Designing Assessments to Track Student Progress

Namsoo Shin, Shawn Y. Stevens, & Joseph Krajcik
University of Michigan, Ann Arbor, MI
Email: namsoo@umich. edu, sstevens @umich. edu, krajcik@umich. edu

Learning scientists have embraced the idea of using learning progressions for the development of assessments to monitor learning over time. Using LPs to develop assessments requires an iterative, process-oriented approach, and involves design products that work in real contexts. This proposal illustrates a design research process, Construct-Centered Design, for the development of assessment tasks to track the long-term development of student learning for a core idea in science – the transformation of matter. Fifty-six items of various types were designed to measure student understanding along the lower levels of a LP. We iteratively collected cross-sectional data from 500 middle students for developing, piloting, and revising the assessment items. We conclude the paper by discussing the strengths of and weaknesses of this process for measuring students' understanding across levels of the LP. The ultimate value of this work will propose the design process of assessment to track student learning.

Learning scientists have emphasized the importance of intensive research for addressing theoretical questions about the nature of learning in context and producing *evidence-based claims* to support their theoretical questions (Collins, Joseph, & Bielaczcy, 2004). Moreover, for valid evidence-based claims, researchers need to consider student learning over an extended period of time because learning challenging content takes years to develop (Stevens, Delgado, & Krajcik, 2009; Duschl, Schweingruber, & Shouse, 2007). Such research requires the development of assessment tasks (e.g., interview protocol, test item, observation instrument, survey) that can track student progress over time in real contexts rather than simply examining isolated variables within laboratory contexts (Brown, 1992; Barab & Squire, 2004) or memorized fragmented pieces of knowledge. However, to develop such assessment tasks is challenging. The learning science community needs models that illustrate principled and systematic ways of doing so. Recently, the science education and learning science communities have embraced the idea of learning progressions (LP) to provide a guide for monitoring student learning over time (see for example, Smith et al., 2006).

Learning progressions are research-based descriptions of how students may build their knowledge, and gain more expertise within and across a core idea over a broad span of time (Duschl, Schweingruber & Shouse, 2007; Smith et al., 2006). They illuminate how learners can develop and connect concepts within and across disciplines as they progress towards a more sophisticated understanding of the key concepts and skills necessary. As such, LPs can provide a potential path for students to develop understanding of core ideas over time, and can guide the alignment of instructional materials, instruction and assessment in a principled way to support the development of integrated learning (Duschl, Schweingruber, & Shouse, 2007; NRC, 2006 & 2007; National Assessment Governing Board, 2006a & 2006b). A LP contains three key factors: a lower and an upper anchor to define the range of content within a core idea and defined levels of understanding between the lower and upper anchors (Smith et al., 2006; Stevens, Delgado, & Krajcik, accepted). The levels of a LP specify not only the order in which students develop understanding of the important concepts, but also how they *interconnect* and *reason with* the important concepts between related ideas.

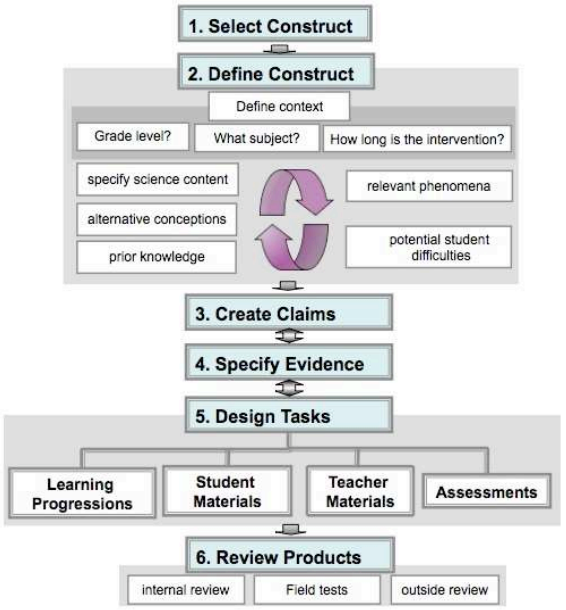
However, developing assessments that align with a LP requires an iterative, process-oriented approach, and involves designing products that work in real contexts. Because there are no such ready-made LPs, an iterative process of building, validating, and revising LPs and associated assessments using exemplary instructional materials is critical. Because learning is a complex process, many factors affect the path that students may follow as they build understanding, including the learning context, instructional materials, instruction, and students' prior knowledge and experiences. Thus, in order to build LPs and associated assessments, empirical data need to be collected from students who have experienced curriculum materials that were developed following LPs and learning principles. This helps us to ensure that students' lack of understanding is not because of inadequate learning experiences, but because of the developmentally challenging ideas students are expected to learn. Well-developed coherent curriculum materials and associated assessments based on a LP should be designed, implemented, and tested iteratively throughout the process of refining a LP. In addition, intensive research is needed to better characterize student understanding because gaps in learning research still exists. These gaps need to be filled prior to building LPs. In sum, developing assessment tasks using LPs is a complicated process and requires thorough, longitudinal studies related to how students learn core ideas over time in diverse contexts. A principled research design process should guide the complex, iterative process of developing assessments along an LP.

