Paper presented at the Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA

ASSESSING STUDENTS' PROGRESSING ABILITIES TO CONSTRUCT SCIENTIFIC EXPLANATIONS

Abstract: Assessments must be thoughtfully constructed in order to gain empirical evidence about the validity of a hypothesized learning progression. This paper describes how our project has used a definition of learning progressions as including both content and inquiry reasoning as a template and translated this into an assessment system for gathering empirical evidence of students' abilities to formulate evidence-based explanations about biodiversity and ecology. Specifically, this paper outlines distinguishing characteristics of tasks used to gather evidence of students' learning, what makes these tasks difficult (including task format as well as content and context) and how we use our learning progression templates to score responses to these tasks. All tasks are matched to particular locations on both the content and the inquiry reasoning progressions. Task design is followed by empirical data collection, both cognitive interviews and item response modeling, then empirical results are used in an iterative redesign process. This paper describes how the design and evaluation of items of different levels of complexity matched to different locations on both the content and inquiry reasoning progressions and the empirical results from these assessments can provide information on how students of many ability levels fall along our learning progression. In addition, these results provide us with insights to guide the revision of learning progression templates and associated curricular units.

Amelia Wenk Gotwals, Michigan State University
Nancy Butler Songer, University of Michigan
Lea Bullard, University of Michigan

Learning Progressions & Assessment in an Inquiry-Based Biodiversity Context

Learning science does not consist of just memorizing a body of facts; rather, the development of scientific knowledge is a dynamic endeavor that represents a complex interplay of content and reasoning skills, also called scientific practices (National Research Council, 2007). National science standards advocate the development of complex thinking and reasoning skills in science content areas for all students (American Association for the Advancement of Science, 1993; National Research Council, 1996). Similarly, our previous work has demonstrated that students who work with curricular units that integrate the development of complex reasoning skills with the development of content knowledge outperform their peers following a more traditional textbook-based program focused on the same content area (Songer, Kelcey and Gotwals, in press).

Giving students opportunities to engage in scientific inquiry, such as asking questions, gathering data, interpreting data, and explaining results is essential if we are to meet the standards set forth to help all students learn science (American Association for the Advancement of Science, 1993; National Research Council, 1996). Various inquiry methods have been shown to promote greater student achievement gains in content and inquiry reasoning knowledge (Krajcik et al., 1998; McNeill, Lizotte, Krajcik, & Marx, 2006; Songer, Lee, & McDonald, 2003). In addition, inquiry programs have been correlated with enhanced metacognitive skills (White & Frederiksen, 1998) and motivation towards science (Mistler-Jackson & Songer, 2000).

However, research documents that developing the abilities to productively engage in scientific inquiry-based practices takes time, repeated exposures, and an organized scaffolding approach (National Research Council, 2000). Even after targeted instruction or particular scaffolds, students often struggle to demonstrate complex reasoning about science concepts (Songer, Kelcey and Gotwals, in press; Krajcik et al., 1998; Lee, 2003; Lee & Songer, 2003; White & Frederiksen, 1998). Increasingly, the research suggests that the details of the particular scaffolds, the fading of the scaffolds, and the manner in which complexity is developed over time need greater investigation.

Learning Progressions

The *BioKIDS: Kids' Inquiry of Diverse Species* project focuses on the idea that developing the abilities to reason complexly about science takes time, repeated encounters with scientific phenomena, and careful guidance. Our project conducts research on the development of complex reasoning in biodiversity and ecology with cohorts of 4th, 5th, and 6th grade Detroit Public School students. In our work, we consider learning progressions as templates for the design of curricular, assessment, and professional development products.

> Learning progressions take a stance about both the nature and the sequence of content and inquiry reasoning skills that students should develop over multiple curricular units and years. Learning progressions are successively more sophisticated ways of thinking about a topic that can be used as templates for the development of curricular and assessment products. Learning progressions-driven curricular and assessment products are one of several possible manifestations of a given learning progression. The learning progression can only be evaluated indirectly, through the evaluation of the curricular products, professional development modules, and assessment instruments that are constructed from the learning progressions template. (Songer, Kelcey, & Gotwals, in press).

Following this definition, our early discussions with scientists and teachers and our consultation with National and state standards led to a preliminary content progression and a preliminary inquiry reasoning progression that we tested through empirical studies and then revised. The upper levels of the content progression have as the learning goal that students will be able to knowledgeably explain how human activity and other natural factors affect the biodiversity of ecosystems and also how biodiversity in an area can help to buffer ecosystems against change. The content progression has three interconnected strands of classification, ecology, and biodiversity topics. Figure 1 presents our current three-year content progression that was developed after the first round of empirical testing (see Songer, Kelcey and Gotwals, in press, for more details on this process).

