

# Reasoning Up and Down a Food Chain: Using an Assessment Framework to Investigate Students' Middle Knowledge

AMELIA WENK GOTWALS

*College of Education, Michigan State University, East Lansing, MI 48824-1034, USA*

NANCY BUTLER SONGER

*School of Education, The University of Michigan, Ann Arbor, MI 48109-1259, USA*

*Received 26 March 2009; revised 27 July 2009; accepted 10 August 2009*

*DOI 10.1002/sce.20368*

*Published online 20 October 2009 in Wiley InterScience (www.interscience.wiley.com).*

**ABSTRACT:** Being able to make claims about what students know and can do in science involves gathering systematic evidence of students' knowledge and abilities. This paper describes an assessment system designed to elicit information from students at many placements along developmental trajectories and demonstrates how this system was used to gather principled evidence of how students reason about food web and food chain disturbances. Specifically, this assessment system was designed to gather information about students' intermediary or middle knowledge on a pathway toward more sophisticated understanding. Findings indicate that in association with a curricular intervention, student gains were significant. However, despite overall gains, some students still struggled to explain what might happen during a disturbance to an ecosystem. In addition, this paper discusses the importance of having a cognitive framework to guide task design and interpretation of evidence. This framework allowed for the gathering of detailed information, which provided insights into the intricacies of how students reason about scientific scenarios. In particular, this assessment system allowed for the illustration of multiple types of middle knowledge that students may possess, indicating that there are multiple "messy middles" students may move through as they develop the ability to reason about complex scientific situations. © 2009 Wiley Periodicals, Inc. *Sci Ed* **94**:259–281, 2010

*Correspondence to:* Amelia Wenk Gotwals; e-mail: Gotwals@msu.edu

Contract grant sponsor: National Science Foundation.

Contract grant numbers: REC-0089283 and REC-0129331.

Any opinions, findings, and conclusion or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Spencer Organization or the National Science Foundation.

## INTRODUCTION

Goals for science education include deep conceptual understandings of key scientific ideas that integrate knowledge of focal science concepts with the ability to utilize this content in inquiry reasoning situations (National Research Council [NRC], 2001b, 2007). However, on international tests that focus on complex reasoning and scientific literacy (e.g., PISA), American students score well below students in other industrialized nations (Organisation for Economic Co-operation and Development [OECD], 2007). While these standings are disappointing, it is important to acknowledge that the development of deep conceptual understandings about key scientific topics takes time. Most current assessment systems are not designed to examine more than the end points of students' understandings: whether students do not understand (the lower end point) or whether students do understand key scientific content (the upper end point).

Learning progressions in science articulate trajectories for how students build more sophisticated and coherent understandings of the big ideas in science (NRC, 2007; Songer, Kelcey, & Gotwals, 2009). Having assessment systems that can provide opportunities for students to demonstrate more than the end points of their knowledge are important if we value the pathways that students take as they develop deep conceptual understandings. In particular, as research on learning progressions is relatively new, it is important to have assessment systems that can help to gather evidence of the intermediary knowledge or "middle knowledge" that students have as they develop more sophisticated and deep understandings and abilities to reason complexly about key scientific ideas. These articulations of students' middle knowledge will allow us to build better and more targeted curricular, pedagogical, and assessment resources to help students move toward sophisticated understandings.

With current international media interest in global warming and decreasing biodiversity (e.g., <http://topics.nytimes.com/top/news/science/topics/globalwarming/index.html>), a need for student scientific literacy about ecological and environmental ideas is suggested. However American science students demonstrate poor performance on international assessments on these topics (OECD, 2009). While we may not expect that elementary and middle school students will fully understand the intricacies of these scientific ideas, ensuring that they have a base of knowledge from which to build toward a more sophisticated understanding of ecological issues is very important. One of these keys to scientific literacy how energy and matter flow through ecosystems as represented by food chains and food webs (Alexander, 1982). In addition, understanding how disturbances in ecosystems can influence food chain and food web interactions is an integral concept for students to understand as they build toward scientific literacy.

This paper outlines the development of an assessment system to systematically gather evidence of students' understandings about and abilities to reason with concepts in ecology, including food chain and food web disturbances. Specifically, this assessment system was designed to gather evidence of how students develop more sophisticated understandings and abilities to reason about ecological scenarios—focusing especially on the intermediary understandings or middle knowledge that students move through as they develop higher levels of knowledge and reasoning abilities. The analysis in this paper provides information into the nuances of students' middle knowledge in these content areas. The research questions guiding the study were as follows:

- What guiding principles are needed to develop an assessment system that illuminates middle knowledge along the path toward the development of complex reasoning about ecology concepts?

- What kinds of information about students' development of complex reasoning about ecological concepts can be illustrated through a systematic approach to assessment that illuminates middle knowledge in science?
- What implications for teaching and learning complex ecology concepts can be drawn from this work?

### Student Knowledge of Food Webs

Complex understandings of ecological issues are extremely important in this time when issues in the environment are in the news and policy realm (OECD, 2009). Inherent in the ideas about how global warming impacts the earth is the understanding of ecosystem interactions. The concepts associated with food webs and food chains are central to understanding more complex environmental issues such as population management, the effects of pollution, and the influence of global warming on biodiversity (Alexander, 1982). However, research has shown that students at many different levels hold misunderstandings about aspects of food chains and food webs (Barman & Mayer, 1994; Gallegos, Jerezano, & Flores, 1994; Hogan, 2000).

More specifically, Gallegos and colleagues (1994) found that fourth-, fifth-, and sixth-grade students often think of predator-prey relationships as larger animals eating smaller animals. In addition, their work reports that children often mix up the direction of arrows in a food chain or food web, thinking of arrows as moving from prey to predator instead of following the energy in a system. Even when students correctly begin their food chains with a producer, they often do it because the plant is the smallest organism, rather than it being a producer. High school students generally understand that food chains and food webs can illustrate feeding relationships but do not make the step to see these relationships as a means of energy transfer between organisms (Barman & Mayer, 1994; Smith & Anderson, 1996). Another related difficulty is focused on recognizing the organisms in the ecosystems. Sometimes students do not recognize the organisms (specifically, the names of organisms) in the ecosystem or do not realize that parts of plants (such as nectar) that are consumed count as producers. These misunderstandings can influence students' abilities to reason about specific food chain and food web structures and disturbances (Reiner, 2001).

When examining how students think about food chain and food web disturbances, Hogan (2000) found that sixth-grade students generally traced food web and food chain disturbances as having effects only in one direction and did not take into account multiple pathways in food webs even after a 1-month instructional unit on ecosystems. Similarly, Barman and Mayer (1994) found that many high school students could not provide an accurate explanation of how a change in one population in a food web would influence other populations in the food web. Several students in this study also thought that a change in one population in a food web would affect another population only if it were directly connected in a predator-prey relationship. In addition, Reiner (2001) found that students thought that if one organism in a food chain ceased to exist then all organisms upward of that organism would be eliminated, the populations below that organism would be increased, and organisms on other chains in the food web would feed on alternative food sources.

Research results identifying specific difficulties provide an inconsistent characterization. Leach, Driver, Scott, and Wood-Robinson (1996) found that students tend to trace linkages and disturbances upward more easily from producers to predators than downward in trophic levels. They hypothesize that it is because it is easier to understand lack of food and starvation as reasons for why an animal population would decline. In contrast, Hogan (2000) found that students had an easier time understanding that populations of prey would increase in size when predators decreased and attributed this to a lack of conceptual models

for complex systems interactions. Despite research on how students reason about food chains and food webs, there are still many aspects of students' abilities that are not fully understood. What is needed is an assessment system specifically designed to unpack the complexity associated with reasoning about food chain and food web disturbances leading to empirical data that can be applied to teaching and learning contexts and materials.