In this paper, we illustrate a research design process that can be used to develop assessment tasks to track the long-term development of student learning for a core idea in science – the transformation of matter.

We explore the following research question: *How can assessment tasks be developed that monitor student learning over time and that assign where student understanding lies along the LP?* Based on our previous work, we propose using Construct-Centered Design (CCD) (Shin, Stevens, & Krajcik, 2010; Pellegrino, et al., 2008) to develop assessments based on LPs. Because CCD focuses on the construct that students are expected to learn as well as what researchers and teachers want to measure, the CCD process provides a flexible and systematic approach for guiding product development, monitoring the development process, and examining the effects on learning outcomes. Next, we describe the foundation of CCD and present how the CCD process applies to the development of assessment to track student understanding in a LP. Finally, we conclude by discussing the strengths of and weaknesses of the process.

Research Design Process: Construct-Centered Design

A research design process to guide learning research and the development of products should provide flexibility for (a) mapping out the constructs associated with core ideas and (b) developing assessments and instructional materials that support and measure how student understanding develops over time. Thus, a design model should emphasize both defining constructs for instructional material development and specifying evidence for assessment development. By modifying and adapting the learning-goal-driven design (LGD) process for developing construct-focused curriculum materials (Krajcik, McNeill, & Reiser, 2008) and the evidence-centered design (ECD) model for developing assessments (Mislevy & Riconscente, 2005), the CCD process provides such a model. The CCD process begins with specifically defining the focus of the construct. We define the construct as the core ideas that students are expected to learn and researchers and teachers want to measure (Messick, 1994; Wilson, 2005). Because the foundation of the process focuses on the description and explicit specification of content that lies within constructs, the process is termed construct-centered design (CCD). In describing the process, we do not mean to imply this is a linear process. In practice, the process is interactive and highly recursive, with information specified at one step clarifying and often modifying what was specified earlier. Figure 1 illustrates the CCD process. A detailed description follows.



Claim: Student should be able to distinguish between matter and non-matter and justify their conclusions

Evidence: Students' evaluations and justifications should include the following criteria:

- Matter takes up space and has mass (has weight in a gravitational field)
- Matter can exist in three different states–solid, liquid and gas. The same substance can exist in all three forms under specific conditions

Assessment task (multiple True or False item)

Do you consider the following to be matter?

	Yes	Not sure	No
a rock			
an idea			

(…grain of salt, heat, a wish, air, a tree, a shadow, particle of chalk dust, beam of light, image in a mirror, smoke, helium inside a balloon, a cloud)

Which criteria are true of all matter?

Criteria	True	Not Sure	False
Whether you can see it			
Whether you can touch it			
Whether it takes up space			
Whether it is alive			
Whether it is a solid			

Figure 1. CCD process, Claim, Evidence, and multiple True or False Item example

Select and Define the Construct

The first step in CCD is to choose the construct and define the target learners (see Figure 1, step 1). The construct is essential as it identifies the set of ideas for which learners will study and develop understanding. Because students at different grade ranges have different knowledge and experiences that influence their learning, defining the target students helps define the construct appropriately, and also guides preparation of level-appropriate instructional materials, instruction and assessment. The construct of our LP is the transformation of matter and includes core ideas associate with the atomic model and the interactions between atoms and molecules as they undergo various transformations. To help define the range of content that needs to be unpacked, the lower and upper anchors for the LP were defined. In this case, the lower anchor was defined using the learning progression for atomic molecular theory for grades K-8 (Smith, et al., 2006), and additional empirical research. The upper anchor of the LP was defined based upon national standards (AAAS, 1993; NRC, 1996), ideas required as a foundation for nanoscale science and engineering learning for grade 7-12 students (Stevens, Sutherland, & Krajcik, 2009) and current learning research related to those ideas.

