Finding the Expert Knowledge that Everyone Can Use: Year 2 and 3 DSA Progress Report

Authors: Louise Yarnall, Larry Gallagher, & Geneva Haertel

December 2010

As part of the nationwide effort to increase the numbers of American students completing college, measuring student learning is considered key. Yet policy makers have barely begun to define the forms of knowledge that are most informative and important to measure. Researchers at SRI International are studying how to enhance outcomes measurement with a greater understanding of expert knowledge and reasoning in a field. In theory, such understanding can support accountability goals, help instructors prioritize lessons, and reveal the links between coursework and real world skills. To ground this work, SRI researchers applied theory and basic research in educational psychology that defines and measures the forms of knowledge and reasoning that contribute to expertise.

Domain-Specific Assessment Early Findings

SRI International has used this framework to develop a prototype assessment that measures schematic forms of knowledge and domain-specific forms of reasoning in two areas that many undergraduates study, biology and economics. The schematic form of knowledge differs from general critical reasoning. It defines how key ideas in a field relate to one another—serving as a kind of conceptual map for future learning and problem solving. The four-year Domain-Specific Assessment (DSA) project involves prototyping and pilot testing an assessment that measures these schematic forms of knowledge and the forms of reasoning valued in the fields of biology and economics. SRI researchers have spent the past 2.5 years modeling the conceptual complexity and reasoning in the two fields and documenting how to elicit such knowledge and reasoning through the principled design of assessment items.

In spring 2010, SRI administered the resulting prototype assessments in biology and economics to 296 community college students. As part of an initial validation study of the assessments, SRI asked four experts in biology and economics to rate the forms of knowledge and reasoning elicited by the DSA test items and those of two other existing assessments. Experts, who were blind to which assessment SRI had designed, reported that DSA assessments in both fields assessed knowledge equally well and assessed reasoning better than the other existing assessments. In addition, the experts rated the forms of knowledge measured by the DSA assessments as more relevant to use in everyday life.

These findings indicate that it is possible to test students on forms of knowledge that are both foundational to a domain and have broad applicability to reasoning beyond school. Finding ways to demonstrate to students the real world relevance of what they learn in school has been identified as a strategy for improving student engagement and motivation (National Research Council, 2003). Having reliable measures of such outcomes can provide American college educators with tools to track this important form of student learning and, ultimately, design better instruction to support it. It's potentially a powerful tool for shifting away from "fire-hose" instruction and testing in complex domains to instruction and testing informed by the powerful ideas and reasoning that experts use in these domains.

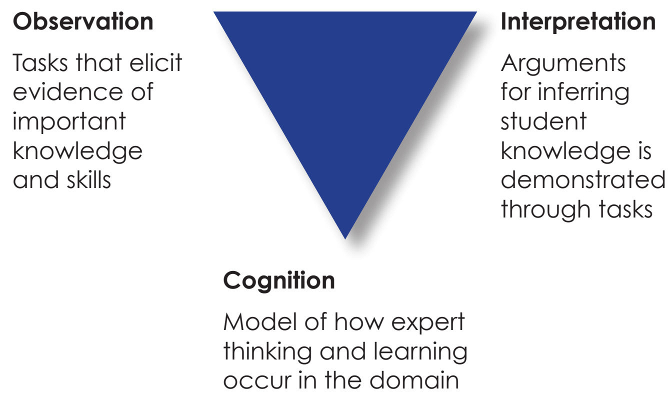
Evidence-Centered Assessment Design Methods

This research has not just created some prototype assessments; it has documented the recipe so that others may create similar powerful assessments. Some elements of this design documentation library are intended to be re-used by both assessment designers and instructors so they can better track how students are learning the core schematic knowledge of a field and the valued forms of reasoning.

The documentation process is central to the evidence-centered design (ECD) approach to assessment (Mislevy & Risconscente, 2006). For example, the team has created "design patterns" for assessments in undergraduate biology and economics. "Design pattern" is a term borrowed from architecture that refers to the basic elements that architects use repeatedly in many different building designs. In an analogous way, this project defines the reusable essential elements that test designers can use when measuring student learning in undergraduate biology and economics—the knowledge, skills and abilities to be measured, the kinds of evidence that needs to be gathered, and the features of tasks that will elicit the evidence. It also involves documenting the hypothesized cognitive structure of that knowledge, such as links and connections among different ideas and concepts in the domain.

