

Model analysis: Representing and assessing the dynamics of student learning

Lei Bao*

*Department of Physics, The Ohio State University, 174 West 18th Ave., Columbus, Ohio 43210, USA*Edward F. Redish[†]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*

(Received 15 May 2005; published 2 February 2006)

Decades of education research have shown that students can simultaneously possess alternate knowledge frameworks and that the development and use of such knowledge are context dependent. As a result of extensive qualitative research, standardized multiple-choice tests such as Force Concept Inventory and Force-Motion Concept Evaluation tests provide instructors tools to probe their students' conceptual knowledge of physics. However, many existing quantitative analysis methods often focus on a binary question of whether a student answers a question correctly or not. This greatly limits the capacity of using the standardized multiple-choice tests in assessing students' alternative knowledge. In addition, the context dependence issue, which suggests that a student may apply the correct knowledge in some situations and revert to use alternative types of knowledge in others, is often treated as random noise in current analyses. In this paper, we present a model analysis, which applies qualitative research to establish a quantitative representation framework. With this method, students' alternative knowledge and the probabilities for students to use such knowledge in a range of equivalent contexts can be quantitatively assessed. This provides a way to analyze research-based multiple choice questions, which can generate much richer information than what is available from score-based analysis.

DOI: [10.1103/PhysRevSTPER.2.010103](https://doi.org/10.1103/PhysRevSTPER.2.010103)

PACS number(s): 01.40.Fk

I. INTRODUCTION

One of the most important things educational researchers have learned over the past few decades is that it is essential for instructors to understand what knowledge students bring into the classroom and how they respond to instruction. Qualitative physics education research on a variety of topics has documented that students bring knowledge from their everyday experience and previous instruction to their introductory physics classes and that this knowledge affects how they interpret what they are taught.¹ Two important facts are critical in any attempt to probe student knowledge.

(i) Student knowledge (ideas, conceptions, interpretations, assumptions) relevant to physics may be only locally coherent. Different contexts can activate different (and contradictory) bits of knowledge.^{2,3}

(ii) On any particular topic, the range of alternative conceptions seen in a particular population tends to be fairly limited. Often, two or three specific ideas account for most observed student responses (though sometimes as many as a half a dozen are needed).⁴ (Another point that is often noted is that these alternative conceptions can be quite firmly held and difficult to transform. Since this paper is about measuring student conceptions and not changing them, that is less relevant here.)

These two ideas have been used by many researchers to create multiple-choice exams that use common alternative student conceptions revealed by qualitative research as “attractive distracters.”^{5,6} The impact of these exams can be both revealing and powerful. Faculty who are not aware of the prevalence and strength of student alternative conceptions fail to see the distracters as reasonable alternatives and may consider the exam as trivial. They can then be surprised

when many of their students choose these distracters, even after instruction.⁷

Careful analysis of the responses to these exams shows that for many populations the responses are not consistent. A student may answer one item correctly, but answer another item, one that an expert might see as equivalent to the first, incorrectly. The assumption that a student “either knows the topic or does not know it” appears to be false, especially for students in a transition state between novice and expert. The level of a student's confusion—how the knowledge the student activates depends on context—becomes extremely important in assessing the students' stage of development.

In small classes, this information can be obtained from careful one-on-one dialogs between student and teacher. In large classes, such as those typically offered in introductory science courses at colleges and universities, such dialogs are all but impossible. Instructors in these venues often resort to pre-post testing using research-based closed-ended diagnostic instruments.

But the results from these instruments tend to be used in a very limited way—through overall scores and average pre-post gains. This approach may miss much valuable information, especially if the instrument has been designed on the basis of strong qualitative research, contains subclusters of questions probing similar issues, and has distracters that represent alternative modes of student reasoning.

In this paper, we present a method of *model analysis* that allows an instructor to extract specific information from a well-designed assessment instrument (test) on the state of a class's knowledge. The method is especially valuable in cases where qualitative research has documented that students enter a class with a small number of strong naive conceptions that conflict with or encourage misinterpretations of

the scientific view. As students begin to learn scientific knowledge that appears to contradict their intuitive conceptions, they may demonstrate confusions, flipping from one approach to another in an inconsistent fashion.

The model analysis method works to assess this level of confusion in a class as follows.

(i) Through systematic research and detailed student interviews, common student models are identified and validated so that these models are reliable for a population of students with a similar background.

(ii) This knowledge is then used in the design of a multiple-choice instrument. The distracters are designed to activate the common student models, and the effectiveness of the questions is validated through research.

(iii) One then characterizes a student's responses with a vector in a linear "model space" representing the (square roots of the) probabilities that the student will apply the different common models.

(iv) The individual student model states are used to create a "density matrix," which is then summed over the class. The off-diagonal elements of this matrix retain information about the confusions (probabilities of using different models) of individual students.

(v) The eigenvalues and eigenvectors of the class density matrix give information not only how many students got correct answers, but about the level of confusion in the state of the class's knowledge.

Our analysis method is mathematically straightforward and can be easily carried out on a standard spreadsheet. The result is a more detailed picture of the effectiveness of instruction in a class than is available with analyses of results that do not consider the implications of the incorrect responses chosen by the students.