### Using Assessment for Cognitive Research

As students' minds are not transparent vessels illustrating what they know, we must make interpretive claims (inferences) about what a student knows and can do based on observations of his or her performances on assessment tasks. Assessment includes the processes of gathering evidence about students' knowledge and abilities as well as making inferences from that evidence about what students know or can do more generally (Mislevy, Wilson, Ercikan, & Chudowsky, 2002; NRC, 2001a). Assessment fulfills the "desire to reason from particular things students say, do, or make, to inferences about what they know or can do more broadly" (Mislevy, Steinberg, & Almond, 2003, p. 6).

All assessments are based in a conception or philosophy about how people learn and what tasks are most likely to elicit observations of knowledge and skills from students. They are also premised on certain assumptions about how best to interpret evidence to make inferences (Mislevy et al., 2003). Making explicit this conception of how people learn and the knowledge and skills that are valued under that philosophy is the first key step in designing an assessment that will gather evidence of students' understandings and abilities. Once the focal knowledge and skills have been articulated, the next step is to design situations where students can demonstrate their abilities. Finally, deciding what to do with these observations, or how to interpret them, will provide the information about students' knowledge and abilities. These three steps have been outlined as corners of the assessment triangle: cognition, observation, and interpretation (NRC, 2001b) and are the basis of the idea of evidence-centered design (ECD) of assessments (Mislevy et al., 2003). A quote from Messick (1994, p. 17) functions as a grounding for understanding the principles underlying ECD and for the design of principled assessments:

A construct-centered approach [to assessment design] would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics.

This paper outlines the development and application of a principled assessment system to illuminate middle knowledge associated with the nuances of what students know and can explain about ecological concepts, specifically reasoning up and down food chains.

## METHODS

### Research Context

This work is conducted in association with the *BioKIDS: Kids' Inquiry of Diverse Species* project funded by the National Science Foundation. BioKIDS research focused on the systematic development and evaluation of technology-rich, middle school curricular units focused on the use of cognitive scaffolds to foster complex reasoning in science

(Songer et al., 2005). The curriculum development work (e.g., Songer, 2006), produced three consecutive 8-week science units that encompassed the entire sixth-grade year and that utilized a system of cognitive scaffolds to foster deep conceptual understandings of three consecutive topics: biodiversity, weather, and simple machines. In each unit, the cognitive scaffolds were designed to foster the integration of scientific content, such as biodiversity, with scientific thinking and reasoning skills (sometimes called process skills, NRC, 2006, or science practices, NRC, 2007) such as constructing evidence-based explanations, toward a deep conceptual understanding of the focus science topic.

The main focus of cognitive scaffolds was on the construction of evidence-based explanations. Similarly to others (e.g., Bell & Linn, 2000; Driver, Newton, & Osborne, 2000; Lee, 2003; McNeill et al., 2006; Sandoval, 2003), we developed a modified version of Toulmin's (1958) model of argumentation to support students in creating scientific explanations. We define a scientific explanation as a response to a scientific question that takes the form of a rhetorical argument and consists of three main parts: a claim (a statement that establishes the proposed answer to the question), evidence (data or observations that support the claim), and reasoning (the scientific principle that links the data to the claim and makes visible the reason why the evidence supports the claim). In short, a scientific explanation is a compilation of evidence elicited through observation and investigation and the explicit links those data have to related scientific knowledge (Kuhn, 1989).

While our work valued the development of evidence-based explanations, research suggested that that development of explanations within real-world contexts can be particularly difficult for novice students, particularly when students are expected to draw from discipline-based knowledge to determine salient from irrelevant variables (Lee & Songer, 2003; McNeill et al., 2006). In each of our curricular units, we developed both structure and problematizing curricular scaffolds (Reiser, 2004) that followed a common template to guide students' development of evidence-based explanations using appropriate scientific evidence. Research results illustrated significant achievement gains by intervention students as compared to control populations on all content evaluation, but the largest gains were on items focused on complex reasoning about the focal topic (e.g., biodiversity; Songer et al., 2009).

### **Research Participants**

This study was conducted in the Detroit Public Schools (DPS), an urban district that has a total district enrollment of approximately 183,000 students in 263 schools. The project has a long history of working with students and teachers in DPS. Detroit Public Schools is characteristic of many urban school districts in the United States in that it contains a concentration of students of color, students from low-income families, and students learning English as a second language (e.g., 94% of DPS students characterize themselves as ethnic minorities and more than 70% of students are eligible for free or reduced lunch). This study focused on 318 students in three DPS schools. School and student demographics are available in Table 1.

### **Data Sources/Instruments**

As our curricular unit was designed to foster students' development of scientific explanations in biodiversity and ecology, we needed an assessment system that would allow us to gather information about students' trajectories as they develop competence with developing scientific explanations in these areas. Using the principles of ECD (Mislevy et al., 2003), we created and empirically evaluated an assessment system that allows us to gather this type of information. In this paper, we will provide a summary of the design aspects of the assessment system and then show how this assessment system has allowed us to better

**TABLE 1**  
**School and Student Characteristics**

	School A	School B	School C
Grades <sup>a</sup>	K–8	6–8	6–8
Racial composition <sup>a</sup>	99.9% African American	50% Hispanic 5% African American 45% White	99.9% African American
Qualify for free/reduced lunch <sup>a</sup>	90%	92%	80%
Numbers	Written assessment: 108 Think aloud interviews: 7	Written assessment: 121 Think aloud interviews: 8	Written assessment: 83 Think aloud interviews: 5

<sup>a</sup> Information gathered from <http://www.cepi.state.mi.us/scm/>.

understand some of the nuances of how students think about and explain disturbances in food chains and food webs.

The purpose of our assessment system is to diagnose students' progression in developing more complex understandings of content and ability to create scientific explanations. While these two competencies are highly correlated (Gotwals, 2006), we chose to separate the two dimensions of content and explanations building in our assessment system to better help us create assessment tasks that assessed a range of difficulties and also to help us better interpret students' responses to these tasks. To address both aspects, we created a content-reasoning matrix that lays out three possible levels for each dimension (Gotwals & Songer, 2006). The content-reasoning matrix for the inquiry reasoning skill formulating scientific explanations is shown in Table 2. In creating this matrix, we first classified science content knowledge into three levels: *minimal*, meaning that minimal content is required to complete the task; *moderate*, meaning that students need a solid understanding of the basic underlying scientific concepts (such as knowledge of what a producer is or what a consumer is); and *complex*, meaning that students need not only an understanding of concepts but also need the ability to link different concepts together and understand more complex scientific phenomena (such as understanding about how a disruption at one point in the food web will influence the rest of the food web). There are not firm boundaries between these levels; rather it is more of a continuum of difficulty and amount of knowledge. However, these distinctions can help to create tasks at different difficulty levels and help to provide criteria for creating and categorizing items. Our present work is investigating further the aspects of complexity of scientific content (Songer et al., 2009).

