P R I N C I P L E D A S S E S S M E N T D E S I G N S F O R I N Q U I R Y
T E C H N I C A L R E P O R T 1 8

An Illustration of PADI Design System Capability with GLOBE Assessments

Prepared by:
Angela Haydel DeBarger, SRI International
Futoshi Yumoto, University of Maryland
Edys Quellmalz, SRI International

DRAFT

Acknowledgment
This material is based on work supported by the National Science Foundation under grant REC-0129331 (PADI Implementation Grant).
Disclaimer
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

FIGURES

Tables

ABSTRACT

The Principled Assessment Designs for Inquiry (PADI) Design System relies on evidence-centered design (Messick, 1994; Mislevy, Steinberg, and Almond, 2002) to guide users in assessment development. As a use-case for the PADI Design System, assessment resources developed for the Global Learning to Benefit the Environment (GLOBE) program were *reverse engineered* into Design Patterns and Task Templates. These Design Patterns and Task Templates can be used to support the development of GLOBE assessments and other assessments of science inquiry. Results from simulated data used to test the GLOBE Student Models represented in the Task Templates are presented, and the lessons learned from using the PADI Design System to develop GLOBE Task Templates are discussed.

1.0 Introduction

Global Learning to Benefit the Environment (GLOBE) is a "worldwide hands-on, primary and secondary school-based education and science program" (www.globe.gov). Students in GLOBE science classes take measurements related to earth science content investigation areas. These data are archived and can be used for research purposes by students and teachers who participate in the GLOBE program. While working with the GLOBE program, SRI International developed several assessment resources, such as a template for classroom assessment and a generic rubric, to support teachers in assessing their students' understanding of scientific inquiry using GLOBE data. As a use-case for the Principled Assessment Designs for Inquiry (PADI) approach to assessment design, the assessment resources developed for the GLOBE program were *reverse engineered* or translated into PADI Design System components.

GLOBE is a particularly important use-case for PADI because it helps to demonstrate the capability of the PADI Design System to accommodate assessments that address multiple phases 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 based in part on existing assessment resources that describe the GLOBE assessment design approach, as opposed to reverse engineering from one or two specific assessments (e.g., Haynie et al., in press).

This report begins by providing background on GLOBE assessments and the PADI approach. We present the components of the GLOBE Task Template and discuss the process by which GLOBE assessment resources were reverse engineered into a PADI Task Template. We then present findings from a study using simulated data to test the Student Models represented in the GLOBE Task Template. The final section of the paper includes a discussion of the lessons learned from using the PADI Design System to develop GLOBE Task Templates.

2.0 GLOBE Assessments

GLOBE assessments are uniquely complex compared to assessments that are typically administered in science classes or on state and national science examinations. GLOBE assessments address multiple aspects of science inquiry and require students to integrate inquiry skills as they move from data analysis to data interpretation and communication. GLOBE assessments also draw on a large dataset collected by students. As a result, analyses of data trends can be explored more readily, and students can compare data from different sites around the country. The GLOBE Assessment Framework, Classroom Assessment Template, and generic rubric developed by SRI International help to articulate the approach used in the development of assessments for the GLOBE program.

2.1 GLOBE Assessment Framework

The GLOBE Assessment Framework includes six investigation areas 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 several assessment exemplars. These resources were intended to assist teachers in developing tasks to assess their students' understanding of science inquiry using GLOBE data and can be found at http://www.globeassessment.sri.com.

Table 1. GLOBE Assessment Framework

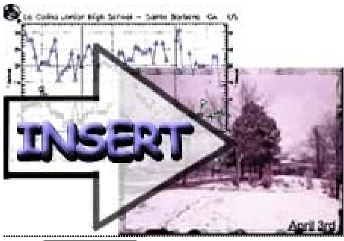
GLOBE Investigation Areas	GLOBE Investigation Strategies
Atmosphere	Plan investigations
Hydrology	Set up a new, appropriate problem/application
Landcover	Design an experiment
Soils	Specify measurements/variables to investigate
Earth Systems	Pose relevant questions
Visualizations	Take GLOBE measurements
	Measurements are accurate and appropriate
	Detect errors
	Use quality assurance procedures (multiple 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 findings

2.2 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 for particular school sites, 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 that illustrates relationships between cloud cover and temperature, students may be asked to use the data to make a recommendation to the school beach party planning committee about when the party should take place. Students must then 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 from (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 summarize 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.

2.3 GLOBE Generic Rubric

The GLOBE generic rubric provides holistic criteria for scoring students' conceptual understanding and inquiry strategy use, as shown in 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

Table 3. GLOBE Generic Rubric (continued)

	Criteria	1 Absent, inaccurate, confused	2 Partially developed	3 Adequately developed	4 Fully developed
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.

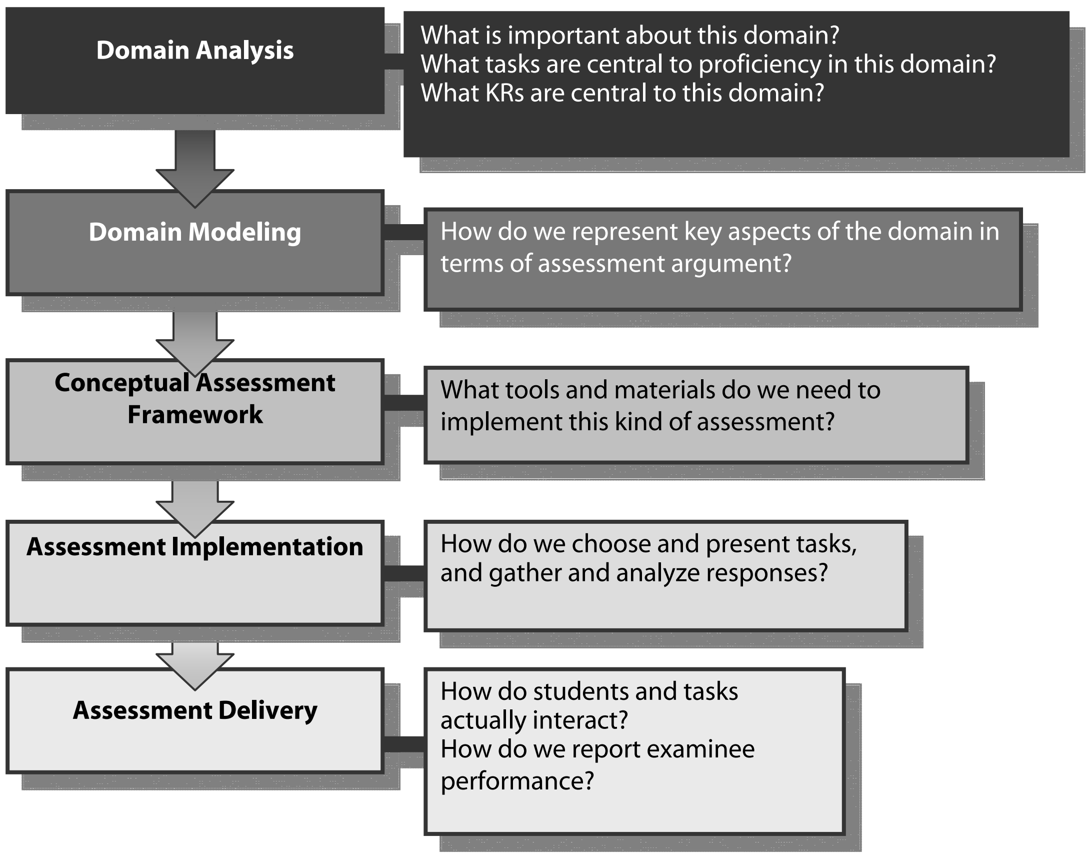
2.4 Sample GLOBE Task

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 a component 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.

3.0 The PADI Approach

The PADI project is defining and implementing an evidence-centered design (ECD) approach to support assessment developers' ability to create complex assessments of student learning. Evidence-centered assessment articulates the interrelationships among substantive arguments, assessment designs, and operational processes and 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). Figure 1 illustrates the layers of evidence-centered assessment design from domain analysis to assessment delivery.

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



Assessment design begins with domain analysis—the identification of important issues, tasks, and knowledge representations within a domain. In domain modeling, the assessment developer begins to lay out the assessment argument by identifying the key knowledge, skills, and abilities within the domain and the features of tasks that can provide evidence about them. In specifying the conceptual Assessment Framework, the assessment developer articulates details about the constructs to be tested (Student Model), the presentation materials and work products produced (Task Model), and the evaluation procedures that link scores on the work products to estimates of proficiency on the constructs of interest (Evidence

Model). In the assessment implementation phase, the assessment is created, and in the assessment delivery phase, it is presented to students.

The PADI Design System supports the domain modeling and conceptual Assessment Framework phases of assessment development. In the Design System, the assessment argument is instantiated initially through *Design Patterns*. PADI *Task Templates* facilitate the ECD approach in the articulation of the more detailed nuts and bolts of conceptual Assessment Framework.

3.1 Design Patterns

The role of Design Patterns is to articulate an assessment argument by identifying, in narrative form, the student knowledge, skills and abilities, potential observations, work products and rubrics that test designers may want to use, as well as characteristics and variable features of potential assessment tasks. The rationale for the use of Design Patterns in assessment starts from a need to extend thinking from individual assessment tasks to prototypical ways of obtaining evidence about the acquisition of various aspects of knowledge. Design Patterns help to ground the subsequent detailing of the "nuts and bolts" of assessments, the operational elements such as psychometric models, evaluation procedures, and specific stimulus materials. In addition to supporting the identification of aspects of knowledge that are similar across content areas or skill levels, Design Patterns provide reusable schemas for obtaining evidence about such knowledge.

Thinking at the level of Design Patterns is integral in the development of complex assessments because it enables content and measurement experts to share a coherent view about the substantive argument needed to create a principled assessment. Design Patterns can advance the current approaches used by subject area specialists in designing assessments by incorporating the advances in cognitive psychology and learning, statistics, measurement, and technology. Subject matter specialists using such Design Patterns gain access to these newer approaches rather than being constrained to familiar item formats and simple measurement models.

Each PADI Design Pattern has a title and summary, which provides an overview of the target inferences addressed by this design pattern, as well as a rationale for using certain kinds of information about student performance as evidence of the targeted knowledge, skills, and abilities. Table 4 contains an exhaustive set of Design Pattern attributes with brief descriptions of each. For more information about Design Patterns, see DeBarger & Riconscente, 2005 and Mislevy et al., 2003.

