Using PADI Templates as an Alternative Structure for Specifying GLOBE Investigation Strategies

Angela Haydel DeBarger
SRI International

Futoshi Yumoto
University of Maryland

Edys Quellmalz
SRI International

Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April 2005.

DRAFT: DO NOT CITE OR QUOTE

Using PADI Templates as an Alternative Structure for Specifying GLOBE Investigation Strategies

The Principled Assessment Designs for Inquiry (PADI) design system employs evidence-centered design (Messick, 1994; Mislevy, Steinberg, and Almond, 2003) as the foundation for guiding assessment development. As a use-case for the PADI design system, classroom assessment resources developed for the Global Learning to Benefit the Environment (GLOBE) program were *reverse-engineered* or translated into PADI design system components. Multiple design patterns and a task template were developed to shape assessments of the inquiry skills addressed in GLOBE. This paper briefly describes the process by which the GLOBE task template was reverse-engineered from the extant design and samples for assessing the GLOBE earth science concepts and investigation strategies.

GLOBE is a "worldwide hands-on, primary and secondary school-based education and science program" (http:www.globe.gov.). Students in GLOBE science classes take measurements related to earth science content investigation areas such as hydrology, atmosphere, and soils. These data are submitted to global data archives maintained by agencies such as NASA, NOAA, and EPA. The data are intended to be useable by scientists and also by students and teachers who participate in the GLOBE program. While working with the GLOBE program, SRI International conducted a project funded by the National Science Foundation to development classroom assessment tools for teachers. The assessment project developed an assessment framework and procedures for aligning GLOBE content and inquiry to national, international and state science standards. In addition the project designed a template to guide teachers' development of classroom assessments and sets of sample assessments for use at the elementary, middle, and secondary levels to test students' use of GLOBE data collection protocols, concepts, and investigation strategies focused on using GLOBE data to solve authentic problems. The project developed generic scoring rubrics and task specific scoring keys for the sample assessments. The GLOBE task template developed in the PADI system can be used to support both the design of GLOBE assessments and the design of other science inquiry assessments.

GLOBE is a particularly important use-case because it helps to demonstrate the capability of the PADI design system to accommodate assessments that address multiple components of science inquiry such as planning, data analysis, data interpretation, and communication. The GLOBE use-case is also a unique illustration of task template development because the reverse-engineering was derived in part from a multi-layered assessment design system developed for the GLOBE program that included an assessment framework, template, and sample performance assessment tasks and rubrics.

This paper begins by providing background on the PADI approach and GLOBE assessments. We then discuss the process by which GLOBE assessment resources were reverse-engineered into a PADI task template. The final section of the paper presents the next steps for GLOBE assessment design using the PADI.

PADI Approach

The PADI project is defining and implementing a set of schemas in order to enhance assessment developers' ability to create complex assessments of student learning. The evidenced-centered design (ECD) approach that guides PADI articulates the interrelationships among substantive arguments, assessment designs, and operational processes. ECD facilitates reasoning about the knowledge, skills, and abilities that should be assessed to the behaviors that would reveal evidence about those constructs and the tasks that will elicit those behaviors (Messick, 1994). PADI goes further to identify structures that capture commonalities across a set of problems or situations at different stages of the assessment process. These structures support the design and implementation of assessments that may vary greatly in terms of surface features, but retain an underlying assessment argument that links the inference of interest to the evidence garnered in its support. This assessment argument is instantiated through *design patterns*. PADI *task templates* facilitate the ECD approach in the articulation of the more detailed nuts and bolts of assessment components. Figure 1 illustrates the layers of evidence-centered assessment design.