Although the desire to "understand what our students know" is an honorable one, we cannot make much progress until we both develop a good understanding of the characteristics of the system we are trying to influence (the student's knowledge structure) and have a language and theoretical frame with which to talk about it. Fortunately, much has been learned over the past few decades about how students think and learn and many theoretical models of human cognition have been developed and are beginning to show some evidence of coalescing into a single coherent model.^{8,9} In this model, knowledge corresponds to the activation of a network of neurons. These networks can be linked so that activation of one bit of knowledge is coordinated with the activation of other bits. This model treats knowledge in a highly dynamic fashion and supports the idea that an individual may have alternative contradictory models that can be activated by different contexts without their being particularly aware of the contradiction. We discuss this theoretical framework briefly in Sec. II.

Despite the progress in cognitive science, most educational researchers analyzing real-world classrooms make little use of this knowledge. Many of the mathematical tools commonly used to extract information from educational observations rely on statistical methods that (often tacitly) assume that quantitative probes of student thinking measure a system in a unique true state. We believe that this model of assessing student learning is not the most appropriate one for

analyzing a student's progress through goal-oriented instruction and is inconsistent with current models of cognition. (Examples will be given in the body of the paper.) As a result, the analysis of quantitative educational data can draw incomplete or incorrect conclusions even from large samples.

In Secs. III and IV, we describe in detail model analysis, a method that represents the student's mental state as a vector in a "model space" spanned by a set of basis vectors, each representing a unique type of reasoning that has been identified through qualitative research. In Sec. V, we apply a model analysis to Force Concept Inventory results and show how we can get new insights into the state of students' knowledge. In Sec. VI, we compare our method to other more traditional approaches. Section VII gives our conclusions and suggestions as to how the approach can be used.

II. THEORETICAL FRAME: A MODEL OF COGNITION

The theoretical framework we use to describe various models of thinking and learning is based on a triangulation of three kinds of scientific research: phenomenological observations of normal behavior, especially in educational environments^{1,10} (often carried out by educational researchers); careful studies of responses to highly simplified experiments designed to get fundamental cognitive structures^{11,12} (mostly carried out by psychologists); and studies of the structure and functioning of the brain (mostly carried out by neuroscientists).^{9,13} In each of these areas of research, numerous models of cognition have been built. Although there is still much uncertainty about what the final model will look like, there is now a significant overlap. We particularly rely on elements that have significant support in all three areas.

This evolving model is described in some detail in other papers.¹⁴⁻¹⁶ For our work here, we need to specifically understand three elements of the model: the nature of the storage and access to elements of knowledge in memory, how this leads to context dependence, and how these features of learning can be represented and assessed.

A. Long-term memory

The critical issue for teaching and learning is long-term memory. We are interested not only in what our students know, but in their access to that knowledge: what contexts activate it for them. We use the term *knowledge element* to refer to something a student knows that seem irreducible to them (that is, containing no obvious component parts). It could be something that they believe or a simple procedure.

A few principles briefly describe some characteristics of long-term memory that can help us better understand the responses of students.

(i) *Long-term memory is associative and productive. Activating one knowledge element typically leads, with some probability, to the activation of other associated elements.*

(ii) *Activation and association of knowledge elements are context dependent. What is activated and subsequent activations depend on the context, both external and internal (other activated elements).*

These principles are supported by a wide variety of studies ranging from the ecological to the neurological.

An example from the physics education literature illustrates the implications of the context dependence of recall from long-term memory in problem solving. Steinberg and Sabella asked two equivalent questions on Newton's first law to students in engineering physics at the University of Maryland.¹⁷ In both questions, the students were asked to compare the forces acting on an object moving vertically at a constant velocity. One question was phrased in physics terms using a laboratory example ("A metal sphere is resting on a platform that is being lowered smoothly at a constant velocity..."). The other was phrased in common speech using everyday experience ("An elevator is being lifted by a cable..."). Here, both the wording of the questions and the ways of the motion of the elevator are important contextual features. In both problems, students were instructed to ignore friction and air resistance.

On the first problem, 90% of the students gave the correct answer that the normal force on the sphere is equal to the downward force due to gravity. On the second problem, only 54% chose the correct answer: the upward force on the elevator by the cables equals the downward force due to gravity. More than a third, 36%, chose the answer to this second problem reflecting a common incorrect model: the upward force on the elevator by the cables is greater than the downward force due to gravity.

A strong context dependence in student responses is very common, especially when students are just beginning to learn new material. Students are unsure of the conditions under which rules they have learned apply, and they use them either too broadly or too narrowly. (Understanding how these conditions apply is a crucial part of learning a new concept.) Students often treat problems that look equivalent to an expert quite differently.

B. Organization of long-term memory

A great deal has been learned about the organization of long-term memory and reasoning in a variety of contexts including the interpretation of text,¹⁸ the learning of grammar,¹⁹ approaching problem solving in mathematics,²⁰ and the interpretation of physical phenomena.²¹ Since our primary interest is science learning in general and the learning of physics in particular, we restrict our discussion here to thinking and learning about physical phenomena. Much of what has been learned under this rubric has analogs in other areas. We briefly discuss two commonly used theoretical models of student thinking in physics: the *knowledge-in-pieces model* of diSessa² and Minstrell³ and the *alternate-conceptions model* of Caramazza *et al.*²² and Vosniadou²³ and their collaborators.