Table 4. Design Pattern Attributes (DeBarger & Riconscente, 2005)

Attribute	Description
Title	A short name for referring to the design pattern.
Summary	An overview of the kinds of assessment situations students encounter in tasks that are instantiations of this design pattern and what one wants to know about students' knowledge, skills, and abilities.
Rationale	Explain why this item is an important aspect of scientific inquiry.
Focal KSAs	The primary knowledge/skill/abilities targeted by this design pattern.
Additional KSAs	Other knowledge/skills/abilities (KSAs) that may be required by this design pattern. These could be nuisance skills, for example, background knowledge the student must be provide, or knowledge intended to be assessed jointly with the focal KSAs. Additional KSAs make assessment designer aware that other KSAs beside the focal one are often addressed by an assessment task, and that determining which ones to include is a design choice that should be made purposefully.
Potential observations	Some possible things one could see students doing that would give evidence about the KSAs (knowledge/skills/abilities). Potential observations differ from work products (below) in that work products are what students produce, while observations are qualities that assessors discern and evaluate in work products.
Potential work products	Modes, like a written product or a spoken answer, in which students might produce evidence about KSAs (knowledge/skills/abilities).
Potential rubrics	Some evaluation techniques that may apply. These may include links to relevant scoring rubrics and procedures (algorithms, guidelines, and/or examples of ways to ascertain values of observations from student work products).
Characteristic features	Aspects of assessment situations that are likely to evoke the desired evidence. These are features of situations (tasks) that are required so that students can provide evidence of the KSAs of interest. If a focal KSA is problem-solving with algebraic representations in ill-structured problems, then a characteristic feature of a tasks to assess this KSA would be that the situation must present a problem that is amenable to algebraic representation and solution—possibly several different ones—but the approach and the representation must developed by the student rather than provided by the assessor.
Variable features	Aspects of assessment situations that can be varied in order to shift difficulty or focus. Given that all the tasks that might be generated from a given design pattern are alike at some level in terms of characteristic features, variable features specify ways in which they might vary to increase or decrease difficulty, focus of information,

	put more or less demand on various additional KSAs, etc.
I am a kind of	Associations to other objects ("my parents") which are more abstract or more general than this object.
These are kinds of me	Associations to other objects ("my children") which are more concrete or more specialized than this object.
These are parts of me	Associations to other objects which contain or subsume this one. For example, an automobile contains a windshield.
Educational standards	Associations with (potentially shared) Educational Standard objects.
Templates	Associations with (potentially shared) Template objects.
Exemplar tasks	Associations with (potentially shared) Task Exemplar objects. These may include links to sample assessment tasks that are instances of this design pattern.
Online resources	Items that pertain and can be found online (URLs). These items may illustrate or provide background for this design pattern.
References	Notes about relevant items, such as academic articles.

3.2 Task Templates

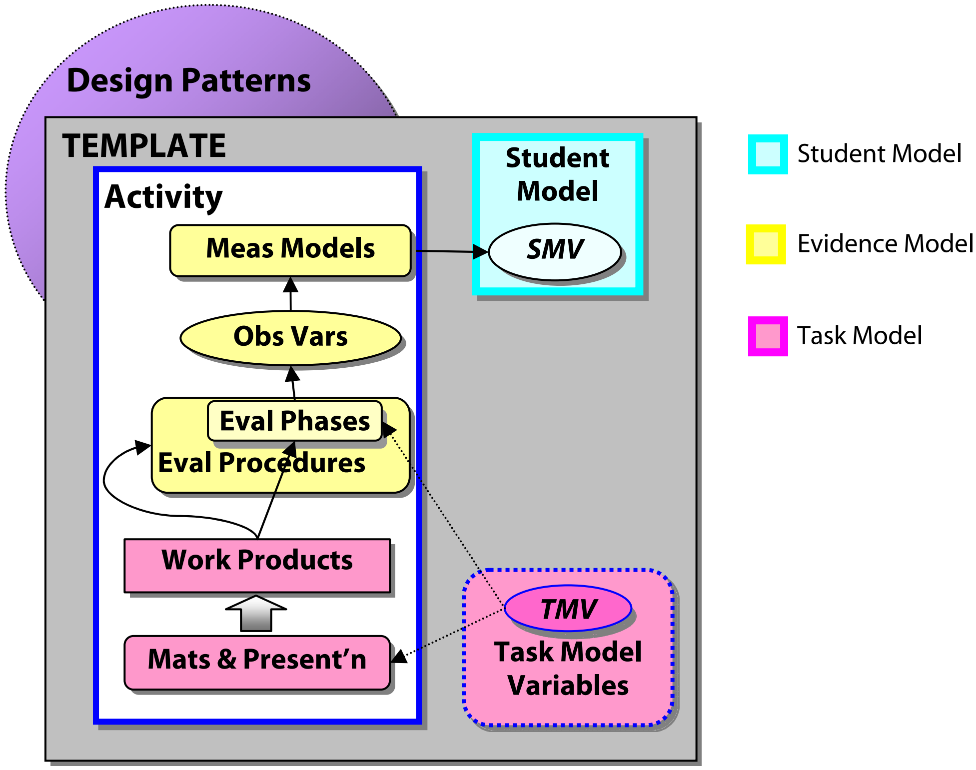
PADI Task Templates provide a structure for helping assessment designers identify the constructs to be tested, stimulus materials, evaluation procedures and measurement models. Task Templates serve as "pre-blueprints" in that they are abstractions of related assessment tasks. A Task Specification is more analogous to an assessment blueprint because it is used to generate a specific assessment task. Thus, components of the Task Template (e.g., task difficulty) are left unspecified until the Task Template becomes a Task Specification. Figure 2 shows the relationship among objects in the Task Template.

Each Task Template includes a Student Model, which contains one or more Student Model Variables. The Student Model Variables are constructs representing the central knowledge, skills, and abilities in the domain to be tested.

Task Templates may contain one or more Activities, which can be conceptualized as components of a task. The parsing of an assessment into Activities is determined by the assessment developer. Some assessment developers may choose to parse the assessment by item, and so each item is represented as an Activity in the Task Template. In other cases, an Activity may represent a phase of inquiry that incorporates several items. The way in which items may be grouped for scoring purposes must also be considered when defining Activities.

Within each Activity, the assessment developer must specify the features of the Materials and Presentation objects (e.g., prompts) that will be presented to students and the Work Products that are produced by students in relation to these Materials and Presentation objects. Task Model Variables define how features of materials presented to students may vary (e.g., amount of data presented to students, type of data presented to students, reading amount). These Task Template components relate to the Task Model.

Figure 2. Relations among Task Template Objects



The assessment developer then specifies the Evidence Model, including the Evaluation Procedures that describe how the Work Products are scored. The Evaluation Procedures may contain a sequence of Evaluation Phases, if multiple aspects of student work are to be scored. These procedures may vary if features of the task (as specified by the Task Model Variables) differ. For instance, scoring may vary as task difficulty varies. Observable Variables that provide information about how to judge the quality of students' Work Products are defined within each Evaluation Phase.

The relationship between Observable Variables and Student Model Variables (i.e., how scores on specific Work Products relate to levels of proficiency on the constructs of interest) is specified in the Measurement Model. The Measurement Model currently supported in the PADI Design System, the Multidimensional Random Coefficients Multinomial Logit (MRCML) model (Adams, Wilson, & Wang, 1997), only supports the input of a single Observable Variable to update one or more Student Model Variables. 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 of this bundled Observable Variable may reflect combinations of the score levels from the Observable Variables used to create the bundled variable.

4.0 Representing GLOBE Assessments in the PADI Design System

As a use-case for the PADI Design System, the components of the GLOBE assessment system (Assessment Framework, Template for Classroom Assessment, and Generic Rubric) were reverse engineered into PADI Design Patterns and Task Templates. Reverse engineering is the process of creating a design or higher level abstraction of an existing product or system. Engaging in this process served to test the usability of the Design System for developing multi-phase assessments of science inquiry. In this section of the paper, we present the GLOBE assessment system components, as they are represented in the PADI Design System.

4.1 GLOBE Design Patterns

Ten Design Patterns related to GLOBE were developed by individuals who bring expertise in educational psychology, science content knowledge, and assessment design. The content of these design patterns was influenced primarily by the investigation (science inquiry) strategies identified GLOBE Assessment Framework. These Design Patterns articulate assessment arguments related to the central knowledge and skills addressed in GLOBE assessments. Table 5 shows the relationship between these Design Patterns and the GLOBE investigation strategies.

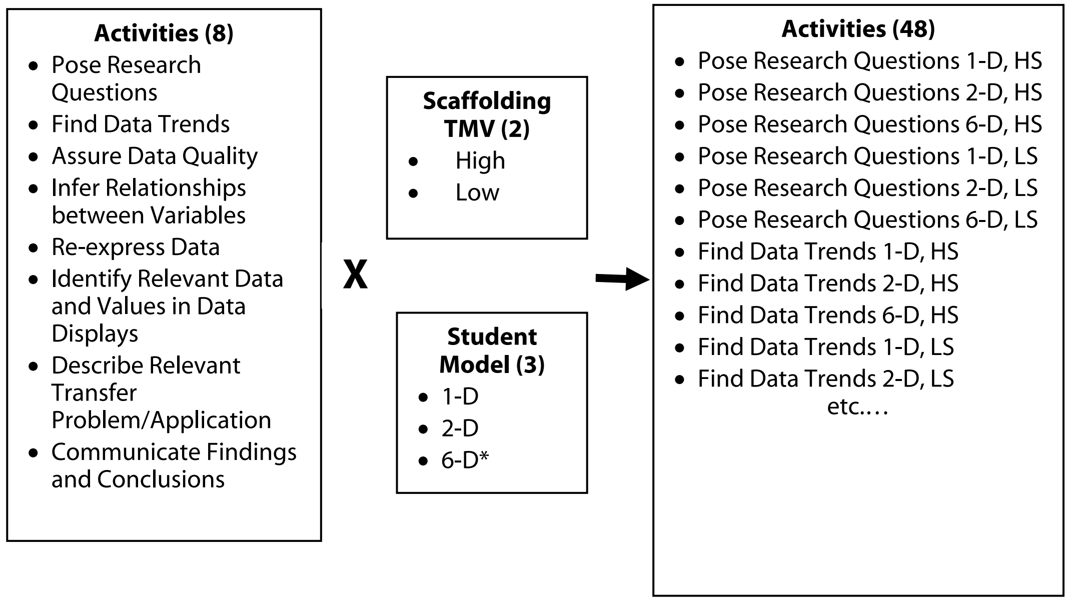
Table 5. Design Patterns Based on GLOBE Investigation Strategies

Design Patterns	GLOBE Investigation Strategies				
	Plan Investigations	**Take GLOBE Measurements**	**Analyze and Compare GLOBE Data**	**Interpret GLOBE Data**	**Communicate**
Analyze Data Relationships			X		
Assurance of Data Quality		X			
Formulating Scientific Explanation from Evidence					X
Generate Explanations Based on Underlying Principles					X
Interpret Data	X	X	X		
Re-expressing Data			X		
Scientific Reasoning	X	X		X	
Use Resources to Conduct Scientific Inquiry				X	
Use the Representational Forms of Science			X		
View Real-World Situations from a Scientific Perspective	X			X	