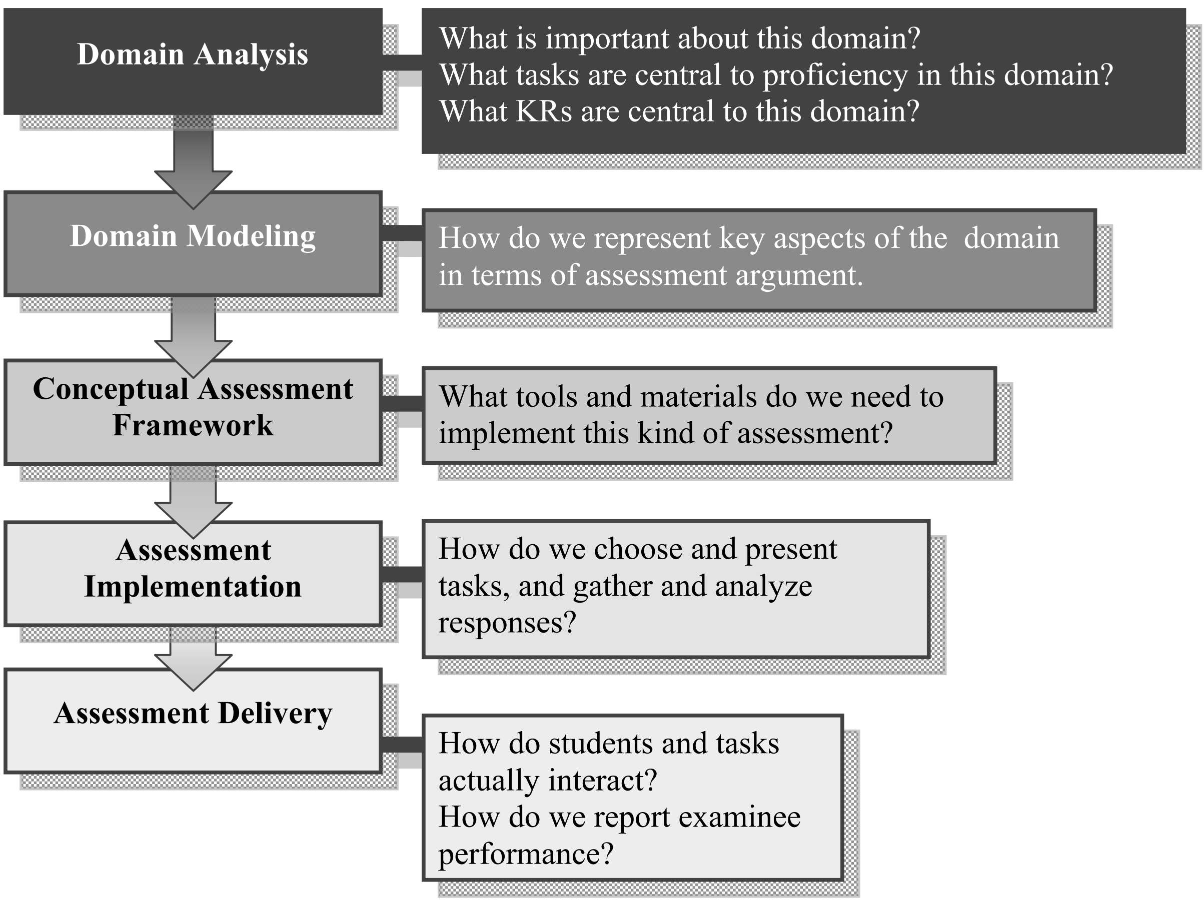


Figure 1. Graphic Representation of ECD Layers (Mislevy & Riconscente, 2005)

Task Templates. Applying the principles of evidence-centered design, PADI task templates link the student model, evidence model, and task model components of the conceptual assessment framework. They provide a structure for helping assessment designers identify the student model variables, stimulus materials, task model variables, evaluation procedures and measurement models. Task templates serve as "pre-blueprints" in that they are abstractions of related assessment tasks. A task specification is more analogous to an assessment blueprint because it is used to generate a specific assessment task. Thus, components of the task template are left unspecified until the task template becomes a task specification. For instance, the levels for a task model variable about scaffolding (e.g., high, medium, low) will be identified in a task template, and this scaffolding variable will be set to one of these levels when the task specification is created. Figure 2 shows the relationship among objects in the template.

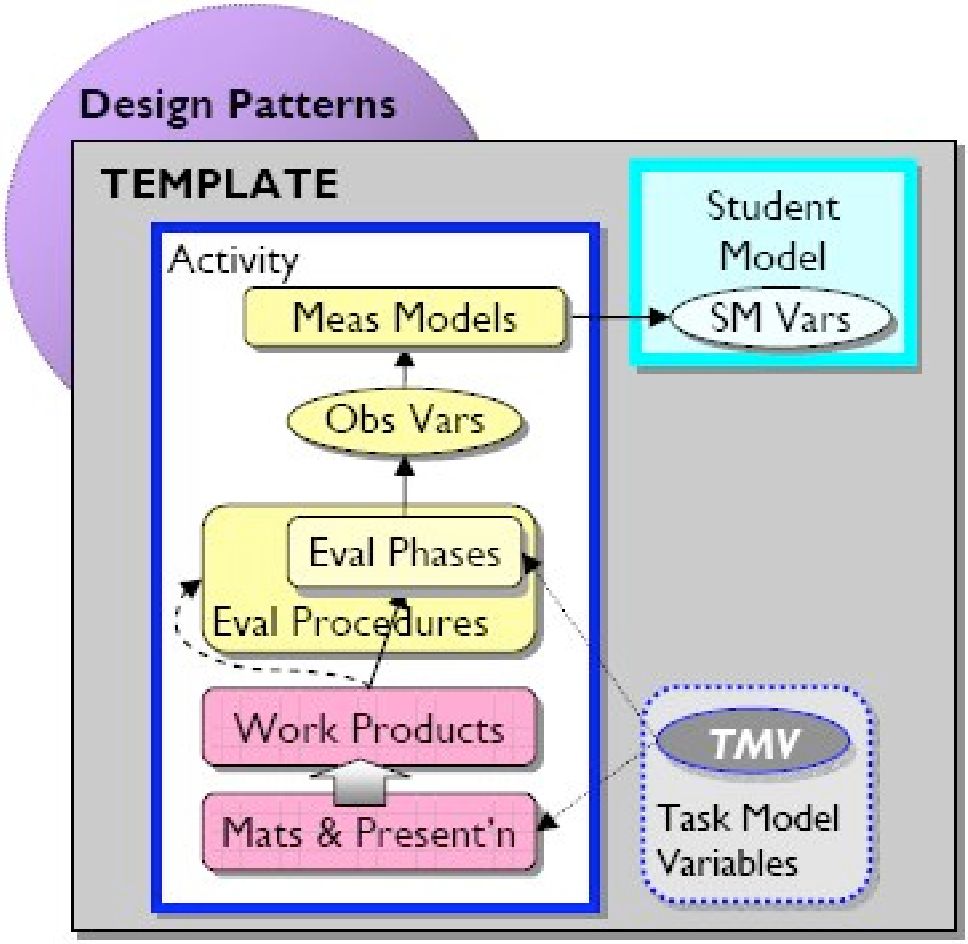


Figure 2. Relations among Template Objects

GLOBE Assessments

The GLOBE performance assessments of investigation strategies are uniquely complex compared to assessments that are typically administered in science classes or on state and national science examinations. GLOBE assessments address multiple components of science inquiry and require students to integrate inquiry skills as they move from data analysis to data interpretation and communication. Designed to assess students' abilities to use GLOBE environmental data to solve authentic problems, the GLOBE assessments draw from large datasets collected by students throughout the world. Thus, analyses of data trends can be explored more readily, and students can compare data from different sites around the world. The GLOBE assessment framework, general template, and generic rubric developed by SRI International help to articulate the approach used in the development of assessments for the GLOBE program.

