Task Templates Based on Misconception Research
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.

Task Templates Based on Misconception Research

Prepared by:
Jennifer G. Cromley, University of Maryland
Robert J. Mislevy, University of Maryland
CONTENTS

1.0 Introduction 1

2.0 Novice-Expert Research 2

3.0 The Challenge for Assessment 4

4.0 PADI Design Patterns and Task Templates 5

5.0 Illustrations from Newtonian Physics 7
   5.1 The Force Concept Inventory 8
   5.2 Student Model 8
   5.3 Evidence Model
      5.3.1 Evaluation Procedures and Evidence Rules 9
      5.3.2 Measurement Model 9
   5.4 Task Models
      5.4.1 G Problems 10
      5.4.2 H Problems 12
      5.4.3 Force Problems—Horizontal and Gravity 13

6.0 Illustrations from Gas Laws 15
   6.1 Student Model 16
   6.2 Evidence Model
      6.2.1 Evaluation Procedures and Evidence Rules 16
      6.2.2 Measurement Model 16
   6.3 Task Model 17

7.0 A General Design Pattern 18

8.0 Final Comments 19

References 20

Appendix A
PADI Template for Force Problems—Gravity Only 25

Appendix B
PADI Template for Force Problems—Horizontal Only 32

Appendix C
PADI Template for Force Problems—Horizontal and Gravity 35

Appendix D
PADI Template for Gas Law Problems 38

Appendix E
PADI Design Pattern for Model Using 41
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The hierarchical structure of PADI templates (from Riconscente et al., 2005)</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>An FCI Problem That Tests Students’ Knowledge of Gravity but Involves No Horizontal Motion</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>An FCI Problem That Tests Students’ Knowledge of Horizontal Motion, but Does Not Test Knowledge of Gravity</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>An FCI Problem That Tests Students’ Knowledge of Both Horizontal Motion and Gravity</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>A TAP Problem That Tests Students’ Knowledge of the Molecular Theory of Gases</td>
<td>15</td>
</tr>
<tr>
<td>A-1</td>
<td>PADI Template “Force Problems—Gravity Only”</td>
<td>25</td>
</tr>
<tr>
<td>A-2</td>
<td>PADI Task Model Variable “Direction of Motion”</td>
<td>27</td>
</tr>
<tr>
<td>A-3</td>
<td>PADI Task Model Variable “Duration”</td>
<td>27</td>
</tr>
<tr>
<td>A-4</td>
<td>PADI Activity “Make Explanations and Predictions from a Physical Situation”</td>
<td>28</td>
</tr>
<tr>
<td>A-5</td>
<td>PADI Student Model “Direction of Motion”</td>
<td>28</td>
</tr>
<tr>
<td>A-6</td>
<td>PADI Evaluation Procedure “Map Multiple-Choice Answers onto Conceptual Model”</td>
<td>29</td>
</tr>
<tr>
<td>A-7</td>
<td>PADI Evaluation Phase “Model Using Multiple Choice”</td>
<td>29</td>
</tr>
<tr>
<td>A-8</td>
<td>PADI Measurement Model “Andersen/Rasch Measurement Model Fragment”</td>
<td>30</td>
</tr>
<tr>
<td>B-1</td>
<td>PADI Template “Force Problems—Horizontal Only”</td>
<td>32</td>
</tr>
<tr>
<td>C-1</td>
<td>PADI Template “Force Problems—Horizontal and Gravity”</td>
<td>35</td>
</tr>
<tr>
<td>D-1</td>
<td>PADI Template “Gas Law Problems”</td>
<td>38</td>
</tr>
<tr>
<td>E-1</td>
<td>PADI Design Pattern “Model Using”</td>
<td>41</td>
</tr>
</tbody>
</table>
Researchers spend much time and effort developing assessments, including assessments of students’ conceptual knowledge. In an effort to make such assessments easier to design, the Principled Assessment Designs for Inquiry (PADI) project has developed a framework for designing tasks and accompanying measurement models. One application of the PADI framework involves “reverse engineering” existing science assessments. This paper reports one such effort, motivated by assessments that elicit students’ qualitative explanations of situations that have been designed to provoke misconceptions and partial understandings. We describe four task-specific templates we created—three based on Hestenes, Wells, and Swackhamer’s (1992) Force Concept Inventory and one based on Novick and Nussbaums’s (1981) Test about Particles in a Gas (TAP). We then describe an overarching framework for these templates, another PADI object called a design pattern, based on Stewart’s concept of “Model Using.” For each template, we describe a multivariate Student Model, a Measurement Model, and a Task Model. We conclude by suggesting how these templates and the design pattern could help researchers (and perhaps teachers) who wish to design new assessments in science domains where students are known to hold misconceptions.
1.0 Introduction

Creating tasks to assess underlying concepts and inquiry processes in science is not easy. The National Science Foundation has funded the Principled Assessment Designs for Inquiry (PADI) project, under the Interagency Educational Research Initiative (IERI), to create a conceptual framework and supporting software to help people design inquiry assessments. Among the data structures PADI has developed to this end are design patterns, which lay out assessment arguments at a conceptual level; task templates, which are schemas for the operational elements of an assessment and support the creation of families of related tasks; and task specs (short for “task specification”), which describe the elements of individual tasks in transportable formats—specifically, the Question and Test Interoperability (QTI) standards for electronic learning and assessment materials developed by the IMS Global Learning Consortium (n.d.) and extensions thereof.

One type of activity within the PADI project involves taking existing assessments used for science inquiry research and writing templates and task specification for those assessments. We refer to this process as “reverse engineering,” in that the templates and task specification that are developed from existing tasks could be used to reproduce the same assessments or to produce new or analogous questions in the same or a similar domain. This report presents the results of applying reverse engineering to two conceptual assessments in science, the Force Concept Inventory (FCI; Hestenes et al., 1992) and the Test about Particles in a Gas (TAP; Novick & Nussbaum, 1981). Both assessments are based on research about student misconceptions, an area of cognitive psychology research on expert and novice performance.

Section 2.0 provides a brief review of the theoretical basis for these assessments in the novice-expert and misconceptions paradigms in cognitive psychology. Section 3.0 discusses the challenges that misconceptions research poses for assessment and the benefits of making available to researchers the type of design patterns and templates that PADI is producing. Section 4.0 summarizes the structure of PADI task templates. Section 5.0 discusses the development of a design pattern and template for the Force Concept Inventory, a conceptual assessment of knowledge about Newtonian physics. The Student Model, Evidence Model, and Task Model for the template are discussed. Section 6.0 describes the process of adapting the template for the Test about Particles in a Gas. Section 7.0 discusses a higher-level, less technical assessment design tool called a design pattern (Mislevy et al., 2003), which describes a general class of tasks like the FCI and TAP tasks, aimed at revealing student misconceptions. Section 8.0 closes with a summary of the potential benefits to users of the design pattern and of the four templates that were developed.
Novice-Expert Research

One of the dominant strains of cognitive psychology research from the mid-1970s through the 1980s was the study of expert performance across numerous domains (for reviews, see Charness & Schultetus, 1999; Ericsson & Charness, 1997). The basic premise of this line of research was that if the characteristics of expert performance could be isolated and identified, perhaps novices could be trained in the same specific knowledge, skills, and abilities to move them closer to expert performance. Moving away from previous notions of expertise as general and inborn, cognitive psychologists conducted research that led them to see expertise as domain specific and acquired through extensive teaching and practice (Ericsson & Smith, 1991; Ericsson, 1996). Whereas early expert-novice research tended to consider only expert performance (e.g., in chess, de Groot, 1946/1978), later research often contrasted experts and novices (e.g., in physics, Larkin, McDermott, Simon, & Simon, 1980). Early research also focused on problem-solving in domains with clearly delimited solutions and limited solution paths, called “well-defined domains” in the literature. Physics, chess, and medicine, in particular, were studied often, but other well-defined domains included occupations such as avionics technicians, waiters, and taxi drivers.

In physics, several researchers contrasted physics professors with typical undergraduate students. For example, Chi, Feltovich, and Glaser (1980) asked both physics professors and undergraduate students to sort various physics problems. Whereas the professors sorted the problems according to the physical laws that would be used to solve the problem (e.g., Newton’s third law), novices tended to sort them according to physical features of the problem (e.g., pulley problems). In chess, Chase and Simon (1973) found that expert players were better able than novices to reconstruct the positions of chess pieces from memory, but only when the pieces were arrayed in actual game positions, not when they were placed randomly on the board. In medicine, Patel and Groen (1991) found developmental trends in medical expertise; whereas first-year medical students were not even aware that a written case contained irrelevant information, second-year students were distracted by that irrelevant information, while physicians recognized it as irrelevant.