4.2 GLOBE Task Templates

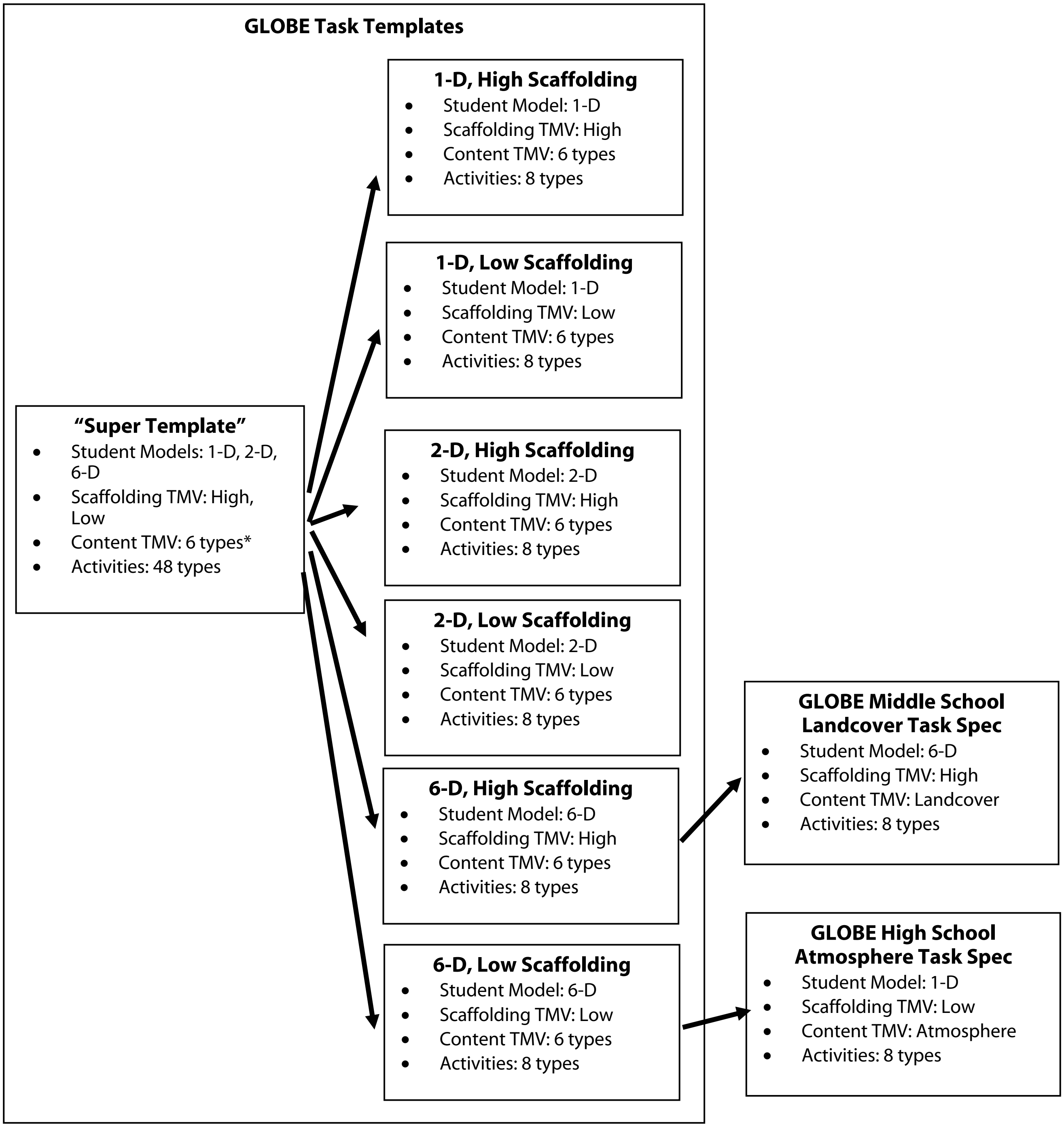
The PADI Design System contains multiple GLOBE Task Templates. We refer to the first of these Task Templates, the *GLOBE Inquiry Template*, as a “Super Template” because it contains all possible Activities and Student Models, and all of the Task Model Variables are unset. The other GLOBE Task Templates are more specialized versions of the “Super Template.” These specialized variants of the *GLOBE Inquiry Template* contain some but not all of the Activities and/or Student Models, and some Task Model Variables have been set. Figure 3 illustrates how the Student Models and scaffolding levels influence the generation of Activities for the “Super Template.” Each Activity must have a unique Measurement Model that reflects a particular Student Model and scaffolding level because the Measurement Model scoring matrices (relationships among Observable Variables and Student Model Variables) are impacted by the type of Student Model Variables and scaffolding. Figure 4 shows the relationships among the GLOBE “Super Template” and more specialized GLOBE Task Templates. Figure 5 is the 6-D Low Scaffolding GLOBE Task Template.

Figure 3. Activities within the GLOBE Inquiry Template (the "Super Template")



**The GLOBE Super Template also includes a 4-dimensional Student Model, which is identical to 6-dimensional Student Model, but does not include 2 of the 6 Student Model Variables.*

Figure 4. Relationships among GLOBE Task Templates and Task Specifications



**The Content Task Model Variable, when unspecified, contains 6 content types: Atmosphere, Hydrology, Soils, Earth as a System, Landcover, and GPS.*

Figure 5. GLOBE 6-D Low Scaffolding Task Template

GLOBE Inquiry Template 6D Low Scaffolding | Template 1881 [View Tree | Convert to Task Spec | Duplicate | Export | Delete]

Title:	[Edit]	GLOBE Inquiry Template 6D Low Scaffolding	
Summary	[Edit]	GLOBE inquiry investigations take a student through phases of an investigation, using data from the Globe data-collection protocols. A typical assessment has 5 to 7 activity phases, all of which are components of the GLOBE assessment framework (e.g., posing questions, analyzing, interpreting, communicating).	
Type	[Edit]	[View]	
Student Model Summary	[Edit]	Multivariate model	Choose among the alternative Student Models by select GLOBE investigation (inquiry) components, and then identify desired conceptual understanding
Student Models	[Edit]	GLOBE Investigation Strategies- Multidimensional (6-D: multiple inquiry, single content). This SM includes multiple SMVs related to inquiry and one SMV related to content.	
Measurement Model Summary	[Edit]	Multidimensional item response modeling, using MRCML measurement models.	
Evaluation Procedures Summary	[Edit]	GLOBE rubrics are employed by teachers to evaluate the students' work products.	Products that result from group cooperation may need to be evaluated differently than the products of individuals.
Work Product Summary	[Edit]	Students products range from written text to graphs, tables, or other visual representations. At end of assignment, they present their findings to their class or post their report on the web.	
Task Model Variable Summary	[Edit]	Template-level task model variables (TMVs) and activity-level TMVs are included in this template. Many TMVs are related to each other in an hierarchical way (e.g., at the template level, a high level TMV is "GLOBE Data Scaffolding" and a low level TMV is "Nature of Association between GLOBE Variables". At some point we may be able to use the design system to set higher-level TMVs, which will impact the settings of lower-level TMVs.	
Template-level Task Model Variables	[Edit]	GLOBE General Scaffolding. This high level TMV can be used to control the settings of all activity scaffolding-related TMVs and... GLOBE Data Scaffolding. The amount of scaffolding of GLOBE data. This high level data scaffolding TMV will be used to adjust... Amount of Data. The number of data points presented to students in graphs, tables and maps. Data Representation Format. The format of data as it is presented to students (bar graph, line graph, scatter plot, map, data ta... GLOBE Data Site(s). The school sites or regions used in data tables, graphs, and maps. Mathematical Form of Association between GLOBE Variables. The mathematical form of association (linear, exponential, curvelinear) between GLOBE variables (hyd... Nature of Association between GLOBE Variables. The nature of the association (causal, correlational, no relationship) between GLOBE variables (hydr... Number of GLOBE Data Representational Formats. The number of GLOBE data tables, graphs, maps, etc. presented to students. GLOBE Datasets (Measurements/Variables). Students are presented with data displays of measurements taken as part of the GLOBE program (e.g., ... Types of Data Anomalies Included in Graphs, Tables, and Figures. In order to create a more challenging task, the assessment designer may choose to select data that c... Grade Level Appropriateness- Vocabulary, Computations and Reading. Extent to which the specialized vocabulary, computations, and reading load are appropriate for the g... GLOBE Content area. Specific domain content under consideration Problem Complexity. Familiarity with data sites, kinds of reasoning required, complexity of situation/system Grade Level. Level of target audience. GLOBE Time Period for Dataset. The beginning and end month, day, and year for GLOBE data presented to students.	
Task Model Variable Settings	[Edit]	[View]	
Materials and Presentation Requirements	[Edit]	Students are presented with paper stimuli in classrooms. They work on assignments individually. No web database or data collection equipment is required, since stimuli have all necessary information.	Group work is handled in a different template (TBD).
Template-level Materials and Presentation	[Edit]	GLOBE Data. Students are presented with data graphs or tables. Problem Situation Description. Text description of the driving problem and resources to be used.	
Materials and Presentation Settings	[Edit]	[View]	
Activities Summary	[Edit]	Present GLOBE activity phases in this order: Problem Presentation, Investigation Planning, Data Quality Assessment, Data Analysis, Planning of New/Related Investigation, Presentation of Results	
Activities	[Edit]	Pose Research Questions LS 6-D. Student is asked to look at the GLOBE data/graphs and develop possible questions about the data. A s... Find Data Trends LS 6-D/4-D. Student is asked to find observable trends in the GLOBE data. The measurement model associated with ... Assure Data Quality LS 6-D/4-D. Student is asked to look through GLOBE data for possible measurement or data entry errors and sugges... Infer Relationships Between Variables LS 6-D/4-D. Student is asked about the relationship between two variables in a table or graph with GLOBE data. A... Re-express Data (GLOBE) LS 6-D/4-D. Student is asked to use the GLOBE data provided in order to generate new data representations. The p... Identify Relevant Data and Values in Data Displays LS 6-D/4-D. Student is asked about specific features of graphs and about what indications there are for various ... Describe Relevant Transfer Problem/Application LS 6-D. Student is asked to choose another school from the GLOBE database that has features similar to the s... Communicate Findings and Conclusions LS 6-D. Student is asked to summarize their analysis of the original school(s) data and to write a short rep...	
Tools for Examinee	[Edit]	paper and pencil Computer for accessing GLOBE Web site (www.globe.gov)	If transfer problem requires identification of new GLOBE data.
Exemplars	[Edit]		
Educational Standards	[Edit]	NSES 8ASI1.1. Identify questions that can be answered through scientific investigations. Students should develop t... NSES 8ASI1.2. Design and conduct a scientific investigation. Students should develop general abilities, such as sy... NSES 8ASI1.3. Use appropriate tools and techniques to gather, analyze, and interpret data. The use of tools and te... NSES 8ASI1.4. Develop descriptions, explanations, predictions, and models using evidence. Students should base the... NSES 8ASI1.5. Think critically and logically to make the relationships between evidence and explanations. Thinking... NSES 8ASI1.7. Communicate scientific procedures and explanations. With practice, students should become competent ... NSES 8ASI1.8. Use mathematics in all aspects of scientific inquiry. Mathematics is essential to asking and answeri...	
Design Patterns	[Edit]	Analyze data relationships. A student encounters two or more sets of data organized into one or more representations, and must d... Interpret data. Students are presented with a set of data or observations and are asked to formulate an explanation ... Re-express data. A student encounters data organized in one or more representational forms (RFs), and must re-express... Use data to support scientific argument. A student must use data, either collected or provided, to support a scientific argument. Does the s... View real-world situations from a scientific perspective. A student encounters a real-world situation that lends itself to being framed from a scientific pers...	
I am a kind of	[Edit]	GLOBE Inquiry Template. GLOBE inquiry investigations take a student through phases of an investigation, using data from the ... GLOBE Inquiry Template Low Scaffolding. GLOBE inquiry investigations take a student through phases of an investigation, using data from the ...	
These are kinds of me	[Edit]	GLOBE Middle School Landcover Task Spec. This task spec represents the reverse engineering of a middle school GLOBE assessment that addresses...	
These are parts of me	[Edit]		
Online resources	[Edit]	www.globe.gov www.globeassessment.sr...	
References	[Edit]		
I am a part of			