Second, we separated the inquiry reasoning skill of formulating scientific explanations into three levels: *minimal*, *moderate*, and *complex*. We made use of the scaffolding structure from our curricular units and created levels of reasoning tasks based on the degree of support or scaffolding the task provides for explanation formation. *Minimal* tasks provide evidence and a claim, and students simply need to match the appropriate evidence to the claim (or vice versa). While this type of item measures only a low level of inquiry reasoning, specifically the ability to match relevant evidence to a claim (or a claim to given evidence), this is still an important step in students' development of the ability to formulate a scientific explanation. A *moderate* task involves a scaffold that provides students with a hint, telling them the key components of a scientific explanation (a claim, use of evidence, and reasoning). This

**TABLE 2**  
**Content-Reasoning Matrix for Tasks Associated With Constructing Evidence-Based Explanations About a Key Science Topic**

		Scientific Content		
		Minimal Content	Intermediate Content	Complex Content
Scientific Reasoning	Minimal Reasoning	Students are given evidence and claim. <b>Students match relevant evidence to a given claim.</b> Minimal content knowledge is required.	Students are given evidence and claim. <b>Students match relevant evidence to a given claim.</b> To do this, students must apply a moderate amount of additional (not given) content knowledge.	Students are given evidence and claim. <b>Students match relevant evidence to a given claim.</b> To do this, students must apply a substantial amount of additional (not given) content knowledge.
	Intermediate Reasoning	<b>Students construct a simple explanation using prompts or partially completed explanations to guide, for example, the development of a claim and the use of relevant evidence.</b> Minimal content knowledge is required.	<b>Students construct a simple explanation using prompts or partially completed explanations to guide, for example, the development of a claim and the use of relevant evidence.</b> Construction draws on a moderate amount of content knowledge.	<b>Students construct a simple explanation using prompts or partially completed explanations to guide, for example, the development of a claim and the use of relevant evidence.</b> Construction draws on a substantial amount of additional (not given) content knowledge.
Complex Reasoning	Complex Reasoning	<b>Students construct a scientific explanation without any prompts or guidance.</b> Minimal content knowledge is required.	<b>Students construct a scientific explanation without any prompts or guidance.</b> Construction draws on a moderate amount of content knowledge.	<b>Students construct a scientific explanation without any prompts or guidance.</b> Construction draws on a substantial amount of additional (not given) content knowledge in order to, for example, determine salient from irrelevant evidence and to justify claim through scientific reasoning.

involves more inquiry-reasoning ability than the *minimal* task of matching, given that students must construct a scientific explanation, but there is still support to guide them in the important aspects of a scientific explanation. Finally, a *complex* task is the most challenging in that it does not provide support in either the creation of a claim, choice of evidence, or the articulation of reasoning. Students able to successfully complete *complex* tasks demonstrate the ability to not only recognize but also create a scientific explanation.

These two dimensions (content and formulating a scientific explanation) come together in the matrix when one chooses a cell of the matrix to focus on. For example, in the top left shaded box, students would be given both evidence and claim, and they would be asked to match the relevant evidence to a given claim. We would describe tasks associated with this box as providing a high amount of scaffolding, in both content and explanation-building areas. In contrast, tasks associated with “complex” boxes would contain virtually no scaffolding. The bottom right shaded box states, “Students construct a scientific explanation without any prompts or guidance.”

One of the goals of the development of our matrix and associated assessment system was to not only articulate various complexity and scaffolding levels but to map assessment items across the range of boxes. To obtain a suite of tasks mapped to the matrix, we utilized both a reverse engineering of tasks, where we mapped and modified already created tasks to cells of our content-reasoning matrix, as well as a forward design process, where we created new tasks to fill in the gaps. The reverse design process entailed identifying, sometimes modifying, and finally mapping assessment items that had been used on previous assessments or released items from the National Assessment of Educational Progress (NAEP) and the Michigan Educational Assessment Program (MEAP) assessments to the our matrix. Many of these tasks fell into the minimal cells and the complex cells. The intermediate reasoning cells were largely blank. The forward design process entailed utilizing scaffolding formats from our curriculum to design tasks to gather evidence of intermediate levels of student reasoning. Examples of items that address food chain and food web content areas from the shaded cells are in Table 3.

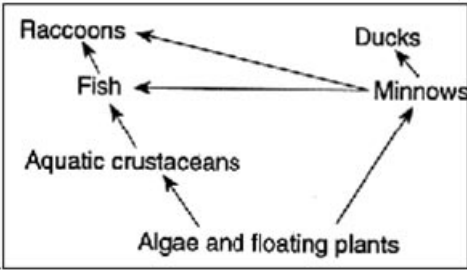
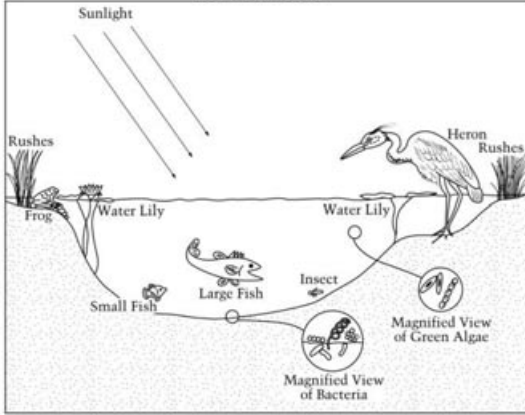
In a pilot study (Gotwals, 2006), we evaluated the predictive nature of the content-reasoning matrix to determine whether item difficulties (calculated using an item response model), as we hypothesized, corresponded to the cell of the matrix that they were mapped to. This study focused on items on the downward diagonal of our matrix—minimal, moderate, and complex for both content and reasoning. The results of the item difficulty analysis demonstrated a pattern in which students found the increasing levels of inquiry reasoning and content more difficult. For example, the simple items tended to have lower difficulty levels than the moderate item, and the moderate items tended to have lower difficulty levels than the complex items. This consistency between the mapped item levels and student interactions with the items shows that the cognitive theory underlying our assessment system is indeed a good predictor of how students interact with the items. In the past, we made educated guesses about the difficulty and appropriateness of our assessment tasks for our students; however, with a suite of tasks based on an articulated theory of cognition (including the important knowledge and reasoning to assess and how this knowledge and reasoning can increase in complexity) and mapped to an assessment framework, like our content-reasoning matrix, relatively accurate predictions can be made about the difficulty levels of items.

### **Dual-Pronged Approach to Gathering and Analyzing Information**

To explore the research questions about how this assessment system could help discover the nuances in how students reason about food web and food chain disturbances, we gathered



**TABLE 3**  
**Sample Assessment Questions at Three Complexity Levels for**  
**“Constructing Evidence-Based Explanations”**

<p><b>Minimal</b></p> <p>2. Which claim is best supported by the evidence found in the food web below?</p>  <pre> graph TD     AC[Aquatic crustaceans] --&gt; F[Fish]     AFP[Algae and floating plants] --&gt; M[Minnows]     M --&gt; F     M --&gt; D[Ducks]     M --&gt; R[Raccoons]   </pre> <p>A. Minnows and fish are producers.          B. Algae and floating plants are consumers.          C. Aquatic crustaceans are producers.          D. Raccoons, fish and ducks are consumers</p>
<p><b>Intermediate</b></p> <p>6. Write a scientific explanation for the following question.</p> <p>Given the food chain: <b>Seeds → Mice → Snakes</b>  <b>Scientific Question:</b> What will happen to the snakes when there are a lot of seeds?  <i>(Make sure your explanation has a claim, 2 pieces of evidence, and reasoning)</i></p>
<p><b>Complex<sup>a</sup></b></p>  <p><b>Write a scientific explanation for the following questions.</b></p> <p>12. <b>Scientific Question:</b> If all the small fish in the pond system died one year from a disease that killed only small fish, what would happen to the algae in the pond?</p> <p>13. <b>Scientific Question:</b> If all the small fish in the pond system died one year from a disease that killed only small fish, what would happen to large fish in the pond?</p>

<sup>a</sup>Both complex items modified from NAEP (<http://nces.ed.gov/NATIONSREPORTCARD/>).

two types of data: written assessments and interviews. We gathered written assessment data from 312 students and conducted think aloud and cognitive interviews with 20 students from three different schools.

