

Designing Diagnostic Assessments

Pamela A. Kraus and Jim Minstrell,
Talaria Inc., 821 2nd Ave. Suite 1150, Seattle, WA, 98104

In this paper we describe the process of creating diagnostic assessments to assist teachers in formatively assessing their students. The process begins with the learning targets and ends with the creation of web-delivered sets of questions designed to diagnose students' facets of thinking. Early analysis from our first year of implementation indicates students are reading and thinking about the questions in their assignment. In addition, we are finding that for certain topics students' facets of thinking are highly context dependent.

Middle and high school physical science teachers are faced with many challenges implementing physics education research informed curricula and instructional practices. These include, but are not limited to, very large student-to-teacher ratio, limited knowledge of and access to research on student conceptions, and little experience formatively assessing students. In this paper, we will describe the process of designing diagnostic assessments to assist teachers with these challenges.

Formative assessment is an integral part of most research-based instruction. By formative assessment we mean gathering data on student thinking for the expressed purpose of informing subsequent instruction. Formative assessment ideally informs both the teacher and student about their ideas with regard to the learning targets for the course.

Most assessments readily available to teachers either online or on paper are scored with a two point rubric—either the student got it right or they got it wrong. While sufficient for summative assessment (giving a grade at the end of a course or unit), a two-point scale is not so useful in determining what students who got it “wrong” are actually thinking or what to do about it. To learn about student thinking with these exams, teachers need to study each of the wrong choices with respect to what a student might be thinking if they chose that answer. This assumes that the incorrect answers are aligned with a possible student idea for the concept. This can be a chore even for teachers motivated and interested in assessing

students understanding of important concepts in physics.

Alternatively, teachers can give students well-constructed open-ended questions. Getting detailed student-thinking data from these assessments requires that students have the skill of expressing their ideas in writing. While this skill is extremely important and should be cultivated, this takes time and practice. An additional drawback to open-ended questions is the time it takes to read and analyze all the responses.

With the increased access to computers in most schools, it makes sense to move beyond the “right” and “wrong” approach to assessment and begin to use technology to help teachers diagnose student thinking with respect to the research base on student conceptions. The Diagnoser Project [1] is doing just that. What follows is a description of the process by which a DIAGNOSER set of questions for a particular topic is constructed and used.

The project team of researchers and resource teachers begins with a learning target either from the state or national standards. For example, one of Washington State's Science Essential Academic Learning Requirements for grades 6 through 10 says that all students should be able to “*know the factors that determine the strength of various forces*” and “*identify various forces and their relative magnitudes, and explain everyday situations in terms of force.*” [2] These statements are clearly made up of several smaller instructional units that would be taught over several years. In this particular instance we have

deconstructed these very general statements into five clusters that both address this standard and align with many instructional units. They are Identifying Forces, Forces as Interactions, Electric Forces, Magnetic Forces, and Gravitational Forces.

Once topic clusters are determined we begin the process of organizing *facets* within each cluster that are associated with these learning targets. Facets and facet clusters are a framework for organizing the research on student conceptions such that it is understandable to both discipline experts and teachers. Facet clusters include the explicit learning goals in addition to various sorts of reasoning, conceptual, and procedural difficulties. Each cluster builds up from the intuitive ideas students have toward the targets of learning.

Table 1: Forces as Interactions facet cluster

00	All forces are the result of interactions between two objects. Each object in the pair interacts with the other object in the pair. Each influences the other.
01	All interactions involve equal magnitude and oppositely directed action and reaction forces that are on the two separate interacting bodies.
40	Equal force pairs are identified as action and reaction but are on the same object. For the example of a book at rest on a table, the gravitational force on the book and the force by the table on the book are identified as an action-reaction pair.
50	Effects (such as damage or resulting motion) dictate relative magnitudes of forces during interaction.
51	At rest, therefore interaction forces balance.
52	"Moves", therefore interacting forces unbalanced.
53	Objects accelerate, therefore interacting forces unbalanced.
60	Force pairs are not identified as having equal magnitude because the objects are somehow different.
61	The "stronger" object exerts a greater force.
62	The moving object or the one moving faster exerts a greater force.
63	More active/energetic exerts more force.
64	Bigger/heavier exerts more force.
90	Inanimate objects cannot exert a force.