DiSessa investigated people's *sense of physical mechanism*—that is, their understanding of why things work the way they do.^{2,21} What he found was that many students, even after instruction in physics, often come up with simple statements that describe the way they think things function in the real world. They often consider these statements to be "irreducible"—as the obvious or ultimate answer; that is,

they cannot give a "why" beyond it. "That's just the way things work," is a typical response. DiSessa refers to such statements as *phenomenological primitives* or *p-prims*. Minstrell observed that students' responses to questions about physical situations can frequently be classified in terms of reasonably simple and explicit statements about the physical world or the combination of such statements. He refers to such statements as *facets*.³

In the modular model of diSessa and Minstrell, students are assumed to have their knowledge "in pieces." Different bits of knowledge tend to be weakly connected. As a result, different contexts can easily cue different responses, although in this model it is possible that a particular "piece" can be robust and activated with a high probability in a variety of situations.

An alternative view of student thinking in physics is the one espoused explicitly by Caramazza *et al.* and Vosniadou.^{22,23} In this view, students possess a coherent and organized "alternative" or "naive" theory of a particular physical topic or situation. Despite describing student responses as theory like, Vosniadou cites cases in which students appear to be mixing elements of contradictory models.

In the theory described in Redish¹⁴ and Hammer *et al.*,¹⁵ these two theories can be seen as extreme assumptions about the nature of knowledge structures most likely to be found among naive students. The difference between the two theoretical models is largely in the expectation of whether one will observe responses that can be interpreted as consistent across many contexts or depending more sensitively on context. The question as to which model is correct becomes an empirical one. The answer as to which model should be preferred could depend on both the populations involved and the circumstances that one wants to consider as an appropriate range of contexts. In one study, a careful multifaceted observation of the behavior of preservice teachers learning topics in physics over long periods (many weeks) revealed shifts in student choices of reasoning from consistent (and wrong), through mixed, and back to consistent (this time, agreeing with the more scientific conceptions).²⁴ If this turns out to be general, assessing the state of the students' choice of reasoning patterns could have important instructional implications.

To be able to discuss the cognitive issues clearly and without prejudice towards one model or another, we use the general term *mental model*: a robust and coherent knowledge element or strongly associated set of knowledge elements. For example, in the contexts involving motion, students often believe that there is always a force in the direction of motion. This represents a robustly established association between motion and force and thus is characterized as a mental model. We use this term in a broad and inclusive sense. A mental model may be simple or complex, correct or incorrect, activated as a whole or generated spontaneously in response to a situation. Note that this term appears frequently in the cognitive and educational literature, often in undefined and inconsistent ways. Our use of the term is probably closest to that used by Norman.²⁵

The popular (and sometimes debated) term *misconception* can be viewed as reasoning involving mental models that have problematic elements for the student's creation of an expert view and that appear in a given population with sig-

nificant probabilities (though not necessarily consistently in a given student). We stress that our use of this term implies no assumption about the structure or cognitive mental creation of this response. In particular, we do not need to assume either that it is irreducible (has no component parts) or that it is stored and recalled rather than generated on the spot.

In assessing the state of students' knowledge, what one needs to determine is both the models the students possess and can use and the context dependence of their use of these models.

C. Context dependence and the state of the student from the point of view of an expert

The context dependence of the cognitive response may be considered in a variety of ways. From the point of view of the student, his or her mental system may feel perfectly consistent, despite appearing inconsistent to an expert. The student might use a mental model inappropriately because he or she has failed to attach appropriate conditions to its application,²⁶ the student might fail to associate a mental model with a circumstance in which it is appropriate, or the student may associate with different mental models in equivalent circumstances, cueing on irrelevant elements of the situation and not noticing that the circumstances are equivalent.

From the point of view of the cognitive researcher, it may be of great interest to consider the student as always being in a consistent mental state or as flipping from one mental state to another in response to a variety of cues. However, from the point of view of the educational researcher or of the instructor interested in goal-oriented instruction—that is, in acculturating students to understand particular community-developed viewpoints—we suggest that there is considerable value in analyzing the student thinking as projected against an expert view. The “expert” here needs to be both a subject expert and an expert in education research so as not to undervalue or misunderstand the view of the student. For example, in considering the motion of compact objects, a naive physics student might view objects in terms of a generic concept of “motion” with inappropriately entangled ideas of position, velocity, acceleration, and force. The mental models used by the student must be understood in terms of their own internal consistencies, not as “errors” when projected against the expert view.

Suppose we prepare a sequence of questions or situations in which an expert would use a single, coherent mental model. We refer to these as *a set of expert-equivalent questions*. Further, suppose that when presented with some questions from such a set, a particular student can use a variety of mental models instead of a single coherent expert model. Such a situation is extremely common in many learning situations and is well documented to occur frequently in introductory physics.²⁷ How each question in the set cues a student to choose a particular model (or a set of models) depends not only on the student's educational history, but even on the student's mental state at the particular instant the question is probed. Since both the educational history and the student's mental state are difficult to determine, we pro-

pose that the most appropriate way of treating this situation is probabilistically.