To briefly review, ECD lays out the three core concepts of building asssessment arguments from evidence—cognition, observation, and interpretation (see Figure 1).

Figure 1. The Assessment Triangle



Source: Pellegrino, Chudowsky, & Glaser, 2001, p.44

ECD also identifies five layers of assessment design and implementation (Mislevy & Riconscente, 2006): domain analysis, domain modeling, the conceptual assessment framework, assessment implementation, and assessment delivery. We have completed the first four layers as part of the design and pilot testing process, with the bulk of the new work in Years 2 and 3 focused on defining the conceptual assessment framework and implementing the assessments. As noted earlier, the validation phase has just begun.

Using ECD to Define the Relevant Forms of Knowledge

Defining the relevant forms of knowledge began in Year 1 with domain analysis and modeling, which was based on input from subject matter experts. This work is described in an earlier technical report (Yarnall, Haertel, & Gallagher, 2009).

In Years 2 and 3, defining the conceptual assessment framework involved initial pilot testing and was followed with a specification of a student model. After the pilot test results were collected, SRI researchers and subject matter specialists from community college compared the scores and cognitive think-alouds of students with three different degrees of experience with college biology or economics courses (none, minimal, completed basics). (This analytic review was also used to identify item hygiene problems with initial assessments requiring revision.) The review of the pilot results led to the development of a logical student model. This model included: simple discrete knowledge (e.g, facts, terms), simple relational knowledge (e.g., a couple ideas in relation), and more complex schematic knowledge (e.g., three or more ideas in relation), and strategic reasoning (e.g., knowing when and how to apply knowledge to a situation). The student model represents the hypothetical structure of domain knowledge that the assessment seeks to measure. It also lays out an argument for the scale that will be used to differentiate levels of performance according to knowledge complexity and reasoning quality. It is the heart of the DSA assessment.

After the student model was defined, SRI researchers turned to defining the features of the assessment tasks that would elicit evidence of the target knowledge and reasoning. Researchers developed an initial set of characteristic features for each assessment prompt as well as rubrics to score student work. The SRI team drew on the work of Nagashima et al. (2008) and Wilson (2005) to differentiate different levels in the rubrics. In the final phase of the work, SRI researchers will interpret empirical evidence of the validity of each of the biology and economics assessments and will refine the student, evidence and task models specified for each of the assessments as part of the evidence-centered design process.

Determining the Scope of the Prototype Assessment

The study examined two big ideas in biology, carbon cycle and aspects of evolution, and three big ideas in economics: the supply and demand model, individual decision-making around opportunity cost, and resource allocation. It also examined a range of valued forms of reasoning in both subjects. Biology reasoning items focused on argumentation/explanation; summarization; computation; representation creation; and lab/field procedures, and economics reasoning items focused on argumentation and model-based reasoning, such as identifying the appropriate model for solving a problem, using graphs to articulate the model, and reasoning with the model to explain phenomena.

The purpose of ECD is to document a wide potential range of knowledge elements relating to the student model, and then select a subset of knowledge elements for measurement. As may be seen in Table 1, knowledge elements are listed in two ways: "potential," which refers to a wide range of knowledge elements relevant to the core big ideas, and "measured," which refers to the knowledge elements actually tested in the pilot.

Table 1. Conceptual Complexity and Reasoning Elements Documented and Measured by the Domain-Specific Prototype Assessment