In the course of developing a computer-based Intelligent Tutoring System (ITS) for military aircraft electronics technicians, Gitomer and colleagues found that expert technicians used a specific problem-solving strategy they termed “space splitting” (Steinberg & Gitomer, 1996). If there was an electrical fault in a line, expert avionics technicians would pick a halfway point between the power source and the inoperable part (e.g., a wing flap or aileron), and would test the circuit on both sides of the line. By continuing this process, they could quickly identify the malfunctioning electrical component. Novice technicians, by contrast, would use an inefficient process of checking each component in turn, one at a time. Ericsson and Polson (1988) found that the expert waiter whom they studied had developed detailed heuristics (e.g., males with certain builds were likely to order certain types of steaks) and mnemonics for remembering diners’ orders. Expert taxi drivers studied by Chase (1982) used mental imagery to determine the quickest route from the current location to their destination.
These studies in well-defined domains established that expertise takes many years of
guided practice to develop, that experts have a larger base of declarative knowledge than
do novices, that their knowledge is better integrated and organized according to key
principles in the domain, that experts know more domain-specific strategies, that their use
of these strategies is automated, and that experts have more conditional knowledge about
strategies (that is, they know in what situations to enact a particular strategy).

A later body of expert-novice research applied these findings to domains that do not have
a single solution or finite set of solution strategies (called “ill-defined domains”), such as
reading (e.g., reading law cases, Lundeberg, 1987; reading poetry, Peskin, 1998), history
(e.g., Wineburg, 1991), teaching (e.g., Sabers, Cushing, & Berliner, 1991), and writing (e.g.,
Bruleux, 1991), with similar results.

Typically, researchers in novice-expert studies collect some sort of verbal report from
participants, such as a think-aloud protocol (participants verbalize everything they are
doing while performing a task), retrospective protocol (participants report after the fact
what they think they were doing), or conduct an interview. A few methodologists have
suggested that expertise researchers should simultaneously collect other process data,
such as recording computer keystrokes, what participants are looking at (using eye-
tracking devices), or reaction times (Ericsson & Smith, 1991; Magliano & Graesser, 1991).
Researchers also have used other methods, such as the sorting task described above for
Chi et al.’s (1980) physics expertise study.

One common finding across many expert-novice studies, especially in the sciences, is that
novices often have specific mistaken ideas about particular domains. In one highly
publicized example, 88% of graduating Harvard seniors surveyed believed that the seasons
were caused by Earth’s elliptical orbit around the sun, rather than by the tilt of the Earth on
its axis (Gardner, 1999). Termed “misconceptions,” these are ideas derived from daily
experience that students bring to their learning experiences and that contradict scientific
understandings and are often resistant to change.

Some other examples of misconceptions include: in biology, exercise makes breathing
faster but shallower (e.g., Michael, 1998); in subtraction, subtract the smaller number from
the larger, regardless of which number is being subtracted from (e.g., 725 – 569 = 244,
since 7 – 5 = 2, 6 – 4 = 4, and 9 – 5 = 4, e.g., Brown & van Lehn, 1980); in history, textbooks
are always correct and we can know “what happened” (e.g., Rouet, Britt, Mason, & Perfetti,
1996); in evolution, animal behavior changes offspring biology (that is, a Lamarckian belief;
e.g., Bishop & Anderson, 1990); and in Earth science, the Earth is a globe with a flat top
(e.g., Vosniadou & Brewer, 1992). Research about misconceptions thus has been a wide-
ranging and productive field for cognitive science research, and one that has been
frequently used in science inquiry research. This misconceptions research, however, has
rarely been used to design assessments in a principled manner.
3.0 The Challenge for Assessment

Assessing students’ learning in a way that reveals misconceptions poses considerable challenges for test designers. Students have many conceptions about any given domain, and some of these conceptions (i.e., misconceptions) can interfere with subject matter learning. However, research shows that students with misconceptions can nonetheless frequently correctly answer questions on traditional assessments. Physics students, for example, may be able to use algorithms to calculate answers to problems about which they have no conceptual understanding (e.g., Clement, 1983; Hunt & Minstrell, 1994). Assessments that are designed in light of the misconceptions literature, therefore, usually eschew calculation problems, on the grounds that these can be solved without having any conceptual understanding of the problem. In addition, because the purpose of assessing students’ conceptions may be more diagnostic than normative, researchers (and teachers) may want to know which specific misconceptions students have (see Frederiksen & White, 1988). This type of use for assessment therefore requires different types of Measurement Models (and perhaps also Student Models) than does an end-of-semester physics examination used for assigning grades.

Researchers who are studying science inquiry also may be particularly interested in measuring conceptions, since inquiry activities (more so than didactic or cookbook lab approaches) may reveal students’ misconceptions (e.g., Dalton, Morocco, Tivnan, & Mead, 1997; White & Frederiksen, 1998). However, developing psychometrically sound measurement instruments of science learning that reveal students’ misconceptions is a difficult and time-consuming task, and one at which test developers are not always successful (see Cornely-Moss, 1995). Researchers who study science inquiry may be spending much time, energy, and resources “reinventing the wheel” as they develop measures for individual projects. In addition, researchers rarely have access to the kind of sophisticated statistics and technology that make both Student Models and statistical models more accurate and efficient. The goal of the PADI project is to address these needs by creating structures, including design patterns and templates, to help assessors organize their thinking about science learning into the shape of assessment arguments and tasks, and to illustrate their use in a variety of contexts. In this technical report, we present four templates and one design pattern that encompass misconception-based assessment in two science domains: Newtonian physics and gas laws.
4.0  PADI Design Patterns and Task Templates

PADI design patterns and task templates build on the evidence-centered assessment design (ECD) models of Mislevy, Steinberg, and Almond (2003). A good starting point is a quote from Messick (1994):

A construct-centered approach [to assessment design] would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics. (p. 16)

A PADI design pattern lays out, at a conceptual level, coherent sets of possibilities for the elements of the assessment argument outlined in the Messick quotation, organized around some aspect of scientific inquiry or conceptual knowledge. For example, there are the targeted knowledge, skills or abilities (targeted KSAs); things that students might say, do, or make that provide evidence of these KSAs (potential observations); and characteristic features of task situations in which these observations might be made. The design pattern structure is described and illustrated in detail in PADI Technical Report 1, Design Patterns for Assessing Science Inquiry (Mislevy et al., 2003). This brief description, however, should suffice to understand the example in Section 7.0.

More attention is focused in this presentation on task templates. The reader is referred to PADI Technical Report 3, An Introduction to PADI Task Templates (Riconscente, Mislevy, Hamel, & PADI Research Group, 2005). Task templates are organized around three basic models in the Mislevy et al. (2003) evidenced-centered assessment framework—namely, the Student Model, the Evidence Model, and the Task Model:

- The Student Model contains variables that correspond to knowledge, skills, and abilities of an examinee about which inferences will be made—decisions about selection, placement, certification, instruction, task selection, and so on.

- The Evidence Model is a set of instructions for interpreting the response (Work Product) to a specific task. The Evidence Model contains two parts. The first is a series of Evaluation Procedures that describe how to identify and evaluate essential features of the Work Product. The second is a Measurement Model that tells how the belief about the Student Model Variables for a given student should be updated in light of the observed features of that student’s responses.

- The Task Model is a generic description of a family of tasks. A Task Model contains (1) a list of variables that are used to describe key features of the tasks, such as their content, difficulty, and conditions under which they are presented; (2) a collection of Presentation Material specifications that describe the structure and format of material that will be presented to the participant as directions, stimulus, prompt, or instruction; and (3) a collection of Work Product specifications that describe the structure and format of traces of what the student will say, do, or make that will be
evaluated (e.g., item choices, sequence of solution steps, written explanations, videotapes of performances).

Figure 1 shows the constituent objects in a PADI template, which together encompass the Student, Evidence, and Task Models described above. The interested reader is referred to Riconscente et al. (2005) for detailed explanations of the structure of templates. This diagram, however, will help in understanding the examples of templates presented in the appendices, by indicating the hierarchical structure of the pieces that make up a template.