If a student always uses a particular mental model in a reasonably coherent way in response to a set of expert-equivalent questions, we say they are in a *pure model state*. If the student uses a mixture of distinct mental models in response to the set of questions, we say the student is in a *mixed model state*. We view the individual student who is in a mixed state as simultaneously occupying a number of distinct models with different probabilities in applying these models in expert equivalent contexts. The distribution pattern of the probabilities gives a representation for a student model state.

When the student's state is probed by the presentation of a particular question or scenario, the student will often respond by activating a single mental model. We view the student's mental state as having been *momentarily collapsed by the probe into the model state* selected.

The process by which this selection is made can be quite complex. In some cases, only a single model is activated. In others, multiple models are activated and an “executive process” is assumed to make a choice of one, suppressing other models. When such a choice is difficult to make, a student can get into an explicit state of confusion where several models appear to be equally plausible (but generating contradictory results) and the student cannot determine which one is more appropriate to use. Depending on the design of the probing instruments, such states may or may not be extracted. For example, multiple-choice single-response questions often force students to pick one answer and thus can only measure the existence of one of the models, while multiple-choice multiple-response questions can extract information about such mixing states. Although this topic has not been studied extensively (to our knowledge) in physics education, there is extensive research on the issue in cognitive and neuroscience.²⁸

Note that the probabilistic character of the student model state arises from the presentation of a large number of questions or scenarios, not from the probing of multiple students. We view the context dependence of mental model generation as a fundamental probabilistic character built into the individual. The probabilistic treatment is a way of treating many “hidden variables” in the problem that are both uncontrollable and possibly unmeasurable even in principle.

This approach, which will be developed mathematically below, provides an alternative assumption to the one traditionally made,²⁹ that a probe of the state of a student yields the “true value” plus some random error: $M = T + X$. Although it may be appropriate to consider a student mental state as having a “true value” on a very short time scale (a few seconds or less), this may not be appropriate when thinking about a student's knowledge state over a period of minutes, hours, or days. We propose that a more appropriate model for analyzing student thinking is to consider the distribution of a student's inconsistent results on a set of expert equivalent questions as a measure of a property of the student, not as “random error.” (Of course random errors do occur and must be taken into account through a statistical consideration of the effect of random fluctuations on the state probabilities introduced in this paper. Consideration of these effects is

beyond the scope of this paper and is discussed in Ref. 30.) In this paper, the probabilistic distribution is interpreted as fundamental and representing the characteristics of students.

As discussed earlier, mental models are productive structures that can be applied to a variety of different physical contexts to generate explanatory results. Mental models can be either complex or simple. For this work, to clarify the nature of our model and method, we have chosen to restrict our considerations to simple models—essentially single facets. We do not intend to imply that all student reasoning is describable by such a simple situation. There are, however, numerous examples of such situations, and we intend to demonstrate the value of our approach by applying it to this simple, highly restrictive situation.

The mixed use of models or competing concepts appears to be a typical and important stage in student learning of physics.^{6,27,30,31} To study the dynamical process of students' applying their models, we first define two important concepts: *common models* and *student model states*.

D. Common models

When the learning of a particular physics concept is explored through systematic qualitative research (these researches should always involve detailed individual student interviews and the results should also be verifiable by other researchers), researchers are often able to identify a small, finite set of commonly recognized models.³² These models often consist of one correct expert model and a few incorrect or partially correct student models. Note that different populations of students may have different sets of models that are activated by the presentation of a new situation or problem. When presented with novel situations, students can activate a previously well-formed model or, when no existing models are appropriate, they can also create a model on the spot using a mapping of a reasoning primitive or by association to salient (but possibly irrelevant) features in the problem's presentation. The identified common student models can be formed in both ways. Although the actual process is not significant in the research of this paper, the specific structure of the models involved may have important implications for the design of instruction.

E. Student model state

When a student is presented with a set of questions related to a single physics concept (a set of expert equivalent questions), two situations commonly occur.

(i) The student consistently uses one of the common models to answer all questions.

(ii) The student uses different common models and is inconsistent in using them; i.e., the student can use one of the common models on some expert-equivalent questions and a different common model on other questions.

The different situations of the student's use of models are described as student *model states*. The first case corresponds to a pure model state and the second case to a mixed model state.