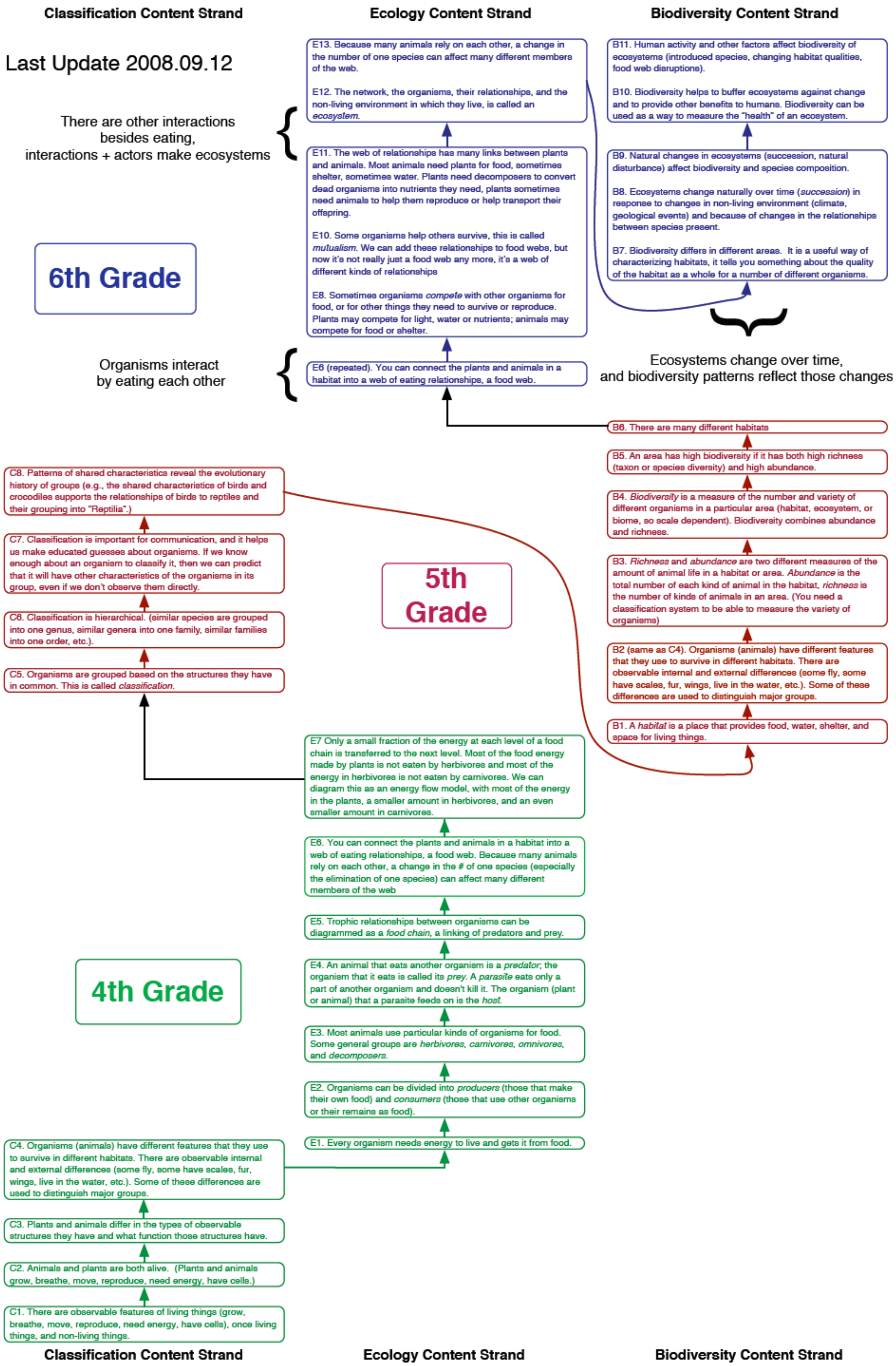


Figure 1: Three-year Content Progression

Scientific Explanations

Our inquiry reasoning progression focuses on a particular dimension of complex reasoning: the development of evidence-based explanations. Research demonstrates that students who engage in the development of evidence-based explanations about science content can significantly improve their understanding of scientific concepts, their understanding of appropriate use of evidence, and their ability to provide coherent and logical arguments (Bell & Linn, 2000; Scardemalia & Bereiter, 1991; White & Frederiksen, 1998; Toth, Suthers, & Lesgold, 2002). In addition, an examination of the quality and coherence of students' explanations can serve as a valuable means to assess student understanding (Metz, 1991). For these reasons and others, the National Science Education Standards state that, "…students should experience science in a form that engages them in the active construction of ideas and explanations" (National Research Council, 1996, p. 121).