Conceptual Complexity and Reasoning Elements Documented and Measured								
Domain	**Level 3**		**Level 4**		**Level 5**		**Level 6**	
Big Ideas	**Having discrete knowledge**		**Relating 2 ideas**		**Relating 3 or more ideas**		**Knowing when to apply ideas strategically**	
	# discrete knowledge elements		*# interrelated knowledge elements*		*# interrelated knowledge elements*		*# situations for knowledge application (reasoning built in)*	
	Potential* 27	Measured 7	Potential 18	Measured 7	Potential 22	Measured 5	Potential 19	Measured 4
Biology		# Relevant Items 8		# Relevant Items 16		# Relevant Items 9		# Relevant Items 8
Reasoning		1		8		1		
Carbon Cycle		3		4		4		4
Mutation/ Natural Selection/ Speciation		4		4		4		4
	Potential* 21	Measured 4	Potential 25	Measured 6	Potential 26	Measured 10	Potential 10	Measured 4
Economics		# Relevant Items 11		# Relevant Items 12		# Relevant Items 15		# Relevant Items 9
Supply & Demand Model (reasoning built into all)		7		8		7		5
Decision Making (reasoning built into all)		4		3		4		2
Resource Allocation (reasoning built into all)		0		1		4		2

* "Potential" refers to the full range of knowledge elements specified in the student model; "measured" refers to the knowledge elements specified in the student model that were included in the prototype assessment.

The table also illustrates the DSA pilot testing approach, which included creating a number of tasks measuring the following levels, which represent an increasing progression of schematic understanding:

(1) relatively simple knowledge elements (Level 3)

(2) simple relational knowledge elements (Level 4)

(3) more complex relational knowledge elements (Level 5)

(4) strategic knowledge elements of how to apply ideas to a real world situation (Level 6).

(5) reasoning elements, also to be measured on a 6-level scale

The assessment instrument included tasks focused on these progressive levels for two key validation purposes. First, for the purposes of the correlational study, the SRI team hypothesized that the student performance on the simpler DSA assessment tasks would highly correlate with their performance on assessment tasks from existing domain specific assessment instruments. Second, the SRI team also hypothesized that the simpler tasks would be sensitive to the knowledge of students with fewer courses in biology and economics, while the more complex tasks would be sensitive to the knowledge of students who had completed more courses. This progression is critical to the DSA assessment's underlying theory that schematic knowledge is essentially relational in structure. If a student scores well on the simple knowledge items, but not on the more complex knowledge items, it provides evidence that these forms of knowledge are distinct and may require different conditions for instruction and learning.

The Level 6 items were called "strategic" because they did not provide any hints to the student about the knowledge element or big idea that needed to be applied. The SRI team drew on the knowledge of Li (2002) to define the characteristic features of strategic knowledge tasks. To distance the strategic items from the related schematic items—which might have given students a hint about the relevant concepts—the strategic items were presented in separate test sections called "Additional Items".

Characteristic Features of DSA Assessment Tasks

The characteristic features are potentially the most re-usable elements of the ECD documentation process. "Characteristic task features" are aspects of the assessment situation that are likely to evoke the desired response. These features help ensure that the construct of interest will be measured and not "underrepresented." These are the ingredients that instructors may follow or use to construct assessments to measure the different forms of knowledge in their own classes. Or, they may compare their existing assessments to a list of these features to see what kinds of knowledge they are emphasizing through their classroom tests. As might be expected, the characteristic features of the DSA assessment tasks that measure simple discrete knowledge are likely to be familiar to most instructors and assessment designers—they usually require students to recall a discrete term or fact.

One of the DSA project's primary goals focused on defining the characteristic features of assessments that measure complex schematic knowledge and strategic knowledge. There were several distinguishing characteristic features of these tasks. One was mentioned above: Not hinting what specific big idea was in play in a prompt to elicit strategic knowledge. Some other features may be seen in the examples in Table 2. This table provides some context for prompt features: a sample item and a listing the knowledge and skills measured.

Table 2. Sample DSA Items, Target Knowledge Descriptions, and Prompt Features

Sample DSA Biology Item	Knowledge Measured	Characteristic Features
An old wooden shed in the park is falling apart and must be removed to prevent accidents. Two park rangers discuss how they could dispose of the shed while minimizing the carbon dioxide emitted into the environment. One park ranger says they should just take all of the old wood and bury it. The other ranger says they should burn the wood. Who has the better idea for disposing the shed with minimum impact on carbon emissions? Make a scientific argument to justify your choice.	Ability to predict how carbon-based life forms transform in the soil through metabolic processes of organisms. Student can identify products, substances used, cellular processes of transformation. Students understand these processes and interactions and their interactions on the organism level, the environmental level and the cellular level.	**Stimulus:** Describe two opposing solutions to a real world problem. One solution builds on natural processes and the other on parallel human-made processes, which often accelerate the natural processes. **Prompt:** Support a scientific recommendation for either one plan or the other.