**Figure 1: The Hierarchical Structure of PADI Templates**
5.0 *Illustrations from Newtonian Physics*

One domain in which much work on misconceptions has been done is Newtonian physics. Several studies have shown that students can solve typical classroom quantitative problems in Newtonian physics, while failing to understand basic Newtonian principles (see, for example, DiSessa, 1993; White & Frederiksen, 1998; Clement, 1983). Newtonian motion problems can be divided into problems that consider an object moving in a horizontal plane (e.g., a hockey puck moving across ice), problems that concern only gravity acting on an object (either a still object or a moving object), and problems concerning horizontal motion and the vertical force of gravity acting simultaneously. A typical undergraduate physics assessment question asks whether students can correctly compute the time, distance, acceleration, or force needed to move an object:

---

*A projectile is fired horizontally from a flare gun located 45.0 m above the ground. The projectile’s speed as it leaves the gun is 250 m/s.*

(a) How long does the projectile remain in the air?

(b) What horizontal distance does the projectile travel before striking the ground?

(c) What is its speed as it strikes the ground?

(d) If the projectile were simply dropped from a height of 45.0 m, instead of fired horizontally from that height, how much time would it take to reach the ground? How does this compare with your answer to part (a)?

---

The following are common misconceptions in Newtonian physics:

- “Impetus,” the pre-Newtonian notion that as long as a body is in motion, a force must be acting on it (McCloskey, 1983). Students may be able to correctly calculate the amount of force required to set a body in motion—e.g., to move a 10-kg box from rest—but they might believe that force is required to keep the box moving. In actuality, once the box has accelerated to its final velocity (assuming a frictionless world), no force is required to keep it in motion.

- Heavier objects exert more force (disSessa, 1993). (It does, of course, require more force to start a heavier object in motion.)

- Objects not at rest must have no forces acting on them (Clement, 1983). In actuality, when an object such as a book sits on a table, gravity (a force) pulls down on the book, and the table pushes back up with exactly the same amount of force against gravity.

- Heavier objects fall faster than lighter ones (the belief tested in Galileo’s perhaps apocryphal experiment at the leaning tower of Pisa).

- The "Wile E. Coyote" misconception—the notion that “Any body suspended in space will remain in space until made aware of its situation" (Hope, 1994).

- Falling objects that are also moving horizontally fall straight down or in right-angle or circular arcs. A student might be able to calculate the horizontal distance that an
object travels but misunderstand that objects that move both horizontally at a constant speed and vertically under the force of gravity always move in a parabola.

5.1 **The Force Concept Inventory**

In 1992, David Hestenes and colleagues published a multiple-choice measure of students’ conceptual knowledge of Newtonian physics called the Force Concept Inventory, or FCI (Hestenes et al., 1992). The FCI has been widely used as an undergraduate physics pre- and posttest to measure whether students truly understand motion or whether they can simply do calculations but lack a fundamental understanding. The FCI consists of 30 multiple-choice Newtonian physics problems that do not require any calculation but are designed to tap students’ understanding of various aspects of Newtonian mechanics and circular motion. The measure was constructed by reviewing prior research on correct Newtonian conceptions and specific student misunderstandings. In the original publication, 23 specific Newtonian force concepts were listed (Hestenes et al., 1992, Table I) and used to construct correct options for the FCI. Additionally, 30 specific misconceptions were listed (Table II) and used to construct incorrect options for the same items. The Force and Motion Conceptual Evaluation (FMCE; Thornton & Sokoloff, 1998) is a similar assessment.

The following sections describe elements necessary for task templates that describe FCI-type tasks—that is, reverse-engineered templates that might be thought of as more general structures from which the actual tasks could have been derived. More substantive descriptions follow; PADI design system screen shots of actual PADI templates appear in the appendices.

5.2 **Student Model**

Hestenes and colleagues (1992) originally conceived of a univariate Student Model—students are Newtonian to either a greater or lesser extent (i.e., they answer more or fewer FCI questions correctly). Bao and Redish (2001) suggested an alternative interpretation: A student can be in multiple belief states (e.g., Newtonian reasoning; Galilean, or “impetus” reasoning; and nonscientific, including Aristotelian reasoning) simultaneously, applying one in any given situation with a probability that depends partly on the student’s propensity to reason through each of the perspectives and partly on the features of the situation. The Student Model Variables characterize the students’ propensities to reason through each of the perspectives. Task parameters, discussed below, characterize a task’s tendency to evoke thinking through the same perspectives.

This approach is similar to other multivariate models used in developmental and cognitive psychology. For example, Robert Siegler (1976) has analyzed children’s developing ability to solve problems in which different weights are put on either side of a balance beam. Children at the same stage of development may give different answers to the same balance beam problem, and children at different stages may give similar answers. However, overall, children in the same stage show a similar distribution of answers that are characteristic of their developmental stage. Thus, with the FCI we could imagine a student who answers like a Newtonian 70% of the time, like a Galilean 20% of the time, and in a
nonscientific manner 10% of the time, recognizing that which kind of answer a student gives also depends on the features of the task at hand.

In the PADI design system, this Student Model has been named the “FCI-ish Student Model.” Students have propensities to respond to tasks in a certain class (e.g., FCI and FMCE items) in terms of three specific conception models: Newtonian, Galilean, and nonscientific. Formally, each student \( i \) is characterized by a vector of three real numbers, \( (\theta_i^1, \theta_i^2, \theta_i^3) \), where higher numbers indicate a greater propensity toward a given response class—in this case, Newtonian, Galilean, and nonscientific. Statistically, the model can be made identified by implicitly fixing the sum of each examinee’s three parameters at zero, or fixing the first parameter to zero. The latter approach is taken in the example illustrated here, so there are only two Student Model Variables to be estimated, those for the Galilean and nonscientific propensities.

## 5.3 Evidence Model

### 5.3.1 Evaluation Procedures and Evidence Rules

The FCI was originally a multiple-choice measure; evidence about the Student Model consisted of which multiple-choice option was chosen. With the multivariate Student Model approach we have taken in PADI, these evidence rules are expanded. Specifically, each multiple-choice option has been mapped to known Newtonian conceptions (Hestenes et al., 1992, Table I) and non-Newtonian misconceptions (which are further subdivided into the Galilean and nonscientific misconceptions contained in Table II). Although we have chosen to analyze the FCI in its original multiple-choice format, it would be easy to adapt it for open-ended responses, which would be coded. These codes would then be mapped to specific conceptions. We will see later that just this approach was taken by the designers of an analogous measure, the Test about Particles in a Gas (TAP; Novick & Nussbaum, 1981).

### 5.3.2 Measurement Model

The Measurement Model for the FCI was originally a simple additive model. A multivariate statistical model for the FCI was recently investigated by Huang (2003) in dissertation research. Huang used an Andersen/Rasch (A/R) multivariate model (Andersen, 1973), in which a student is seen as being in several belief states, each with a certain propensity, and each FCI question is seen as evoking belief states, each question with a certain tendency to evoke responses of the different types.

The A/R model takes the following form for a situation in which there are \( m \) response types and student and task parameters that correspond to them. Let \( X_{ij} (i = 1, \ldots, n; j = 1, \ldots, k) \) be independent observable random variables (\( i \) is the index for examinees, \( j \) is the index for items), where \( X_{ij} \) can be any integer between 1 and \( m \). The probability that the response \( X_{ij} \) that student \( i \) produces for task \( j \) is of type \( p \) is given as

\[
P(X_{ij} = p) = \frac{\exp(\theta_{ip} + \beta_{jp})}{\sum_{p=1}^{m} \exp(\theta_{ip} + \beta_{jp})},
\]
where:
\( p \) is an integer between 1 and \( m \), indicating response class;
\( \theta_i^p \) is the \( p \)th element in person \( i \)'s vector-valued parameter; and
\( \beta_j^p \) is the \( p \)th element in item \( j \)'s vector-valued parameter.

Note that there are \( m \) probabilities for each examinee on a given item, representing the probability of that person's making any particular choice on that item.

The Andersen/Rasch model can be written as a special case of the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) used by the PADI project (Adams, Wilson, & Wang, 1997). For more information, see Huang (2003).

5.4 Task Models

In the PADI design system, we have created four task templates related to misconception research. All have the same Student and Evidence Model structures but differ as to the Task Models. Three Task Models are for particular types of FCI problems (Hestenes et al., 1992), and one is for the TAP (Novick & Nussbaum, 1983). In these Task Models, we do not create new items or content that goes beyond the work done by the authors of the measures. Rather, we write a more general set of task template structures (similar to item specifications) that can enable researchers—and perhaps teachers—to create analogous measures in their own domains of interest.