5.0 GLOBE Template Development Process

This section describes how and why particular design decisions were made as we reverse engineered GLOBE assessment documents into a Task Template in the PADI Design System. Our reverse engineering process can be distilled into six distinct stages: (1) review GLOBE documents and determine which features of the GLOBE assessment system relate to Student Model, Task Model, and Evidence Model components of the Task Template; (2) complete the Materials and Presentation and Work Product components for each Activity; (3) identify features of the task that may vary (Task Model Variables); (4) define the Student Model Variables and Student Model; (5) design Evidence Model components of the Task Template (e.g., Evaluation Procedures, Evaluation Phases, Measurement Model), and (6) organize and integrate the overall structure of the Task Template. This development process was iterative, which frequently required us to review and modify previous steps as we developed new aspects of the Task Template. These stages of development illustrate only one approach to reverse engineering an assessment or assessment system into PADI objects. Other approaches may be equally appropriate depending on the needs of the assessment designer and the nature of the materials on which the reverse engineered design is based.

5.1 Define Template Structure

The first step in developing the GLOBE Task 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 Task 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).

We began by trying to identify the key constructs (Student Model Variables) addressed in GLOBE assessments. We determined that this information is present in the GLOBE Assessment Framework (Table 1) and that the Student Model would be informed by the Investigation Areas and Investigation Strategies. However, initially, we were uncertain about determining the grain size of the Student Model Variables. As a result, we decided to wait before defining the GLOBE Student Model in the Task Template, but we were confident that we knew the range of knowledge and skills that could be addressed by GLOBE assessments.

Because Activities are one of the central organizing Task Template structures, we worked on specifying Activities next. The GLOBE Classroom Assessment Template (Table 2) identifies eight investigation strategies or phases of inquiry. Because these eight phases define the core tasks in most GLOBE assessments, regardless of the grade level or science content focus of the assessment, we represented these phases as Activities. Thus, the GLOBE Task Templates instantiated in the PADI Design System contain 8 types of 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.[1]

[1] The *GLOBE Inquiry Template* (the "Super Template") contains 48 Activities. Activities are the container for the Measurement Models, which link Observable Variables to Student Model Variables. Thus, each Activity is related to a particular Student

This conceptualization of the GLOBE Task Template will suggest to an assessment developer that a complete GLOBE assessment for any grade level or content area requires each of these Activities.

5.2 Complete Materials and Presentation and Work Products

Because the items and expected student work were defined concretely within the GLOBE assessment system, and there were several examples of these within the sample GLOBE assessments, we began to map these into Materials and Presentation and Work Product objects in the Task Template. For each Activity we created Materials and Presentation objects that reflected the types of prompts used to elicit student responses. We then generated objects that describe the types of Work Products produced by students in relation to the Materials and Presentation objects.

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 Task Template. The Materials and Presentation objects for each Activity are presented in Table 6. In addition, to including prompts as stimulus materials in the Task 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.

Sample GLOBE assessments, and to some extent the GLOBE rubrics and the GLOBE Classroom Assessment Template, were used to specify the Work Products in the Task Template. On the basis of an 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 6 with their associated Materials and Presentation objects.

Model (e.g., 1-dimensional, 2-dimensional, or 6-dimensional) and to a particular scaffolding level (high or low), and a particular Measurement Model and Student Model must be associated with each Activity. The more specific GLOBE Task Templates only include one Student Model and some of their Task Model Variables set to high or low scaffolding. As a result, these more specified Task Templates only include 8 of the Activities from the *GLOBE Inquiry Template*.

Table 6. Materials and Presentation Objects and Work Products for GLOBE Activities

Activity	Materials and Presentation Objects	Work Products
Pose Research Questions	Prompt to Pose Question(s) Sample Response for Posing 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 Prompt to Evaluate/Discuss Data Collection Protocol Sample Response for Assure Data Quality Work Products and Findings from Previous GLOBE Activities	Identified Errors in GLOBE Data Description of Error Finding Procedures Evaluation of Implementation of Data Collection Protocol
Infer Relationships between Variables	Prompt to Infer Relationship between Two Variables Work Products and Findings from Previous GLOBE Activities	Explanation of Relationship between Variables and Measurements
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 Similarities and Differences in Data Prompt to Identify Values of Data or Data Components Work Products and Findings from Previous GLOBE Activities	Identify and Describe GLOBE Data Components
Describe Relevant Transfer Problem/ Application	Prompt to Identify Additional Dataset(s) Prompt to Find Data Trends and Relationships Sample Response for Transfer Problem Work Products and Findings from Previous GLOBE Activities	Propose Additional GLOBE Problem/Application
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

5.3 Identify 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). Template-level Task Model Variables related to GLOBE data were then created. This process resulted in the generation of a large number of Task Model Variables and created a disjunctive and complex overall Task Model Variable structure (see Figure 6).

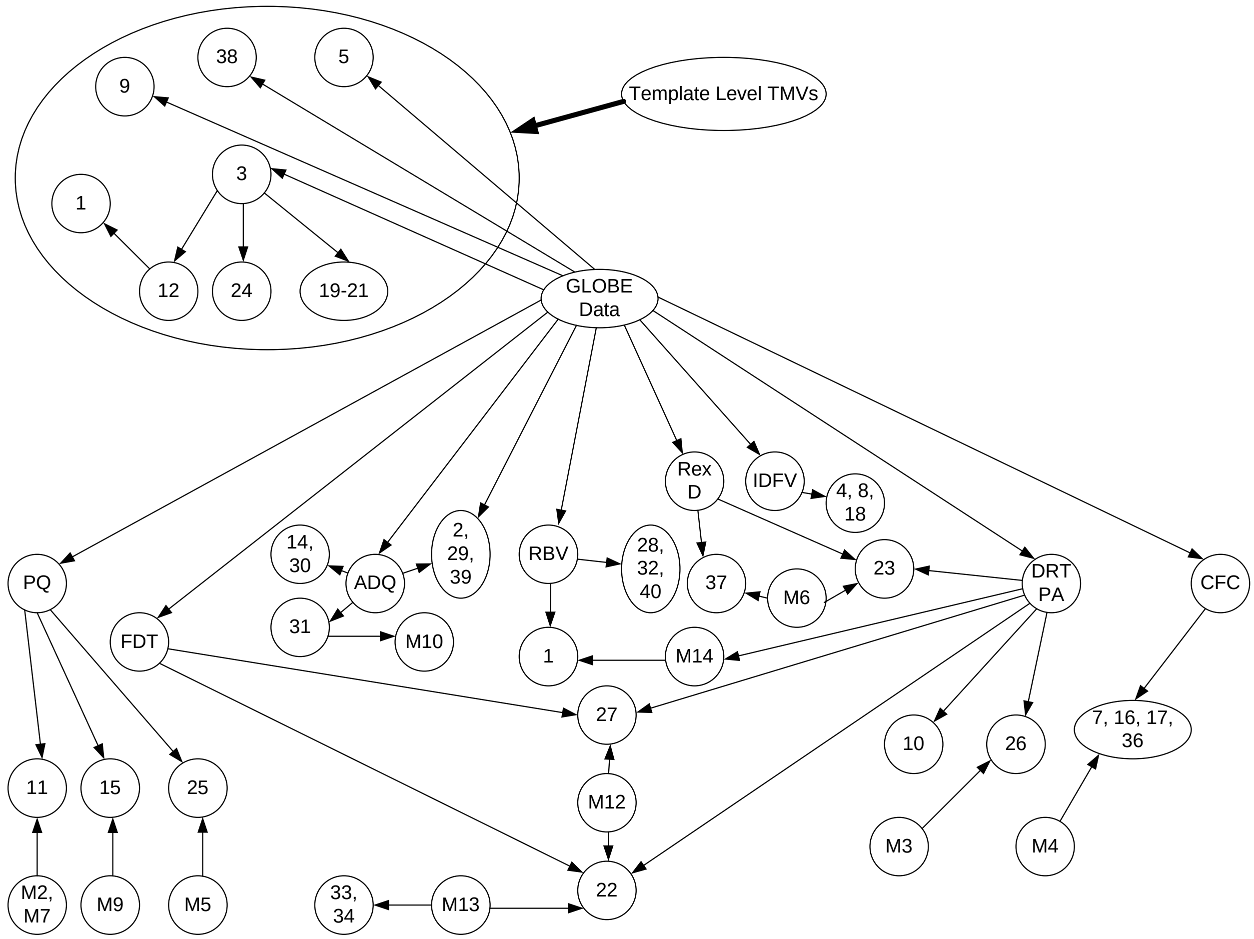
To simplify the Task Model Variable structure, we examined all Task Model Variables again and eliminated redundant or confusing Task Model Variables. After making these changes, we were still left with over 30 Template-level and Activity-level Task Model Variables. Given that there are eight activities 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 that is new to GLOBE or PADI, 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 (see Figure 7). These relationships are currently specified in the Task 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. The highest level Task Model Variable is GLOBE general scaffolding/grade level (shown in the center of the box on the left in Figure 7). As illustrated by the arrows, this Task Model Variable is intended to 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) are intended to be influenced by a higher-level Task Model Variables, but also impact lower level Task Model Variables. Lowest-level Task Model Variables (e.g., Type of Data Transformation in the Re-express Data Activity) do not influence any other Task Model Variables. There are a few independent higher level Task Model Variables such as GLOBE content area and grade level appropriateness Task Model Variables.

The hierarchical structure provides flexibility to test designers by allowing for different levels of Task Template modification. For example, someone who wants to create a Task Template quickly only needs to define the highest level Task Model Variables. Although, the PADI Design System does not support Task Model Variable hierarchies in which Task Model Variables can influence each other automatically, a Wizard[2] could be designed to assist users in making adjustments to Template-level and Activity-level Task Model Variables at different levels in the hierarchy.