Sample DSA Economics Item	Knowledge Measured	Characteristic Features
As Carla considers what she should do, she talks with her father who is an economist. "You are confronting the essential economics problem," her father tells her. What does her father mean? Explain, using the economic terms "scarcity" or "scarce."	Is able to explain why one faces tradeoffs in relation to the concept of scarcity	**Stimulus:** Description of a decision making situation with mutually exclusive action options **Prompt:** Why is X facing an essential economics problem? Explain, using the economic terms "scarcity" or "scarce"

Validation Study Plans

The final phase of the DSA project has begun. The status report on each element is briefly described below:

Expert Panel Study: The expert panel has been conducted to see how the assessments tasks within each instrument—the prototype domain-specific assessment, a college biology content knowledge test, a college economics content knowledge test, and a college critical reasoning test—measured different kinds of knowledge: simple discrete knowledge (e.g, facts, terms), simple relational knowledge (e.g., a couple ideas in relation), and more complex schematic knowledge (e.g., three or more ideas in relation), and strategic reasoning (e.g., knowing when and how to apply knowledge to a situation). Each instrument was presented in its entirety. The DSA includes 8 task scenarios (4 biology, 4 economics), which comprise 41 biology items and 47 economics items. The team published test booklets that included the domain-specific assessment tasks, mostly short-answer scenario-based tasks, with multiple choice items from the three other tests: 67 items from the biology test (Test C), 30 items from the economics test (Test C), and 40 items from the general critical thinking test (Test B). These tests were included to examine and contrast the forms of knowledge that each measures. The SRI team sought to recruit 300 students, and recruited 296 students (148 biology, 148 economics) from a single Northern California community college to participate in the validity-testing phase.

The items were presented in blocks that were alternated. Each expert received a different ordering of the blocks. During two one-day rating sessions, the four experts in each domain were asked to rate each of the three assessment instruments' tasks according to the following dimensions (see Appendix A for rating protocol):

1) What forms of knowledge students would use to answer each assessment task: simple discrete knowledge, multiple discrete knowledge, simple relational knowledge, and complex relational knowledge

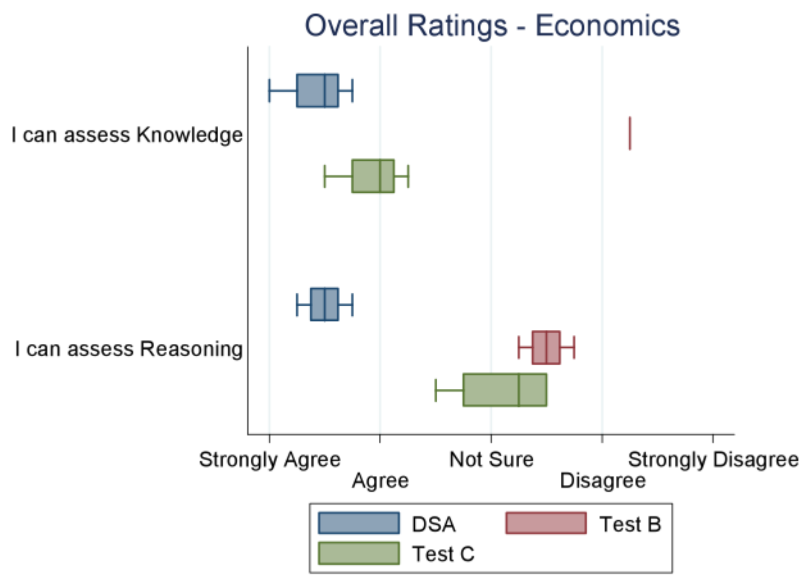
(2) What forms of reasoning students would need to answer each assessment task: Biology (Argumentation; Data interpretation; Field/Lab procedures; Computation; Data representation creation) and Economics (Use of narrative; graphs; models; argumentation)

After a block of assessment tasks from a single instrument were rated, the raters were asked for a holistic judgment of what they could learn about student knowledge from the set of assessment tasks: "I would have a good sense of how much a student knows in the domain," and "I would have a good sense of how much a student thinks through problems in the domain."