Similar to others, we have adopted a modified version of Toulmin's (1958) model of argumentation to support teachers and students in creating scientific explanations (Bell & Linn, 2000; Driver, Newton, & Osborne, 2000; Erudan, Simon, & Osborne, 2004; Lee, 2003; McNeill et al., 2006; Sandoval, 2003; Songer & Gotwals, 2006). In our work, we emphasis three essential aspects of evidence-based explanations:

(1) Articulations of causal <u>claims</u>;

(2) Use of appropriate and sufficient <u>evidence</u> to support these claims; and

(3) Use of <u>reasoning</u> that draws on scientific principles to explicitly link the evidence to the claim.

Scaffolding

Despite a reasonable amount of early research on the value of student work with evidence-based explanations, our research, and that of others, testifies that guiding students towards an understanding of appropriate evidence and the development of sound arguments in the form of explanations is not a straightforward task. For example, students often struggle with articulating clear claims based on the data that they have and often do not fully understand what counts as evidence (Sadler, 2004), or how to incorporate appropriate (Sandoval, 2003; Lee and Songer, 2003) and sufficient (Sandoval & Millwood, 2005) evidence into their explanations. Middle school students also have been shown to have particular difficulty with the reasoning component of scientific explanations (Lizotte, Harris, McNeill, Marx, & Krajcik, 2003; Gotwals, 2006). In addition, German and Aram (1996) found that students had a hard time presenting evidence in a convincing way which indicates that students did not provide reasoning to indicate why the evidence was appropriate. Often students will make claims, but not back up the claims with evidence or reasoning (Jimenez-Aleizandre, Rodriguez, & Duschl, 2000).

Despite the issues that students face when attempting to build evidence-based explanations, studies have shown that when students work through repeated exposures within parallel examples in order to guide their construction of evidence-based explanations, they make significant gains in content knowledge and their ability to create clear and coherent explanations (Songer, Kelcey and Gotwals, in press; Lee, 2003; McNeill et al., 2006). Educational scaffolds have also been found to help younger students work with complex scientific information and participate in scientific inquiry activities (Metz, 2000).

Inquiry Reasoning Progression

Thus, based on the research around scaffolding of explanations, our empirical studies (e.g., Songer, Kelcey and Gotwals, in press) and the modeling of prior data (Gotwals, 2006; Gotwals & Songer, in preparation), we designed our inquiry reasoning progression focused on the development of evidence-based explanations as seen in Figure 2.

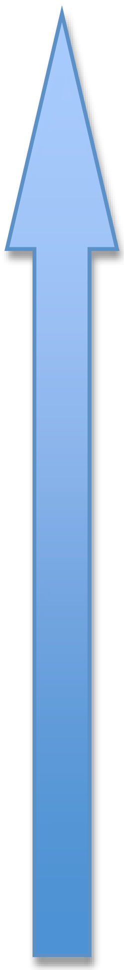
	**Level 4** Student constructs scientific explanation – consisting of a claim, sufficient and appropriate evidence, and reasoning tying the evidence to the claim (*without* scaffolding)
	Level 4s Student constructs scientific explanation – consisting of a claim, sufficient and appropriate evidence, and reasoning tying the evidence to the claim (*with* scaffolding)
	Level 3 Student makes a claim and backs it up with sufficient and appropriate evidence but does not use reasoning to tie the two together (*without* scaffolding)
	Level 3s Student makes a claim and backs it up with sufficient and appropriate evidence but does not use reasoning to tie the two together (*with* scaffolding)
	Level 2 Student makes a claim and backs it up with evidence, but evidence is insufficient /inappropriate (*without* scaffolding)
	Level 2s Student makes a claim and backs it up with evidence, but evidence is insufficient /inappropriate (*with* scaffolding)z
	Level 1 Student makes a claim but does not back it up with evidence (*without* scaffolding)
	Level 1s Student makes a claim but does not back it up with evidence (*with* scaffolding)

Figure 2: Inquiry Reasoning Progression for Building Evidence-based Explanations

Our early empirical results demonstrated that creating claims was the most straightforward aspect of constructing an explanation, while generating an explicit reasoning link between claim and evidence was the most challenging aspect of explanation building. Therefore, the lower end of our reasoning progression consists of students making claims that are unsubstantiated or not backed with any evidence or reasoning. While students may make claims that respond to the question, at the lower end of our reasoning progression, students do not provide any backing of their claims. The middle two levels in our reasoning progression consist of students making claims and providing evidence to back their claims. At level 2, students provide evidence, but the evidence is either insufficient or part of the evidence that students provide is inappropriate. At level 3, students provide sufficient evidence, but do not tie the evidence to the claim with reasoning. At the highest level, level 4, students construct a complete scientific explanation with claim, sufficient and appropriate evidence, and reasoning tying the claim and evidence together. There are two sublevels for each level in our reasoning progression based on the structure of the task that students must complete. One of the sublevels indicates that students are able to do the task (perform the practice of constructing evidence based explanations) with some sort of guidance or scaffolding. At the higher sublevel, students are able to do the task on their own, without any help.