**Written Item Analysis.** Students took an identical pretest and posttest. These assessments had 20 items. There were a total of seven minimal, seven intermediate, and six complex items. Of these, we had two items at each of the three levels that dealt with food chains or food webs. For the analysis of the explanation construction items, we coded the items based on the three parts of the explanation—claim, evidence, and reasoning. Claims and reasoning were coded dichotomously (correct or incorrect), and evidence was coded using a partial credit scale where students could receive credit for having correct evidence (2 points), correct but incomplete evidence (1 point), or incorrect evidence (0 points). In addition, we assigned a content code based on whether students had correct content in their response (regardless of whether their response took the form of a scientific explanation). Raters trained using rubrics that were specified for the content of each item. Before coding, interrater reliability was established between two coders at 89%. Subsequent to the coding of all tests, checks were done on 10% of the tests to ensure that raters remained consistent throughout the coding process. Any differences in scores were discussed, and the raters agreed upon a final code.

To examine the intricacies of students' understandings and the difficulty of assessment items on the diagonal of the matrix, we utilized a multidimensional item response measurement model that describes the relationship between students' abilities and the probability of a certain response on an item. This analysis was used to provide information as to whether students interacted with the items in ways that we would have predicted based on which cell of our matrix we had mapped the item. A simple Rasch model (Rasch, 1960) includes one-person ability parameter and one item difficulty parameter in its formulation. Models within the Rasch family can be articulated using the random coefficients multinomial logit model (RCML) formulation (Adams & Wilson, 1996). This model can be represented as shown below where  $\Pr(X_i = j)$  represents the probability of a response  $j$  to an item  $X_i$ .

$$\Pr(X_i = j) = \frac{\exp(b_{ij}\theta + a'_{ij}\xi)}{\sum_{k=1}^{K_i} \exp(b_{ik}\theta + a'_{ik}\xi)}$$

where  $b_i = (b_{i1}, b_{i2}, \dots, b_{ik})$  is the scoring vector,  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$  is a vector of  $n$  free parameters, and  $a'_{ik}$  denotes the linear combinations for  $i = 1, \dots, I; k = 1, \dots, K_i$ .

For this study, we used the multidimensional random coefficients multinomial logit model (MRCML; Adams, Wilson, & Wang, 1997), which is a multidimensional extension of the random coefficients multinomial logit model (Adams & Wilson, 1996). The MRCML model can be applied to multidimensional polytomous test items. Mathematically, the differences between the RCML and MRCML models are that (1) the ability parameter is a vector with an ability parameter estimated for each dimension, and (2) the scoring, free parameter, and design vectors become matrices. The MRCML therefore allows for estimation of multiple abilities.

Since the coding of our items included multiple codes for the same item (a claim, evidence, reasoning, and content code for all items), we had to account for the assumption of local independence of items in all Rasch models (1960; that a student's probability of getting one item correct is dependent only on their ability level and the difficulty of the particular item, not on their ability to respond to other items). Thus we utilized item bundling for items that were coded for multiple abilities, such as the claim, evidence and

reasoning codes, and content codes, since these codes are likely dependent on each other (it is probable that knowing the content will influence a students' ability to make a correct claim; Wilson, 1988; Wilson & Adams, 1995). There is always a trade-off in bundling items since some information is lost when we combine scores (Wilson & Adams, 1995). However, we utilized two levels of bundling (in addition to an unbundled situation) and examined each model for its fit to the data. The first level of bundling combined the aspects of building an explanation, summing the claim, evidence, and reasoning scores to create a total explanation score but kept the content code separate. We then used the explanation "superscore" and content codes in further analyses. The second level of bundling created a total item code, combining bundles of all codes relevant for a given item (including explanation superscore and content codes). For example, if an item measured content knowledge and explanation ability, we summed the content score with the explanation superscore (obtained from the first-level bundling analysis) to get a total item score. An item bundle model nested in a multidimensional random coefficients multinomial logit model (Adams et al., 1997) provides a way to handle multiple dimensions and local dependence simultaneously. This is done by carefully modeling the expected patterns of dependence, based on our substantive knowledge about the structures and the demands of the tasks.

The main areas that we focus on for this study are content knowledge and ability to formulate scientific explanations; however, many items also required students to interpret data. Therefore, we used a three-dimensional model to account for these three types of abilities (content knowledge, formulating a scientific explanation, and interpreting data) that students needed to respond to the items on the test. A comparison of the three-dimensional model to a two-dimensional model and a unidimensional model indicates that the three-dimensional model fits the data the best. In addition, in all three types of bundling situations described above (no bundling, first-level bundling, and second-level bundling), the data were well fitted by a three-dimensional model, with all item and student fit statistics falling between 0.75 and 1.25 (which according to Bond and Fox, 2001, indicate good fits to the model). In addition, an examination of the test characteristic curves for these models indicated our test provides good information for students with ability levels both below and above average (between  $-3$  and  $+3$ : a range of ability level into which almost all of our students fell). This means that the test provides adequate information about all students who took this test.

**Interview Analysis.** In addition to the examination of written assessments, the first author also conducted interviews with 20 low-, medium-, and high-performing students (using teachers' reports of students' abilities) after they completed the curriculum. The interviews had two parts: a think-aloud section and a cognitive interview section. Common think aloud procedures were used to examine students' thought processes as they worked on the assessment tasks (Ericsson & Simon, 1993). After being instructed about the thinking aloud procedure, the interviewer modeled how to think aloud on one practice problem. Then the student practiced thinking aloud on a second practice problem. Following the practice, students thought aloud as they completed the assessment. The interviewer did not interact with the student as he or she completed the assessment except to remind the student to keep talking or to speak louder. After the student completed the assessment, the interviewer went back over the assessment with the student asking the student to clarify responses on items, to explain how they reasoned about an item, and/or and to talk about their perceptions of the items, for example, which they found difficult or easy.

Following standard procedures (DeBarger, Quellmalz, Fried, & Fujii, 2006; Ericsson & Simon, 1993), interviews were transcribed, segmented, and coded the responses for

evidence of content and reasoning. We analyzed the coded data for presence of the focal content ideas and we examined the data for the types and level of reasoning that students employed when interacting with the tasks.

## RESULTS

### Written Assessments

To better understand how students were reasoning about ecological content knowledge, we examined the characteristics of each of the items and how students responded to these items. To do this, we examined the difficulty parameters of the items from the item response model. The Item Response Theory (IRT) difficulty continuum is set up as a logit scale with 0 set as the mean of the item difficulty parameters. Each integer above 0 represents items that are more difficult, and each integer below 0 represents items that are less difficult. On the basis of the cognitive underpinnings of our matrices and our prior evaluation of our matrix and its ability to help us gather a range of item difficulties, we predicted that the empirical difficulty analysis of our suite of items would find that the complex items have the highest difficulty parameters, the intermediate items have moderate difficulty parameters, and the minimal items have the lowest difficulty parameters.

