012). It is important to note, that students were not instructed on how to construct their MBAs. As our example student MBAs show, students had some understanding that components of an ecosystem were connected, but not all students (e.g., Student B) initially constructed their MBAs to represent those connections.

Our results demonstrate that student MBAs expressed their understandings of how individual components, or chain of components, were related (sequences), but they rarely communicated generalized understandings (explanatory processes). While some students did have scores of 3 or 4 on the *Explanatory Processes* rubric, most students tended to score a 0 or 1. While it is uncertain if the low scores were influenced by the curriculum or the MBA task, it does draw our attention to this facet of understanding as something that should be explored more deeply and addressed in future redesigns. The results demonstrate that students were just beginning to synthesize and think more deeply about the



relationships that exist in an ecosystem (Orgill et al., 2019). This is reminiscent of Hmelo-Silver and associates (2007) work that found students had an easier time comprehending the structures of a complex system than the behaviors and functions. Students could easily identify the component, but initially had a difficult time demonstrating accurate relationships and providing explanations for their thinking. Further, our results support Zangori and colleagues (2020) notion that there is an intermediate level of reasoning, known as relational. Based solely on student MBA data (e.g., what students depicted in their MBAs), students did not assign causality in their MBAs; however, when students were questioned about their MBAs, they provided more detailed explanations and thought processes than what they wrote down.

The rubrics designed by Zangori and colleagues (2017) built upon each other and we too, followed the same development when modifying the rubrics to fit our study. Our students followed a similar trajectory in that their models first contained essential components, then appropriate and plausible sequences, and finally some students were able to include the underlying mechanisms of an ecosystem in their MBA. Thus, in order for students to express deeper understanding of causal mechanisms in a complex system they must first understand what belongs (Zangori et al., 2017). Student development of first fully understanding the components, and then the more complex areas of an ecosystem adhere to the Resources perspective that influenced the work. Learning occurs in phases as students construct their understanding based on their experiences and the ‘reweaving’ and incorporation of new knowledge into an existing often ‘naïve’ understanding (Hammer et al., 2005).

The preliminary testing of the three-tier rubric indicates that this method of analyzing student created models is viable and an improvement upon the initial 11 concept coding scheme that was used in early iterations. The individual scores and observed patterns in the scores provide a more nuanced understanding of the complexity of ecosystems and the intricacies of how students visually represent their understanding while also demonstrating their abilities to engage in model-based learning (Constantinou et al., 2019). As our analysis showed, the three-tier rubric provides a multi-faceted understanding of student conception of ecosystems and also demonstrated that it can capture changes in student systems thinking over time. While further testing still needs to be conducted with the rubrics, they represent a step in the right direction for capturing student understanding of a complex system.

## Conclusion

This paper focused on the development of the MBA and three-tier rubric, not on student outcome data. However, student data informed design changes to both the MBA and the three-tier rubric which reflects the iterative DBR process of Conceive, Build, and Test. Our goal was to create an authentic assessment that allowed for students to demonstrate their understanding of ecological processes at a deeper level in order to fully enact a constructivist approach to learning. The design of the MBA allowed students to express their conceptual understanding and the rubrics identified the change in student understanding from pre- to post-MBA.

The development of the MBA and rubric is preliminary and extends the work by Zangori and associates (2017) and Forbes and colleagues (2015) to other areas of science content. It also builds on the work of Harrison and Treagust (2000), Hmelo-Silver and Azevedo (2006), Passmore and colleagues (2014), Yoon and associates (2015), and

Constantinou and colleagues (2019) in using models to examine complex systems thinking, in addition to understanding the causal linkages and complexities that arise within scientific phenomenon such as the flow of energy and cycling of matter within ecosystems. We again recognize the limitations of this study, with its small population and rural context; however, the iterative design and development of the MBA and rubrics make the assessment strong for examination of student understanding of complex systems.