The next step is to define the construct based on expert knowledge of the discipline and related learning research (see Figure 1, step 2). This process, called unpacking, involves defining the ideas contained within the construct. By unpacking, we mean breaking up the construct into smaller components to explicitly specifying the concepts that are crucial for developing an understanding of the construct. Being related to the construct is not enough; the concept must be necessary for building understanding of the construct. The depth of understanding that is expected from students is also clearly defined in this step. As a step towards defining *how* students should know the content, the prior knowledge that is required both within and from other related constructs is also specified. The unpacking process also includes: identifying potential difficulties students might have learning the content; specifying and clarifying non-normative ideas that might interfere with students learning the content; providing possible phenomena that may help student learn ideas and develop their understanding; and identifying strategies for effectively representing the concepts based on previous learning research. The concepts within the constructs were then unpacked to define what it means to understand them at levels appropriate for grade 7-14 students.

Characterizing How Students Develop Understanding

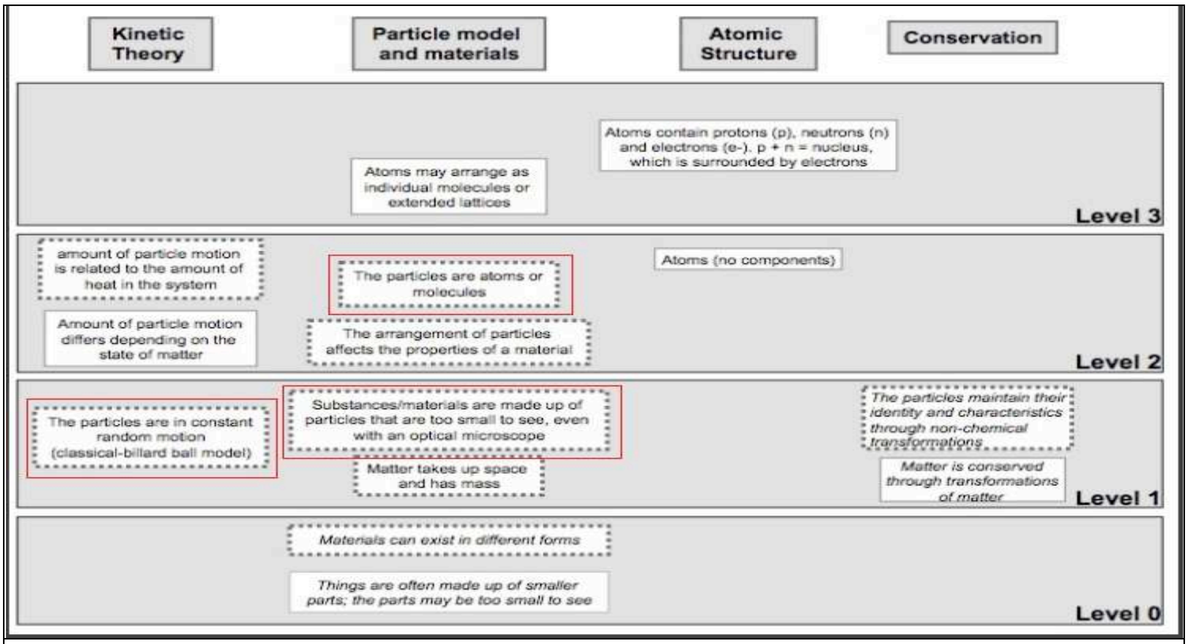
In the process of unpacking the construct, the current evidence-based learning research was not sufficient to completely define the levels of student understanding, possible non-normative ideas, and difficulties related to the developing understanding of the construct. To help fill gaps in the learning research related to the construct, an interview protocol was developed based on the CCD claim and evidence steps to characterize how ideas related to the transformation of matter developed as grade 7-14 students passed through the current curriculum (see claim and evidence sections for further details). A cross-sectional sample of students (N=79) representing the range of grades in our target population was interviewed. In this case, seventh grade students and high school students from the same district or school system were interviewed individually. At each level, we chose students who provided a mix of gender and a range of achievement levels. A 20-30 minute semi-structured interview was conducted with each student to characterize understanding related to sub-constructs of the transformation of matter. The interview questions required students to apply their knowledge to explain real world phenomena. The interviews were conducted in several phases. After each phase, student responses were evaluated and the protocol was revised to better characterize student understanding of the construct. The results of this study inform the strategies that may help students move along the LP. They provide information about how students develop understanding of the construct, and where they have incomplete knowledge. We used the results to describe prior knowledge about student learning using current instructional materials, and identify difficulties and non-normative ideas students hold regarding the content. Understanding students' ideas is critical in determining when and how it might be appropriate to introduce the concepts to students. In addition, knowledge of non-normative ideas aids the development of assessments that measure students' progress (see Stevens, Delgado & Krajcik, 2009 for detailed results).