Table 4 presents the items, their mappings onto the downward diagonal of our content-reasoning matrix, the total item difficulty parameter (using the second level of item bundling to give a single difficulty score to an item), and the item fit. As seen in Table 4, for the most part, our hypothesis of the empirical difficulties and their mapping to the matrix held up. Minimal items tended to have the lowest difficulty parameters, intermediate items tended to have moderate difficulty parameters, and the complex items tended to have the highest difficulty parameters. However for two items (#6 and #13, highlighted in the table), the

**TABLE 4**  
**Model Parameter Estimates for Total Item ( $N = 312$ )**

Item	Difficulty ( $b$ )	Weighted Mean Square fit
1 (minimal)	-0.769	0.95
2 (minimal)	-0.211	0.97
3 (intermediate)	0.291	0.98
4 (intermediate)	-0.338	0.96
5 (minimal)	-0.869	0.97
5d (complex)	1.239	1.23
6 (intermediate)	0.781	0.99
7 (intermediate)	0.266	0.97
8 (complex)	1.160	0.86
9 (minimal)	-1.345	0.98
10 (minimal)	-1.246	0.92
11 (complex)	2.032	1.23
12 (complex)	1.203	0.89
13 (complex)	0.381	1.04
14 (intermediate)	-0.155	0.98
15 (intermediate)	0.242	1.07
16 (minimal)	-0.375	1.00
17 (intermediate)	0.389	0.96
18 (minimal)	0.039	0.97
19 (complex)	1.068	1.25

empirical data did not match our predictions. Item 6 (presented in Table 2) is an intermediate item but had a difficulty parameter above where we would expect it. The difficulty parameter for item 6 ( $b = 0.781$ ) was higher than any other intermediate item and was slightly closer to the average difficulty of complex items ( $b_{\text{avg}(\text{complex})} = 1.135$ ). Item 13 (presented in Table 1) is a complex item, but had a difficulty parameter that fell much below where we would have expected it. The difficulty parameter for item 13 ( $b = 0.381$ ) was more similar to intermediate items ( $b_{\text{avg}(\text{intermediate})} = 0.211$ ) than the complex items.

Even more perplexing than the below predicted difficulty level of item 13, is that items 12 and 13 (both presented in Table 1) have to do with the same scenario of a disease killing all of the small fish in the pond ecosystem. However, item 12 has a high difficulty level (what we would expect for a complex item) at  $b = 1.257$ . Both questions involve predicting and explaining what would happen given this disturbance to the pond ecosystem. Item 12 asks what would happen to the algae, and item 13 asks what would happen to the large fish. To answer both of these questions, students must understand the dynamics of a pond ecosystem and the food web interactions involved and be able to construct a scientific explanation (a claim about what would happen in this scenario, evidence having to do with what organisms eat and are eaten in this system, and reasoning having to do with the effects of disturbances to food chains and food webs) without any scaffolding for creating a scientific explanation. Despite being seemingly similar in both content and scaffolding of explanation formation and with such similar cognitive skills involved, it is not easy to reason why these two items have such different difficulties.

To further examine the differences between items 12 and 13 (as well further examine the intricacies of item 6), we looked at the first-level bundled item response model, where we can separate out the difficulty of the items due to their content and the difficulty due to construction of a scientific explanation (with a claim, evidence, and reasoning). While, according to the model for the overall test, these dimensions of item difficulty are highly correlated ( $r = 0.862$ ), there are some differential components of these items that we can examine to determine which part of the items caused there to be such a difference in overall difficulty level. Table 5 shows the difficulties of these items due to the two dimensions (please note that because this is a different model, the difficulty parameters are on a different scale than the model above).

In item 6, the item that asked what would happen to the snake if there were a lot of grain, the content dimension of the item had a higher difficulty ( $b = 0.628$ ) than the explanation component ( $b = 0.136$ ) of the item. The explanation difficulty dimension of this item was the lowest of all three items, however, not by a large margin. This was an item that provided scaffolding hints about what to include in an explanation and that may have influenced the difficulty of this dimension of the item. Since the explanation component of the item was

**TABLE 5**  
**Content and Explanation Difficulty Estimates for Items 6, 12, and 13**  
**( $N = 312$ )**

Item	Latent Ability	Difficulty ( $b$ )
6	Content	0.628
	Explanation	0.136
12	Content	0.796
	Explanation	0.362
13	Content	-0.781
	Explanation	0.267

not very difficult, it appears that something having to do with the content provided difficulty for students in this item.

For item 12, the item that asked what would happen to the algae, the content dimension of the item had a higher difficulty ( $b = 0.796$ ) than the explanation dimension ( $b = 0.362$ ) of the item. Conversely, in item 13, the item that asked what would happen to the large fish, the content dimension of the item had a lower difficulty parameter ( $b = -0.781$ ) than the explanation dimension ( $b = 0.267$ ). In fact, the difficulty dimensions of the explanation component of the items were similar, but the content dimension difficulties of the items were quite far apart. This would seem to indicate that students are better able to reason “up” a food chain (that if small fish die, big fish will not have food and will die) rather than “down” a food chain in item 10 (if small fish die, nothing will eat the algae and it will grow more quickly because it is missing a predator). However, despite the content difficulties being quite different, students found backing their proposed claim with evidence and reasoning only slightly more difficult in item 12 than in item 13.

### Think Aloud Analysis

To gain more insight into how students reasoned through these items, we examined our think aloud and cognitive interviews with 20 students. We compared students’ responses to the think alouds and cognitive interviews with their ability (expected a posteriori, EAP) estimates for each dimension (content and explanation) given by the modeling of written responses. Overall, the results of the think aloud and cognitive interviews correlated with students’ ability estimates from the written assessments. Students who had higher ability estimates tended to have high-level verbal responses to the items—using the correct content to formulate a claim, back it up with evidence, and use scientific principles to tie the two together. Next, we examine the think aloud and cognitive interview responses for the food web questions for six students. The students’ EAP estimates appear in Table 6.

**Item 6.** In item 6, students were given a food chain and asked what would happen to the secondary consumer (snakes) if there were a lot of grain (a producer). Most of the students demonstrated understanding that if there were a lot of grain, there would be more food for the mice and so it would allow for the mice population to grow; and with more mice, there would be more for the snakes to eat and thus the snake population would increase. For example, Student A, who had well above average ability estimates on both content and explanations, said,

The snakes would get more and more. Snakes eat the mices and there will be more mices because they eat the grain and there is more grain so the snakes is going to have more to

**TABLE 6**  
**Six Students’ EAP Estimates on Content and Explanation Dimensions**

Student	Content EAP	Explanation EAP
A	1.323	1.076
B	-0.235	-0.1966
C	0.169	0.266
D	0.024	0.541
E	-0.681	-0.381
F	1.023	0.753

eat and then they can just get more and more. More food for one animal will mean more food for another animal. . . sometimes. . .

However, of the 20 students interviewed, four students responded that there would be no effect on the snakes. For example, Student B, who had slightly below average ability estimates on both dimensions, said, “The snakes would just keep living in the grain. They eat the mice and they is the same so the snakes just keep on living.” This student was correctly able to read the food chain and see that the snakes eat mice and use this as evidence in her response, but she was unable to make the connection between an increase in the grain and the influence of this on a secondary consumer, the snakes, and therefore her claim was incorrect. Student C, who had slightly above average ability estimates for both content and explanations, said, “The snakes wouldn’t care ’cause they are carnivores and eat mice. . . .” This student showed that he knew the vocabulary of carnivore and was able to read the food chain to provide evidence that the snake ate the mice, but similarly to Student B, was unable to see how an increase in grain would affect the snakes. The other two responses were similar to Students B and C, where they were able to show some indication that they could read the food chain and knew that snakes ate mice, but were unable to see how a change in producers (the grain) would influence the snakes who do not directly eat the grain.

In the cognitive interviews, which occurred *after* the think alouds, the interviewer probed students to further explain their verbal think-aloud responses to items. When probed about this question in the cognitive interview, the students who did not make the connection between an increase in grain and the snakes continued to provide evidence of this lack of connection. For example,

Interviewer: So, tell me more about how you thought about this problem.

Student B: I thought that snakes eat the mice and so they would just keep eating the mice.

Interviewer: So, what would happen here. . . umm. . . if there was more grain?

Student B: There would just be a lot of grain and the snakes could just . . . ummm . . . go around in it.