The GLOBE assessment framework includes the six earth science investigation areas addressed in the GLOBE program and five investigation strategies. This framework is shown in Table 1 and was used to guide the development of the GLOBE template for classroom assessment, generic rubrics, and sample elementary, middle, and secondary level assessment in each area (e.g., hydrology, land cover). These assessment resources were intended to assist teachers in developing tasks to assess their students' science inquiry skills using GLOBE data and can be found at http://www.globeassessment.sri.com and http://www.globe.gov.

Table 1
GLOBE Assessment Framework

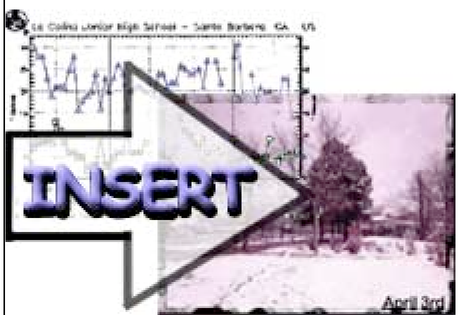
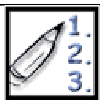
GLOBE Investigation Areas	GLOBE Investigation Strategies
Atmosphere Hydrology Landcover Soils Earth Systems Visualizations	*Plan investigations* • Set up a new, appropriate problem/application • Design an experiment • Specify measurements/variables to investigate • Pose relevant questions
	Take GLOBE measurements • Measurements are accurate and appropriate • Detect errors • Use quality assurance procedures (multiple, repeated readings, re-calibration)
	Analyze and compare GLOBE data • Identify data components • Identify similarities and differences • Explain reasons for differences • Use appropriate mathematical procedures
	Interpret GLOBE data • Infer patterns, trends • Explain data and relationships • Create multiple formats to represent data
	Communicate • Compose informal and informal reports to explain or persuade • Create and make presentations of key conclusions and finding

GLOBE Template for Classroom Assessment

The GLOBE template for classroom assessment identifies eight tasks that can be used to assess students' understanding of the GLOBE investigation strategies for any of the investigation areas. This template is shown in Table 2. Each GLOBE assessment begins with the presentation of GLOBE data drawn from the GLOBE data archives, and students are presented with a scenario including a problem situation or driving question that relates to the data provided. For example, if students are given data tables about soil moisture and temperature, they may be asked to use the data to make a recommendation for where to introduce a native plant species. Students are asked to develop one or more research questions that they could investigate with the data. Students are asked to find trends in the data, identify measurement errors and explain how to minimize them, re-express the data in a new representational form (e.g., use data presented in a table to create a graph), and interpret the data. Students also may be asked to find relevant data from another school in the GLOBE database and perform similar analyses. Finally, students are asked to communicate their findings related to the problem situation in a brief report or presentation.

Table 2

GLOBE Template for Classroom Assessment

List School Information	
	**Insert GLOBE Data or Graphs for the Schools Listed Above**
	**Present problem requiring use of GLOBE data archives** Present problem situation/driving question with background and role of the student in the investigation.
	**Planning Investigations: Ask students to pose relevant questions** Ask the student to look at the GLOBE data/graphs provided above and come up with possible questions that she/he might ask regarding the data. Provide a sample question to help guide the student.
	**Analyzing and interpreting GLOBE data: Ask questions about data in the table** Ask the student to find observable trends in the data.
	**Conducting Investigations: Assuring data quality** Ask student to look through the data for possible measurement or data entry errors and suggest ways to avoid these types of errors in the future.
	**Analyzing and Interpreting Data: Ask questions requiring interpretation of data** Ask the student what the relationship is between the two variables given. Provide the student with an example of a trend.
	**Analyzing and Interpreting Data: Ask to represent data in a graph or table** Ask the student to use the data provided to generate new data representations to analyze trends.
	**Analyzing and Interpreting Data: Ask for interpretation of data on the graph** Ask student about specific features of graphs and what indications there are for various maximums, minimums, etc. Ask them to explain their answer.
	**Planning Investigations: Ask to set up another problem** Ask student to choose another school from the GLOBE database that has some related feature as the school(s) they just analyzed. Have them copy a relevant data set for this new school and to perform a similar analysis on this new data set. Ask them what other variables they would be interested in looking at and why.
	**Summarizing Data: Ask to summarize and report findings** Ask the student to summarize their analysis of the original school(s) and to write a short report or to prepare a short presentation of their findings and recommendations, supporting their conclusions with the analysis they have done and to suggest other data that might be helpful for further study of the situation.

GLOBE Generic Rubric

The GLOBE generic rubric provides holistic criteria for scoring students' conceptual understanding and inquiry strategy use (see Table 3). The leftmost column lists the investigation content areas and investigation strategies. The columns to the right provide general descriptions of the scoring categories.