As mentioned earlier, we see the role of our content and our inquiry reasoning progressions as templates for the development of curricular units, assessment tasks, and professional development materials. Each assessment task and curricular activity maps to a node on both our content and inquiry reasoning progressions. However, we do not conceive of student movement along the two progressions to be a simple linear path. While we do envision a sequence of curricular activities that proceed from bottom to top of our content progression in a linear fashion, we envision a sequence of curricular activities that cycle through our entire inquiry reasoning progression multiple times within any one eight week unit. For example, our progressions suggest that students should perform activities that emphasize the development of scaffolded evidence-based explanations about food chains before they emphasize the development of evidence-based explanations about disruptions to food webs (Songer, Kelcey, & Gotwals, in press). In order to evaluate the pathways that students take as they progress through our curricular units, we needed an assessment system based on our learning progression templates. Below we outline some of our assessment design decisions, empirical results, and challenges that we faced in designing a coherent assessment system mapped to our learning progression templates.

Assessment Task Structures

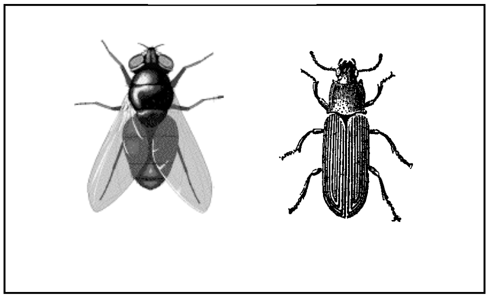
Assessment includes the processes of gathering evidence about students' knowledge and abilities as well as making inferences from that evidence about what students know or can do more generally (Mislevy, Steinberg & Almond, 2003; National Research Council, 2001). All assessments are based in a conception or philosophy about how people learn and what tasks are most likely to elicit observations of knowledge and skills from students. In addition, assessments are premised on certain assumptions about how best to interpret evidence to make inferences about what and how people know (Mislevy, Almond, & Lukas, 2004). Using our content and inquiry reasoning learning progressions as templates, we outline below the process of the development of assessment tasks. We designed a set of tasks that were mapped to the range of content and inquiry reasoning progression nodes so as to gather evidence of how students were using their content knowledge to formulate scientific explanations associated with a range of ecology, classification, and biodiversity content topics. In addition to gathering this evidence, we also wanted to be able to use a developmental approach to assessment (Wilson & Sloane, 2000) in order to determine how students progress over time in their abilities to formulate evidence-based explanations.

Task Design – What types of tasks could be used to elicit student understanding and skills?

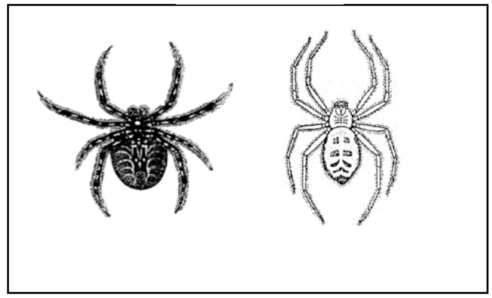
In order to gather the most information about our full sample of students and the growth trajectories that they may take as they engage in our curricular units, we desired a suite of assessment tasks matched to students with a variety of ability levels. Given the range of ability levels in a classroom and the vast range of prior experiences that students bring to the classroom, being able to provide well matched tasks for *all* students at many levels in their learning pathways may not be possible. However, having tasks that target lower, middle, and higher levels of skill in crafting evidence-based explanations can provide evidence of where students fall along these major ability markers as well as provide opportunities for students of numerous ability levels to demonstrate what they know. We know that students tend to struggle with formulating evidence-based explanations (Sadler, 2004; Sandoval, 2003; Lee & Songer, 2003), but that scaffolding students in their creation of the key components of explanations (claim, evidence, and reasoning) tends to allow students to create better and more coherent explanations (Lee & Songer, 2003; Gotwals & Songer, in preparation). Therefore, we chose to draw on our previous work (Gotwals & Songer, 2006), and proposed three levels of explanation items for our assessments that have different levels of inquiry-reasoning scaffolding. In items at the minimal level students are given the claim and evidence and asked to match appropriate evidence to the claim; in intermediate items students construct an explanation with structural inquiry-reasoning scaffolds (Figure 3); and complex explanation construction items have students construct scientific explanations with no scaffolds (Figure 4). By including three levels of explanation-building items on our assessments, we aimed to provide students across a range of ability levels opportunities to demonstrate their knowledge and skills associated with a range of biodiversity and ecology topics. In addition, these tasks are a direct manifestation of using our inquiry-reasoning progression as a template, as the progression indicates that for each level of progressing competence in creating evidence-based explanations, students should be provided the opportunity to demonstrate their abilities with and without scaffolding.