Create Claim(s) and Specify Evidence

A set of claims is generated based upon the unpacked construct and student data. Claims specify the nature of knowledge and understanding expected of students regarding a particular construct (see Figure 1, step 3). In constructing a claim, vague terms like *to know* and *to understand* are not used. Rather, claims specifically define what cognitive activities students will need to apply to respond to the task (e.g., Bloom's taxonomy; Bloom, 1956). For example, students should be able to provide examples of phenomena, explain patterns in data, construct explanations, or develop and test hypotheses. An important part of student learning involves the ability to connect ideas and apply knowledge to new situations (Bransford, Brown & Cocking, 1999). Therefore, it is important that the claims specify how students should connect ideas both within individual sub-constructs, and among related sub-constructs in order to describe how students build integrated understanding of the construct. The evidence specifies the aspects of student work (e.g., behaviors, performances) that would be indicative that a student has the desired knowledge to support a specific claim or set of claims (see Figure 1, step 4). In particular, this step explicitly defines the expected level and depth of understanding that the target learners should demonstrate. The understanding defined by the evidence provides a guide for specifying levels in the LP.

Design Tasks

The tasks, which are generated based on the claims and evidence, provide a response that offers appropriate evidence to support the relevant claim (see Figure 1, step 5). The tasks can be either learning products that will help learners develop the knowledge in the claim, or assessment products that measure whether learners hold the knowledge stated in the claim. The assessment tasks are designed to elicit or generate students' performances to allow for a judgment to be made about whether sufficient evidence exists to support the learning *claim*. A single assessment task may provide evidence for more than one claim; multiple tasks may be necessary to assess a single claim. A single task or set of tasks can be associated with a claim or a set of claims assigned to multiple levels on the LP. An individual claim, its evidence and corresponding task may link to a single level on the progression. The right column in Figure 1 provides an example of a claim, its corresponding evidence and assessment tasks. Based upon the claims and evidence, we developed the first version of a LP (see Figure 2 for a portion of the LP).



Likert-type item: The following are possible characteristics of atoms and molecules. Mark whether the statements are true or false.

Statement	T	PT	NS	PNT	NT
1. At 75°F (~24°C) molecules and atoms are always moving.					
2. Molecules are all the same size.					
3. Molecules that make up ice cream are very cold.					
.					
21. A balloon is filled with water. When squeezed the balloon changes shape. The pressure from squeezing also changed the shape of the water molecules.					
22. The atoms in solid copper metal are hard.					

T: true, PT: pretty sure it is true, NS: not sure, PNT: pretty sure it is not true, and NT: not true

Figure 2. A portion of LP and Likert-type item for measuring Level 1 and 2 of the LP

Develop Assessment Items to Measure Levels of Understanding in a Learning Progression

The claims and evidence specify how students should be able to connect ideas both within a sub-construct, and among related sub-constructs in order to describe how students build integrated understanding. The claims and evidence can be used to refine the levels. Figure 2 illustrates part of the LP for the transformation of matter. In figure 2, the dashed boxes represent the content for which we are developing assessment tasks. The red color boxes correspond to the assessment item in the figure. The levels in the LP represent a set of ideas that describe a path towards developing a more complex understanding of the construct. The set of ideas within a level connects to explain a range of phenomena; higher levels describe the phenomena with greater scientific sophistication and completeness. In this way, the levels of the LP describe increasing levels of understanding. Further research is required to examine whether the order in which students learn concepts within a single level is important.

Based on the LP, we developed assessment items to locate the level of student understanding in the LP. Because the development of assessment items is a challenging activity, our design process will take several iterations to develop assessment tasks that will track student from the lower anchor to the upper anchor. In the first iteration, we focused on the development of assessment items for the lower level including levels 1 and 2. When possible, we selected previously published items that supported claims and evidence appropriate for the target students and modified them as necessary; otherwise, we designed and developed new items. We evaluated the items using the criteria of sufficiency, necessity and age appropriateness (Deboer, et al., 2008). We then revised the item stems, answer choices, and associated representations to ensure they support the claims and evidence associated with the lower levels of the LP and to ensure that text and representations were level appropriate. Fifty-six items using various item types (e.g., two-tiered, multiple true and false, complex multiple-choice, and Likert-type scale) were designed to measure how well students apply ideas within and across the sub-constructs. Although in future iteration, we will develop open-ended items, in the first Iteration, various item types as listed above were employed to develop assessment items that measure carefully and efficiently student understanding across grades 7-16 (Scalise, & Gifford, 2006) (see Figure 1, 2 and 3 for example).