Table 3
GLOBE Generic Rubric

	Criteria	**1** Absent, inaccurate, confused	**2** Partially developed	**3** Adequately developed	**4** Fully developed
GLOBE Investigation Content					
Atmosphere Hydrology Landcover Soils Earth Systems Visualizations	Appropriate Accurate Sufficient	Concept is described incorrectly.	Content is depicted incompletely or with some errors. Definition may be vague, unexplained.	Conceptual definition and explanation is appropriate with some level of detail.	Describes and explains concept/principle accurately, explains relationships with the concept and with relevant other concepts.
GLOBE Investigation Strategies					
Plan investigations	Appropriate Accurate Sufficient	Design is inappropriate, cannot be followed, impractical.	Design is mostly appropriate, general, somewhat flawed. Plan is incomplete, addresses only part of the problem.	Design is appropriate, may have a minor flaw, lack some detail.	Proposes an organized design and logical procedures appropriate for answering the question.
Take GLOBE measurements	Appropriate Accurate Sufficient	Data readings and records are incorrect or disorganized. Does not detect errors. Does not identify appropriate procedures for quality assurance.	Minor errors in reading or recording data. Misses some errors.	Data readings are accurate, relevant to the question, may have a minor problem. Appropriate selection and use of quality assurance procedures.	Data readings are accurate and thorough, no errors. Quality assurance is thorough.

Table 3
GLOBE Generic Rubric (continued)

	Criteria	**1** Absent, inaccurate, confused	**2** Partially developed	**3** Adequately developed	**4** Fully developed
Analyze and compare GLOBE data	Appropriate Accurate Sufficient	Components may not be identified or may be inappropriate. Comparisons may be inappropriate or confusing.	Analysis may omit some components. Some features compared may be irrelevant or inappropriate. Explanations may be general, vague, or partially incorrect.	Summarizes data accurately and sufficiently. For comparisons, identifies relevant features and explanations for most similarities and/or differences.	Components of analysis are accurate and thorough Features compared are important and accurate. Explanations thoroughly justify conclusion.
Interpret GLOBE data	Appropriate Accurate Sufficient	Pattern identified is incorrect. Many errors in data. Attempts to represent data are inappropriate or confusing.	Pattern may be partially correct. Some, but not sufficient data. Some inaccurate or inappropriate data and representations. Tables, graphs have numerous errors.	Identifies sufficient data to support trend. Explains relationships among variables. Some explanation may be general. Creates appropriate data table or other format to represent data.	Identifies more than enough data to support trend or conclusion. Supplies thorough explanation of relationship of data to conclusion. Creates multiple representations of data
Communicate	Clear main point. Appropriate, accurate, sufficient supporting data and explanations. Conclusions and evidence are logically connected and sequenced. Appropriate use of scientific terminology. Relevant use of representations and graphics. Adherence to conventions of Standard Written English.	Conclusion is incorrect or confusing. Data cited are mostly inaccurate or irrelevant. Organization and logic are confusing. Inappropriate use of terminology.	Conclusion may be general or somewhat flawed. Data may be cited but inadequate to support the conclusion. Explanations may be vague, incomplete. Some errors in terminology and data displays.	Report summarizes main point or conclusion, investigation procedures, how data support conclusion. Most data is relevant. Explanations are appropriate, some may be general. Representations of data are appropriate. Logical connections are made from data to conclusions.	Report identifies the issue or research problem, conclusion and presents ample, accurate data to support the conclusion. Appropriate and effective representations of data to support conclusion.

Sample GLOBE Task

A sample GLOBE task that addresses the hydrology investigation content area at the middle school level is presented in the Appendix. Each item relates to an aspect of the GLOBE classroom assessment template, as indicated by boldface text in parentheses. Additional samples of GLOBE assessments in different content areas and at different grade levels can be found at http://www.globeassessment.sri.com.

Representing GLOBE Assessments in the PADI Design System

Using the GLOBE assessment resources described in the previous section, a task template was developed using the PADI Design System. In this section of the paper we present the GLOBE task template and discuss the decision-making process that guided the reverse-engineering of these GLOBE resources into a PADI task template. We then discuss the affordances and limitations of the PADI system for designing GLOBE assessments and similar multi-component inquiry assessments.

GLOBE Template Development Process

The GLOBE template development team included experts in educational psychology, assessment design, and measurement. The first step in developing the GLOBE template was to review the GLOBE assessment framework, classroom assessment template, sample tasks, and rubrics. In some instances, we found it necessary to create diagrams and tables to determine how GLOBE assessment elements might be translated into components of a PADI template. These diagrams and tables helped us to clarify how elements of the GLOBE assessment resources (e.g., item, rubric, investigation phases in classroom assessment template) could be represented as objects in the PADI design system (e.g., materials and presentation, work product, observable variable). When the GLOBE template design group reached consensus about how GLOBE assessment elements should be represented, we generated the objects in the PADI design system.

Specification of Activities. The GLOBE classroom assessment template (see Table 2) identifies eight phases related to the investigation strategies. Because these eight phases define the core tasks in most GLOBE assessments, regardless of grade level or the content of the assessment (e.g., atmosphere, hydrology), we represented these phases as activities in the PADI system. Thus, the GLOBE template instantiated in the PADI design system contains 8 activities: (1) Pose Research Questions; (2) Find Data Trends; (3) Assure Data Quality; (4) Infer Relationships Between Variables; (5) Re-express Data; (6) Identify Relevant Data and Values in Data Displays; (7) Describe Relevant Transfer Problem/Application; (8) Communicate Findings and Conclusions. This conceptualization of the GLOBE template, which includes these investigation phases represented as activities, will suggest to an assessment developer that a complete GLOBE assessment of integrated investigation strategies for any grade level or content area requires each of these activities. Each of these activities contains descriptions of stimulus materials, work products to be generated, evaluation procedures for scoring the work products and a measurement model for specifying the relationship between student model variables and the observable variables produced as part the evaluation of the student work product. These components of the task model, student model, and measurement model are discussed below.