Shan and Niki collected four animals from their schoolyard. They divided the animals into two groups based on the physical characteristics of the animals:

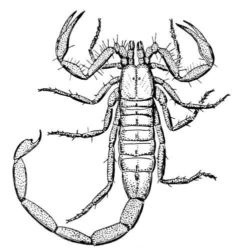
Group A



Group B



Below is a picture of a scorpion. Scorpions have several pairs of legs and a pair of claws. Scorpions also have a body with two parts (segments) and a tough outer covering.



Give a scientific explanation for the following question.

Scientific Question: Are scorpions more closely related to the animals in Group A or Group B?

Claim: Scorpions are more closely related to the animals in ______________________.

Evidence: Give 2 pieces of evidence to support the claim.

Reasoning: Provide the scientific reason that your evidence supports your claim.

Figure 3: Explanation-building assessment item with reasoning scaffolds

David is in Mr. Leon's science class. For homework, David has to make a table of all the living things he sees at the playground and what each of the living things eats. Below is the table that David made.

Name of living thing	What it eats	Number seen
Snake	frogs, mice	3
Grass and seeds	--	too much to count
Grasshopper	leaves, grass	14
Frog	grasshoppers	10

Give a scientific explanation for the following question.

Scientific question: Which 2 living things on David's list are predators?

Figure 4: Explanation-building assessment item with no reasoning scaffolds

We were careful to match the format and scaffolds in the assessments with what was used in the curriculum so that the curricular activities and assessment items paralleled one another. In doing this, we sought to imbue our assessments with diagnostic characteristics that would allow us to closely and thoroughly evaluate the effectiveness of our learning progression. Additionally, matching the structure of curricular activities and assessment items removed a source of possible distortion in our assessment results by eliminating student confusion that might result from unfamiliar question format.

However, despite the benefits of making these items proximal (Ruiz-Primo, et al, 2002) to the curriculum, the specificity of the scaffolding format also limits the generalizabilty of the results of our study. However, given that the purpose of our project is to develop a coherent package of learning progression-driven curricular, assessment, and professional development products, the benefits of the coherence between the components of the project seem to outweigh the limited generalizability of our results beyond our project.

Using reasoning progression to score items

We utilized the four levels of our reasoning progression to guide the coding rubric development for the intermediate and complex constructed-response explanation items. Mapping the coding rubric to the reasoning progression template adds another link to the reasoning progression that will help gather evidence of students' developing abilities to formulate evidence-based explanations. Not only are students provided with opportunities to demonstrate their ability to formulate evidence-based explanations in a range of supported situations (minimal, intermediate, and complex items), their responses are also coded with multiple possible increasingly "partially correct" responses to allow students to provide us with evidence of their competence in using their scientific knowledge to formulate evidence-based explanations. Table 1 illustrates the generalized coding rubric for our constructed response items. For each item, the rubric was customized to the particular content and item scenario.

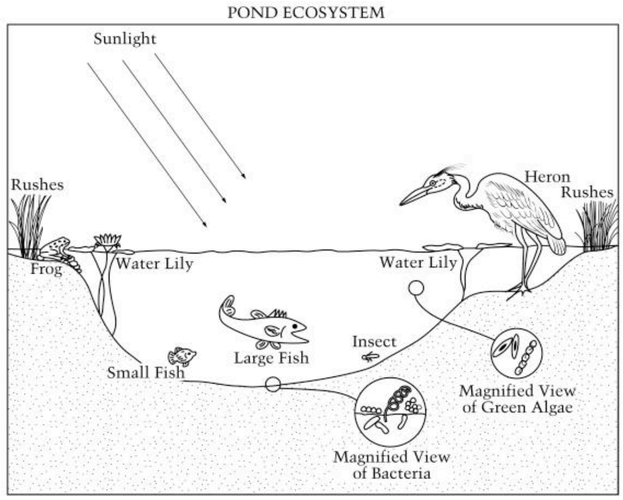
Level 4	Student constructs scientific explanation
Level 3	Student makes a claim and backs it up with sufficient and appropriate evidence but does not use reasoning to tie the two together
Level 2	Student makes a claim and backs it up with evidence, but evidence is insufficient /inappropriate
Level 1	Student makes a claim but does not back it up with evidence

Table 1: General scoring rubric for writing scientific explanations, informed by our reasoning progression

The combination of three levels of items and coding for answer sophistication level resulted in allowing students with a wide range of ability levels opportunities to provide us with evidence of their abilities mapped to our reasoning progression.