When analyzing the use of common models, it is necessary to allow an additional dimension to include other less

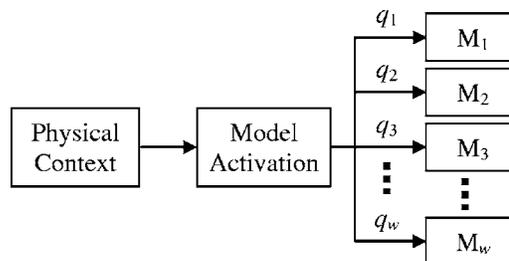


FIG. 1. Using a set of questions designed for a particular physics concept, we can measure the probability for a single student to use different physical models in solving these problems. In the figure, M_1, \dots, M_w represent the different physical models (there are a total of w physical models including a null model) and q_1, \dots, q_w represent the probabilities for a student being triggered into activating the corresponding models.

common and/or irrelevant ideas that student might come up with. To collect this set of responses we identify a *null* model—one not describable by a well-understood common model. With the null model included, the set of models becomes a complete set; i.e., any student response can be categorized. (Of course, in addition to collecting random and incoherent student responses, coherent models that have not yet been understood as coherent by researchers may well be classified initially as “null.” When a significant fraction of student responses on a particular question winds up being classified as null, it is possible that a better understanding of the range of student responses needs to be developed through qualitative research. In this way, we also have a quantitative tool to alert the needs of further qualitative research.) Specific examples of common models and student model states will be discussed in later sections.

Using a set of questions designed to probe a single concept, we can measure the probability for a single student to activate the different common models in response to these questions. We can use these probabilities to represent the student model state. Thus, a student's model state can be represented by a specific configuration of the probabilities for using different common models in a given set of situations related to a particular concept.

Figure 1 shows a schematic of the process of cueing and activating a student's model, where M_1, \dots, M_w represent the different common models (assuming a total of w common models including a null model) and q_1, \dots, q_w represent the probabilities that a particular situation will result in a student activating the corresponding model. (Note that given different sets of questions, the measured probabilities can be different. The measured student model state is a result of the interaction between the individual student and the instrument used in the measurement and should not be taken as a property of the student alone. This is discussed in detail in the next section.) For convenience, we consistently define M_1 to be the expert model and M_w to be the null model. The possible incorrect models are represented with M_2, \dots, M_{w-1} .

III. STUDENT MODEL SPACE: A MATHEMATICAL REPRESENTATION

We represent the mental state of the student with respect to a set of common models in a linear vector space. Each

common model is associated with an element of an orthonormal basis, \mathbf{e}_η :

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \mathbf{e}_w = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \quad (1)$$

where w is the total number of common models being considered (including a null model) associated with the concept being probed. It can be argued that different mental models can have common and overlapping components.

The use of orthogonal vectors in representing the different common models is inspired by studies in biologically plausible neural networks; the brain can distinguish overlapping inputs into distinctive categories, which are represented in terms of sparsely distributed orthogonal neural activation patterns.³³ Suppose a set of concepts is developed over a range of dimensions of features. Between any two concepts, there will be certain dimensions that are identical and certain dimensions that are different. For example, one can imagine a list of identical and different features between the concept of a bird and the concept of a bat. In conceptual space, birds and bats are two distinctive categories, whereas in feature space, they have many overlapping features. Here, orthogonality was employed in conceptual space only, rather than in feature space, to represent the distinctive conceptual categories. Another example can be found in image processing for symbol recognition. The letters “B” and “P” have many overlapping features in “pixel” space as seen by a computer through a digital camera. Once recognized, the two letters are orthogonal categories in symbolic space. Such a treatment is a standard method in pattern recognition and signal processing. The orthogonal basis in Eq. (1) is employed in a similar manner to label the distinctive categories of student knowledge (models).

We refer to the space spanned by these model vectors as the *model space*. As discussed in Sec. II, in general, the student can be expected to be in a mixed model state. For a given instrument, we represent this state using the probabilities for a student to be cued into using each of the different models. In principle, these probabilities can be probed in experiments; however, a precise determination is often difficult to achieve even with extensive interviews. But in practice we can obtain estimations of this probability with properly designed measurement instruments.

A convenient instrument is a set of research-based multiple-choice questions. Suppose we give a population of students m multiple-choice single-response (MCSR) questions on a single concept for which this population uses w common models. Define \vec{Q}_k as the k th student’s probability distribution vector measured with the m questions. Then we can write

$$\vec{Q}_k = \begin{pmatrix} q_1^k \\ q_2^k \\ \vdots \\ q_w^k \end{pmatrix} = \frac{1}{m} \begin{pmatrix} n_1^k \\ n_2^k \\ \vdots \\ n_w^k \end{pmatrix}, \quad (2)$$

where q_η^k represents the probability for the k th student to use the η th model in solving these questions and n_η^k represents

the number of questions in which the k th student applied the η th common model. We also have

$$\sum_{\eta=1}^w n_\eta^k = m. \quad (3)$$

In Eq. (2) we have taken the probability that the k th student is in the η th model state to be $q_\eta^k = n_\eta^k/m$. Note that q_η^k is affected by the specific question set chosen. The student model state represents the result of an interaction between the student and particular instrument chosen.

To see why this is the case, consider an infinite set of expert equivalent questions concerning a particular concept that an individual student might consider as requiring two different models, model A or model B, depending on the presence or absence of a particular (actually irrelevant) element in the problem. Assume that if the element is present, the student strongly tends to choose model A; otherwise, they will choose model B. Since the set of questions can contain infinitely many items that have the element and infinitely many items that do not, the instrument designer may create an instrument that has any proportion of the items containing the irrelevant element. The percentage of student choices of model A or B thus depends on the number of items on the test containing A.