Specification of Task Model. Template components related to the task model include materials and presentation, task model variables, and work products. Materials and presentation and task model variable objects appear both at the template level and at the activity level in the GLOBE template. Work products exist only at the activity-level. The template-level, or template-wide, objects relate in some way to all activities within the template. The activity-level objects relate only to a specific activity.

Materials and Presentation. As shown in the GLOBE classroom assessment template (Table 2) and sample assessment (Appendix), the GLOBE data and problem situation description are presented first, and students are asked to use them or refer to them as they complete each investigation phase. Because they are used by students for each activity, these became template-level materials and presentation in the PADI version of the GLOBE template. Activity-level materials and presentation were determined by reviewing sample GLOBE assessments to identify how students were prompted to complete the assessment tasks. The materials and presentation objects for each activity are presented in Table 5. In addition, to including prompts as stimulus materials in the template, activities 2-8 include a stimulus material called *Work Products and Findings from Previous GLOBE Activities*. This stimulus material was added for each of these activities because students may be using work products they have generated from previous activities as they complete subsequent activities.

Work Products. Sample GLOBE assessments, and to some extent the generic GLOBE rubric and GLOBE classroom assessment template, were used to specify the work products in the template. Based on this analysis of GLOBE assessment resources, we determined that students may generate one or more work products for each activity, depending on what they are prompted to do (i.e., the specification of materials and presentation objects). Work products for each activity are presented in Table 5, with their associated materials and presentation objects.

Table 5
Materials and Presentation and Work Products Associated with Each GLOBE Activity

Activity	Materials and Presentation	Work Products
Pose Research Questions	• Prompt to Pose Research Question(s) • Sample Response for Posing Research Questions	• Question(s) Based on GLOBE Data
Find Data Trends	• Prompt to Find Data Trends • Work Products and Findings from Previous GLOBE Activities	• Infer GLOBE Data Trends • Explanation of GLOBE Data Trends
Assure Data Quality	• Prompt to Review Data for Errors • Prompt to Describe How Identified Errors • Prompt to Describe Procedures for Collecting Quality Data • Sample Response for Assure Data Quality • Work Products and Findings from Previous GLOBE Activities	• Identified Errors in GLOBE Data • Description of Error Finding Procedures • Description of Procedures for Collecting Quality Data
Infer Relationships between Variables	• Prompt to Infer Relationship between Two Variables • Prompt to Explain the Mathematical or Scientific Basis for Why the Relationship Exists • Work Products and Findings from Previous GLOBE Activities	• Infer Relationship between Variables/Measurements in GLOBE Data • Explanation for Why the Relationship Exists
Re-express Data	• Prompt to Re-express Data • Work Products and Findings from Previous GLOBE Activities	• Transformation of GLOBE Data from One Format to Another Format
Identify Relevant Data and Values in Data Displays	• Prompt to Identify and Describe Values of Data • Work Products and Findings from Previous GLOBE Activities	• Identify and Describe GLOBE Data Components (e.g., read graphs)
Describe Relevant Transfer Problem/ Application	• Prompt to Identify Additional Dataset(s) • Prompt to Identify Variables to Investigate • Prompt to Find Data Trends and Relationships • Prompt to Explain the Mathematical or Scientific Basis for Why the Relationship Exists • Sample Response for Transfer Problem • Work Products and Findings from Previous GLOBE Activities	• Identification of Additional GLOBE Dataset(s) • Identification of Additional GLOBE Variables to Investigate • Identification of Relationship in GLOBE Data • Explanation for Why the Relationship Exists
Communicate Findings and Conclusions	• Prompt to Communicate Findings for GLOBE Investigation • Specification of Required Components or Criteria (for judging the quality of student response) • Work Products and Findings from Previous GLOBE Activities	• Communication of Findings Using GLOBE Data

Task Model Variables. Initially, we began developing task model variables that relate to specific materials and presentation objects for each activity. To develop these activity-level task model variables, multiple sample GLOBE assessments were examined to determine how features of the stimulus materials varied for tasks in different content areas (e.g. hydrology and atmosphere) and at different grade levels (i.e., elementary school, middle school, and high school). After generating activity-level task model variables, we began thinking about template-level task model variables, those that are likely to impact all activities similarly, such as overall scaffolding and task model variables related to the presentation of the GLOBE data.

The process of developing task model variables required multiple iterations to reduce redundancies among similar task model variables at the template-wide level and within each activity and to ensure clear and relevant relationships among materials and presentation objects and task model variables. Even after this revision and refining process, we still had over 30 task model variables. Given that there are eight activities and multiple features of the data that can be manipulated, this number of task model variables is fairly reasonable. However, a more novice assessment designer, or an assessment designer who is new to GLOBE, may be overwhelmed by having to set this many task model variables.

In order to address this concern, we created a hierarchy among the GLOBE task model variables to provide flexibility in supporting assessment designers with different levels of experience. With the support of a Wizard to guide users in the construction of GLOBE task specifications, a more novice designer may choose to set only the higher level task model (e.g., scaffolding) variables, which will automatically impact the settings of the lower-level task model variables (e.g., providing a sample question). A more experienced designer may choose to adjust all task model variables individually. These hierarchical relationships exist from template-to-template level task model variables, from template-to-activity level task model variables, and from activity-to-activity level task model variables. This is currently specified in the template as *kinds of me* relationships, in which the lower-level task model variables are identified as a *kinds of* the higher-level task model variables.