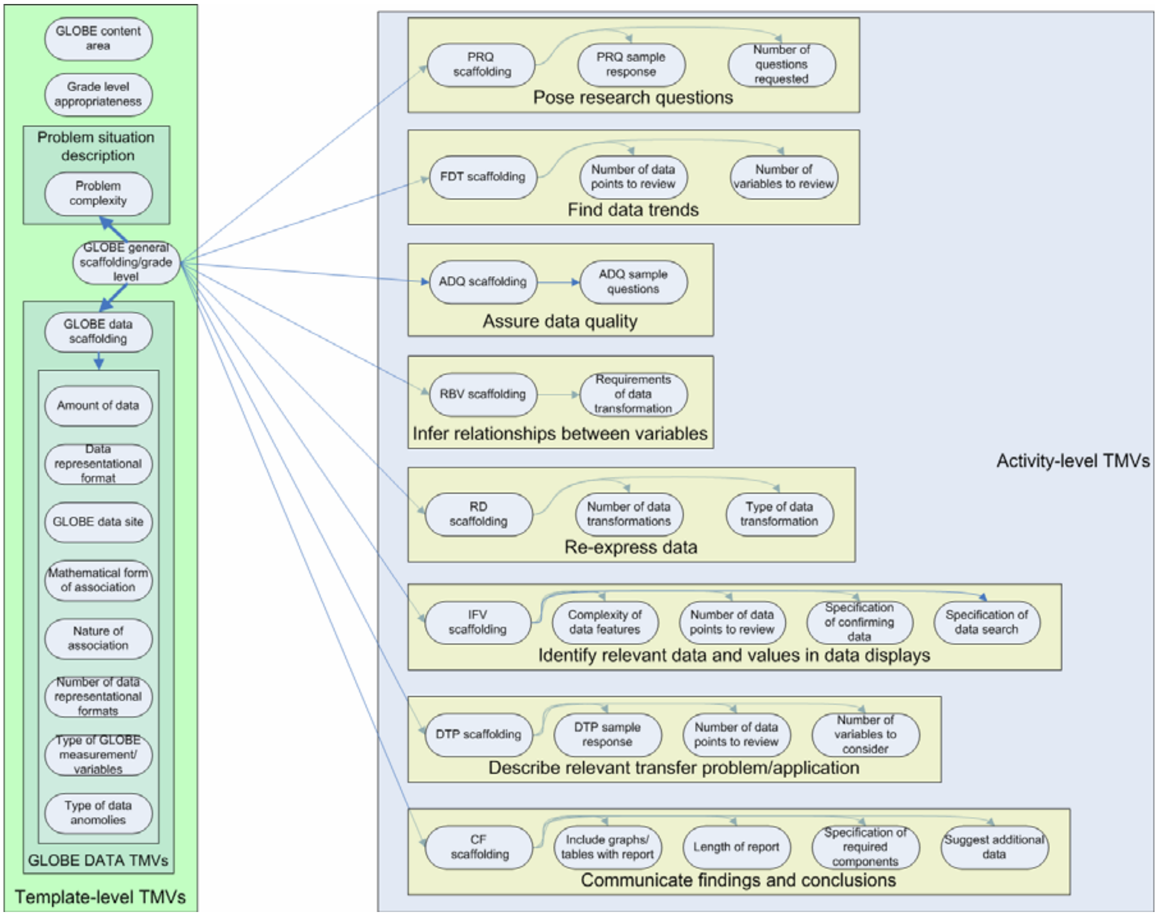
[2] A Wizard is an interactive computer program that guides users through complex tasks.

Figure 6. Relationships among GLOBE Task Model Variables



Note. Numbers within circles refer to Task Model Variables. PQ, FDT, ADQ, RBV, RexD, IDFV, DRTPA, CFC refer to the 8 GLOBE Activities. M refers to Materials and Presentation objects that are related to the Task Model Variables within each Activity.

Figure 7. Hierarchical Relationships among Task Model Variables

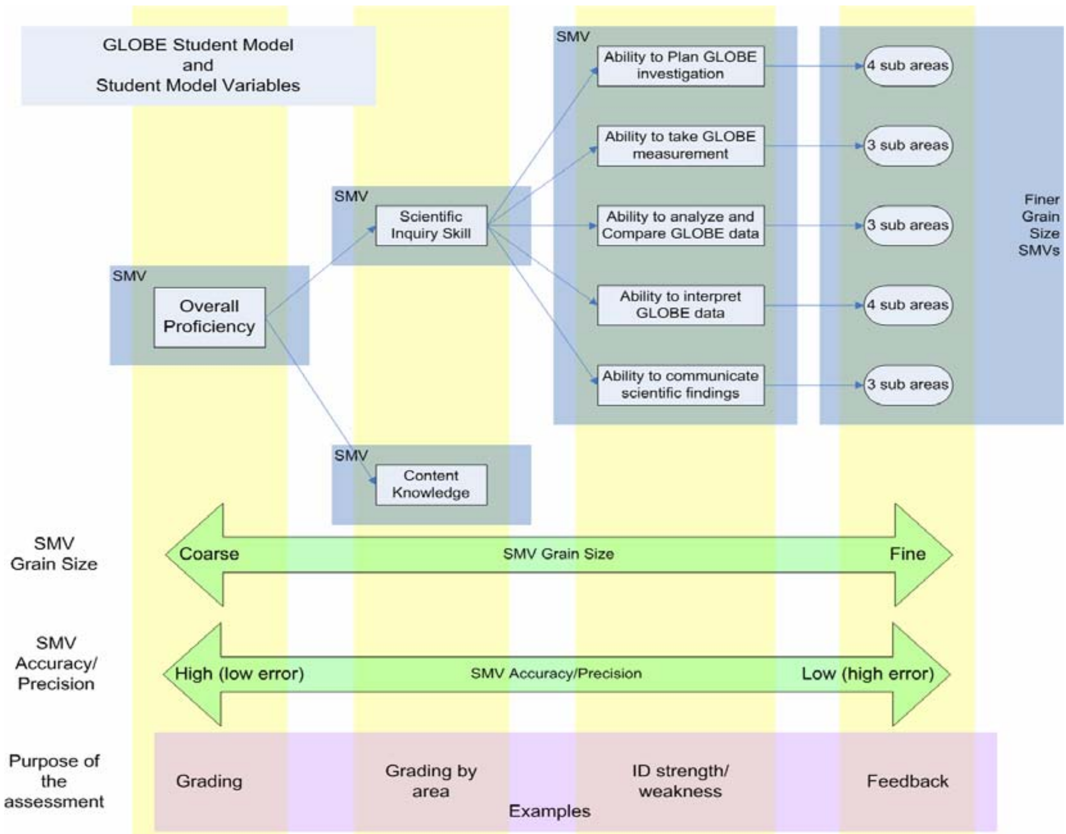


5.4 *Define Student Model*

The next stage of GLOBE development focused on defining the Student Models and Student Model Variables. As we determined in the first stage of the Task Template development process, the GLOBE Assessment Framework (Table 1) and Generic Rubric (Table 3) defined five distinctive components of scientific inquiry, which are central in GLOBE tasks: Ability to Plan GLOBE Investigations, Ability to Take GLOBE Measurements, Ability to Analyze and Compare GLOBE Data, Ability to Interpret GLOBE Data, and Ability to Communicate Findings. Each of these components of inquiry has subcomponents that further specify the abilities required in GLOBE assessments.

We chose to develop multiple Student Models that contain Student Model Variables at different grain sizes. It is realistic to expect that a GLOBE assessment can be used for different purposes, and so we determined that Student Models that provide differing amounts of information related to the key components of inquiry in GLOBE assessments would be a valuable addition to the Task Template. Some instructors may use a GLOBE assessment as a part of summative test (e.g., final grade) whereas others may use it for more formative or diagnostic purposes. In the first case an overall indicator of student proficiency (i.e., fewer Student Model Variables) with the highest precision of measurement (i.e., low standard errors) may be preferred. The second case requires more detailed information about student understanding and warrants a greater number of Student Model Variables. In this second case, detail about student scores related to the finer grain size components of inquiry is more critical than the precision of measurement because the main purpose of the assessment is to provide formative feedback to students about their strengths and weaknesses in the domain. The relationships among Student Model Variable grain size, accuracy/precision, and assessment purpose are summarized in Figure 8.

Figure 8. Relationships among Student Model Variable Grain Size, Student Model Variable Accuracy, and Assessment Purpose



The *GLOBE Inquiry Template*, the "Super Template," contains four Student Models: a 6-dimensional model, a 4-dimensional model, a 2-dimensional model, and a 1-dimensional model. In the 6-dimensional Student Model each of the inquiry skills (Ability to Plan GLOBE Investigations, Ability to Take GLOBE Measurements, Ability to Analyze and Compare GLOBE Data, Ability to Interpret GLOBE Data, Ability to Communicate Scientific Findings) are expressed as separate Student Model Variables. Content Knowledge is the sixth Student Model Variable. The meaning of the Content Knowledge Student Model Variable changes depending on the content of the assessment(s) administered (e.g., Atmosphere, Hydrology, Land Cover).

The 4-dimensional Student Model contains Student Model Variables that are at the same grain-size as those in the 6-dimensional model, but the Ability to Plan GLOBE Investigations and Ability to Communicate Scientific Findings Student Model Variables are not included. This affords the assessment developer the flexibility of creating GLOBE assessments that provide detailed information about student proficiency on all of the GLOBE inquiry skills, except Planning and Communicating.

The 2-dimensional Student Model includes a Science Inquiry Student Model Variable and a Content Knowledge Student Model Variable. The unidimensional Student Model contains a single Student Model Variable to provide information about students' inquiry skills and content knowledge combined. Although these Student Models do not provide detailed feedback about student proficiency related to specific components of science inquiry, the coarser grain-size Student Model Variables may be a more accurate measure of student proficiency, if the amount of data available to update the Student Model is limited.

5.5 Define Evidence Model

The Evidence Model helps to define connections among Activities, Work Products, Observable Variables, Measurement Models, and Student Model Variables. A GLOBE assessment is composed of eight different Activities, which produce multiple scores that update one or more areas of scientific inquiry or Student Model Variables. Figure 9 shows the relationships among the GLOBE Activities and Student Model Variables, where the heavy line indicates an activity provides significant amount of information to Student Model Variable and the thin line show a minor connection providing less information to Student Model Variable.