We used Item Response Theory (IRT) (Hambleton, Swaminathan, & Rogers, 1991) and think alouds (Ericsson & Simon, 1993) to analyze students' responses to the assessment items. The results, in general, support our hypothesized coding levels and indicate that the presence of scaffolds tended to make explanation items easier for students, relative to the non-scaffolded items (Gotwals and Songer, in preparation). However, while we were able to discern general trends in student responses that corresponded to our learning progression, not all student responses fit neatly into our progression. For example, some students gave answers that were correct but that could not be scored on the highest level of our coding rubric: some students would provide a scientific principle (reasoning) to support their claim, but did not provide specific evidence.

One example of this trend occurred with students completed a complex explanation item asking them to formulate an explanation about what would happen to the algae in the given pond ecosystem if all of the small fish died (Figure 5). Prior to completing the complex item shown below, students were asked to create a food web showing the interactions between organisms in the pond ecosystem. A complete (level 4) response to this question would include a claim that the algae should increase or multiply, evidence consisting of the fact that small fish eat algae, and a statement of reasoning relating the claim to the evidence showing that disturbances in one part of an ecosystem can have effects on other parts of the ecosystem.



Write a scientific explanation for the following question.

<u>Scientific Question</u>: 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?

Item modified from NAEP (http://nces.ed.gov/NATIONSREPORTCARD/)

Figure 5: Pond Ecosystem Complex Item

For the most part, students generally followed the trend of the reasoning progression, where, at the lower end, students would have an unsubstantiated claim as to what would happen to the algae, "*It would increase.*" At the middle levels, students would also include the evidence that small fish eat algae, and upper end responses included the reasoning statements tying these two statements together. However, there were several students who did not fit into this pattern. During the think aloud interviews, one student (*DTF0853010)* responded to this question by saying, "*The algae would increase because, when something is removed from a ecosystem it affects everything…"*. Similarly, another student (*DTF08530111)* stated, "*Algae would increase because if not any predator when one part of ecosystem is removed it affects the whole ecosystem."* Both students here had a correct claim statement that algae would increase, and both students included reasoning stating that when one part of an ecosystem is affected, other parts are affected also. However, neither of these students included specific evidence that small fish eat algae. One student hinted at it by saying, "if not any predator" however, did not include any specific evidence that the small fish eats algae. While the majority of students' responses fell into our coding scheme based on the reasoning progression, there were a few cases like these that did not.

While we saw a few responses that did not follow the pattern that the reasoning progression projected, the majority of the responses did follow this sequence. Given the diversity of learners, it is likely that there will always be responses that do not neatly fit into a given reasoning progression. However, it is important to take these "non-conforming" responses into account to determine how to structure activities and assessments so that all students can make progression to having more sophisticated understandings and abilities in science. In a section below, we discuss how we took these non-conforming responses into account in a subsequent restructuring of the assessments and curriculum.

Difficulty of Tasks: What makes a task more or less difficult? How do the context and other characteristics of a task affect its difficulty?

Scaffolding: As discussed above, our assessment system has specifically designed tasks to target different ability students with a wide range of item difficulties. The format of the item (specifically, the different levels of scaffolding) gives students a range of opportunities to demonstrate what they are able to do with and without support. The minimal items, that provide students with a claim and possible pieces of evidence and ask students to match the claim to the evidence, allow students to show if they have the ability to pick out what is appropriate and sufficient evidence to back a claim. These items, since they are highly scaffolded, tend to be less difficult than items that require students to construct an evidence-based explanation. The intermediate items, that contain scaffolding in the form of prompts for the components of explanations, tend to be more difficult than the minimal items as students must construct a response as opposed to choosing a response from a given list. However, intermediate items tend to be less difficult than the complex items that do not provide any form of scaffolding, as the intermediate items provide students with guidance as to what is required of a high level response. Using the reasoning progression as a template for both the design of items with multiple levels of scaffolding as well as a template for the coding of items into four levels of increasing competence, allows for a range of difficulty in tasks that can be matched to students at many different levels in a learning trajectory.

Complexity of Question: In addition to the different levels of scaffolding of assessment items, the type of question and the response that it requires also can influence the difficulty of the task. For example, several of the explanation items have claims that are relatively straightforward, such as choosing a group (A or B; Figure 1) or choosing a direction as to what will happen to the algae (increase or decrease; Figure 3). However, other items require more complex claims, such as an item that asks students to evaluate the suitability of a given habitat for a given organism or an item that asks for ways in which to increase the biodiversity of a given area. When the claim requires more than a simple choice, this increases the difficulty and complexity of the task.