The three templates for FCI-type problems correspond to (a) problems involving only gravity and no horizontal motion, which we refer to as G Problems; (b) problems involving only horizontal motion, which do not tap students' knowledge of gravity, which we refer to as H Problems; and (c) problems that tap students' knowledge of both horizontal motion and gravity, which we call H & G Problems.

5.4.1 G Problems

A typical G Problem from the FCI is shown in Figure 2. The template titled Force Problems—Gravity Only refers to the state of motion of the object (up, down, or no motion), the duration of motion, friction, the identity of the vertical force (e.g., gravity, the force of an elevator cable pulling up), whether or not there is an illustration used in the problem, the mass and identity of the object being moved, the speed at which the object is being moved, and the time period of interest (e.g., how long it takes an object to fall). (Appendix A presents templates and related task specification for Force Problems—Gravity Only.) For example, the problem shown in Figure 2 involves the continuous upward motion of an elevator at an unspecified speed, no friction, the upward force of the elevator cable, the downward force of gravity, and an illustration.
Some of these Task Model Variables are important in describing incidental features of the task situation, which may be varied to provide a range of tasks that are different on the surface but similar as to the thinking they tend to evoke. The identity of the object is such a variable. Other Task Model Variables are important because they characterize features that are linked to common misconceptions. State of motion—whether an object is at rest, moving up, or moving down—is an example of this type. Many students believe that when an object is at rest, there are no forces acting on it.

G Problems in the FCI ask about the forces acting on an elevator being pulled up a shaft, a steel ball that has been tossed straight up, a stone dropped from the roof of a building, a tennis ball at an instant after it has been hit, a metal ball at an instant inside a tube, and an office chair at rest on a floor. This particular template helps a researcher to create more FCI-like G Problems. The researcher creates a situation of the kind described in the Task Model and describable by the Task Model Variables. For multiple-choice tasks like those on the FCI, options are constructed by giving one or more predictions or explanations for the situation, each consistent with Newtonian reasoning, with Galilean impetus reasoning, or with nonscientific (e.g., Aristotelian) reasoning. As illustrated in Section 7.0, one can reason by analogy to another science domain in which students have misconceptions to create one or more new templates from which to develop an entirely new measure.
5.4.2 H Problems

A typical H Problem from the FCI is shown in Figure 3. The template titled Force Problems—Horizontal refers to the identity, direction, amount, and duration of force applied; the mass and identity of the object being moved; whether an illustration is included; the speed and direction of the object being moved; and the time period of interest (before, during, or after the force is applied). (See Appendix B for the Force Problems—Horizontal template.) Note that the G Problem and H Problem templates have seven Task Model Variables in common: the direction of motion, duration, friction, the identity of the force applied, whether an illustration is included, the mass of the object being moved, and the time period of interest. For example, the problem shown in Figure 3 involves an instantaneous “kick” of unspecified force delivered at right angles to the direction of motion of a hockey puck; the puck’s speed and mass are not specified (although we can assume it is lighter than, say, a rocket); we are told to imagine a frictionless situation; and the problem is illustrated.

Figure 3: An FCI Problem That Tests Students’ Knowledge of Horizontal Motion But Does Not Involve Gravity


Task model variables that are particularly important in eliciting misconceptions are the direction and duration of the force. Table 1 shows examples of how certain combinations of task model variables for a Newton’s third law problem (“for every action there is an equal and opposite reaction”) can tend to elicit specific conceptions or misconceptions (see Hammer & Elby, 2003).
Table 1: Conceptions That Tend to Be Provoked by Combinations of Task Model Variables

<table>
<thead>
<tr>
<th>Object That Does the Hitting</th>
<th>Light</th>
<th>Heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object That Gets Hit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>Correct Newtonian conception</td>
<td>Misconception of more force from the heavier object</td>
</tr>
<tr>
<td>Heavy</td>
<td>Misconception of more force from the heavier object</td>
<td>Correct Newtonian conception</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object That Does the Hitting</th>
<th>Slow</th>
<th>Fast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object That Gets Hit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow</td>
<td>Correct Newtonian conception</td>
<td>Misconception of more force from the faster object</td>
</tr>
<tr>
<td>Fast</td>
<td>Misconception of more force from the faster object</td>
<td>Correct Newtonian conception</td>
</tr>
</tbody>
</table>

H problems in the FCI ask about a hockey puck that receives a kick, a rocket in space whose engine turns on, a car that is pushing a truck, a truck that collides with a car, a woman pushing a box, and two students on office chairs who push away from each other. This template likewise would help a researcher create more FCI-like H Problems or to create a new measure.

5.4.3 Force Problems—Horizontal and Gravity

The third Task Model we created was for more complex problems that require students to apply their knowledge of both horizontal motion and the effect of gravity. A typical H & G Problem from the FCI is shown in Figure 4. The template titled Force Problems—Horizontal and Gravity (see Appendix C) refers to the identity, direction, and amount of force applied; the mass and identity of the object being moved; whether an illustration is included; the direction of motion; and the shape of the path of motion of the object being moved. Note that several task model variables from the G Problem and H Problem templates are missing from this more complex template: the duration of force applied, the speed and direction of the object; the mass, size, and identity of the object; and the time period of interest (before, during, or after force is applied). For example, the problem shown in Figure 4 involves the instantaneous force of a cannon (the amount of force is not specified) shooting a (presumably heavy) cannonball; the shot is made in a forward direction parallel to the ground; the question asks examinees to determine the shape of the path of motion; and an illustration is included.
Figure 4: An FCI Problem That Tests Students’ Knowledge of Both Horizontal Motion and Gravity

12. A ball is fired by a cannon from the top of a cliff as shown in the figure below. Which of the paths would the cannon ball most closely follow?


H & G Problems in the FCI ask about a cannonball shot out of a cannon, balls rolling off a cliff, a bowling ball dropped from a flying airplane, and a tennis ball hit into the wind. Similar to the other two templates, this template would allow a researcher to create more FCI-like H & G Problems or to create a new measure.
A second scientific domain in which students have been shown to hold misconceptions is
the gas laws—systematic relationships among the pressure, volume, and temperature of
gases in closed containers. Some misconceptions that have been observed include the
notion that in a partial vacuum gases “rise to the top” or “sink to the bottom” of their
container (see Benson, Wittrock, & Baur, 1993; Lin, Cheng, & Lawrenz, 2000; Mas, Perez,
& Harris, 1987; Meheut, 1997). On the basis of several interview studies, Novick and
Nussbaum (1978, 1981) identified five core beliefs that make up a scientific conception of
gas behavior: that gases are made up of particles; that there is empty space between the
particles; that the particles are uniformly distributed in a closed container; that the
particles are in constant motion, and that when a gas becomes a liquid, there is a change in
the density of the collection of particles.

Novick and Nussbaum (1981) constructed the Test about Particles in a Gas (TAP), a
noncomputational measure of students’ conceptual knowledge about gas behavior. It is
an eight-question measure combining multiple-choice questions with drawing tasks and
other constructed-response questions that are scored by coders into conceptual
categories. A sample problem from the TAP is shown in Figure 5.

**Figure 5: A TAP Problem That Tests Students’ Knowledge of the Molecular Theory of
Gases**

*Note. From “Pupils’ understanding of the particulate nature of matter: A cross-age study,” by S. Novick & J. Nussbaum,
Like the FCI, the TAP is designed to reveal students’ conceptions about how gases behave, not to test their ability to compute typical gas law problems. For contrast, a typical high school assessment question asks whether students can correctly compute the mass, volume, or pressure of a gas under certain conditions, or to identify a gas given information about its temperature, volume, and pressure:

105. 1.09 g of H₂ is contained in a 2.00 L container at 20.0 °C. What is the pressure in this container in mm Hg?

118. Determine the number of grams of carbon dioxide in a 450.6 mL tank at 1.80 atm and minus 50.5 °C. Determine the number of grams of oxygen that the same container will contain under the same temperature and pressure.

132. If 9.006 grams of a gas are enclosed in a 50.00 liter vessel at 273.15 K and 2.000 atmospheres of pressure, what is the molar mass of the gas? What gas is this?

Note: Selected gas law problems from Diamond Bar High School, Walnut Valley Unified School District, CA (Park, 1996)

As with the FCI, we “reverse engineered” a design template for constructing a TAP-like assessment that uses the research base about student misconceptions to develop a principled assessment (see Appendix D).