Figure 9. Relationships among GLOBE Activities and Student Model Variables

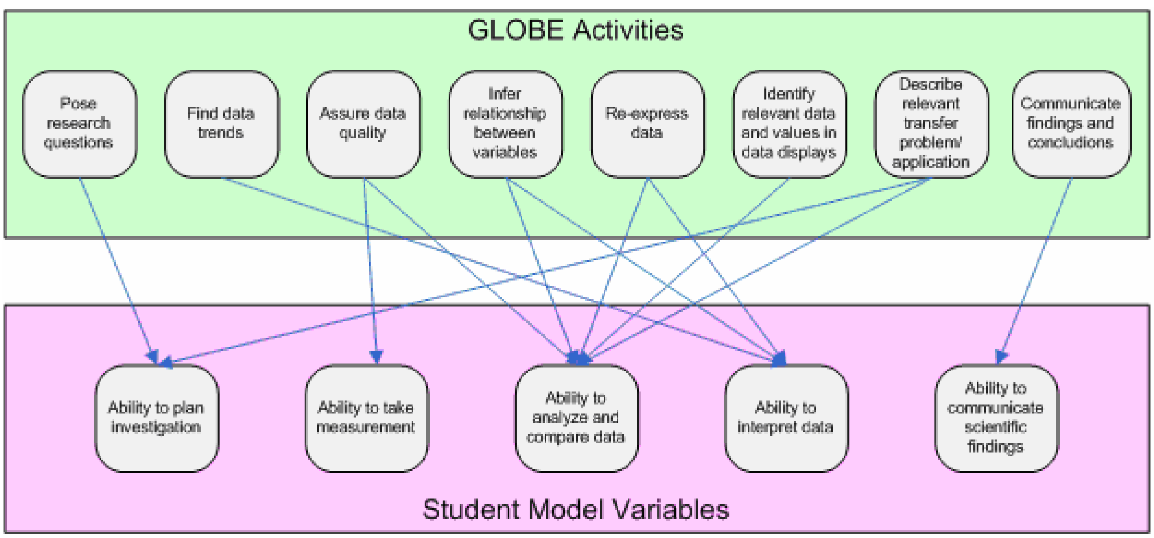


Figure 10 illustrates the process by which Student Model Variables are updated by showing the flow of information from Work Products to Student Model Variable. Some Observable Variables may be conditionally dependent and require additional evaluation procedures for item bundling.

Figure 10. Information Flow in Updating a Student Model Variable

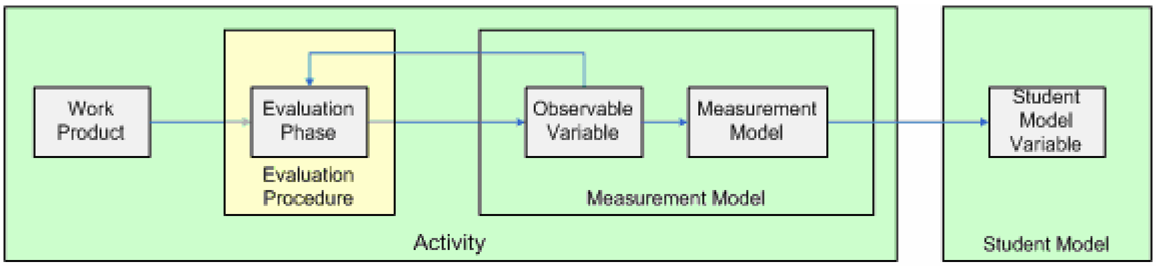


Table 7 provides information about the specific Observable Variables for each Activity, how they are bundled, and which Student Model Variables they relate to in the 6-dimensional Student Model.

Table 7. GLOBE Measurement Models: Relationships among Observable Variables and Student Model Variables in the 6-Dimensional Student Model

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: Appropriateness of Identified Trends and Explanation	Ability to Interpret GLOBE Data Content Knowledge
Assure Data Quality	Bundled OV: Error Analysis	Ability to Take GLOBE Measurements Ability to Analyze and Compare GLOBE Data Content Knowledge
	Strategy for Finding Errors	Ability to Take GLOBE Measurements Ability to Analyze and Compare GLOBE Data
	Evaluation of Data Collection Protocol Score	Ability to Take GLOBE Measurements
Infer Relationships between Variables	Bundled OV: Data Interpretation, Description and Explanation of Relationships between Variables	Ability to Analyze and Compare GLOBE Data Ability to Interpret GLOBE Data Content Knowledge
Re-express Data	Mathematical Transformation Score	Ability to Analyze and Compare GLOBE Data
	Bundled OV: Appropriateness of Representational Format and Data Placement	Ability to Interpret GLOBE Data GLOBE Content Knowledge
Identify Relevant Data and Values in Data Displays	Accuracy of Identified Measurements and Values	Ability to Analyze and Compare GLOBE Data
Describe Relevant Transfer Problem	Bundled OV: Appropriateness of GLOBE Problem, Dataset, Variables, and Explanation	Ability to Plan GLOBE Investigations Ability to Analyze and Compare GLOBE Data Content Knowledge
Communicate Findings and Conclusions	Use of Representation to Communicate Findings	Ability to Communicate Scientific Findings
	Use of Data and Explanation of Findings	Ability to Communicate Scientific Findings Content Knowledge
	Argument Coherence	Ability to Communicate Scientific Findings
	Content	Content Knowledge

We used the subcomponents of each investigation strategy from the GLOBE Assessment Framework to help us think more deeply about the relationships among Work Products, Observable Variables and Student Model Variables. Thinking about which specific component of the Student Model Variable the Observable Variable and Work Product might inform helped to strengthen the relationships among these Task Template objects.

Figure 11 shows these relationships for the *Describe Relevant Transfer Problem* Activity. The components of each investigation strategy are represented as the “finer grain size Student Model Variables.” After several iterations, the Evidence Model for this Activity was simplified to include one Work Product and one bundled Observable Variable including information from three separate intermediate Observable Variables (see Figure 12). One advantage of the approach 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 students for formative purposes. In addition, teachers or scorers are forced to think about the dimensions of student responses in more detail. The relationship between Observable Variables and Student Model Variables is further specified with the scoring matrices for the Measurement Models within each Activity.

Figure 11. Relationships among Work Products, Evaluation Phases, Observable Variables, Student Model Variables (SMV) and SMV Components for *Describe Relevant Transfer Problem* Activity

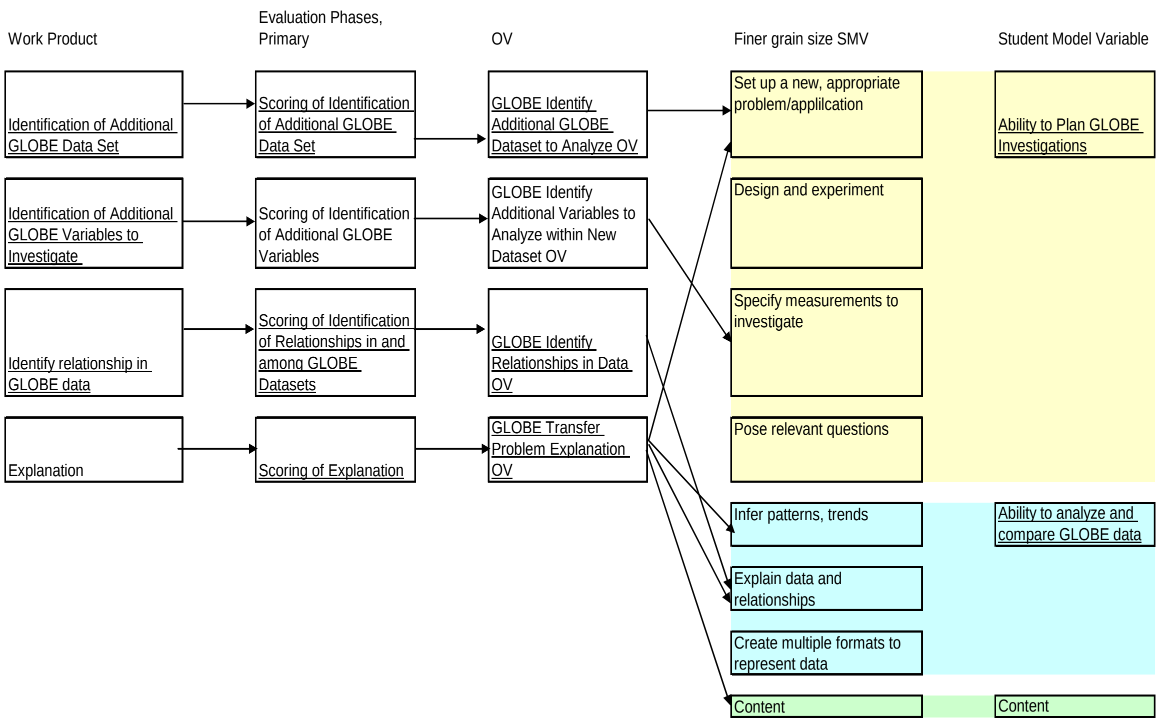
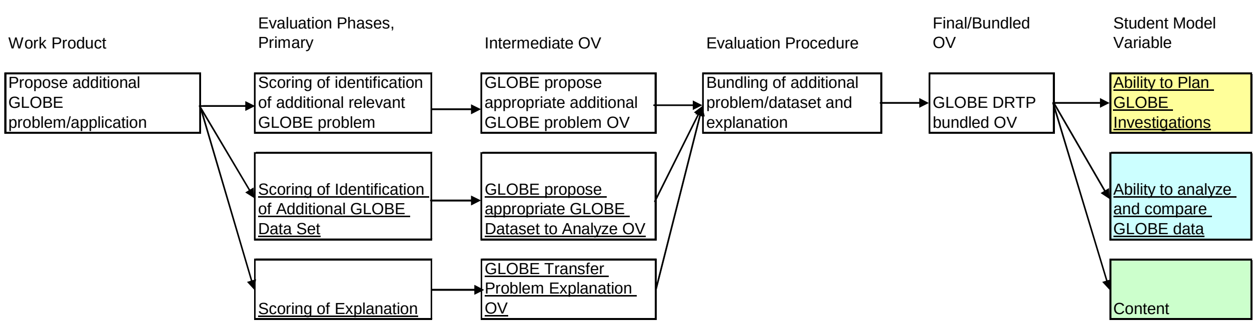


Figure 12. Revised Relationships among Work Products, Evaluation Phases, Observable Variables, Student Model Variables for *Describe Relevant Transfer Problem* Activity



As we determined the procedures for evaluating GLOBE student work and reviewed examples of GLOBE assessments, we determined that Task Model Variable settings may influence Evaluation Procedures. This is particularly true for the Task Model Variables related to scaffolding. Greater scaffolding may limit the amount of content knowledge required or may provide additional information about which data to use for the investigation. As a result, the Work Products produced would differ, as would the Evaluation Procedures that specify how student work should be scored. In addition, Student Model Variable proficiency estimates may be more limited for highly scaffolded items. Table 8 illustrates the effect of the scaffolding on the Measurement Models, as represented in the scoring matrices, for the Pose Research Questions and Find Data Trends Activities.

Table 8. Effect of Scaffolding on the Measurement Model

			Low Scaffolding Level						High Scaffolding Level					
Item	Activity	Score	Plan	Take MM	Interpret	Analyze	Comm.	Content	Plan	Take MM	Interpret	Analyze	Comm.	Content
1	PRQ	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	1	0	0	0	0	1	**0**	0	0	0	0	1
		2	2	0	0	0	0	2	**1**	0	0	0	0	2
		3	3	0	0	0	0	2	**2**	0	0	0	0	2
2	FDT	0	0	0	0	0	0	0	0	0	0	0	0	0
		1	0	0	1	0	0	1	0	0	1	0	0	**1**
		2	0	0	2	0	0	2	0	0	2	0	0	**1**
		3	0	0	3	0	0	3	0	0	3	0	0	**2**

5.6 Review and Refine Task Template

After the Task Template was completed, we reviewed the Task Template for consistency and coherence among objects. We determined whether objects were sufficiently defined and whether all linkages among related objects were specified.