The type of data that students need to interpret and use as evidence also influences the difficulty of the task. Some items, such as the item in Figure 6 (which is in a newer scaffolded format than our previous items – see the next section for an explanation), in which students need to identify which zone has the highest richness, have straightforward evidence. In this item, students would be required to know the definition of richness as the number of different types of organisms in an area. One zone, Zone B, clearly has the highest animal richness; the evidence is straightforward and uncomplicated. However, a follow-up question about this scenario asks students which zone has the highest biodiversity. While the correct response is also Zone B, selecting the correct zone for this item and providing appropriate and sufficient evidence is not as straightforward.

Zone C actually has a higher animal abundance (total number of animals) than Zone B. However, students need to weigh the importance of richness of animals in an area versus the abundance of animals in a given area and then identify that while one zone has a higher abundance of animals, the richness of animals, in this case, is more important for determining the highest biodiversity. Thus, the type of evidence and how students discuss any contradicting evidence also will influence the difficulty of the task.

Content and context: Clearly, the content of the item and the familiarity of students with the context will also influence the difficulty of the item. When students were able to talk through their problem-solving strategies in the think alouds and follow-up cognitive interviews, nuanced levels of student understanding of content were revealed. Below is a transcribed think aloud excerpt that demonstrates this. Here, the student responds to an item (Fig. 3) that asks whether a scorpion fits better in group A (insects) or group B (arachnids):

> *...scorpion are more closely related to the animals in group B, group B because it has, one two three four five and eight legs and group B there's two arachnids and they also have eight legs and it has almost the same picture figure and they are more closely alike because alike because they have um poison and they all um they all look like the same and because give two pieces of evidence to support the claim, um I think it belongs to group B because they are most related because of the legs and it has the same like the same mouth and both poison and they are all dangerous predators and reasoning provide the scientific reason that your evidence support your claim. I think the scorpion is also related to the arachnids because mhm, provide the scientific reasoning, okay, I don't know this one. I don't know about most about scorpions I know more about arachnids than scorpions.*

This think aloud provides insight into *why* the student found the item difficult: while she is familiar with one of the animal groups in the question (arachnids), she is unfamiliar with the animal featured in the item (the scorpion); the student identifies this unfamiliarity as a hindrance in her ability to answer the question. The student's participation in the think aloud allowed us to learn about why she did not fully answer this question correctly. This type of information is important in informing our future assessment and curricular revisions. While such information is rich and necessary for building a valid learning progression, the sorts of interviews described above are time- and labor-intensive. Therefore, it is important to conduct these types of think alouds and interviews at the beginning of the learning progression design process so that the information gathered from these experiences can help to inform the learning progression templates and subsequent assessments and curricula.

Iterative Design of Assessments

We began our assessment and curricular design process with hypothesized content and reasoning progressions as templates. We used these templates to design our first round of curricula and assessments and conducted empirical studies to gather evidence of the validity of our progressions and the effectiveness of our curricula and assessments.

Based on data that we gathered, we made multiple iterative revisions to our learning progression templates, our curricula, and our assessments based teacher and student interactions with our learning progression driven products. Below we outline one of the main revisions that we made to our assessment system.

In response to learning that students are most successful at tasks involving novel content when provided scaffolds, we added more content-based scaffolding to our curricular units and another level of item between our original minimal items and intermediate items that include both content and reasoning scaffolding. This revision is manifested in our formative and summative assessment items as four levels of explanation items:

1. Minimal items (highly scaffolded multiple choice items)
2. **New level (content- and reasoning-scaffolded explanations; Figure 6)**
3. Intermediate items (reasoning-scaffolded explanations)
4. Complex items (unscaffolded explanations)

Figure 6, below, shows the new level of scaffolding in our latest round of assessments. Students are provided with both content and reasoning scaffolds to guide them in crafting their explanations. The other intermediate-level items with the reasoning scaffolds only, also use this new format style of providing boxes in which students can provide their responses to each part of the explanation. This coincides with new formatting changes in our curricular units. In addition, we have changed the sequence in which we prompt students to complete the parts of the explanation. While doing think alouds and in examining the structure of students' written complex explanations, we noticed that students tended to verbalize or write the scientific principle as to why they created their claim before they provided the specific evidence for the item. Therefore, we now provide students with the prompt to provide reasoning for their claim before the prompt to provide specific evidence for the given explanation.

This table shows school yard animal data collected using CyberTracker. Use the table to help you answer the question.

<u>School Yard Animal Data</u>

Animal Name	Zone A	Zone B	Zone C
Pillbugs	1	3	4
Ants	4	6	10
Robins	0	2	0
Squirrels	0	2	2
Pigeons	1	1	0

Write a scientific explanation for the following question.
<u>Scientific Question</u>: Which zone has the highest richness?