Review Products

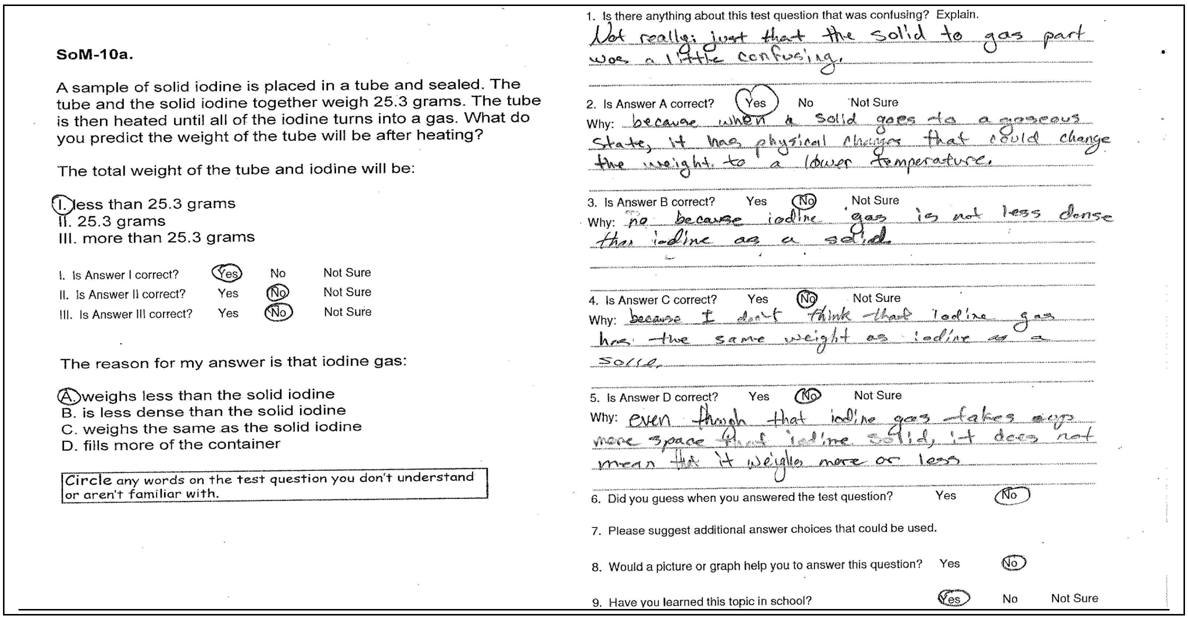
The next CCD step is to review the products. For each step within this iterative process, the products are reviewed internally and when appropriate, externally (see Figure 1, step 6). The internal review focuses on critique and revision of the products to ensure that they align with the claims and evidence. External review includes feedback from teachers of the target students or from content or assessment experts. Conducting pilot tests and field trials with target students is an essential component that provides invaluable information about the products.

Pilot Study

In order to ensure that students interpret the questions and associated representations as intended, we piloted the items with 479 middle school students (grades 6-8) from three schools in two distinct communities representing a range of race, ethnicity and SES. The cross-sectional design ensures that items measure the understanding of students with a range of knowledge and abilities. A range of different types of assessment items can provide information about students' understanding. The different types of items vary in complexity (e.g., matching < categorizing < ranking and sequencing < assembling proof) and in the amount in which learners' responses are limited or *constrained* (e.g., traditional multiple choice vs. open-ended essay; Scalise & Gifford, 2006). Items that are more constrained are easier to score, but likely provide less information about student understanding and higher-order reasoning. In this pilot study, we investigated "*what are the advantages and limitations of different item types for measuring student understanding across Levels 1 and 2 of the LP?*" Each student was given a test form containing four to five items. We piloted each test item with 30-35 students. The piloted items were each accompanied by a set of questions to explore how students interpreted the item (Deboer, et al., 2008). Figure 3 illustrates the questions that accompanied the two-tier multiple-choice questions; slightly different sets of questions accompanied different item types.