Figure 3 was developed by the GLOBE template development team to provide assistance in conceptualizing the hierarchical relationships among GLOBE task model variables. The task model variables in the box on the left are at the template level, and the task model variables in the in the large box on the right are at the activity level. (There are two to five task model variables associated with each activity.) The highest level task model variable is GLOBE general scaffolding/grade level (shown in the center of the box on the left). As illustrated by the arrows, this task model variable can influence both template-level task model variables, such as those related to the GLOBE data, and activity-level task model variables. Middle-level task model variables (e.g., GLOBE Data Scaffolding, Posing Research Questions (PRQ) Scaffolding) may be influenced by a higher-level task model, but also impact lower level task model variables. Lowest-level TMVs (e.g., Type of Data Transformation in the Re-express Data activity) do not influence any other TMVs.

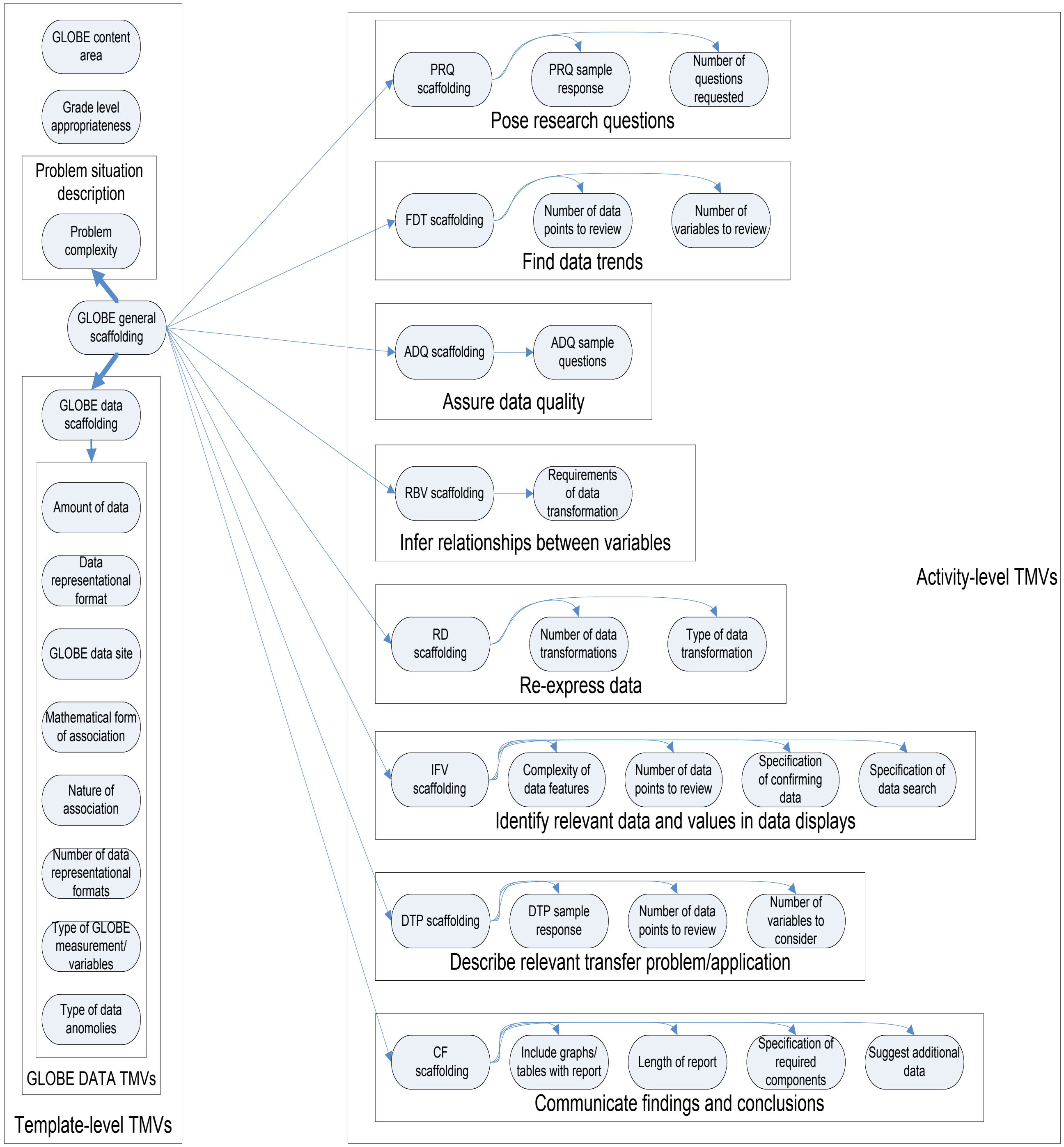


Figure 3. Hierarchical Relationships among Task Model Variables

Specification of Student Model

The student model organizes one or more student model variables that provide information about student proficiency in the content and skills addressed by the assessment task. The student model for the GLOBE template is based on the GLOBE assessment framework

(Table 1). The assessment framework identified five investigation strategies or inquiry skills (e.g., interpret trends, communicate) and six content/investigation areas (e.g., soils, land cover). Each of these inquiry skills and content areas are expressed as student model variables in the GLOBE template: (1) Ability to Plan GLOBE Investigations; (2) Ability to Take GLOBE Measurements; (3) Ability to Analyze and Compare GLOBE Data; (4) Ability to Interpret GLOBE Data; (5) Ability to Communicate Scientific Findings; (6) Ability to Understand GLOBE Atmosphere Concepts; (7) Ability to Understand GLOBE Hydrology Concepts; (8) Ability to Understand GLOBE Land Cover Concepts; (9) Ability to Understand GLOBE Soils Concepts; (10) Ability to Understand GLOBE Earth Systems Concepts; and (11) Ability to Understand GLOBE Visualization Concepts. These 11 student model variables can be grouped to form a multidimensional student model, which can provide relatively detailed diagnostic information about student proficiency on the inquiry and content dimensions. We created second multidimensional student model that includes a single student model variable to provide information about students' proficiency in inquiry and a single student model variable to provide information about students' proficiency in the content. We also created a unidimensional student model that includes one student model variable, which provides information about students' inquiry skill and content knowledge combined.