6.0 Testing the GLOBE Student Model with Simulated Data

The GLOBE team conducted a study using simulated data to examine how test information functions change as a result of bundling and scaffolding level (high or low) for the 1-D, 2-D, and 6-D Student Models. The results from these simulated data analyses were then used to develop an Activity Selection Wizard that informs assessment designers about the quality of information (e.g., standard error) related to the Student Model Variable proficiency estimates for the 6-D Student Model when different Activities are included.

6.1 Simulated Data Analyses and Findings

Item responses were generated based on the Student Models and Measurement Models specified in the GLOBE Task Template. We used one approach to generate the bundled data and a second approach to generate the unbundled data. To create the bundled data, the simulation program first generated scores for each Intermediate Observable Variable under the assumption of local independence of item responses. Assuming local independence means that scores were not correlated, given the proficiencies of the students. Next, the simulation program implemented the item bundling procedures defined in the GLOBE Task Template to generate *bundled* data. To create the unbundled data, we used a rejection sampling method. This approach generates scores for each Observable Variable, examines scores that could be bundled, and eliminates invalid responses. For instance, if on two items, a correct response on the first item is required for a correct response the second item, cases that contain an incorrect response on the first item and a correct response on the second are eliminated. We generated data from these two methods under two different scaffolding levels, which affect either or both item difficulty and the design of the scoring matrices. The test information functions from the bundled data and the data resulting from the rejection sampling were compared to evaluate the effect of the two procedures in terms of the amount of information (or precision) each scale dimension has.

Figures 13 and 14 show the test information functions for the bundled and unbundled (rejection method) data for the 2-D Student Model set at low scaffolding. Both Figures show higher peaks for test information for the bundled model, and the bundled model has narrower score range. Findings related to bundling and scaffolding level were similar when the 6-D Student Model was examined with the simulated data. (Because the x-axes in these figures show observed scores, statements about test difficulty in relation to bundled data or unbundled data can not be made on the basis of these figures.)

Figure 13. Test Information Function for Inquiry Student Model Variable in 2-D Student Model with Low Scaffolding

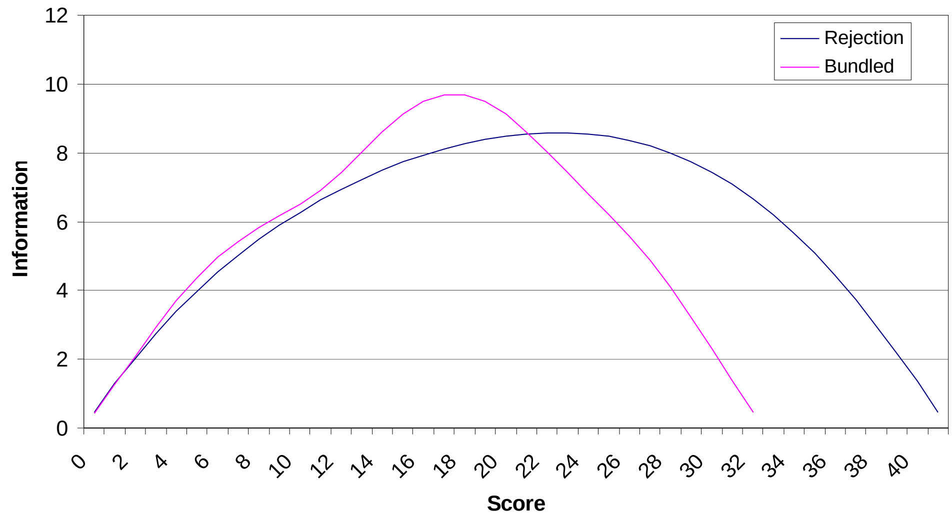
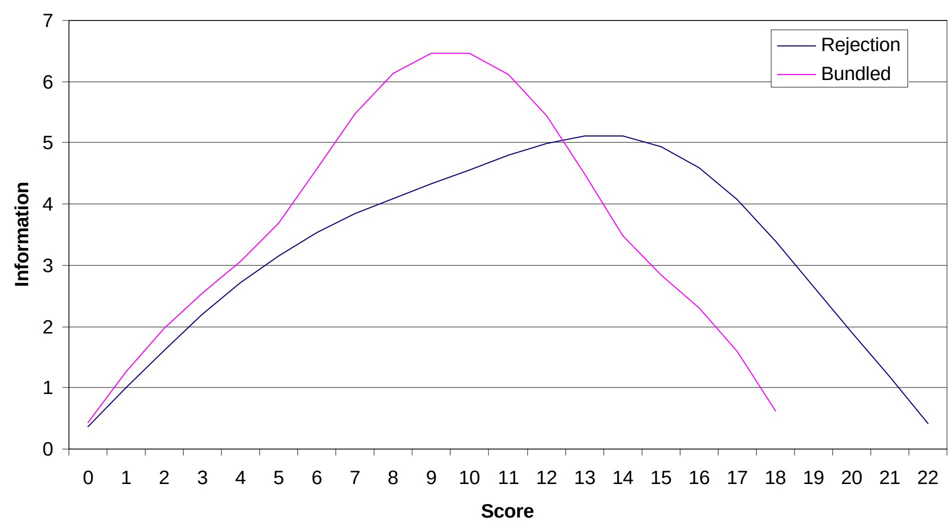


Figure 14. Test Information Function for Content Student Model Variable in 2-D Student Model with Low Scaffolding



6.2 Activity Selection Wizard

We developed a Microsoft Excel-based program that provides test information functions and standard error curves interactively when assessment designers want to determine how the use of different Activities and scaffolding levels affects the precision of the Student Model Variable proficiency estimates. The interactive generation of test information functions and standard error curves is based on the simulated data analyses (bundled approach), but parameter settings such as item difficulty and the scoring and design matrices can be modified within the program.

Figure 15 shows a screenshot of the GLOBE Activity Selection Wizard with the default setting, in which all Student Model Variables are selected. First, the user needs to identify which inquiry skills to investigate by selecting "1" to indicate "Yes" or "0" to indicate "No" within the yellow section on the left side of the screen. The green column to the right of the list of Activities is then automatically updated and indicates which Activities should be included, given the Student Model Variables that were selected. The pink column provides information about how critical the Activities are for providing sufficient information to update the selected Student Model Variables. Any Activity with "high" importance should be included in the assessment to maintain sufficient accuracy of the Student Model Variable estimates.

Figure 15. GLOBE Activity Selection Wizard: Identification of Activities and Level of Importance

SMV	Yes=1, No=0
Plan	1
Take MM	1
Interpret	1
Analyze	1
Comm.	1
Content	1
	Select Skills to Measure

Activities	Yes=1, No=0	Setting	Importance
Pose Research Questions		1	High
Find Data Trends		1	High
Assure Data Quality		1	High
Infer Relationships		1	High
Re-express Data		1	High
Identify Relevant Data and Values		1	Middle
Describe Relevant Transfer Problems		1	High
Communicate Findings and Conclusions		1	High
	Select Activity if you want to select specific activities	This is the activity level TMV setting to be used	Importance of activity given SMV setting

Figure 16 shows the other part of the program, which allows the user to specify scaffolding level. The user can adjust the general scaffolding level, which then updates the setting column associated with each Activity, or users can specify the scaffolding level individually for each Activity. The Activity-level scaffolding is adjusted by entering scaffolding setting in the yellow column next to the Activities.

Figure 16. GLOBE Activity Selection Wizard: Identification of Scaffolding Level

General Scaffolding Level	Low=L Middle=M High=H
This TMV sets activity level scaffolding level settings and task difficulties and information level of	L
	Adjustment to Difficulty
	High= - 1.0
	Middle = - 0.5

Activity Level Scaffolding Level	Low=L, Middle=M, High=H	Setting
PRQ		L
FDT		L
ADQ		L
RBV		L
REXD		L
IFV		L
DRTP		L
CFC		L
	Select Activity Level Scaffolding Level if you want to use cutomize setting	This is the activity level Scaffolding TMV setting to be used

Based on the settings shown in Figures 15 and 16, the Activity Selection Wizard generates the test information functions for each Student Model Variable. Figure 17 shows the test information functions when all Activities are included, and the scaffolding level is low for all Activities. Figure 16 shows the test information functions when all Activities are included, and the scaffolding level is at high for all Activities. The effect of scaffolding can be seen by comparing the maximum height of each curve and the location of the maximum point. The curve with low scaffolding level has higher peak point at right side of theta continuum, so proficiency estimates are more accurate for higher achieving individuals when scaffolding is low.

Figure 17. Test Information Functions for Student Model Variables with Low Scaffolding