**Acknowledgements** The authors would like to thank Dr. Sharon J. Derry for her support and critical evaluation of the MBA and rubric design, Dr. Kelly J. Barber-Lester for her critical work on evaluating the MBAs and rubrics, and finally Molly Ewing for her work analyzing the data.

**Funding** Funding was supported by National Science Foundation (Grant No. DRL1418044).

## References

- Aduriz-Bravo, A., (2019). Semantic views on models: An appraisal for science education. In zu Belzen, K. & van Diel (Eds.). *Toward a competence-based view on models and modeling in science education*. Springer. <https://doi.org/10.1007/978-3-030-30255-9>
- Assaraf, O.B.-Z., & Orion, N. (2005). Development of system thinking skills in the context of Earth system education. *Journal in Research in Science Teaching*, 42, 518–560.
- Baek, H., Schwarz, C., Chen, J., Hokayem, H., & Zhan, L. (2011). Engaging elementary students in scientific modeling: The MoDeLS fifth-grade approach and findings. *Models and modeling* (pp. 195–218). Springer.
- Barak, J., Gorodetsky, M., & Chipman, D. (1997). Understanding of energy in biology and vitalistic conceptions. *Journal of Science Education*, 19(1), 21–30. <https://doi.org/10.1080/0950069970190102>
- Barak, J., Sheva, B., & Gorodetsky, M. (1999). As ‘process’ as it can get: Students’ understanding of biological processes. *International Journal of Science Education*, 21(12), 1281–1292.
- Bell-Basca, B. S., Grotzer, T. A., Donis, K., & Shaw, S. (2000, April). Using domino and relational causality to analyze ecosystems: Realizing what goes around comes around. In *Paper presented at the annual meeting of the National Association of Research in Science Teaching*.
- Ben-Zvi-Assaraf, O., & Orion, N. (2010). Four case studies, six years later: Developing system thinking skills in junior high school and sustaining them over time. *Journal in Research in Science Teaching*, 47, 1253–1280.
- Berland, L. K., & Reiser, B. J. (2009). Classroom communities’ adaptations of the practice of scientific argumentation. *Science Education*, 95(2), 191–216.
- Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, 95(4), 639–669. <https://doi.org/10.1002/sce.20449>
- Clark, D., & Linn, M. (2003). Designing for knowledge integration: The impact of instruction on time. *The Journal of the Learning Sciences*, 12(4), 451–493.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Collins, A., Joseph, D., & Bielczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15–42.
- Constantinou, C. P., Nicolau, C. Th., & Papaevripidou, M., (2019). A framework for modeling-based learning, teaching and assessment. In K. zu Belzen & van Diel (Eds.). *Toward a competence-based view on models and modeling in science education*. Springer. <https://doi.org/10.1007/978-3-030-30255-9>
- Crawford, B., & Cullin, M. (2004). Supporting prospective teachers’ conceptions of modelling in science. *International Journal of Science Education*, 26(11), 1379–1401.
- diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Lawrence Erlbaum Associates.
- diSessa, A. A. (2002). Why “conceptual ecology” is a good idea. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change. Issues in theory and practice* (pp. 29–60). Kluwer Academic Publishers.
- diSessa, A. A. (2014). A history of conceptual change research: Threads and fault lines. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 88–108). Cambridge University Press.