We analyzed each item using a simple descriptive analysis, a classical item analysis and an Item Characteristic Curve analysis that focused on identifying item clarity. For the descriptive analysis, we created five categories: confusing words in an item, confusing item stem, reasons for choosing their answer, percent who correctly answered an item, and helpfulness of representation. We carefully reviewed the items that students answered correctly more than 80% and approximately less than 30% in order to investigate whether the items are too easy or difficult for target students, or are written poorly. Figure 3 provides an example of student responses to an item. In this case, students answered correctly 24% in the first-tier question and 33% in the second-tier question. This item posed conceptual and literacy difficulties for students. Because the item involved not only the idea of conservation of matter but sublimation as well, we judged that it might be a level 3 item, as students did not seem to understand the phenomenon of sublimation described in the item. Other students did not know that "iodine" could be a solid. In addition to ensuring that our pool of items were written clearly and at an appropriate level, student responses helped ensure that there is only one correct answer and provided an evaluation of the distractors. After the initial descriptive analysis, we then conducted an additional review of the two-tier multiple-choice and multiple-choice items using classical item analysis and Item Characteristic Curve (ICC). For the classical item analysis, we focused on the correlation between the individual item score and the total score of the test. Then, we reviewed the ICC of each item to investigate how well an item differentiates between students having ability below and above the item location using dichotomous data (1=correct, 0=incorrect), and how students respond to the distractors of the items using polytomous data (raw responses = 1, 2, 3, 4). These analyses helped us to revise each item stem, as well as the distractors for each item. For example, the results indicate that the two-tier item type was not appropriate for measuring middle-school students understanding because the students were confused about how they should answer the question, and the item did not provide additional information than when only using the second part of the two-tier question. We speculate that middle school students are not developmentally ready to interpret the structure of such an item type. We decided to conduct a future investigation for this item type because our finding is inconsistent with Treagust's study using high-school students (Treagust & Chandrasegaran, 2007)

We conducted an additional analysis to investigate the characteristics of the item types on measuring student understanding along the LP. First, we developed items using two different item types (multiple-choice and multiple True and False (T/F)) to compare how students perform differently on the two item types and whether the two item types provide similar or different information. We assessed student understanding using a Multiple T/F list as opposed to a series of multiple-choice questions for gaining insight into connections and/or the complexity of student understanding (see Figure 2). The results show that most high-ability students responded that the multiple T/F item type is more difficult than multiple-choice items. They reported that they have to think carefully before responding to the item because they need to respond to every answer rather than to select a single correct answer. Therefore, the multiple T/F item type may better assess student understanding because they cannot rely on test-taking skills such as answering a multiple-choice item. In contrast, a majority of low-ability students responded that the multiple T/F is easier than the multiple-choice item because they is not just one correct answer.


SoM-10a.

A sample of solid iodine is placed in a tube and sealed. The tube and the solid iodine together weigh 25.3 grams. The tube is then heated until all of the iodine turns into a gas. What do you predict the weight of the tube will be after heating?

The total weight of the tube and iodine will be:

I. less than 25.3 grams
II. 25.3 grams
III. more than 25.3 grams

I. Is Answer I correct? Yes No Not Sure
II. Is Answer II correct? Yes No Not Sure
III. Is Answer III correct? Yes No Not Sure

The reason for my answer is that iodine gas:

A. weighs less than the solid iodine
B. is less dense than the solid iodine
C. weighs the same as the solid iodine
D. fills more of the container

Circle any words on the test question you don't understand or aren't familiar with.

1. Is there anything about this test question that was confusing? Explain.
Not really; just that the solid to gas part was a little confusing.

2. Is Answer A correct? Yes No Not Sure
Why: because when a solid goes to a gaseous state, it has physical change that could change the weight to a lower temperature.

3. Is Answer B correct? Yes No Not Sure
Why: no because iodine gas is not less dense than iodine as a solid.

4. Is Answer C correct? Yes No Not Sure
Why: because I don't think that iodine gas has the same weight as iodine as a solie.

5. Is Answer D correct? Yes No Not Sure
Why: even though that iodine gas takes up more space that iodine solid, it does not mean that it weighs more or less

6. Did you guess when you answered the test question? Yes No

7. Please suggest additional answer choices that could be used.

8. Would a picture or graph help you to answer this question? Yes No

9. Have you learned this topic in school? Yes No Not Sure

Figure 3. Two-tier type item and accompanying questionnaire with student response.