6.1 Student Model

As with the FCI, we propose that students can be seen as being in multiple belief states, each with a certain propensity. For example, students might simultaneously believe that particles in a gas are uniformly distributed in a closed container and also believe that particles are non-uniformly distributed, each with a certain propensity. As with the FCI, the original Student Model was a simple univariate model (students are characterized as either scientific or nonscientific).

6.2 Evidence Model

Because the TAP includes both multiple-choice and constructed-response questions, its Evidence Model is more complicated than those of the three FCI templates.

6.2.1 Evaluation Procedures and Evidence Rules.

The evidence rules for multiple-choice items on the TAP are analogous to those for the FCI Evidence Models—evidence about the student corresponds to which multiple-choice option was selected. For constructed-response items, however, the responses must be coded by a coder into a priori categories that correspond to model classifications that are mapped onto Novick and Nussbaum’s (1981) five principles—that is, coded in terms of correct understandings or misunderstandings that are keyed to one or more of the five targeted principles. For example, when students make a drawing of gas in a container, coders have to rate the drawing as showing a particulate vs. continuous conception of gases (see also Benson et al., 1993).

6.2.2 Measurement Model

We propose that the Andersen/Rasch Measurement Model (Andersen, 1995) could likewise be applied to the TAP. Each student would be modeled as being in several belief states,
each with a certain propensity; each problem would also be seen as provoking certain belief states, each with a certain propensity. As with the FCI, the original statistical model for the TAP was a simple additive model.

6.3 Task Model

The Task Model for TAP questions has two variables in common with the FCI Task Model: whether an illustration is used and the nature of the substances involved in the problem. Many of the task model variables are similar to those that would be used in a conventional calculation problem (e.g., pressure, volume, and temperature in a closed system); however, TAP problems never give numerical quantities (similar to FCI problems). As in the FCI, students make predictions about or explanations of gas behavior. Interestingly, Novick and Nussbaum (1981) found that problem number 8 (shown in Figure 5) was often answered incorrectly, even by students who demonstrated a correct conception of uniform distribution of gases in a rigid container. An additional task model variable is therefore the rigidity of the container (e.g., a rigid glass container vs. a flexible balloon). TAP problems also vary the response format (e.g., multiple choice vs. drawing). As with the FCI and mechanics principles, each TAP problem links to only one of the five scientific gas principles.
7.0 A General Design Pattern

We see the three FCI templates (H, G, and H & G) and the TAP as instantiations of a general type of scientific reasoning that Stewart refers to as “Model Using” (Stewart & Hafner, 1994; Stewart, Hafner, Johnson, & Finkel, 1992; see Appendix E). In Model Using, students reason through a given model about a scientific problem. These problems do not present anomalous data that challenge students’ existing models; rather, they provide practice on applying targeted scientific models. In Model Using, students may consolidate their understanding of a model and how to apply it to problems.

Model Using is itself an instantiation of a general class of scientific reasoning involving models, called “model-based reasoning.” In addition to Model Using, Stewart and colleagues have identified Model Elaboration and Model Revising (Stewart & Hafner, 1994). Model Elaboration and Model Revising differ from Model Using primarily in that these types of reasoning involve anomalous data that do not fit students’ existing models. Model Using, on the other hand, involves practice in applying a model to a situation that does not contradict the targeted model.

We see Model Using as a flexible design pattern that can encompass student reasoning with both correct (i.e., scientific) and partially correct (e.g., DiSessa’s [1993] “p-prims”) models. Model Using requires knowledge of a model, but especially conditional knowledge about when the model can be applied to a situation (see also Larkin & Simon, 1987). In addition, students need domain-specific or general knowledge to use models. Students can use knowledge of the model and their domain-specific knowledge together to make predictions about and explanations of phenomena that the model applies to. For example, given a model of gravity as a force that pulls objects downward and general knowledge about elevators (e.g., that they are heavy, that they are attached to cables that are moved by pulleys), students can reason through their gravity model to make predictions about or explanations of the motion of an elevator (e.g., they could predict that if the elevator cable were cut, the elevator would fall because of gravity).

The FCI and TAP templates, in addition to using Model Using, also depend on a preexisting body of research that has identified student misconceptions for a particular domain. These are required in order to construct distractors that will appeal to students who do not hold scientific conceptions. For example, the TAP distractors were chosen on the basis of Novick and Nussbaum’s (1978) interview study; the FCI distractors were built from a large body of prior studies by researchers such as Clement (1983), Frederiksen and White (1988), Hunt and Minstrell (1994), McCloskey (1983), and others. Templates could specifically point to misconceptions that could be used to construct distractors.
8.0 Final Comments

The ultimate goal of the design pattern and task templates is that researchers, teachers, and test developers can have worked-out samples of assessments that are theory based, psychometrically sound, and less burdensome than a “reinvent the wheel” approach, and that make optimal use of technology. For example, the Model Using design pattern and the four associated templates developed here are based on research on students’ misconceptions in science—one application of the novice-expert paradigm in cognitive psychology. These filled-in PADI task design structures are able to take a significant part of the assessment development burden off of task designers, whether they choose to use the templates as is or adapt them.

Tasks developed from these templates, such as the illustrations above or new ones built around the same design structures, are grounded in a branch of cognitive research in science learning. They could be developed either for informal use in the classroom or for larger-scale and more formal use with psychometric models. In assessment systems that follow this latter route, the templates’ theory-based Student Models and aligned Evidence Models are psychometrically sound. They take advantage of sophisticated statistical models (e.g., Rasch models, Bayes nets) that need not be developed afresh and the details of which can be made invisible to the task designer and end user. Finally, because PADI templates can be expressed in transportable form (specifically, they can be represented in XML format), they are simple to share and access via the Web and to use in systems that take advantage of automated student record keeping, adaptivity, and other technological features.
References


PADI Template for Force Problems—Gravity Only
The PADI template for “Force Problems—Gravity Only” is reproduced below. The first illustration shows the top layer of the template. It provides an overview of all of the objects in the hierarchically organized schema (see Figure 1 of the text). The seven following illustrations show lower layers in the template that are linked to the top layer. They specify, in turn, information about the Student Model, the Measurement Model, and the Task Model. The Student Model information is briefly described in the “Student Model Summary” section in the top layer, and detailed in the subsequent linked section that gives fuller details about the Student Model. The summary explains the multivariate Student Model proposed above, while the linked Student Model section defines in detail the model identification and the covariance matrix.

Measurement Model information is briefly described in the top layer’s “Evaluation Procedures Summary” (evidence rules) and “Measurement Model Summary” (statistical model) sections. In more detailed sections of the template, the Evaluation Procedures explain the mapping of the multiple-choice items to Newtonian, Galilean, and nonscientific conception categories. The Measurement Model explains the Andersen/Rasch model.

Task Model information is briefly described in the top layer’s “Activities,” “Activity Sequencing,” “Template-level Task Model Variables,” and “Task Model Variable Settings.” The primary activity for this template is “Make explanations and predictions from a physical situation,” which is linked to the template. The Task Model Variables are linked to this particular template, with template-specific comments (e.g., the “Direction of Motion” Task Model Variable in this template is always vertical).

In addition, the template includes links to the Model Using design pattern, relevant research, Web resources (e.g., the Web site for the FCI), and other print resources, such as the Bao and Redish (2001) paper.

Figure A-1: PADI Template “Force Problems—Gravity Only”
### Evaluation Procedures Summary

Correct answer always corresponds to the notion of reasoning from a Newtonian conception. In principle, could have open-ended answers, which would be mapped into model categories—e.g., [Minstrell, 1990](#).

### Work Product Summary

Multiple choice approaches have been constructed so that each one a priori maps to exactly one of the model classifications. Correct answers apply Newtonian principles correctly to the problem. Incorrect answers reflect non-Newtonian principles; specifically, misconceptions identified in [prior literature](#), which might be Galilean, Aristotelian, or nonscientific. The list of Newtonian concepts tested in the FCI is listed in [Harshbarger, Wells & Swaffhauser (1992), Table I](#). The list of common misconceptions used to construct distractors is listed in Table III of the same paper.

### Task Model Variable Summary

Task model variables include description of motion (e.g., down, or no motion); time elapsed; mass of object; forces other than gravity.