The facets for the Forces of Interactions cluster are shown in Table 1. The 0X facets are the learning targets for this particular topic. The facets that begin with the numbers 2X through 9X represent more problematic ideas. The X0's represent more general student ideas. Often these are followed by more specific examples, which are coded X1 through X9.

Facet clusters are not exhaustive of all the student conceptions on this topic, but each grouping of problematic facets (X0-X9) is seen in the classroom with at least 10% of the students. The problematic ideas in a cluster of facets are a combination of student difficulties culled from the literature on student conceptions [3], discussions with teachers, and results of action research in the classroom.

The facet cluster is used as a tool to write questions that target at least one of the goal facets and one or more problematic facets in the cluster. Questions can be either multiple choice or numeric response to get a facet diagnosis with this program. We often adapt open-ended questions that were used in researching the facet cluster.

Figure 1: Example of paired questions in the Forces as Interactions set

<p>During soccer practice, Sarah kicks the ball to the other end of the field. Which statement describes the force acting on Sarah's foot by the ball <u>during the kick</u>?</p> <p>A. The ball does not exert a force on Sarah's foot. [Facet 90]*</p> <p>B. The force by the ball is less than the force of Sarah's kick. [Facet 5X or 6X]</p> <p>C. The force by the ball is equal to Sarah's kick. [Facet 00]</p> <p>D. The force by the ball is greater than Sarah's kick. [Facet 5X]</p> <hr/> <p>Which reason best fits your answer to the previous question?</p> <p>A. Sarah is stronger than the ball. [Facet 61]</p> <p>B. Sarah's kick made the ball move, but the ball did not move Sarah. [Facet 52]</p> <p>C. Only Sarah can exert a force--the ball is not alive. [Facet 90]</p> <p>D. All interacting objects exert equal forces on each other. [Facet 00]</p> <p>E. The ball hurt Sarah's foot more than she hurt the ball. [Facet 50]</p>

*Facets shown in brackets above are not displayed to students.

Occasionally, questions are written in pairs to look at whether or not a student shows consistency between their prediction to a situation (a phenomenological question) and their reasoning of the situation. (See Figure 1 on the previous page.) In these cases, students are given feedback after the second question on whether or not their answers were consistent and on the ideas they used to answer the questions.

As one can see by the number of facets in a given cluster, there are too many ideas for a single question to accurately diagnose all the ideas. This is why we design sets of questions to fully diagnose both the learning targets and the problematic facets within a particular cluster. A set is typically between 5 and 10 questions long.

We apply several rules to determine which individual questions are put into a set. The first rule is that all the facets in a cluster should be represented by at least one response within the set of questions that all students see. A second rule is that the context of the questions within a set should vary in accord with the different surface features known to be important to students. In the case of the Forces as Interactions set of questions, writing questions specifically to diagnose the different sub facets in the 5X and 6X categories was sufficient to vary the context of the questions. A third rule is that students receive feedback after each question or pair of questions. This feedback is designed to help students think about their thinking. There are additional rules that have evolved as sets were constructed. One newly enacted rule is that each set contain at least one paired question designed to check students' consistency between their predictions and their reasoning.

An additional benefit of using computers to deliver questions (beyond being able to diagnose facets) is that not all students need to be asked all the same questions. In several of our clusters, different answers will direct students down different paths in the logic of the program. For example, in the Electric Forces sets, there are several questions that ask about forces between a charged object and an uncharged object. In this context, students will often choose a response that is consistent with facets from the Forces as Interactions cluster. Choosing an answer aligned

with a problematic facet from the Interactions cluster will direct students to more questions about interactions. This is done in order to fully diagnose this difficulty before returning the student to the final questions related to electric forces in that set.