Providing different student models that contain student model variables at different grain-sizes helps to meet the diagnostic and evaluative needs of a variety of assessment developers. When generating a GLOBE task specification, the assessment designer will need to determine which of these three student models to use. Finer grain-size student model variables can provide more detailed information about students' performance. Coarser grain-size student model variables may be more suitable for summative evaluative purposes. It is also important to note that the precision of finer grain-size student model variables can be less accurate than coarser grain-size student model variables depending on the amount of data available to update each student model variable.

Specification of Evaluation Procedures and Measurement Model

The scoring of students' work products is specified in the evaluation procedures. The evaluation procedures may contain a sequence of evaluation phases, if multiple aspects of student work are to be scored. Within each evaluation phase, observable variables are identified. These are the scores that provide information about the quality of students' work products. An observable variable or combinations of observable variables will be used to update the student model variables. The relationship between a single observable variable and one or more student model variables is specified in the measurement model. If multiple observable variables are produced within an activity, and these responses are determined to be dependent, they must be *bundled*. The process of bundling will result in the specification of a single observable variable that will be linked to one or more student model variables with the psychometric model. The score levels on this bundled observable variable reflect combinations of the score levels from the observable variables used to create the bundle.

Within GLOBE, several activities produce multiple work products, as shown in Table 5. These work products are scored separately. However, these work products are generated from the same stimulus materials and may require similar knowledge and skills. For instance, in the Assure Data Quality activity, three work products are produced: Identified Errors in GLOBE Data, Description of Error Finding Procedures, and Description of Procedures for Collecting Quality Data. Students receive scores on: Identifying Errors, Identifying the Type of Errors, Explanation of Errors, and Describing a Strategy for Finding Errors, Describing Data Quality Procedures, and Describing Measurement Procedures. Because describing the type of errors, explaining why they are errors, and describing the strategy for finding the errors are dependent on being able to identify the errors, these scores must be bundled into a single observable variable or score. Similarly, describing data quality procedures and measurement procedures are related, and as a result, are bundled.

The advantage of producing finer grain-size observable variables and then bundling, rather than developing a more holistic, coarser grain-size observable variable initially, is that finer grain-size observable variables can be reported to teachers and students for formative purposes. In addition, teachers or scorers are forced to think about the dimensions of student responses in more detail.

Table 6 provides information about the observable variables for each activity, how they may be bundled, and the student model variables to which they relate. This table represents a simplified version of multiple spreadsheets that were used by GLOBE template development team to articulate relationships among work products, evaluation phases, observable variables, and student model variables prior to creating the measurement model objects in the PADI design system.

Using Task Model Variables to Influence Evaluation Procedures. As we thought through the procedures for evaluating student work for GLOBE and reviewed examples of GLOBE assessments, we determined that task model variable settings may influence evaluation procedures. This is particularly true for the scaffolding task model variable. Greater scaffolding may limit the amount of content knowledge required of student or may provide additional information about which data to use for the investigation. As a result, the work products generated would differ, as would the evaluation procedures that specify how student work should be scored. Figure 4 shows the effect of the scaffolding task model variable where the connection with ‘L’ exists only when the scaffolding level is low. In this figure, the connection to the content knowledge SMV is added for the low scaffolding level.

Table 6
GLOBE Measurement Models: Relationships among Observable Variables and Student Model Variables

Activity	**Observable Variables**	**Student Model Variables**
Posing Research Questions	Score for GLOBE Posing Research Questions	Ability to Plan GLOBE Investigations Content Knowledge
Find Data Trends	Bundled OV (combine Score for Identification of Trends and Score Explanation of Data Trends)	Ability to Interpret GLOBE Data Content Knowledge
Assure Data Quality	Bundled OV1 (combine Score for Identifying Errors, Score for Identifying Type of Errors, Score for Explanation of Errors, and Score for Strategy of Finding Errors)v	Ability to Take GLOBE Measurements Ability to Analyze and Compare GLOBE Data Content Knowledge
	Bundled OV2 (combine Score for Describing Data Quality Procedures and Score for Describing Measurement Procedures)	Ability to Take GLOBE Measurements
Infer Relationships between Variables	Bundled OV (combine Score for Data Interpretation of Each Variable, Score for Description of Relationship, and Score for Scientific or Mathematical Explanation of Relationship)	Ability to Analyze and Compare GLOBE Data Ability to Interpret GLOBE Data Content Knowledge
Re-express Data	Bundled OV (combine Score for Appropriateness of Data Representational Format, Score for Data Placement on New Format, and Score for Mathematical Transformations of Data)	Ability to Analyze and Compare GLOBE Data Ability to Interpret GLOBE Data Content Knowledge
Identify Relevant Data and Values in Data Displays	Score for Accuracy of Identified Measurements and Values	Ability to Analyze and Compare GLOBE Data
Describe Relevant Transfer Problem	Bundled OV (combine Score for Identification of Additional GLOBE Dataset, Score for Identification of Additional Variables to Analyze, Score for Identification of Relationships in Data, and Score for Explanation of Relationships in Data)	Ability to Plan GLOBE Investigations Ability to Analyze and Compare GLOBE Data Content Knowledge
Communicate Findings and Conclusions	Bundled OV (combine Score for Use of Representations to Communicate Findings, Score for Use of Data to Explain Findings, Score for Argument Coherence, and Score for Application of Content Knowledge)	Ability to Communicate Scientific Findings Content Knowledge