Students who struggled with this item did not appear to struggle reading the food chain. In fact, they were able to use their ability to read the food chain to provide evidence of snakes eating mice in their responses. However, students were not able to see how an increase in the grain (a producer) could influence the snakes that are two trophic levels above the grain.

**Item 12.** In item 12, students were asked, given the scenario presented of a disease killing all of the small fish, “What would happen to the algae?” In the difficulty analysis above, this complex item had a difficulty level ( $b = 1.203$ ) that we would have expected given its mapping to our content-reasoning matrix. Most students demonstrated an understanding that if the small fish were gone, then they would not eat the algae and the algae would increase. For example, Student A, who was above average on both dimensions, said, “The algae will increase. Small fish eats it and the disease got rid of the small fish, so it will just get more.” However, several students appeared to be unclear about either what algae are and/or the directionality of the food chain, saying that the algae would have nothing to eat. In fact, of the 20 students who were interviewed, five students possessed some kind of confusion about what algae were and/or the directionality of the food web. For example, Student B, who had below average ability estimates on both dimensions, said, “It would not have any food to eat. . . Algae eat, mostly eats everything in the pond or water and

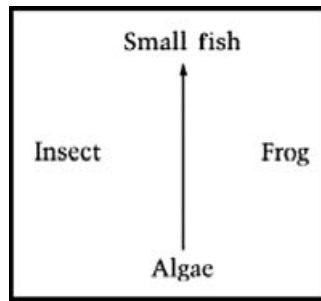


Figure 1. Partial food web provided.

stuff.” Another student (Student D, who had slightly above average ability in content and above average ability in formulating explanations) misinterpreted the word algae, perhaps by reading too quickly and said,

If all the small fishes in the pond system died one year from a disease that killed only the small fish, what would happen to the alligator in the pond? . . . I don’t think like the cause if all the small fishes die, I wouldn’t think the alligator would die either because it probably could eat other things other than the small fish. . . The alligator. Wait. Yeah that. That probably like can eat other things more then like the small fishes. It probably wouldn’t just eat small fishes because if the small fishes die, it would probably find something else to eat.

Many students who possessed the mistaken belief about the direction of the food chain or about what type of organism algae was, however, continued with this mistaken identification of what algae was and proceeded to create a coherent (if often incomplete) explanation as if the algae was a consumer of small fish, showing that they knew what was required to create a scientific explanation, but they just did not have the content knowledge to come up with an accurate prediction. For example, Student C, who had slightly above average ability on both dimensions, said, “Algae will all die. They have nothing to eat since it eat small fish and they are all gone from the pond. Or it will get sick and die from eating infected fish. . . .” So this student was able to create a claim, while incorrect, that algae will die and provide evidence, again incorrect, that the algae eat small fish and reasoning that the small fish were gone from the pond. Thus, while the content of the response was incorrect, the structure of the explanation was coherent.

What was consistent in the five responses of students who did not have the content correct, was that they thought that algae ate small fish and would not have anything to eat. It is possible that if students knew what algae was, they would realize that it is a producer and therefore would not eat anything, including the small fish. However, even if students did not know what algae was, if they used the food web provided to them in an earlier item that had one arrow drawn in for them (from algae to the small fish; see Figure 1), and understood what that arrow meant, they would have had the information that the energy from algae is passed to small fish and that algae do not eat small fish. While the food web is not traditional evidence, like data from observations or experiments, it does provide students with information to make a claim or prediction about what would happen if there was a disturbance in the ecosystem. Thus, for our purposes, we count the food web and the information that it contains as evidence for students to make predictions about ecosystem disturbances.

In the cognitive interviews, the interviewer probed students to further explain their think aloud responses to items. Even when probed about this question in the cognitive interview,



these five students stuck with their response that the algae eat small fish. Two of the five students who got this answer incorrect provided some evidence that they misinterpreted the food web picture (Figure 1). For example in this exchange with Student D, who had about average content ability, but above average ability to create explanations,

Interviewer: So, tell me more about how you thought about this problem.

Student D: The algae is gonna die from not having food.

Interviewer: How do you know that the algae don't have any food.

Student D: It eat the small fish and the small fish is dead from the sick . . . umm . . . disease

Interviewer: How do you know that the algae eat the small fish?

Student D: From the picture [points to the arrow in the food web].

Student D did not have the content knowledge (or ability to interpret the information presented in the food web) to correctly respond to this item. However, he was able to utilize evidence of what he thought ate what in his response. This is consistent with his above average ability to create evidence-based explanations but his lower content knowledge ability.

The other three students who got this question incorrect, however, did not reference the food web when speaking about their response to this question, but still continued to reason as though the algae ate small fish. Student B, for example, when probed about this item, said, “. . . algae eats small fish and everything, so it will have to eat other stuff or die. That's it.” When probed further about how she knew that the algae ate small fish, Student B said, “I just knows that algae eat small fish and so that is how I know this question.” Thus, it is unclear where the student got this information from—the food web picture or from some other source. Other students, similarly, were not specific in how they knew that algae ate small fish. Student E who had below average content and explanation abilities, even said, “I don't know how I know this, I just do.”

Even students who responded correctly to this question, sometimes began with a mistake of what algae was or the directionality of the food chain but corrected themselves. For example, Student F, who had above average content and explanation abilities, said,

. . . What would happen to the algae. . . I think they'd all die too. They would try to eat something else because they like. . . . No. . . wait a minute. No they live. The algae would live and they live more because they get eaten by the small fish then if there's no small fish then they'd be alive. They'd be more of them too. They'll be abundant, higher abundance of them algae leaves in the pond. . . . So. . . if all the small fish die then the algae be more abundant. . . they. . . because they don't get eaten by the small fish . . . the food chain at work.

This student first had to work out his understanding of the directionality of the food chain, however, once he did this, he was able to apply his understanding of what would happen if there were a disturbance in the food chain/web to another organism in this system, which is consistent with his above average abilities on both dimensions.

Many students who did get the question correct began with defining what algae is. For example, one student began thinking about this question by saying, “. . . Algae is plants. . . so if the small fish died. . . .” In examining the responses to these items, it seemed that many students, like this one, thought out loud about the content and what algae was before attempting to formulate their explanation. This shows that first the students drew on content knowledge, that algae are plants, and after establishing that fact, proceeded to formulate their explanation. Perhaps, this signals that when some students are interacting with items that require them to draw on complex content they first ensure that they understand the content

knowledge that they must utilize and then proceed with creating a scientific explanation to answer the scientific question.

**Item 13.** Item 13 asked students, given the same scenario, “What would happen to the large fish?” Item 13, which was a complex item, had a difficulty parameter much below what we would have expected given its mapping on our content-reasoning matrix ( $b = 0.381$ ). Unlike item 12, with the 20 students interviewed, there were no instances of mistaking the directionality of the food chain or the type of organism that large fish are. All students understood that the large fish ate the small fish and used this content knowledge to formulate a scientific explanation. Students did not always create complete explanations, though. For example, Student B, who had below average abilities on both dimensions, said, “The large fish would die because it has nothing to eat.” While this is correct in terms of content, the student did not articulate the evidence that he used (that large fish eat small fish), which is consistent with this question having easier content (so even a student with lower content ability could get it correct), but that this student still could not formulate a full evidence-based explanation. Many students, however, did create a claim statement and utilize evidence of the directionality of the food chain and also provide evidence and reasoning to back up their claim. For example, Student F, who had above average abilities on both dimensions, said,

... What will happen to the large fish... they will die. If all the small fish die, the large fish will die too. They will die because they eat the small fish and there is no more small fish 'cause they died... they will run out of food... that's how the food chain... uhh... web... works.