Second, we analyzed the relative difficulty of the Likert-type and multiple T/F items to explore how they measure the range of understanding for each level of the LP using Item Response Theory (IRT). We used a series of 14 T/F questions to measure how well students distinguish between matter and non-matter (See Figure 1). We treated each statement as a separate individual item of the multiple T/F item for the analysis. We reviewed the difficulty level of the items by examining a Wright map (Wilson, 2005) and an item fit output to ensure the items measure across Level 1 and 2 of the LP. The Wright map showed that item difficulty ranged between logit -2 and 1 out of a range from – 6 to 6. The – 2 to 1 range is consistent with our intention of measuring Level 1 and 2. A logit takes into consideration both the ability of respondents as well as the item difficulty in assigning a task to a particular location on the Wright map. In addition, we found that virtually all students classified the solid forms of matter–rock, a grain of salt, a tree and a particle of chalk dust–as matter. The item difficulty levels are very similar between logit -1 and 0. In the revision of the item, we streamlined the list by removing three of these objects since they were not providing additional information. The other part of the question was designed to gain information on students' reasoning as they distinguish between matter and non-matter. Students were asked about the criteria necessary for something to be classified as matter. Students in the first pilot suggested the "whether it's a solid" as a possible criteria. We found in subsequent administration of the item that it proved to be an effective distractor. We created a Likert-type item that surveys a range of ideas related to the atomic and kinetic theories based on a survey created by Harrison and colleagues (Harrison, Kease, & Voss, 2006) to efficiently measure students' models of the structure of matter (see Figure 2). Their difficulty levels are between logit -1.5 and 1.5, which is the range for Levels 1 and 2. Twenty-one statements were spread evenly across three logits on the Wright map. From the findings, we conclude that the Likert-type and multiple T/F type items can be used to measure a range of student understanding along the LP without spending significant testing administration time or giving up probing student reasoning.

Interview Data

After data analysis, we selected problematic items where the difficulty was not clear (e.g., whether the problem resulted from students' lack of understanding or the item itself). The five selected items included a multiple-choice item with a picture, a two-tier item, a Likert-type item, a graphic representation item, and a model representation item. We developed a semi-structured interview focusing on the selected items and interviewed 19 middle school students to ensure the accompanying items adequately measured students' understanding. We analyzed the student data using the five categories described above.

In the case of the two-tier item, the interview confirmed that most students did not know what was "iodine." Their responses were similar to students in the pilot study. However, they could guess that it is a chemical substance and that it can be liquid and solid. We initially decided to use "iodine" in the question because the students could answer the question correctly without knowing is the meaning of "iodine." In addition, the interview results about the two-tier item confirmed that middle school students have difficulty understanding the structure of the item. They had to spend time trying to understand how to answer the item rather than how to apply their knowledge to answer it. From the pilot and interview data, we conclude that we will not employ the two-tier item type for future development of assessment items for middle school students. For the Likert-type item, students felt comfortable responding to the item, but felt it was more difficult than a multiple-choice item because they had to think about the degree of the correctness of each statement. Based on the pilot and interview data, the results confirmed that the Likert-type item is appropriate for measuring middle school students' level of understanding. The results from the graphic and model representation items provided detailed information about student's interpretations of the item stems, distractors' representations, and graphical representations to guide revision of the items for the next iteration. Overall, the results from the survey instrument and interview data were consistent, which means that in the future, we can use a survey instrument to collect data about items instead of collecting intensive, time-consuming qualitative data.

From the first iteration of data collection and analysis, and further expert and internal review, we are revising 11 problematic items for the next iteration. The important task of the next iteration is to analyze the relative difficulty of the items to gain insight into the relationship between the students and the assessment items using Item Response Theory (IRT). We will collect a student data set, approximately 100 students per item to analyze the relative difficulty of the items. This item analysis will confirm that the items measure student understanding corresponding to Levels 1 and 2 of the LP.

Conclusion

We have described the iterative CCD research design process that researchers can use to develop assessment items that align with a LP to track learning across time. These items will provide a scale that can locate students on the lower levels of the LP. Ultimately, we will use these items to collect longitudinal data to evaluate student learning as they experience three years of coherent instruction that supports learning of the nature of matter.

A number of researchers have discussed the value of developing learning progressions to track students learning (Wilson, & Berenthal, 2006; Smith et al., 2006; Duschl, Schweingruber, & Shouse, 2007). However, there are still critical challenges to overcome in the process of developing and refining LPs and valid associated assessments to measure the level of student understanding across time (Pellegrino, Chudowsky, & Glaser, 2001; Smith, et al., 2006). First, because of the complexity of building LPs, we need to use a research design process that is iterative, process-oriented, and involves designing products that work in real contexts that extend our understandings of the nature and condition of learning and development as well as promote student learning (Barab & Squire, 2004; Brown & Collins, 1992; Collins, Joseph, & Bielaczcy, 2004). However, the typical strategy for this type of learning research employs naturalistic methodologies to investigate how learning occurs and the product development process for building evidence-based claims (Barab & Squire, 2004). A fundamental challenge for such research is the extensive quantity of qualitative and quantitative data that must be collected and organized in order to provide appropriate evidence to support the research claims (Collins, Joseph, & Bielaczcy, 2004). Second, although LPs offer a promising framework, limited validated assessment items exist which connect assessment items and the developmental progress of student understanding to illustrate conceptual growth. Prior to using the LP to track student progress, extensive research is needed to validate assessment items.