Expert Panel Findings

In biology, experts saw key differences between DSA and the other domain assessment instrument, which mainly measured simple or multiple discrete knowledge, and the DSA, which mainly measured simple or complex relational knowledge. In economics, the experts saw more similarities between the knowledge complexity measured by the DSA and the other domain assessment instrument. In biology, experts rated the DSA as testing student reasoning more than the other instrument. In economics, the experts rated both DSA ad the other domain assessment instrument as testing student reasoning within the model—an area where the alternative instrument surpassed the DSA—but gave the DSA higher ratings for measuring student capacity to generate and create representations and arguments and to extract information from graphs. The tables below show the comparisons experts made about the utility of the different assessments for future study in the fields and everyday life, and their ratings on what the assessments told them about student understanding. The experts judged the DSA to be particularly informative about student knowledge with applicability to everyday life and about student reasoning skills (see Figures 2, 3, 4 and 5).

Figure 2. Overall ratings of economics domain knowledge and reasoning measured by three tests*



* For all figures, DSA denotes Domain Specific Assessment, Test B denotes the general critical reasoning test, and Test C denotes an existing domain knowledge assessment instrument

Figure 3. Overall ratings of biology domain knowledge and reasoning measured by three tests

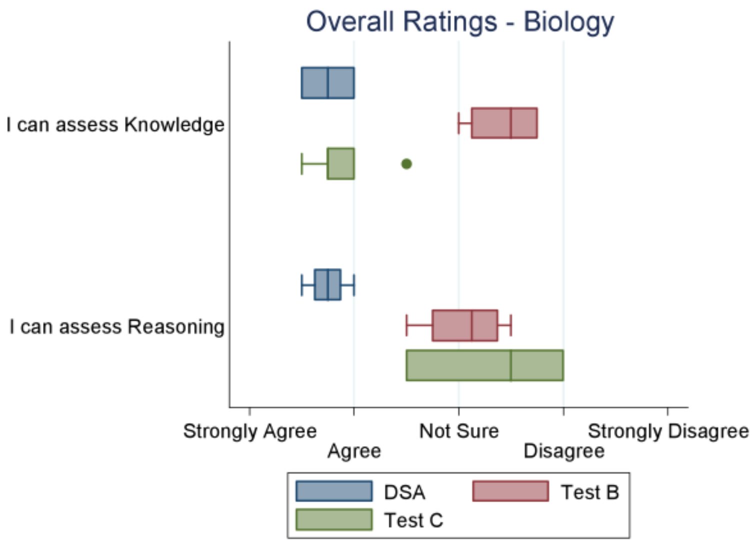


Figure 4. Overall ratings of usefulness of biology domain knowledge measured by three tests