The student model state as measured by a particular instrument therefore depends on both the student and instrument. Since we are concerned with evaluating normative instruction, in which the student is being taught a particular model or set of models, the choice of the proportion of questions depends on normative goals—what the instrument designer considers important for the student to know. *The student model state should therefore be thought of as a projection of student knowledge against a set of normative instructional goals, not as an abstract property belonging to the student alone.* For the purpose of assessment, researchers can develop (through systematic research on student models) a rather standardized set of questions based on the normative goals. These questions can then be used to provide a comparative evaluation of situations of student models for different populations.

We do not choose the probability vector \vec{Q}_k to represent the model state of the k th student. Rather, we choose a vector consisting of the *square roots* of the probabilities. We refer to these square roots as the *probability amplitudes*. In principle, either approach might be considered. In practice, there are considerable advantages to the square root choice, as it naturally leads to a convenient structure, the density matrix, as we will see below. [We choose to define the square root vector so that when the inner and outer products of this vector are taken with itself it yields useful and straightforward relationships. The inner product leads to the sum of probabilities constraint, and the outer product produces the density matrix defined in Eq. (7). Although there could be many ways of constructing a density matrix from probabilities and their joint products, we choose to build with the square root vector. This construction respects the symmetry of the space with respect to the exchange of the models, and the use of a matrix built by outer products permits useful manipulative

5. A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):
- (A) a downward force of gravity along with a steadily decreasing upward force.
 - (B) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth.
 - (C) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity.
 - (D) an almost constant downward force of gravity only.
 - (E) none of the above. The ball falls back to ground because of its natural tendency to rest on the surface of the earth.

FIG. 2. Question 5 of the FCI test.

techniques.] We therefore choose to represent the model state for the k th student in a class with a vector of unit length in the model space, \mathbf{u}_k :

$$\mathbf{u}_k = \begin{pmatrix} \sqrt{q_1^k} \\ \sqrt{q_2^k} \\ \vdots \\ \sqrt{q_w^k} \end{pmatrix} = \frac{1}{\sqrt{m}} \begin{pmatrix} \sqrt{n_1^k} \\ \sqrt{n_2^k} \\ \vdots \\ \sqrt{n_w^k} \end{pmatrix}, \quad (4)$$

where

$$\mathbf{u}_k^T \mathbf{u}_k = \sum_{\eta=1}^w q_\eta^k = 1. \quad (5)$$

IV. ANALYZING STUDENT MODELS WITH MULTIPLE-CHOICE INSTRUMENTS

Using our mathematical representation, we can analyze student responses to multiple-choice questions to measure student model states and study the evolution of a class's learning. The development of an effective instrument should always begin with systematic investigations of student difficulties in understanding a particular concept. Such research often relies on detailed interviews to identify common models that students may form before, during, and after instruction. Using the results from this research, multiple-choice questions can be developed where the choices of the questions is designed to probe the different common student models. (For some tools to help design effective distracters and to see how different design may affect the measurement, see Refs. 30 and 34.) Then interviews are again used to confirm the validity of the instrument, elaborate what can be learned from the data, and start the cyclic process to further develop the research.

In physics education, researchers have developed research-based multiple-choice instruments on a variety of topics. The two most popular instruments available on concepts in Newtonian mechanics are the FCI and FMCE.^{5,6} The questions were designed to probe critical conceptual knowledge, and their distracters are chosen to activate common naive conceptions. As a result, many of the questions on these tests are suitable for use with the model analysis method. In this paper, we use the data of the FCI test from engineering students in the calculus-based physics class at

the University of Maryland. Results of the FMCE test with students from other schools are discussed in Ref. 30.

A. Force-motion model

An example in Newtonian mechanics where students commonly have a clearly defined and reasonably consistent facet is the relation of force and motion. Student understanding of the force-motion connection has been thoroughly studied for the past two decades, and researchers have been able to develop a good understanding of the most common student models.^{4,34–38} A commonly observed student difficulty is that students often think that a force is always needed to maintain the motion of an object. As a result, students often have the idea that there is always a force in the direction of motion. For the population in our introductory physics class, this is the most common incorrect student model related to the concepts of force and motion. Some even consider that the force is proportional to the velocity. In the physics community model, an unbalanced force is associated with a change in the velocity—an acceleration. Therefore, for this concept, we can define three common models.

Model 1: a nonzero net force results in change of the velocity of motion (correct expert model).

Model 2: there is always a force in the direction of motion (student model, sometimes correct, sometimes incorrect).

Model 3: null model.

In the FCI, five questions activate models associated with the force-motion concept (questions 5, 9, 18, 22, and 28). (In the FCI, two clusters of questions, those on Force-Motion and Newton III, provide most of the FCI's discriminatory power. For details on how we identified these questions using a quantitative argument, see Ref. 34.) As an example, consider question 5 (see Fig. 2). The distracters "a," "b," and "c" represent three different responses associated with the same incorrect student model (model 2). All of the three choices involve a force in the direction of motion. If a student selects one of these three choices, we consider that the student is using model 2. (Here we use a model assignment scheme based on the student response to a single item. More complex situations can be considered. See Ref. 30.) To use this method, we have to assume that if a student is cued into using a particular model, the probability for the student to apply the model inappropriately is small (<10% empirically) compared to random guessing. Such probabilities can often be evaluated with interviews. (More detailed analysis on the

TABLE I. Associations between the physical models and the choices of the five FCI questions on the force-motion concept.