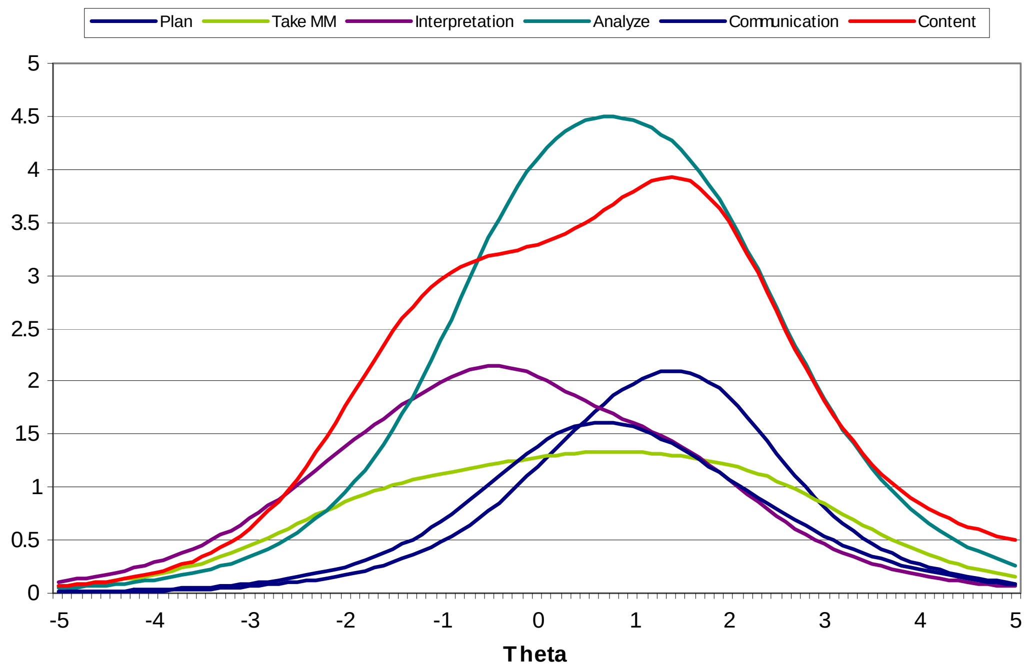
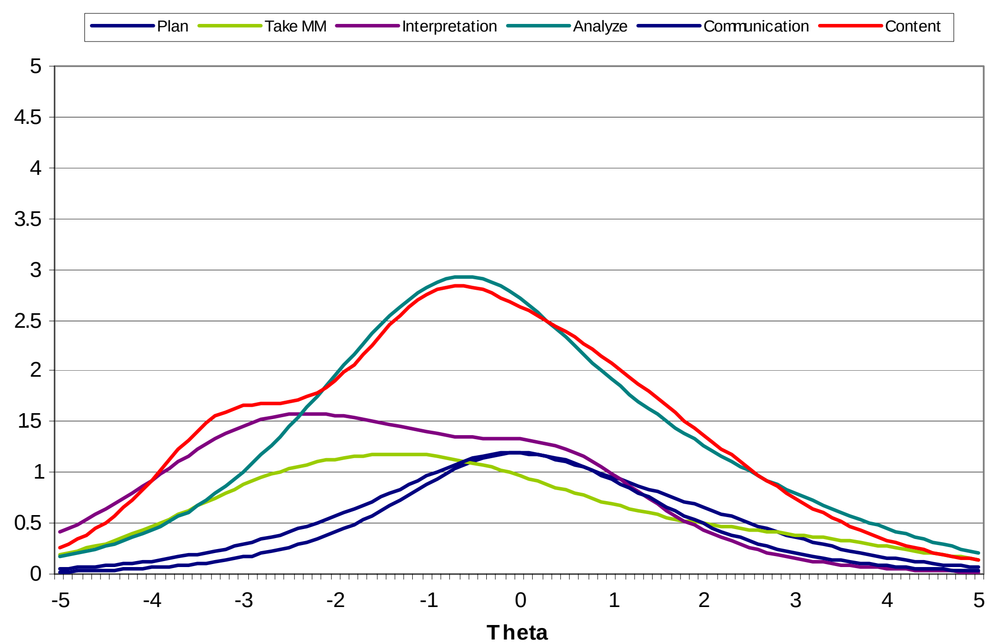
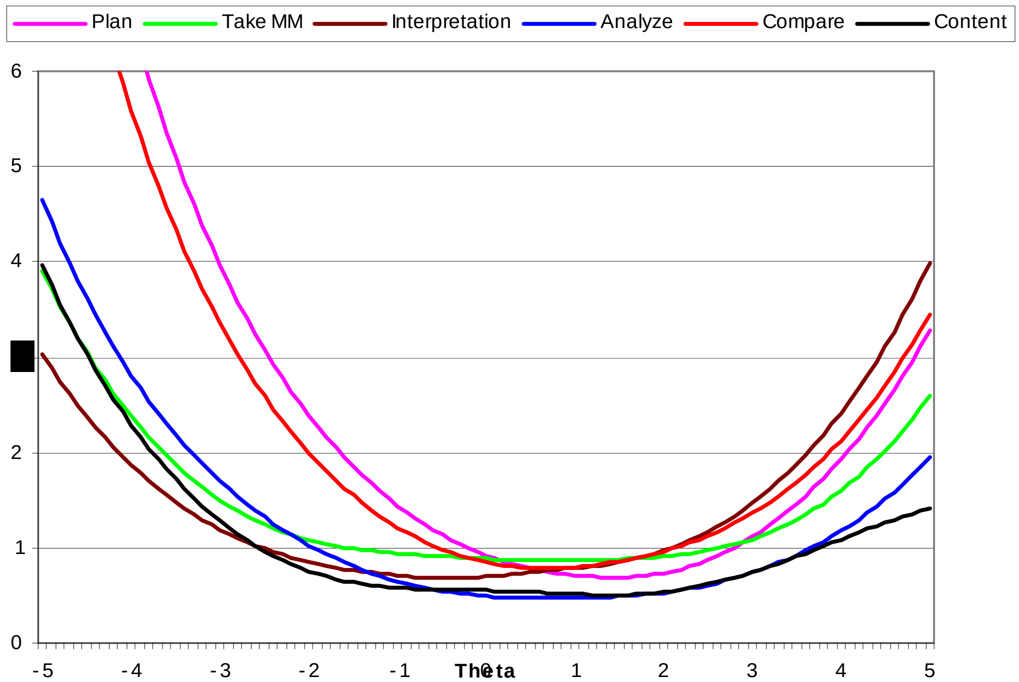


Figure 18. Test Information Functions for Student Model Variables with High Scaffolding



As shown in Figure 19, the Wizard also generates standard error curves to provide additional information to guide users in the selection of Activities.

Figure 19. Standard Error Curves for Student Model Variables with Low Scaffolding



7.0 Conclusions and Implications

Using the PADI Design System to reverse engineer the GLOBE assessment system into Design Patterns and Task Templates helped us to understand the complexity of GLOBE assessments. Generating the Task Template required us to consider the multiple task features that need to be specified (e.g., how the data will be presented and how much scaffolding to provide). The process of creating the GLOBE Task Template also required thinking about the range of information that may be provided about students' understanding of the GLOBE investigation areas and strategies and the advantages and limitations of using finer grain size and coarser grain size Student Model Variables.

While the original GLOBE assessment system contained all of the components needed in order to develop new GLOBE assessments, the PADI Design System pushed us to articulate an assessment argument and strengthen the links among these assessment components. For instance, PADI forced us to think about which specific Observable Variables might be produced, and how scaffolding might impact the Measurement Models. Using the PADI Design System in an iterative way to develop the *GLOBE Inquiry Template* ensured that a coherent and complete assessment argument for GLOBE assessments was instantiated within the Design System.

Through the articulation of an assessment argument, the PADI Design System defines the design space for GLOBE assessments. PADI required us to identify the key components of GLOBE tasks (e.g., Activities), as well as the surface features of these tasks (e.g., representational format of GLOBE data). The *GLOBE Inquiry Template* contains a relatively complete spectrum of components that need to be considered when designing GLOBE assessments.

The development of a complex and comprehensive Task Template for GLOBE assessments facilitates the generation of a family of GLOBE Task Templates and Task Specifications. Simply by setting key Task Model Variables and selecting desired Student Models, the *GLOBE Inquiry Template* was customized into more specified Task Templates. Further specification of Task Model Variable settings resulted in GLOBE Task Specifications. Because the initial *GLOBE Inquiry Template* was thoughtfully developed to ensure a coherent assessment argument, we can be confident that these more specialized GLOBE Task Templates and Task Specifications also reflect a coherent assessment argument.

In conclusion, the PADI Design System is well-equipped to handle the complexities of a multiple-component inquiry task, such as a GLOBE integrated investigation assessment. It 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.

References

Adams, R., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

DeBarger, A. H., & Riconscente, M. M. (2005). *An example-based exploration of design patterns in measurement* (PADI Technical Report 8). Menlo Park, CA: SRI International.

Haynie, K. C., Haertel, G. D., Lash, A. A., Quellmalz, E. S., & DeBarger, A. H. (in press). *Reverse Engineering the NAEP Floating Pencil Task Using the PADI Design System* (PADI Technical Report 16). Menlo Park, CA: SRI International.

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. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (PADI Technical Report 9). Menlo Park, CA: SRI International.

Mislevy, R. J., Hamel, L., Fried, R., Gaffney, T., Haertel, G., Hafter, A., et al. (2003). *Design patterns for assessing science inquiry* (PADI Technical Report 1). Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-66.

Appendix

Sample GLOBE Assessment

Appendix

(Present data from the GLOBE data archives)

GLOBE Data for Holcomb School and Jefferson

School, Fayetteville, Arkansas

Holcomb Elementary
WATER

YY/MM/DD	TEMP (°C)	pH
99/01/24	2.5	8.9
99/01/17	3.2	8.8
99/01/10	3.5	8.8
99/01/03	3.8	8.7
98/12/27	4.0	8.8
98/12/20	4.1	8.8
98/12/13	4.3	8.8
98/12/06	4.2	8.7
98/11/29	4.2	8.7
98/11/22	4.3	8.6

Jefferson Elementary
WATER

YY/MM/DD	TEMP (°C)	pH
99/01/24	4.0	8.7
99/01/17	4.1	8.6
99/01/10	4.5	8.6
99/01/03	4.8	8.6
98/12/27	5.5	8.5
98/12/20	6.0	8.4
98/12/13	6.2	8.4
98/12/06	6.4	8.3
98/11/29	6.4	8.1
98/11/22	6.5	8.1

Holcomb and Jefferson are two schools located within 5 miles of each other in Fayetteville, Arkansas. Both schools sit next to the same river, with Holcomb located upstream from Jefferson. Even though the schools are relatively close to each other, the plant and fish life appears to be different between the two sections of the river. You and several other students have been asked to report to your science class what some of the differences are and why you think they exist. To the left is data from the two schools between late November and late January to help you in your investigation.

1. **(Plan Investigations: Pose relevant questions)** Look at the GLOBE data in the tables. Think of two questions you might ask regarding the data. A sample question might be "What is unusual regarding water temperature between the two schools considering they take measurements from the same river?

2. **(Interpret GLOBE Data: Infer patterns & trends)** One of the students in your research group, Tom, suggests that you can sometimes see patterns in the data. Describe one pattern you see in the pH.

3. **(Take GLOBE Measurements: Use quality assurance procedures)** You have watched some of the students at your school collect GLOBE data and you've noticed that they have done a very good job. Just to check, are there any data that look to you to have errors? How can you tell? What would you tell these students to insure that measurement errors do not happen?

4. **(Interpret GLOBE Data: Explain data & relationships)** One of the students in your science group, Hilda, remembers in her math class showing how different variables relate to each other and suggests doing the same with the GLOBE data. What is the relationship between water temperature and pH for Jefferson Elementary? For example, if water temperature increases, what happens to the pH level?

5. **(Interpret GLOBE Data: Create multiple formats to represent data: explain data & relationships)** Using the data provided for Holcomb, create a graph that has time on the x-axis and pH on the y-axis. Describe what happens to pH as time increases.

6. **(Interpret GLOBE Data: Create multiple formats to represent data; explain data & relationships)** Using the data provided for Jefferson, create another graph with temperature on the x-axis and pH on the y-axis. What happens to pH as the temperature increases?

7. **(Interpret GLOBE Data: Explain data & relationships)** One of the students in your science group, Debbie, mentioned that certain pH levels can affect fish and plant life. If the pH level dropped by 3 in the month of December for either of the schools, what would be the concern regarding life in the river?

8. **(Plan Investigations: Set up another problem)** In the questions so far you have been looking at water temperature and pH. Randomly choose another site in the GLOBE database, pick a different variable that you would investigate, and choose a set of 10 dates in chronological order for the same time period to use as your data. Follow the same steps for this data as you did for Jefferson Elementary and Holcomb Elementary. Why does the new variable you chose seem interesting to you?

9. **(Communicate: Compose reports to explain or persuade)** Using the data analysis you have done, write a short report (1-2 pages) that summarizes your findings and explains why you think the plant and fish life vary between the two schools. Be sure to support your conclusions with data you have analyzed and suggest other data that might be helpful for further study of the river.