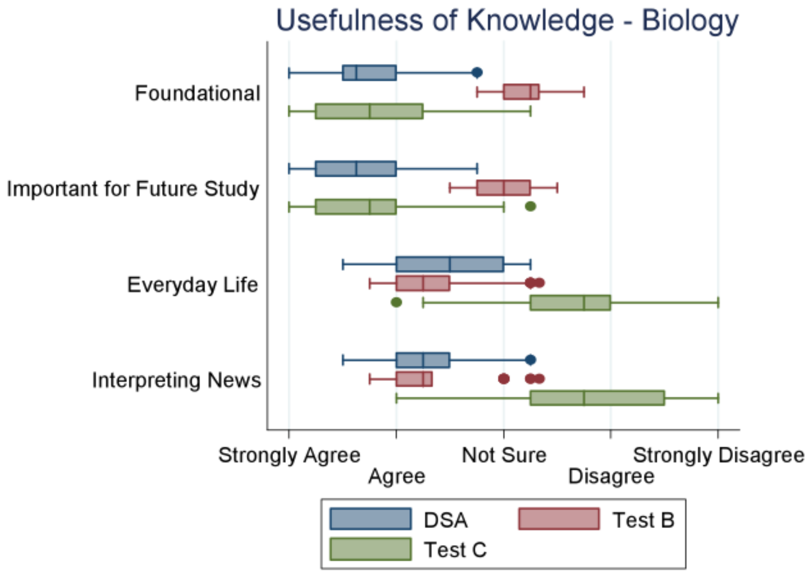
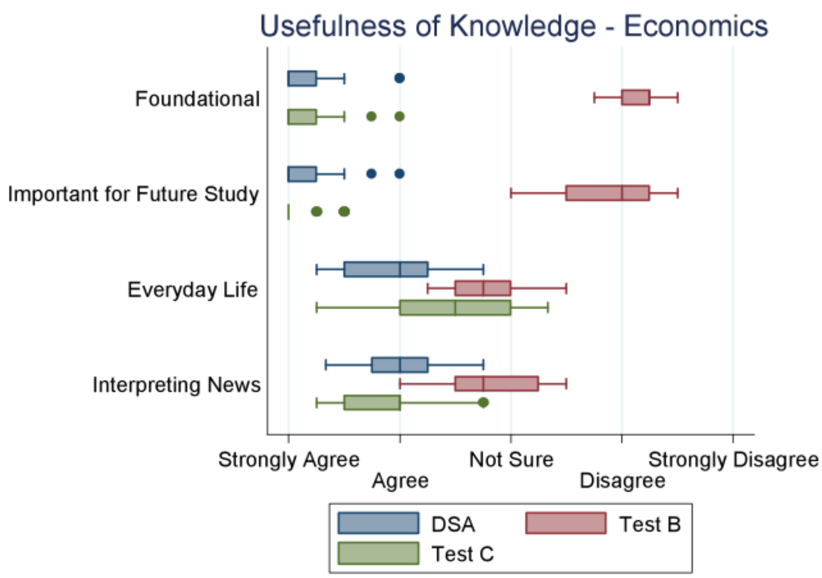


Figure 5. Overall ratings of usefulness of economics domain knowledge measured by three tests



Cognitive Think-aloud Study. The validation studies will continue with a cognitive analysis of student thinking—a complement to the expert panel's ratings of the kind of knowledge they believe each assessment elicits. These studies will examine the reasoning of a sample of 10 community college students (five from economics, five from biology) who have completed the basic sequence of required courses and scored grades of A to B+ in those classes. They are being recorded as they solve problems aloud in the DSA assessment and the other two tests (e.g., B and C). Expert scorers blind to condition will rate the kinds of reasoning and knowledge elicited by the assessment instruments.

Correlational and Instructional Sensitivity Analyses. The validation studies will conclude with correlational analyses of student performances across the three tests and an instructional sensitivity analysis that compares these instruments in their ability to reflect opportunity to learn the subject matter.

Discussion

The early results from the expert panel generally confirmed that the DSA assessments appear to measure the forms of knowledge they have been designed to measure. In addition, the experts judged the DSA assessments, both the biology and the economics exams, as providing important information about a student's knowledge of a subject area and the best information about the quality of students' reasoning as compared to other existing assessments of domain-specific knowledge and critical thinking. It does indeed appear that the domain-specific assessments capture the ways that students relate different big ideas relevant to a field and employ the reasoning skills useful in the domain.

Future validation studies will either support or qualify this initial finding. With respect to the forms of knowledge measured, the SRI team has found that both biology and economics are forms of model-based reasoning (Mislevy & Risconcente, 2006). Each discipline's pedagogical cultures have led to different opportunities to learn about these models and the representations and reasoning procedures around them. For example, in biology, students devote more time to learning the rudiments of highly complex models, but not as much time to applying these models to real world situations in narrative or representational forms. In economics, by contrast, students learn more basic models and devote more time to applying these models to real world situations in narrative or representational forms. Both the prompts and rubrics for student reasoning take into account the differences in opportunity to learn about these models and ways of manipulating them and reasoning with them.