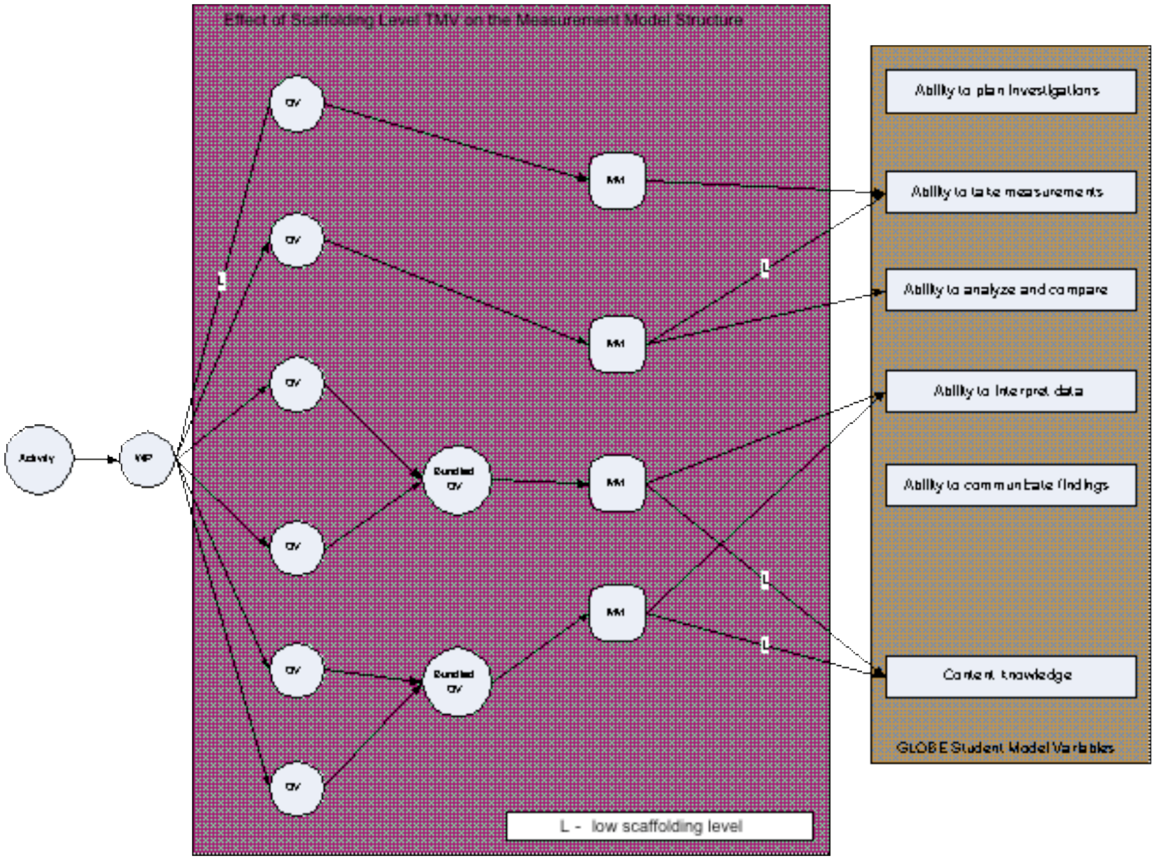


Figure 4. Effect of Scaffolding on Measurement Models

Affordances and Limitations of the PADI System for Designing GLOBE Assessments

The PADI design system is well-equipped to handle the complexities of a multiple-component inquiry task, such as a GLOBE integrated investigation assessment. The design system provides flexibility to allow for the specification of multiple student model variables, hierarchical relationships among task model variables, and multiple activities that contain work products, evaluation procedures and measurement models that vary in important ways to reflect different aspects of science inquiry. In addition, the design system has the added advantage of allowing for immediate updating, and thus, changes to the template can be viewed almost instantaneously by multiple designers who are collaborating from different locations.

One of the limitations of the design system is that it does not track changes to the template components—once a change has been made to an object, any previous versions of the object do not exist unless the object was duplicated prior to the change. In addition, the current layout of the template does not allow users to simultaneously view relationships among objects across different activities (e.g., look at patterns in relationships between observable variables and student model variables for multiple activities). In order to address these limitations, we created working documents, such as Excel spreadsheets (e.g., Table 6) and other diagrams (see Figures 3 and 4) to help us visualize the relationships expressed in the template. Future versions of the

PADI template may include additional visual representations to support assessment designers in conceptualizing the relationships among template components.

Future Directions

Our next steps are to develop GLOBE task specifications and to develop a GLOBE Wizard. GLOBE task specifications will illustrate how template components can be set and used to generate particular GLOBE assessments (e.g., assessments of investigation strategies within hydrology or land cover). The GLOBE Wizard will assist other assessment designers in creating more specialized GLOBE templates and task specifications. As we continue to develop GLOBE templates and task specifications, we will document the design process in order to facilitate the design of other complex inquiry assessments using the PADI design system.

References

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 32(2)*, 13-23.

Mislevy, R. J., & Riconscente, M. M. (in press). Evidence-centered assessment design. In S. Downing & T. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Erlbaum.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-67.

Appendix

Sample GLOBE Assessment