A second example of redirecting students happens when we are not sure what reasoning a student might be using to arrive at a particular answer. This student may be asked to type in his or her reasoning as an open-ended response in a text window before continuing. (Open-ended responses are recorded but not assigned a facet.) Another example of branching in the program occurs if a student is asked to repeat a question after some feedback. This might happen in a set if subsequent questions depend on having a correct understanding of the idea, concept or procedure asked about in the repeated question. If a student continues to be diagnosed with a problematic facet in the repeated question, he or she may be told to ask the teacher or a fellow student for help with those ideas before continuing with the set.

The main purpose of the DIAGNOSER is to provide information to teachers about their students' thinking. Immediately after a student has completed a set of questions, the teacher can view the facets diagnosed by the program via the web. The facet data is displayed in a table with rows representing individual students and columns containing the facet codes for the diagnosis in each question. Tables of data on the report page are organized by set and data on individual student, individual class, or multiple classes can be accessed. The description of the facets diagnosed in the set can be opened as a pop-up window. A text version of each of the questions can also be opened as a pop-up window. Another link allows the teacher to see suggested activities to address each of the problematic facets in the cluster. These and many other instructional resources are also available to teachers in the teacher guidance portion of the website.

Preliminary analysis of student data from the DIAGNOSER sets from this past year has found some interesting patterns. [4] One finding is that in the paired questions students are surprisingly consistent between their predictions and their reasoning. We are speculating that this indicates

that the paired-question reasoning responses are aligned well with student facets of thinking and that students are *actually* reading (and thinking about?) the question and responses before selecting an answer.

Another query of the first year data is whether or not students diagnosed with a particular facet in a non-paired question will also be diagnosed with the same facet the next time it is given as a choice. It appears that each cluster is behaving slightly differently. In the Force as Interactions cluster, students do not necessarily use the same facet of thinking in questions that have different contexts (e.g., a student responding with a facet 61 in one question does not necessarily choose facet 60, 62, 63, or 64 if they are options in another question.) However, when the questions are related closely in context such that specific facets are possible choices in the different questions, then a student who chooses facet 63 in one question will very likely choose facet 63 the next time it is a choice. This second finding is true across many of the clusters. We interpret this pattern in the data as reinforcement that to a student facet 61 and facet 62 are very different ideas even though to teachers and researchers these represent the same general problematic idea. Therefore, we should continue to have questions that change context for the student, even if the questions appear to be the same in the eyes of a physicist.

An additional interpretation of these findings is that students likely hold more than one facet of thinking at a given time. Students who think that the effects of the force dictate the relative magnitudes of the force of interactions (facet 50) may also believe that the stronger object will exert a greater force (facet 61). The facet a student chooses may depend on context features of the question, the placement of the question within the set of questions, and/or the order of the responses within the question. Additionally, students' facets of thinking may be influenced by the prescriptive feedback messages they receive in the course of working through a set. Further analysis of the data will help us continue to refine the design process to create sets that fully diagnose students' thinking.

In closing, designing diagnostic assessments is an iterative process of research, design,

implementation, and analysis. We are presently in the implementation and analysis phase. The results of which will have implications for the research and design of the next generation of the Diagnoser instructional tools.

Acknowledgement: The preparation of this paper and the research and development described in this paper were supported by the National Science Foundation under grants REC9972999 and REC0129406.

- [1] The Diagnoser Project Tools can be currently found at the following URL:
<<http://tutor.psych.washington.edu>>
- [2] Washington State. (1998). *Essential Academic Learning Requirements*. Olympia: Office of the Superintendent of Public Instruction.
- [3] The following publications represent a few of the reference sources:
Camp, C., Clement, J., & et. al. (1994). *Pre-conceptions in mechanics: Lessons dealing with students' conceptual difficulties*. Dubuque: Kendall Hunt.
Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. London: Routledge.
Gabel, D. L. (Ed.). (1994). *Handbook of Research on Science Teaching and Learning*. New York: Macmillan Publishing Company.
McDermott, L. C. and Redish, E. F. (1999). Resource letter on Physics Education Research. *American Journal of Physics*, 67(9), 755.
- [4] Private communication with Prof. Earl Hunt and Anne Thissen-Roe, Dept. of Psychology, University of Washington.