- **Equation**: \( \text{downward force} = g = 32 \text{ ft/sec}^2 \), friction always = 0, and Motion is always in a vertical plane in this subset of problems.
- **Ex. FCI Q4**: direction of motion = down time elapsed = \( t > 0 \), \( m \) mass of object = same for both objects
- **Ex. FCI Q17**: direction of motion = up forces other than gravity = force of cable opposite to force of gravity, irrelevant to this Q; time elapsed, mass of object

### Template-level Task Model Variables

- **Direction of motion**: The path taken by an object in motion. For example, an FCI-type problem, is a force applied instantaneously or over a period of time.
- **Friction**: Amount of friction between two objects such as a puck and ice, a rope and a pulley, a ball and air.
- **Identity of Horizontal Force**: The identity of the force(s) which act(s) on an object(s) for an instant or over a period of time.
- **Illustration**: Whether or not an illustration (e.g., line drawing, photograph, video clip) is part of the task.
- **Mass**: Mass of object(s) on which force is applied. Objects on which forces act. May vary in size, mass, animate/inanimate, etc.
- **Speed**: The speed of an object at the moment when a force is applied to it.
- **Time of interest**: Past, present, or future, e.g., Question asks about motion or speed while a force is applied vs. at

### Task Model Variable Settings

**[Data]**

### Materials and Presentation Requirements

**[Data]**

**[Data]**

### Materials and Presentation Settings

**[Data]**

### Activities Summary

Sets of questions relating to a single situation may be presented or answered in any order. Ordinarily all questions on the FCI are administered at one sitting (it is commonly used as a pretest-posttest measure in introductory physics courses).

### Activities

Make explanations and predictions from a physical situation. Given a physical situation with some underlying regularities (e.g., a Newtonian law or gas law), etc.

### Tools for Examiner

[pen or pencil]

### Exemplars

Force Concept Inventory (FCI), a multiple-choice measure of students’ conceptual knowledge of forces, including Newton’s 1st-3rd law.

### Educational Standards

### Design Patterns

Model Using. This design pattern generates tasks where the student has to reason through a given model using data.

### I am a kind of

### These are kinds of me

### These are parts of me

### Online resources

FCI available at [http://modeling.asu.edu](http://modeling.asu.edu)

### References


### I am a part of

---

### Figure A-1: PADI Template “Force Problems—Gravity Only” (continued)
Appendix A 27

Figure A-2: PADI Task Model Variable “Direction of Motion”

<table>
<thead>
<tr>
<th>Title:</th>
<th>Direction of motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>The path taken by an object in motion. E.g., Answer choices refer to the path of motion of an object (perpendicular vs. acute angle).</td>
</tr>
<tr>
<td>TMV Type</td>
<td>Free-form text entry</td>
</tr>
<tr>
<td>TMV Category (possible value)</td>
<td>Continuing in a straight line</td>
</tr>
<tr>
<td></td>
<td>Down</td>
</tr>
<tr>
<td></td>
<td>No motion</td>
</tr>
<tr>
<td></td>
<td>Perpendicular turn</td>
</tr>
<tr>
<td></td>
<td>Turns at an angle</td>
</tr>
<tr>
<td></td>
<td>Up</td>
</tr>
<tr>
<td>I am a kind of</td>
<td></td>
</tr>
<tr>
<td>These are kinds of me</td>
<td></td>
</tr>
<tr>
<td>Online resources</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
<tr>
<td>I am a part of</td>
<td>Force Problems—Gravity Only (Template)</td>
</tr>
<tr>
<td></td>
<td>Force Problems—Horizontal Only (Template)</td>
</tr>
<tr>
<td></td>
<td>Force Problems—Horizontal and Gravity (Template)</td>
</tr>
</tbody>
</table>

Figure A-3: PADI Task Model Variable “Duration”

<table>
<thead>
<tr>
<th>Title:</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>E.g., in FCI-type problem, is a force applied instantaneously or over a period of time. E.g., for FCI-type problem, units of measurement and measurement instrument may be issues.</td>
</tr>
<tr>
<td>TMV Type</td>
<td>Free-form text entry</td>
</tr>
<tr>
<td>TMV Category (possible value)</td>
<td>For 5 seconds</td>
</tr>
<tr>
<td></td>
<td>For a time</td>
</tr>
<tr>
<td></td>
<td>Instantaneously</td>
</tr>
<tr>
<td>I am a kind of</td>
<td></td>
</tr>
<tr>
<td>These are kinds of me</td>
<td></td>
</tr>
<tr>
<td>Online resources</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
<tr>
<td>I am a part of</td>
<td>Force Problems—Gravity Only (Template)</td>
</tr>
<tr>
<td></td>
<td>Force Problems—Horizontal Only (Template)</td>
</tr>
</tbody>
</table>
**Figure A-4: PADI Activity “Make Explanations and Predictions from a Physical Situation”**

**Title:** Make explanations and predictions from a physical situation

**Summary:** Given a physical situation with some underlying regularities (e.g., a Newtonian or gas law), students explain what is happening in the situation or make a prediction about what will happen.

**Measurement Models:** Anderson-Flash measurement model (flash).

**Evaluation Procedures:** Map multiple-choice answers onto conceptual model (e.g., PCI) answers mapped onto Newtonian, Aristotelian, and nonconventional models.

**Work Products:** Choice from ordered choices. The student is to make a selection from among presented choices right/wrong multiple choice items.

**Materials and Presentation:**
- **Problem 1:** Text. Two objects start at the same position, and at different positions, and travel in the same direction.
- **Problem 2:** Text. Two objects start at different positions, and at same position, and travel in the same direction.
- **Problem answers choices:** For problems that present answer choices either text or illustrations.
- **Problem illustration:** For problems that include an illustration to be used in solving the problem.
- **Problem text:** For problems that include text (e.g., a question, scenario, case study, etc.).

**Presentation Logic:**

**Task Model Variables:**

**Design Patterns:** Model elaboration. This design pattern concerns working with mappings and extensions of given scientific models.

**Model Using:** This design pattern generates tasks where the student has to reason through a given model using data.

**Online resources:**

**References:** Refer to Design Pattern (e.g., Model Using).

**I am a part of:**
- Force Problems—Gravity Only (Template)
- Force Problems—Horizontal Only (Template)
- Force Problems—Horizontal and Gravity (Template)
- Gas Law Problems (Template)

---

**Figure A-5: PADI Student Model “Direction of Motion”**

**Title:** FCISH student model

**Summary:** Properties to respond to tasks in a certain class (e.g., PCI and FRMCE items) in terms of conception models. The conceptions are Newtonian, Galilean, and nonconventional. The model is identified by implicitly fixing all examiners at zero, so there are only two variables, namely for the Galilean and nonconventional properties.

**Distribution Summary:** Binormal distribution for the two SM variables that are explicit.

**Distribution Type:** Multivariate normal

**Covariance Matrix:**

```
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>
```

**Means Matrix:**

```
<table>
<thead>
<tr>
<th></th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
```

**Student Model Variables:** Galilean property, nonconventional property.

**I am a kind of:**

**These are kinds of me:**

**These are parts of me:**

**Online resources:**

**References:**

**I am a part of:**
- Force Problems—Gravity Only (Template)
- Force Problems—Horizontal Only (Template)
Figure A-6: PADI Evaluation Procedure “Map Multiple-Choice Answers onto Conceptual Model”

Map multiple-choice answers onto conceptual model | Evaluation Procedure (rubric) 290

Title: Map multiple-choice answers onto conceptual model

Summary

Evaluation Phases

Online resources

References

I am a part of

Make explanations and predictions from a physical situation. (Activity)

Figure A-7: PADI Evaluation Phase “Model Using Multiple Choice”

Model using multiple choice | Evaluation Phase 327

Title: Model using multiple choice

Summary

Preceding Evaluation Phase

Work Products

Input Observable Variables

Task Model Variables

Output Observable Variables

Evaluation Action Data

Evaluation Action

Online resources

References

I am a part of

Map multiple-choice answers onto conceptual model. (Evaluation Procedure (rubric))
### Figure A-8: PADI Measurement Model “Andersen/Rasch Measurement Model Fragment”

**Andersen/Rasch measurement model fragment | Measurement Model 289**

<table>
<thead>
<tr>
<th>Type of Measurement Model</th>
<th>Multinomial generalization of Rasch model.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>measurement model fragment for the Andersen/Rasch model, which predicts 3 categories: Newtonian, Galilean, and &quot;nonscientific&quot;. In principle each student and each item would have three parameters also, for their propensities for each category. Under the MRCMLM, however, identification is achieved by fixing every student's first parameter at zero and every item's first parameter at zero. So we have 50 and every item with two parameters—in this case, for the Galilean &amp; nonscientific categories.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observable Variable</th>
<th>Conception category, which conceptual category the examinee’s response is in (e.g., Newtonian vs. Galilean for FCD)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Student Model Variables</th>
<th>Galilean propensity, nonscientific propensity.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Scoring Matrix</th>
<th>[View]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design Matrix</td>
<td>[View]</td>
</tr>
<tr>
<td>Calibration Parameters</td>
<td>[View]</td>
</tr>
</tbody>
</table>

| Online resources           |                                                                                                             |
|----------------------------|                                                                                                             |
| References                 |                                                                                                             |
| I am a part of             | Make explanations and predictions from a physical situation. (Activity)                                       |

---

**Appendix A**
APPENDIX B

PADI Template for Force Problems—Horizontal Only
The PADI template for “Force Problems—Horizontal Only” has links similar to those of the template for “Force Problems—Gravity Only,” but has its own Template-level Task Model Variables. Only the top layer of this template is shown.