- Dushl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching in grades K-8* (pp. 163–166). National Academies Press.
- Easterday, M., Rees Lewis, D. & Gerber, E. (2014). Design-based research process: Problems, phases and applications. In J. L. Polman, E. A. Kyze, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, K. O'Connor, T. Lee, & L. D'Amico (Eds.) *11th international conference of the learning sciences* (pp. 317–324).
- Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of the potential of interactive simulations for developing system thinking skills in elementary school: A case study with fifth graders. *International Journal of Science Education, 31*, 655–674.
- Forbes, C. T., Zangori, L., & Schwarz, C. V. (2015). Empirical validation of integration learning performances for hydrologic phenomena: 3rd-grade students' model-driven explanation construction. *Journal in Research in Science Teaching, 52*(7), 895–921. <https://doi.org/10.1002/tea.21226>
- Giere, R. N. (1999). Using models to represent reality. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model based reasoning in scientific discovery* (pp. 41–57). Kluwer/Plenum.
- Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.
- Gogolin, S., & Krüger, D. (2018). Students' understanding of the nature and purpose of models. *Journal of Research in Science Teaching, 55*(9), 1313–1338.
- Grotzer, T. A., & Bell-Basca, B. (2003). How does grasping the underlying causal structures of ecosystems impact students' understanding? *Journal of Biological Education, 38*(1), 16–29.
- Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences, 12*(1), 53–90. [https://doi.org/10.1207/S15327809JLS1201\\_3](https://doi.org/10.1207/S15327809JLS1201_3)
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of Learning from a modern multidisciplinary perspective* (pp. 89–119). Information Age Publishing.
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education, 22*(9), 1011–1026.
- Hayes, M. L., Plumley, C. L., Smith, P. S., & Esch, R. K. (2017). A review of the research literature on teaching about interdependent relationships in ecosystems to elementary students. Horizon Research, Inc.
- Hmelo-Silver, C. E., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *The Journal of the Learning Sciences, 15*(1), 51–61. [https://doi.org/10.1207/s15327809jls1501\\_7](https://doi.org/10.1207/s15327809jls1501_7)
- Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science, 45*(1), 53–72. <https://doi.org/10.1007/s11251-016-9392-y>
- Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-Novice understanding of complex systems. *The Journal of the Learning Sciences, 16*(3), 307–331. <https://doi.org/10.1080/10508400701413401>
- Hogan, K. (2000). Assessing students' systems reasoning in ecology. *Journal of Biological Education, 35*(1), 22–28.
- Hogan, K., & Fisherkeller, J. (1996). Representing students' thinking about nutrient cycling in ecosystems: Bidimensional coding of a complex topic. *Journal of Research in Science Teaching, 33*(9), 941–970.
- Hokayem, H., & Schwarz, C. (2014). Engaging fifth graders in scientific modeling to learn about evaporation and condensation. *International Journal of Science and Mathematics Education, 12*(1), 49–72. <https://doi.org/10.1007/s10763-012-9395-3>
- Honward, S., Hmelo-Silver, C., Jordan, R., Eberbach, C., Gray, S., Sinha, S., Goel, A., Vattam, S., Rugaber, S., & Joyner, D. (2010). Connecting the visible to the invisible: Helping middle school students understand complex ecosystem processes. In *Proceedings of 32nd annual meeting of the cognitive science society* (pp. 133–138)
- Hughes, R. (1997). Models and representation. *Philosophy of Science, 64*, S325–S336.
- Krajcik, J., & Merritt, J. (2012). Engaging students in scientific practices: What does constructing and revising models look like in the science classroom? *The Science Teacher, 79*(3), 38.
- Lehrer, R., & Schauble, L. (2007). A developmental approach to supporting the epistemology of modeling. *Modeling and applications in mathematics education* (pp. 153–160). Springer.
- Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education, 96*(4), 701–724. <https://doi.org/10.1002/sce.20475>
- Lin, C.-Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. *International Journal of Science Education, 25*(12), 1529–1544.
- Manz, E. (2012). Understanding the co-development of modeling practice and ecological knowledge. *Science Education, 96*(6), 1071–1105. <https://doi.org/10.1002/sce.21030>