The faculty members who scored the DSA student tests and sat on our expert panel have expressed endorsements for the overall approach provided additional helpful ideas on how to refine the characteristic features of the assessments. One unexpected byproduct of the work was some insight into how to facilitate discussions with community college faculty members around student assessments and think alouds. In our work, we found individual faculty members responded positively to seeing actual student work and they also found the project work to be illuminating for their own practice. This experience suggests that simplified versions of the processes and tools we are using for formal instrument development might be created to serve the needs of institutional professional development, particularly around assessment and accountability, topics that typically do not arouse enthusiasm or consensus among faculty members.

References

Li, M. (2002). *A framework for science achievement and its link to test items.* (Doctoral dissertation). Retrived from UMI (3038116).

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp. 61-90). Mahwah, NJ: Erlbaum.

Nagashima, S., Brown, N. J. S., Fu, A., Timms, M., & Wilson, M. (2008, March). *A framework for analyzing reasoning in written assessments.* Paper presented at the Annual Meeting of the American Educational Research Association, NY, NY.

National Research Council. (2003). E*ngaging Schools: Fostering High School Students' Motivation to Learn.* Washington, DC: National Academies Press.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6): 716-730.

Yarnall, L., Haertel, G., & Gallager, L. (2009). *Measuring how a college education lays the foundation for "thinking like an expert."* SRI International, Menlo Park, CA. (http://ctl.sri.com/publications/displayPublication.jsp?ID=797)

Appendix A. Expert Panel Rating Protocol

Economics Expert Panel Item Rating Sheet - Page 1

DSA - Economics Expert Panel Item Rating Sheet

1. DSA Economics Item Coding

Domain Specific Assessment Economics Rating Sheet

Rater

- ○ Mark
- ○ Betty
- ○ Sam
- ○ Tony

Item Block

- ○ Training
- ○ 7
- ○ 1
- ○ 8
- ○ 2
- ○ 9
- ○ 3
- ○ 10
- ○ 4
- ○ 11
- ○ 5
- ○ 12
- ○ 6

Item ID

Discrete Domain Knowledge

Choose one:

- ○ No discrete knowledge used
- ○ Simple (1 fact/definition/identification)
- ○ Multiple components of knowledge (multiple facts/definitions/identifications)

Economics Expert Panel Item Rating Sheet - Page 2

DSA - Economics Expert Panel Item Rating Sheet

Relational Domain Knowledge

Choose one:

- ○ No relational knowledge used
- ○ Simple (relates 2 concepts)
- ○ Multiple Simple (multiple pairs of concepts)
- ○ Complex (relates 3 + concepts)

Reasoning

Select all that apply:

- ☐ Extracting relevant information from tables and graphs
- ☐ Creating data representations (accuracy)
- ☐ Making computations (accuracy)
- ☐ Producing an argument or explanation (claim, evidence, theory that links evidence)
- ☐ Reasoning within a model (explaining and predicting without application, in the abstract)
- ☐ Applying a model to a specific situation (framing, explaining and predicting, assessing assumptions behind model)

Overall Importance

Please indicate the extent to which you agree or disagree with the statements below

The knowledge being assessed in this item is:

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
Foundational in the field	○	○	○	○	○
Important for future study in the major	○	○	○	○	○
Useful in everyday life decisions	○	○	○	○	○
Helpful for interpreting the news	○	○	○	○	○

Economics Expert Panel Item Rating Sheet - Page 3

DSA - Economics Expert Panel Item Rating Sheet

Academic Level

Where in the curriculum would a college student typically be learning these ideas?

Select all that apply.

- [ ] First 2 years
- [ ] Major level
- [ ] Graduate school
- [ ] Not typically in the curriculum

If you have any observations you would like to record, please do so here.

When you complete your ratings for a Block, go to Overall Impressions.

Economics Overall Importance - Page 1

DSA Economics Overall Importance

1. Default Section

Rater

- ◯ Mark
- ◯ Betty
- ◯ Sam
- ◯ Tony

Item Block

- ◯ Training
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10
- ◯ 11
- ◯ 12

Overall Importance

After reviewing this block of items, how much do you agree with the following statements.

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
I would have a good sense of how much a student knows in the domain.	◯	◯	◯	◯	◯
I would have a good sense of how a student thinks through problems in the domain.	◯	◯	◯	◯	◯