Make a CLAIM:
Write a sentence that answers the scientific question.

Hint:
Think about the number of different types of animals in each zone.

Give your REASONING:
Write the scientific concept or definition that you thought about to make your claim.

Hint:
Think about the definition of richness.

Give your EVIDENCE:
Look at your data and find two pieces of evidence that help answer the scientific question.
1.

2.

Hint:
Think about the richness of each zone.

Figure 6: Explanation-building assessment item with content and reasoning scaffolds

Conclusions

Utilizing a learning progression-driven assessment design process that was evaluated with empirical data has allowed us to iteratively design and empirically test the validity of our assessment tasks for evaluating students' development of evidence-based explanations about biodiversity and ecology content. Our larger findings and trends are as follows.

First, while our empirical data largely supported our hypothetical projections, exemplified in our content and reasoning progressions, about which content and inquiry reasoning tasks were simpler or more difficult for students, not every student response fit into our envisioned progressions. Analysis of the unexpected student responses leads to valuable insights about the structure, presentations, and support/scaffolding necessary to guide complex thinking about biodiversity and ecology concepts for our target audience. Second, the cognitive interviews and the think aloud activities provided valuable complimentary information to the pre-post achievement and growth curve results. Utilizing cognitive interviews and think alouds allowed us to gain greater insights beyond what written tests can provide into how into how students think and what makes test items difficult.

In summary, assessments must be thoughtfully constructed in order to gain empirical evidence about the validity of a hypothesized learning progression. We have presented one approach to gather empirical evidence to guide the development and refinement of learning progressions and the associated assessments. In aligning the curriculum, assessment tasks, and professional development with the learning progression templates, we are able to draw strong conclusions about the value, sequence, and structure of the learning progression. By designing items of different levels of complexity associated with different nodes on the content and inquiry reasoning progressions, we were able to gain insight into how students of many ability levels develop along that learning progression. Student think alouds provided insight and additional nuances in student thinking towards more specific ways we could support the development of complex reasoning about biodiversity and ecology concepts.

Acknowledgements

This research was funded by the National Science Foundation under Grant ESI-0628151 to the Deep Think project at University of Michigan. The opinions expressed herein are those of the authors and not necessarily those of the NSF.

Resources

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

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.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287-312.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Erudan, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915-933.

Germann, P. J., & Aram, R. J. (1996). Student Performances on the Science Processes of Recording Data, Analyzing Data, Drawing Conclusions, and Providing Evidence. *Journal of Research in Science Teaching, 33*(7), 773-798.

Gotwals, A. W. (2006). Students' science knowledge bases: Using assessment to paint a picture. Unpublished doctoral dissertation. Ann Arbor: University of Michigan.

Hambleton, R., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publications.

Jimenez-Aleizandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education, 84*, 757-792.

Krajcik, J., Blumenfeld, P., Marx, R., Bass, K. M., Fredericks, J., & Soloway, E. (1998). Middle school students' initial attempts at inquiry in a project-based science classroom. *The Journal of the Learning Sciences, 7*(3 & 4), 313-350.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*(4), 674-689.

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., and Songer. N.B. (2003). Making Authentic Science Accessible to Students. International Journal of Science Education. 25(1), 1-26.

Lee, H.-S., & Songer, N. B. (2004, April). *Longitudinal knowledge development: Scaffolds for Inquiry.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Lizotte, D., J., Harris, C., McNeill, K., Marx, R., & Krajcik, J. (2003). *Usable Assessments Aligned with Curriculum Materials: Measuring Explanation as a Scientific Way of Thinking.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

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.

Metz, K. E. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching, 28*(9), 785-797.

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.

Mistler-Jackson, M., & Songer, N. B. (2000). Student motivation and internet technology: Are students empowered to learn science? *Journal of Research in Science Teaching, 37*(5), 459-479.

National Research Council (1996). National science education standards. Washington, DC: National Academy Press.

National Research Council (2000). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

National Research Council. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington D. C.: National Academy Press.

National Research Council (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, D.C.: National Research Council.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369-393.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*, 513-536.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12*(1), 5-51.

Sandoval, William A. and Millwood, Kelli A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Scardemalia, M., & Bereiter, C. (1991). Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of new Knowledge Media. *The Journal of the Learning Sciences,* 1(1), 37-68.

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., Kelcey, B., & Gotwals, A. (in press). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*.

Songer, N. B., Lee, H.-S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education, 87*(4), 490-516.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). Mapping to know: The effects of representation on guidance and reflective assessment on science inquiry. *Science Education, 86*(2), 264-286.

Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.

White, B., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*, 181-208.