- McKenney, S. E., & Reeves, T. C. (2018). *Conducting educational research design: What, why and how* (2nd ed.). Routledge.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- Minshew, L. M., Barber-Lester, K. J., Derry, S. J., & Anderson, J. L. (2017). Leveraging students' knowledge to adapt science curricula to local context. *Educational Technology & Society*, 20(4), 205–218.
- Minshew, L. M., Barber-Lester, K., Anderson, J. L., & Derry, S. (2018). Uncovering students' ecological knowledge in pieces: Advancing our knowledge and informing design. In Proceedings of the International Conference of Learning Sciences – 2018 London.
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- Nicolaou, C. T., & Constantinou, C. P. (2014). Assessment of the modeling competence: A systematic review and synthesis of empirical research. *Educational Research Review*, 13(3), 52–73.
- Nowak, A., Rychwalska, A., & Borkowski, W. (2013). Why simulate? To develop a mental model. *Journal of Artificial Societies and Social Simulation*, 16(3), 12.
- Opitz, S. T., Harms, U., Neumann, K., Kowalzik, K., & Frank, A. (2015). Students' energy concepts at the transition between primary and secondary school. *Research in Science Education*, 45, 691–715. <https://doi.org/10.1007/s11165-014-9444-8>
- Orgill, M., York, S., & MacKellar, J. (2019). Introduction to systems thinking for the chemistry education community. *Journal of Chemical Education*, 96, 2720–2729. <https://doi.org/10.1021/acs.jchemed.9b00169>
- Passmore, C., Gourvea, J. S., & Giere, R. (2014). Models in science and in learning science: Focusing scientific practice on sense-making. In Matthews, M. (Ed.) *International handbook of research in history, philosophy and science teaching*. Springer.
- Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching*, 39(3), 185–204.
- Randolph, J. J. (2005). Free-marginal multirater kappa: An alternative to Fleiss' fixed-marginal multirater kappa. In *Paper presented at the Joensuu University Learning and Instruction Symposium 2005*, Joensuu, October 14–15th, 2005.
- Ritchey, T. (2012). Outline for a morphology of modeling methods: Contribution to a general theory of modelling. *Acta Morphologica Generalis*, 1, 1–20.
- Schwarz, C., Reiser, B., Davis, E., Kenyon, L., Acher, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Svandova, K. (2014). Secondary school students' misconceptions about photosynthesis and plant respiration: Preliminary results. *Eurasia Journal of Mathematics, Science & Technology Education*, 10(1), 59–67.
- Smith, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, 3(2), 115–163. [https://doi.org/10.1207/s15327809jls0302\\_1](https://doi.org/10.1207/s15327809jls0302_1)
- van Joolingen, W. R., Schouten, J., & Leenaars, F. (2019). Drawing-based modeling in teaching elementary biology as a diagnostic tool. In K. zu Belzen, & van Driel (Eds.). *Toward a competence-based view on models and modeling in science education*. Springer. <https://doi.org/10.1007/978-3-030-30255-9>.
- Yoon, S. A., Anderson, E., Koehler-Yom, J., Klopfer, E., Sheldon, J., Wendel, D., et al. (2015). Designing curriculum and instruction for computer-supported complex systems teaching and learning in high school science classrooms. *J-STEM: Journal of Research in STEM Education*, 1(1), 4–14.
- Zangori, L., & Forbes, C. T. (2015). Exploring third-grade student model-based explanations about plant relationship within an ecosystem. *International Journal of Science Education*, 37(18), 2942–2964. <https://doi.org/10.1090/09500693.2015.1118772>
- Zangori, L., Ke, L., Sadler, T. D., & Peel, A. (2020). Exploring primary students causal reasoning about ecosystems. *International Journal of Science Education*, 42(11), 1799–1817. <https://doi.org/10.1080/09500693.2020.1783718>
- Zangori, L., Peel, A., Kinslow, A., Friedrichsen, R., & Sadler, T. D. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching*, 54(10), 1249–1273.
- zu Belzen, A. U., van Driel, J., & Krüger, D. (2019). Introducing a framework for modeling competence. In *Towards a competence-based view on models and modeling in science education* (pp. 3–19). Springer.