As we have illustrated, CCD can be a valuable process for learning research that may overcome these challenges. In particular, the CCD approach focuses on clearly defining the construct to focus the research and development strategies. Another critical characteristic of CCD is the explicitly specified evidence based on the unpacking of the construct that links directly to the claims. Specifying the claims and evidence supports the development and alignment of a range of products. The systematic process outlined by CCD provides guidance for the collection, organization and analysis of data by defining what data is essential for supporting the learning claims we hope to make about student learning. CCD can be considered a component of an iterative development process that is constantly being refined and revised to accommodate the needs of learning researchers. We still have much work to accomplish to make CCD a usable design model for other researchers. We need to further develop the guidelines and examples for each step of CCD to provide guidance on how researchers can use CCD to accomplish a variety of design-based research goals. To do this, researchers need to apply the CCD process to design various research, instructional materials, learning progressions and assessments tasks in order to articulate the subcomponents of the various CCD steps more clearly. As we and other researcher use CCD to guide a greater amount of research and development products, the process will become articulated better.

References

Barab, S., & Squire, K. (2004). 'Design-based research: Putting a stake in the ground', *Journal of the Learning Sciences*, 13(1), 1-14.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How People Learn: brain, mind, experience, and school*, Washington, DC: National Research Council.

Brown, A. L. (1992). 'Design experiments: Theoretical and methodological challenges in creating complex interventions', *Journal of the Learning Sciences*, 2, 141-178.

Cobb, P., & Bowers, P. (1999). 'Cognitive and situated learning perspectives in theory and practice', *Educational Researcher* 28(2), 4-15.

Collins, A., Joseph, D., & Bielaczyc, K (2004). 'Design research; Theoretical and methodological issues', *Journal of the Learning Sciences*, 13(1), 15-42.

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

Harrison, A., Klease, G., & Voss, L. (April, 2006). *Chemical literacy in grades 7-12: Teaching particle concepts for understanding*. A paper presented at the annual conference of the National Association for Research in Science Teaching, San Francisco, CA.

Krajcik, J.S., McNeill, K. L., & Reiser, B. J. (2007). *Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy*. Retrieved Nov. 17, 2007, from www.interscience.wiley.com.

Messick, S. (1994). 'The Interplay of Evidence and Consequences in the Validation of Performance Assessments', *Educational Researcher,* 23(2), 13-23.

Mislevy, R. J., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology*, Menlo Park, CA: SRI International.

National Assessment Governing Board. (2006a). *Science assessment and item specifications for the 2009 National Assessment of Educational Progress* (Prepublication ed.), Online. Available HTTP: <http://www.nagb.org/ pubs/naep science specs 2009.doc> (accessed 11 July 2006).

National Assessment Governing Board. (2006b). *Science framework for the 2009 National Assessment of Educational Progress* (Prepublication ed.), Online. Available HTTP: <http://www.nagb.org/pubs/ naep science framework 2009.doc> (accessed 11 July 2006).

National Research Council (2006). *Systems for state science assessment*, Washington, DC: The National Academies Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*, Washington, DC: National Academies Press.

Scalise, K., & Gifford, B. R. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing "Intermediate Constraint" Questions and Tasks for Technology Platforms. *Journal of Teaching, Learning and Assessment, 4(6).*

Shin, N., Stevens, S. Y., & Krajcik, J. (2010). "Tracking student learning over time using Construct-Centered Design." In S. Routledge (Ed), *Using analytical frameworks for classroom research: collecting data and analysing narrative*. Taylor & Francis, London.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). 'Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory', *Measurement: Interdisciplinary Research and Perspectives,* 4 (1), 1 – 98.

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (in press). 'Developing a theoretical learning progression for atomic structure and inter-atomic interactions', *Journal of Research in Science Teaching*.

Stevens, S. Y., Sutherland, L. M. and Krajcik, J. S. (2009). 'The big ideas of nanoscale science and engineering'. Arlington, VA: NSTA Press.

Treagust, D. F., & Chandrasegaran, A. L. (2007). 'The Taiwan National Science Concept Learning Study in an International Perspective', *International Journal of Science Education*, 29 (4), 391 — 403

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Wilson, M., & Berenthal, M. W. (2006). *Systems for state science assessment*. Washington DC: National Academies Press.

Acknowledgments

This research is funded by a Developing an Empirically-Tested Learning Progression for the Transformation of Matter to Inform Curriculum, Instruction and Assessment Design grant, number 0822038, and a National Center for Learning and Teaching in Nanoscale Science and Engineering grant, number 0426328, from the National Science Foundation. Any opinions expressed in this work are those of the authors and do not necessarily represent those of the funding agency.