**Figure B-1: PADI Template “Force Problems—Horizontal Only”**

<table>
<thead>
<tr>
<th>Design Patterns</th>
<th>Template Variables</th>
<th>Student Model Variables</th>
<th>Task Specifications</th>
</tr>
</thead>
</table>

**Summary**

This template is for multiple-choice conceptual Newtonian force problems that do not involve gravity or circular motion.

This template can produce a measure like a subset of Newtonian force problems from the Force Concept Inventory (FCI).

**Type**

| [ ] |

Student Model Summary

Student is seen as being in several belief states simultaneously (e.g., Newtonian, Galilean, and nonconservative), each with a certain propensity.

In some ways analogous to the Siegler (1976) balance beam study.

Student Models

FCI is a student model. Propensities to respond to tasks in a certain class (e.g., FCI and FMCE items) in terms of concepts...

Measurement Model Summary

Measurement model is multivariate, a la Anderson/Rasch model. Original measurement model was univariate.

Measurement model is an Andersen/Rasch model (Andersen, 1960) being tested by Ken Huang at UMCP. It can be written as a special case of the PROBLOG (Adams, Wilson, & Wang, 1997). Previous models include Bao and Redish (2001) simple summary scores.

Evaluation Procedures Summary

Answer key—only correct answer per question.

Correct answer always corresponds to the result of reasoning from a Newtonian conception. In principle, could have open-ended answers, which would be mapped into model categories (e.g., Menzler, 1989).

Work Product Summary

Answers to multiple-choice problems are main work product.

Multiple-choice responses have been constructed so that each one a priori maps to exactly one of the model classifications. Incorrect answers reflect non-Newtonian principles. Specifically, misconceptions identified in prior literature, which might be Galilean, Aristotelian, or conservative. The list of Newtonian concepts tested in the FCI is listed in Hestenes, Wells & Swackhamer (1992), Table I. The list of common misconceptions used to construct distractors is listed in Table I of the same paper.

Task Model Variable Summary

Task model variables include identity, direction, amount, and duration of force applied; mass and identity of object; speed and direction of object; time period of interest (before, during, or after force is applied).

E.g., FCI Q05 amount of force = “constant horizontal force” mass = “large boy” speed after force is applied = “no” Friction always = 0 and motion is always in a horizontal plane in the subset of problems, therefore gravity is never a force.

Other variables are implicitly or explicitly set constant (e.g., direction not specified, duration not specified, speed of object before force = 0, time period of interest = while the force is applied, path of motion must be a straight line).

(continued)
Figure B-1: PADI Template “Force Problems—Horizontal Only” (continued)

<table>
<thead>
<tr>
<th>Task Model Variable Settings</th>
<th>Paper-and-pencil measure.</th>
<th>Could easily be adapted for computer administration.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template-level Materials and Presentation Settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Materials and Presentation Settings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities Summary</td>
<td>Sets of questions relating to a single situation may be presented or answered in any order. Ordinarily, all questions on the PCI are administered at one sitting (this is commonly used as a pretest-posttest measure in university introductory physics courses).</td>
<td></td>
</tr>
<tr>
<td>Activities</td>
<td>Make explanations and predictions from a physical situation. Given a physical situation with some underlying regularities (e.g., a law of motion or energy conservation), state the regularities and explain why they are expected to hold true.</td>
<td></td>
</tr>
<tr>
<td>Tools for Examinee</td>
<td>Pen and pencil.</td>
<td></td>
</tr>
<tr>
<td>Exemplars</td>
<td>Force Concept Inventory (FCI). A multiple-choice measure of students’ conceptual knowledge of forces, including Newton’s 1st-3rd law.</td>
<td></td>
</tr>
<tr>
<td>Educational Standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design Patterns</td>
<td>Model Using. This design pattern generates tasks where the student has to reason through a given model using data.</td>
<td></td>
</tr>
<tr>
<td>I am a kind of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>These are kinds of me</td>
<td></td>
<td></td>
</tr>
<tr>
<td>These are parts of me</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online resources</td>
<td>FCI available at [<a href="http://modeling.tea.ut">http://modeling.tea.ut</a>].</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

PADI Template for Force Problems—Horizontal and Gravity
Appendix C

The PADI template for “Force Problems—Horizontal and Gravity” has links similar to those of the FCI templates above, but its own Template-level Task Model Variables. Only the top layer of this template is shown.

Figure C-1: PADI Template “Force Problems—Horizontal and Gravity”

<table>
<thead>
<tr>
<th>TITLE:</th>
<th>Force Problems—Horizontal and Gravity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>This template is for multiple-choice conceptual Newtonian force problems that simultaneously involve horizontal motion and the downward pull of gravity, but NO circular motion. This template can produce a measure like a subset of Newtonian force problems from the Force Concept Inventory (FCI), Force Problems—Horizontal Only, and Force Problems—Gravity Only are special cases of this template.</td>
</tr>
<tr>
<td>Type</td>
<td>[ Task ]</td>
</tr>
<tr>
<td>Student Model Summary</td>
<td>Student is seen as being in several belief states simultaneously (e.g., Newtonian, Galilean, and non-scientific), each with a certain propensity.</td>
</tr>
<tr>
<td>Measurement Model Summary</td>
<td>Measurement model is multivariate, a la Anderson/Rasch model. Original measurement model was univariate.</td>
</tr>
<tr>
<td>Evaluation Procedures Summary</td>
<td>Answer key—one correct answer per question.</td>
</tr>
<tr>
<td>Work Product Summary</td>
<td>Multiple choice responses have been constructed so that each one is a prior map to exactly one of the model classifications. Correct answers apply Newtonian principles correctly to the problem. Incorrect answers reflect non-Newtonian principles; specifically, misconceptions identified in prior literature, which might be Galilean, Aristotelian, or non-scientific. The list of Newtonian concepts tested in the FCI is listed in Hestenes, Wells &amp; Swackhammer (1992), Table I. The list of common misconceptions used to construct distractors is listed in Table II of the same paper.</td>
</tr>
<tr>
<td>Task Model Variable Summary</td>
<td>Task model variables include: identity, direction, and amount of force applied; released from moving or non-moving object; mass; size; and identity of object; horizontal distance traveled; path of motion.</td>
</tr>
</tbody>
</table>

(continued)
### Figure C-1: PADI Template “Force Problems—Horizontal and Gravity” (continued)

<table>
<thead>
<tr>
<th>Template-level Task Model Variables</th>
<th>Amount of force applied: Amount of force which one object exerts on another (moving or stationary) object.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direction of force applied: The direction in which a force is applied to an object</td>
</tr>
<tr>
<td></td>
<td>Direction of motion: The path taken by an object in motion</td>
</tr>
<tr>
<td></td>
<td>Forbids: Amount of friction between two objects such as a puck and ice, a rope and a pulley, a ball and air.</td>
</tr>
<tr>
<td></td>
<td>Identity of horizontal force: The identity/ies of a force(s) which act(s) on an object(s) for an instant or over a period of time.</td>
</tr>
<tr>
<td></td>
<td>Illustration: Whether or not an illustration (e.g., line drawing, photograph, video clip) is part of the task.</td>
</tr>
<tr>
<td></td>
<td>Mass: e.g., Mass of object to which force is applied.</td>
</tr>
<tr>
<td></td>
<td>Objects: Objects on which forces act. They may vary in size, mass, animate/inanimate, etc.</td>
</tr>
<tr>
<td></td>
<td>Shape of path of motion: The shape of the path taken by an object in motion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task Model Variable Settings</th>
<th>[Size]</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Materials and Presentation Requirements</th>
<th>Paper-and-pencil measure.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Template-level Materials and Presentation</th>
<th>[Max]</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Materials and Presentation Settings</th>
<th>[Max]</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Activities Summary</th>
<th>Make explanations and predictions from a physical situation. Given a physical situation with some underlying regularities (e.g., a Newton’s law or gas law), etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tools for Examinee</td>
<td>Pencils or pencil.</td>
</tr>
<tr>
<td>Exemplars</td>
<td>Force Concept Inventory (FCI): A multiple-choice measure of students’ conceptual knowledge of forces, including Newton's 3rd law.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Educational Standards</th>
<th>[Model Using]: This design pattern generates tasks where the student has to reason through a given model using data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design Patterns</td>
<td>I am a kind of:</td>
</tr>
<tr>
<td>I am a part of</td>
<td>Online resources: FCI available at: <a href="http://modeling.is.ryerson.ca">http://modeling.is.ryerson.ca</a>.</td>
</tr>
</tbody>
</table>

|----------------------|----------------------------------------------------------------------------------------------------------|
APPENDIX D