This student created a claim statement, “If all the small fish die, the large fish will die too.” This claim was followed up by the use of evidence that the large fish eat the small fish and reasoning, that without small fish, the large fish will not have anything to eat. In this item, students did not struggle at all with the directionality of the food web or with the affect that removing the small fish would have. While some students struggled to create complete explanations, all students provided evidence of they all were able to at least create a claim statement with the correct content indicating that the large fish would decrease, die, or starve.

## WHAT CAN ASSESSMENT SYSTEMS HELP US UNDERSTAND ABOUT STUDENT EXPLANATIONS?

While our research results demonstrate significant improvements among sixth-grade intervention students in their ability to create scientific explanations in ecological content areas such as food chain and food web disturbances (e.g., Songer et al., 2009), these analyses indicate that some students still struggle with aspects of reasoning about food chains and webs such as how to interpret the directionality of food chains or webs, the influence of a change in a population more than one trophic level away, and the identity of algae. In addition, even when students did understand the content, creating a complete scientific explanation with a claim, sufficient evidence, and reasoning was often still difficult. More specifically, many students demonstrate intermediary or middle knowledge when asked to formulate a scientific explanation about food chain and food web disturbances, where parts of their response are correct, but either pieces of the content or the structure of the scientific explanation are not fully accurate. Perhaps more important, however, is the mechanism for discovering these findings, a principled assessment system that allowed us to predict items’

difficulties and when the difficulties were not what we predicted, investigate these findings further. Next we discuss the findings about students' reasoning around food chains and food webs and then we discuss the importance of having an assessment system that allows for nuanced cognitive analysis of student responses, specifically an examination of students' middle knowledge.

### **Reasoning Up and Down: A Food Chain or Web**

Similar to Gallegos and colleagues (1994), we found that some students confused the meaning of arrows in a food chain or food web and interpreted the arrows as going from predator to prey. From classroom observations, we have found that many students often will draw the arrows incorrectly, but when asked to talk about the food web or food chain, they can tell you at least "what eats what" even if they cannot state that the arrows represent energy transfer in a system. Often teachers who implemented our curriculum used the trick that the "arrow's mouth" is open toward what is getting eaten. However, when students are presented with an unfamiliar organism, such as algae, the interpretation of the arrows becomes more confusing. The arrows do not have as much meaning when it is not clear which organism is the predator and which is the prey. For example, Student B used the arrow going from algae to small fish as evidence for the claim that algae eat small fish. But, in item 6, Student B was correctly able to read the food chain and state that snakes eat mice. So, even after being taught the scientific meaning of the arrow (energy transfer) and the "arrow's mouth" trick, this student (and likely others) still did not understand what the arrow represented (perhaps especially when the organisms were not familiar) and could not use the arrows to appropriately interpret food chains and webs. It is unclear whether algae were an organism that students were more familiar with (or if students were given the information that algae were producers or plants) they would have had the same trouble. This indicates that much of students' middle knowledge is "messy" such that they are able to accurately interpret representations in some circumstances, but not all.

Item 13, which was mapped as a complex item for both content and explanation ability, had a lower than predicted difficulty parameter, despite being very similar to item 12. Students did not have difficulty reasoning that large fish ate small fish. While in the interviews, some students did not provide a full explanation with sufficient evidence to back up their claim, all students were able to state that the large fish would decrease, die, or starve and some students created very complete explanations of the situation. It appears that students who completed the curriculum, especially when familiar with the organisms, were easily able to reason about the affect of a disturbance in a prey source for predators.

However, even in this task, many students may have understood the content knowledge but were not able to apply the content knowledge in the creation of a scientific explanation. This again illustrates a type of "messy middle" where students may have some pieces of knowledge and ability to respond to complex science tasks but not all of the pieces. Thus, while it may be relatively clear to illustrate an end point of what we want students to be able to do (e.g., create a scientific explanation about disturbances to a food web), and it is clear when students do not have any knowledge of the content, defining the pathways or middle knowledge that students may have as they move from a beginning point to the end point is much trickier given that the middle is often messy and may not be the same for all students.

Items 6 and 12 had similar content difficulty levels. However, the way in which students responded to the items and the probable cause of the difficulty of the items was very different. In item 12, students struggled with the definition of algae and some students did not interpret the arrow in the food web correctly. It is unclear whether students would have

demonstrated the same difficulty in determining the effect of the decrease in a predator on its food source if algae had been a more familiar organism. In item 6, some students struggled with reasoning about how a change in one level of a food chain would influence another level of the food chain that it was not directly connected to. While we hypothesized that this item would have had an intermediate difficulty level given that it dealt only with a food chain and not a food web and because we provided explanation scaffolding hints, this item had a higher than predicted difficulty level. Many students were still able to create claim statements and provide some accurate evidence that snakes ate mice, but were not able to make the connection to see that an increase in the mice's food would have an influence on snakes. Thus, in both of these situations, students were often able to make claims and provide evidence (even if incorrect or insufficient) to back up their claims, but the content aspect of the question posed the most difficulty for them. Once again, the finding that some students were able to pull together pieces of a scientific explanation even when they did not have the accurate science content points to a different type of messy middle knowledge that students may have as they move toward a high level of sophistication in reasoning about complex ecological scenarios.

These results suggest that students analyzed food webs differently depending on the trophic level disturbed and whether or not the populations were directly connected in a predator–prey relationship. In the pond ecosystem questions, students found reasoning up the food chain about how a change in a prey population (small fish) would influence the predator (large) fish easier than reasoning down the food chain to see the influence on the algae. However, this may have been due to the unfamiliarity of algae itself rather than an inherent nature of one way of reasoning in a food chain being more or less difficult. In item 6, several students did not see how a change in the producers (grain) would influence the second-order consumers (snakes) as these populations were not directly connected together. In addition, all of the findings point toward a “messy middle” where many students, even after instruction, may have pieces of knowledge and skills but cannot integrate them to be able to create a sophisticated response to complex tasks. Depending on the situation, this messy middle may look different for different students.

These findings have implications for curricular design and teaching. Developing opportunities for students to work more closely with the representations of food chains and food webs and seeing the effects of disturbances to ecological systems might help students better able to reason about these situations. Since these situations are difficult for students to directly experience, having simulations that allow students to “experience” working with this content might provide powerful learning opportunities. One such technological system that includes aspects of embedded assessment activities as well as a diagnosis feature is the SimScientists food web construction project from WestEd (Quellmalz & Pellegrino, 2009). This system holds promise because of its ability to allow students to manipulate ecological systems and have them see the implications in many representational forms (graphs, charts, pictures, and so on).

In addition to having more opportunities for students to experience these phenomena through simulations, scaffolding students in working with the representational forms of food chains and webs could allow students to better explain what would happen in ecological disturbance situations. Helping students see arrows as representing matter and energy transfer is a key component of this. For example, given the arrow connecting algae to small fish, even if students did not know what algae was, they could then reason about what was happening in this situation. Newer versions of the BioKIDS curriculum provide explanation scaffolds that guide students in the structural aspects of explanations and content scaffolds such as hints that remind students of what arrows represent and to think about the whole ecological system (rather than just the immediate predator–prey relationship). In addition,

the latest version of our assessment system adds a level of content scaffolding to gather more information about how students' middle knowledge develops. These scaffolds are important to be able to tease apart what knowledge students do and do not have to give us a better picture of the possible messy middles that students may have as they develop more expertise in a topic.

### **Assessment for Cognitive Research**

The first step in creating any assessment should be to fully articulate the theory of cognition that guides the knowledge, skills, and abilities that are of interest to the assessment designer (AERA, APA, & NCME, 1999; Mislevy et al., 2003; Pellegrino, Chudowsky, & Glaser, 2001). Following the articulation of the cognitive theory guiding the assessment, other aspects of the assessment should be made explicit, such as how to design tasks to allow students to demonstrate the desired knowledge, skills, and abilities and how to interpret the observations of student work. An assessment plan that links all of these steps facilitates the types of cognitive research that provide insights into the intricacies of how students understand content and reason about scientific scenarios. The assessment system described in this paper illustrates one method for gathering this type of information.

Having items mapped to the content-reasoning matrix with a planned interpretive framework provided us with three main affordances. First, we were able to systematically design items that would allow students to demonstrate a range of knowledge and ability levels. Developing the ability to formulate scientific explanation in complex scientific scenarios takes time, repeated exposures, and guidance (Songer et al., 2009). Having items that target different levels in this development acknowledges the time, it takes to develop reasoning skills in science and shows that while we hope that all students are able to independently create scientific explanation by the end of our curricular unit, we realize that there will be students who still need support in creating their explanations. In addition, giving students opportunities to show us what they are and are not able to do helps us to illuminate the messy middles that students may have as they develop more sophisticated explanation abilities. Having a suite of items to assess a range of knowledge and ability levels ensures that there are items that are well targeted to students at many different places in their learning progression. This suite of tasks at different levels can help both us and teachers diagnose how far students' learning has progressed in terms of both content knowledge and abilities to formulate coherent scientific explanations and teachers can use this knowledge as the plan further instruction.

The second affordance of this system has to do with the predictive power of the matrix. Having a suite of items mapped to specific cells of the content-reasoning matrix allowed us to make predictions about how and in what way items are difficulty in terms of both the complexity of content involved and the degree of scaffolding for the creation of scientific explanations. When our predictions did not hold up, we were able to go back and reexamine our interpretations of students' responses to discover where and why our prediction was off, which allowed us insights into some of the nuances of students' messy middle knowledge. Assessment is a process of reasoning from evidence (NRC, 2001b). Gathering the best evidence possible is important if we are to use this evidence to diagnose students' abilities and use this evidence to help them learn. Discovering exactly what evidence items can provide about students' knowledge and abilities is essential if we are to use this evidence to make claims about what students know and can do and use this information to help students and teachers develop more complex abilities in science.

Finally, the third affordance of this system is related to the interpretation methods used to make sense of students' responses. Recognizing that many of our assessment items

required multiple types of knowledge and abilities to respond correctly led us to utilize multidimensional item response models to find patterns in students' responses to the items. In addition to providing information about the overall difficulties of items, these models provided information about the difficulties of different aspects of the items, specifically, the content knowledge and the explanation construction dimensions. This information illustrated that the main difference in difficulties between items 12 and 13 was in the content dimension of the items. Having this information is helpful in thinking about how students think about certain scientific situations and what aspects of problems are difficult for them. Also helpful for our interpretation was the dual-pronged approach used for gathering information about how students responded to items. Using both the item response models for seeing big picture patterns of responses as well as the think aloud and cognitive interviews to gather a richer picture of what students knew relative to the items as well as the processes that they used to respond to the items was helpful in seeing the full picture of how students reason in scientific situations. In particular, these complementary methods allowed for a rich unpacking of some aspects of students' messy middles. Overall, having an assessment plan that allows the cognition, observation, and interpretation corners of the assessment triangle (NRC, 2001b) to be fully articulated and connected is key in gathering complex evidence of student understanding.

This material is based in part upon research supported by a Spencer Dissertation Fellowship. The authors gratefully acknowledge Geneva Haertel, Robert Mislevy, Philip Myers, and the entire BioKIDS and PADI research teams for their support. The authors also thank the Detroit Public School teachers and students for their fascinating learning processes, insights, and guidance.

## REFERENCES

- Adams, R. J., & Wilson, M. R. (1996). Formulating the Rasch model as a mixed coefficients multinomial logit model. In G. Engelhard & M. W. Wilson (Eds.), *Objective measurement: Theory into practice* (Vol. III, pp. 143–166). Norwood, NJ: Ablex.
- Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1–23.
- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Alexander, S. K. (1982). Food web analysis: An ecosystem approach. *The American Biology Teacher*, 44, 190–198.
- Barman, C. R., & Mayer, D. A. (1994). An analysis of high school students' concepts & textbook presentations of food chains and food webs. *The American Biology Teacher*, 56(3), 160–163.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797–817.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in human sciences*. Mahwah, NJ: Erlbaum.
- DeBarger, A. H., Quellmalz, E., Fried, R., & Fujii, R. (2006). Examining the validities of science inquiry assessments with cognitive analyses. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Gallegos, L., Jerezano, M. E., & Flores, F. (1994). Preconceptions and relations used by children in the construction of food chains. *Journal of Research in Science Teaching*, 31(3), 259–272.
- Gotwals, A. W. (2006). *The nature of students' science knowledge bases: Using assessment to paint a picture*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.
- Gotwals, A., & Songer, N. B. (2006). Measuring students' scientific content and inquiry reasoning (pp. 196–202). In S. Barab & K. Hay (Eds.), *The Proceedings of the International Conference of the Learning Sciences*. Mahwah, NJ: Erlbaum.

- Hogan, K. (2000). Assessing students' systems reasoning in ecology. *Journal of Biological Education*, 35(1), 22–28.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674–689.
- Leach, J., Driver, R., Scott, P., & Wood-Robinson, C. (1996). Children's ideas about ecology 3: Ideas found in children aged 5–16 about the interdependency of organisms. *International Journal of Science Education*, 18, 129–141.
- Lee, H.-S. (2003). Scaffolding elementary students' authentic inquiry through a written science curriculum. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.
- Lee, H. S., & Songer, N. B. (2003). Making authentic science accessible to students. *International Journal of Science Education*, 25(1), 1–26.
- McNeill, K., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153–191.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62.
- Mislevy, R. J., Wilson, M. R., Ercikan, K., & Chudowsky, N. (2002). *Psychometric principles in student assessment*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- National Research Council. (2001a). *Classroom assessment and the National Science Education Standards*. Washington, DC: Author.
- National Research Council. (2001b). *Knowing what students know: The science and design of educational assessment*. Washington DC: National Academy Press.
- National Research Council. (2006). *Systems for state science assessment*. Washington DC: The National Academy Press.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington DC: The National Academies Press.
- Organisation for Economic Co-operation and Development (2007). *PISA 2006: Science competencies for tomorrow's world Volume 1: Analysis*. Paris, France: Author.
- Organisation for Economic Co-operation and Development (OECD) (2009). *Green at fifteen? How 15-year-olds perform in environmental science and geoscience in PISA 2006*. Paris, France: Author.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323, 75–79.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Reiner, M. (2001). Conceptual classroom environment—A system view of learning. *International Journal of Science Education*, 23(6), 551–568.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5–51.
- Smith, L., & Anderson, C. W. (1996). Alternative student conceptions of matter cycling in ecosystems. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, San Francisco, CA.
- Songer, N. B. (2006). BioKIDS: An animated conversation on the development of curricular activity structures for inquiry science. In R. Keith Sawyer (Ed.), *Cambridge handbook of the learning sciences*. (pp. 355–369). New York: Cambridge University Press.
- Songer, N. B., Huber, A. E., Adams, K., Chang, H. Y., Lee, H. S., & Jones, T. (2005). BioKIDS: Kids' inquiry of diverse species, an eight-week inquiry curriculum using simple, powerful technologies. Ann Arbor: University of Michigan.
- Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning in biodiversity. *Journal of Research in Science Teaching*, 46(6), 610–631.
- Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.
- Wilson, M. (1988). Detecting and interpreting local item dependence using a family of Rasch models. *Applied Psychological Measurement*, 12, 353–364.
- Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, 60, 181–198.