PADI Template for Gas Law Problems
Appendix D

The PADI template for “Gas Law Problems” reproduced below has its own specific information about the Student Model, the Measurement Model (e.g., mapping student-constructed responses to predetermined misconception categories), and Task Model (e.g., for constructing both multiple-choice and constructed-response items), as well as references to the TAP measure and other print resources. The same Activity used to structure the FCI force tasks (displayed in Appendix A) is reused in this template. Only the top layer of the template is shown.

Figure D-1: PADI Template “Gas Law Problems”

<table>
<thead>
<tr>
<th>Title</th>
<th>Gas Law Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>This template is for conceptual Gas Law problems in chemistry. This template can produce a measure like the Test About Particles in a Gas (TAP).</td>
</tr>
<tr>
<td>Type</td>
<td>(Use)</td>
</tr>
<tr>
<td>Student Model Summary</td>
<td>Student is seen as being in one of several states (e.g., continuous vs. particulate model), homogeneous distribution vs. “dumping”). Could apply a multi-state model, as with Force Problems.</td>
</tr>
<tr>
<td>Student Models</td>
<td></td>
</tr>
<tr>
<td>Measurement Model Summary</td>
<td>Measurement model is multivariate, a la Anderson/Rasch model. Original measurement model was univariate.</td>
</tr>
<tr>
<td>Evaluation Procedures Summary</td>
<td>Answer key—one correct answer per question. Correct answer always corresponds to the result of reasoning from Novick &amp; Rasch’s five principles (1981). Novick &amp; Rasch’s model used both multiple-choice and open-ended answers which are mapped onto their model categories.</td>
</tr>
<tr>
<td>Work Product Summary</td>
<td>Answers to multiple-choice and open-ended questions are main work product. Multiple choice responses have been constructed so that each one a priori maps to exactly one of the model classifications. Open-ended questions have one coding category that maps to exactly one of the five principles. Incorrect answers reflect misconceptions identified in prior literature. Novick &amp; Rasch sometimes identify only one misconception for a question, and sometimes identify more than one. Other researchers have also suggested misconceptions.</td>
</tr>
<tr>
<td>Task Model Variable Summary</td>
<td>Tasks may have: illustrations, varying pressure, volume, type of container, temperature, element/gas. These elements of problems are similar to quantitative gas law problems, but they do not involve computation.</td>
</tr>
<tr>
<td>Template-level Task Model Variables</td>
<td>Context area. Specific domain content under consideration. Illustrations. Whether or not an illustration (e.g., line drawing, photograph, video clip) is part of the task. Open/closed system. Pressure. Pressure inside a container for e.g., a gas law problem. Elasticity of container. Whether the container is rigid (e.g., glass) or flexible (e.g., balloon). Substance. The substance in e.g., a gas law problem. Temperature. Temperature for e.g., a gas law problem. Volume. e.g., Volume of container for a gas law problem.</td>
</tr>
<tr>
<td>Task Model Variable Settings</td>
<td></td>
</tr>
<tr>
<td>Materials and Presentation Requirements</td>
<td>Paper-and-pencil measure. Could easily be adapted for computer administration.</td>
</tr>
</tbody>
</table>

(continued)
Figure D-1: PADI Template “Gas Law Problems” (continued)

<table>
<thead>
<tr>
<th>Template Level</th>
<th>Materials and Presentation</th>
<th>Activities Summary</th>
<th>Activities</th>
<th>Tools for Examinee</th>
<th>Exemplars</th>
<th>Educational Standards</th>
<th>Design Patterns</th>
<th>I am a kind of</th>
<th>These are kinds of me</th>
<th>These are parts of me</th>
<th>Online resources</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Questions may be presented or answered in any order. Ordinarily all questions on the TAI are administered at one sitting.</td>
<td>Make explanations and predictions from a physical situation. Given a physical situation with some underlying regularities (e.g., a Newtonian law or gas law, etc.)</td>
<td>Pen or pencil.</td>
<td>Test About Particles In a Gas (TAP). This is a paper-and-pencil measure of students' conceptual understanding of gases. It includes both...</td>
<td></td>
<td>Model Using. This design pattern generates tasks where the student has to reason through a given model using data...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

PADI Design Pattern for Model Using
Appendix E

The PADI “Model Using” design pattern is reproduced below. It includes a Summary, KSAs (knowledge, skills, and abilities), Potential Observations, Potential Work Products, and Characteristic and Variable Features of the design pattern. The KSAs include both Focal KSAs (e.g., the importance of conditional knowledge) and Additional KSAs (e.g., general knowledge). Potential Observations and Work Products are very high-level statements that are operationalized in the templates as Student Model and Task Model Variables (e.g., Work Products in the Activity). Characteristic and Variable Features are likewise abstract statements that are operationalized in the templates as Task Model Variables.

Figure E-1: PADI Design Pattern “Model Using”

<table>
<thead>
<tr>
<th>Model Using</th>
<th>Design Pattern 274</th>
</tr>
</thead>
</table>

**Title:**
Model Using

**Summary:**
The design pattern generates tasks where the student has to reason through a given model using data that do not contradict the model.

**Focal Knowledge, Skills, and Abilities:**
- Being able to reason through the concepts and relationships of a given model.

**Rationale:**
The activities mediated by this pattern have students reasoning through a model, to make explanations, predictions, and inferences such as filling in missing information.

**Additional Knowledge, Skills, and Abilities:**
- Familiarity with task type (e.g., materials, protocols, expectations).
- Subject-area knowledge (declarative and procedural).

**Potential Observations:**
- Given a model and a situation, making explanations, predictions, or retrodictions reasoned through the model.
- Making and explaining predictions through a model.

**Potential Work Products:**
- Correspondence mapping between elements or relationships of model and real-world situation.
- Correspondence mapping between elements or relationships of overlapping models.
- Hypothesis (constructed/selected).
- Predictions (constructed/selected).
- Written/Oral Explanation of reasoning.

**Potential Rubrics:**
- Real-world situation and one or more models appropriate to the situation, for which details of correspondence need to be fleshed out. Focus is on reasoning about the situation, through the schema and relationships embedded in the model. Addresses correspondence between situation and models, and models with one another. Patterns in the data do not contradict the model.

**Characteristic Features:**
- Is problem context familiar?
- Must experimental work or supporting research be carried out in order to ground the elaboration?
- Single model to reason through, vs. establishing correspondence among and reasoning through models at different levels or with different focus?

**Variable Features:**
- Scientific Reasoning: This design pattern concerns a scientific problem to solve or investigate. Do they effectively plan...
- Interpretation of Dynamic Graphs: Students are provided with a model in which they can manipulate model parameters and view the result.

(continued)
## Figure E-1: PADI “Model Using” (continued)

<table>
<thead>
<tr>
<th>Templates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exemplar tasks</td>
<td>Force Concept Inventory (FCI). A multiple-choice measure of students’ conceptual knowledge of forces, including Newton’s 1st-3rd laws.</td>
</tr>
<tr>
<td>Online resources</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>
  - NGSS standards.

I am a part of |  |
Sponsor
The National Science Foundation, Grant REC-0129331

Prime Grantee
SRI International. Center for Technology in Learning

Subgrantees
University of Maryland
University of California, Berkeley. Berkeley Evaluation & Assessment Research (BEAR) Center and The Full Option Science System (FOSS)
University of Michigan. BioKIDS