Questions	Model 1	Model 2	Model 3
5	d	a, b, c	e
9	a, d	b, c	e
18	b	a, e	c, d
22	a, d	b, c, e	
28	c	a, d, e	b

certainty of such a model assignment is discussed in Chap. 4 of Ref. 30. The probability can also be measured with specially designed questions where we can give a cluster of questions based on similar context settings with the leading ones (often simple) to test if the students are triggered into a particular model state and the following ones (somewhat more complex) to test if the students can apply the models correctly. We have also developed modeling schemes based on the student responses patterns on a series of questions. Details can also be found in Chap. 5 of Ref. 30.) More details on the uncertainties of this method can be found in Ref. 39. With this method, if a student answers “d” on this question, we assume that it is very likely for this student to have a correct model. (Note that this is not always the case. With some questions, students can choose the right answer for the wrong reasons. To obtain more accurate representations of student reasoning using our method, the wording of such items needs to be improved and the probability of student model crossover estimated through interviewing.³⁰) Choice “e” reflects the Aristotelian idea and is rarely held by students in our introductory physics class. If a student does choose this option, we consider this student as having a null model.

We assume that there are clear associations between the three models and the responses corresponding to the five FCI questions in the force-motion cluster as listed in Table I. Notice that the mappings between model and item do not have to be one to one. It is appropriate to have multiple choices mapped to a single model but not the opposite.³⁹ Further, note that having the correct model does not imply having the correct answer. The student might have a correct model but employ it incorrectly. If there are known to be common errors in applying a correct model, we might want to include some of these errors as distracters. A good understanding of the most common student errors allows the construction of questions that probe both student model choice and student accuracy. In this analysis we only consider the students’ model choice. This underlines the fact that a model analysis provides different information about student thinking than does a right and wrong analysis.

Using Table I, we can obtain an estimation of individual students’ model states from students’ responses. For example, if a student answers the five questions with “a,” “d,” “a,” “d,” and “b,” the student probability vector is $(2 \ 2 \ 1)^T/5$. Using Eq. (4), the model state for this student is $(\sqrt{2} \ \sqrt{2} \ 1)^T/5$.

B. Class model density matrix

As discussed above, for a particular physical concept, a single student can have a pure (consistent) model state (not necessarily a correct one) when the student consistently uses a single model for all expert-equivalent questions related to the concept, a mixed model state where the student uses several models (correct and incorrect ones) inconsistently, or a null model state where no clear models can be categorized (no known systematic logical reasoning involved in generating the response).

For a class probed by a given instrument, each student has an individual model state. The combined outcomes of the class contain both individuals’ features of their model states and the group’s behavior of the students in the class. Therefore, we are tackling a very complicated system that involves both individual and group effects. Analyses using scores alone often fail to provide useful details on the students’ real understanding of the physics concept (except in the case when most students consistently give correct answers). For example, a low score can be caused by a consistent incorrect model, calculation errors generated while using a correct model, random guessing, or a persistently triggered incorrect model for a student in a mixed model state. These different situations reflect important information on student understanding of physics, but they cannot be distinguished using an analysis based solely on scores. We introduce here a procedure we call *model estimation* that can provide a way to extract such information.

Using a group of questions associated with a single physics concept, we can measure and represent the single student model state with Eq. (4). In the following, we use the example of the force-motion models and the FCI to demonstrate the model estimation algorithm. The FCI has five force-motion questions and involves three models, so $m=5$ and $w=3$. We can rewrite Eq. (4) as

$$\mathbf{u}_k = \frac{1}{\sqrt{5}} \begin{pmatrix} \sqrt{n_1^k} \\ \sqrt{n_2^k} \\ \sqrt{n_3^k} \end{pmatrix}, \quad (6)$$

where n_i^k is the number of questions the k th student answered using the i th model.

We define the *single student model density matrix* for the k th student as ($w=3$):

$$\mathbf{D}_k = \mathbf{u}_k \otimes \mathbf{u}_k^T = \{\rho_{\eta\mu}^k\} = \frac{1}{m} \begin{bmatrix} n_1^k & \sqrt{n_1^k n_2^k} & \sqrt{n_1^k n_3^k} \\ \sqrt{n_2^k n_1^k} & n_2^k & \sqrt{n_2^k n_3^k} \\ \sqrt{n_3^k n_1^k} & \sqrt{n_3^k n_2^k} & n_3^k \end{bmatrix}. \quad (7)$$

Although the single student model density matrix clearly contains no more information about the student than does the model vector (all the elements of the matrix are uniquely determined by the elements of the vector), the situation changes dramatically when we sum over all students in the class. We define the *class model density matrix* as the average of the individual students’ model density matrices:

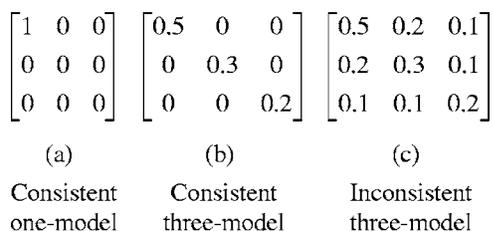


FIG. 3. Examples of the student class model density matrix: (a) an extreme case corresponding to the first type of class model condition where everyone has the same physical model (model 1), (b) the second type of class model condition where the class consists of three different groups of students each with a consistent physical model, and (c) the third type of class model condition where many students have multiple physical models and are inconsistent in using these models.

$$\mathbf{D} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} = \frac{1}{N} \sum_{k=1}^N \mathbf{D}_k$$

$$= \frac{1}{N \cdot m} \sum_{k=1}^N \begin{bmatrix} n_1^k & \sqrt{n_1^k n_2^k} & \sqrt{n_1^k n_3^k} \\ \sqrt{n_2^k n_1^k} & n_2^k & \sqrt{n_2^k n_3^k} \\ \sqrt{n_3^k n_1^k} & \sqrt{n_3^k n_2^k} & n_3^k \end{bmatrix}. \quad (8)$$

The class model density matrix retains important structural information about the individual student models which is otherwise lost if we only sum over the model vectors (this will produce the diagonal elements of the density matrix). By analyzing this matrix, we can study the features of the models used by the students in the class.

Now let us consider a population of students with diverse background. In solving a set of questions on a single concept, students in a class can be in a variety of situations on using their models. Three common situations are the following.

- (i) Most students in a class have the same model (not necessarily a correct one) and are self-consistent in using it.
- (ii) The class population uses several different models but each student only uses one model consistently. Thus the class of students can be partitioned into several groups each with a different but consistent model.
- (iii) Individual students in the class can each have multiple models and use these models inconsistently; i.e., the individual students have mixed model states.

Note that these different situations contain statistical features of the population which are intrinsically different from the probabilistic nature of individual student's model state. Corresponding to these different situations, the class model density matrix will show different structures (see Fig. 3). As indicated from Eq. (8), the diagonal elements of \mathbf{D} reflect the percentage of the responses generated with the corresponding models used by the class. The off-diagonal elements reflect the consistency of the individual students' use of their models. Large off-diagonal elements indicate low consistency (large mixing) for individual students in their model use. Empirically, when the ratio between an off-diagonal element and the multiplication product of the square roots of

the two corresponding diagonal elements is larger than 50%, the mixing between the two corresponding models is regarded as significant.

Using the class model density matrix, we can extract quantitative information on the distribution of student models for the class. One convenient method is to perform an eigenvalue decomposition to extract *class model vectors* (the eigenvectors of \mathbf{D}) and the eigenvalues. A detailed discussion of the eigenvalue analysis is given in the Appendix.

The analysis in the Appendix demonstrates that the μ th eigenvalue is the average of the squares of the overlap (dot product) between the μ th eigenvector and the individual students' model vectors. Consequently, the eigenvalue is affected by both the similarity of the individual students' model vectors and the number of students with similar model state vectors. Thus, if we obtain a large eigenvalue (>0.65 empirically) from a class model density matrix, it implies that many students in the class have similar \mathbf{u}_k 's (i.e., the class has a consistent population). On the other hand, if we obtain several small eigenvalues, it indicates that students in the class behave differently from one another. Therefore, we can use the magnitude of the eigenvalues to evaluate the consistency of a class's population and the applicability of the simple form of the model analysis method.

Using an eigenvalue decomposition to analyze the class model density matrix, we can obtain a quantitative assessment of the structure and popularity of the students' common model states. We can evaluate two types of consistency: the consistency of individual students using different models, which is reflected by the off-diagonal elements of the class model density matrix (mixed or pure), and the consistency among different students which is revealed by the eigenvalues.

As indicated by Eqs. (A5) and (A6), if there is an eigenvector with a large eigenvalue, it contains the dominant features of the single student model vectors. We refer to this as the *primary* eigenvector. The additional eigenvectors act as corrections of less popular features that are not represented by the primary state. When considering the class as a single unit, a primary eigenvector gives a good evaluation of the overall model structure of the class. However, if we regard the class as a composition of individual students, there can exist interesting details that cannot be extracted with a simple eigenvalue decomposition due to the fact that the eigenvalue method necessarily yields orthogonal eigenvectors. [It is also a general problem that will be encountered when attempting to represent the distributive results of a population with several definite items (vectors, values, etc.).]

For example, suppose we have a class that can be divided into several groups of students, where students in each group all have similar model states and students from different groups have significantly different model states. In this situation an eigenvalue decomposition can give good results for the following two cases.

- (i) When the model states from different groups are nearly orthogonal (this limits the number of such groups to be equal to the dimensions of the related model space), the eigenvalue decomposition will produce eigenvectors that are similar to these model states.
- (ii) When one of these student groups has a dominant population, the eigenvalue decomposition will produce a pri-

mary vector, with a large eigenvalue, very close to the model state held by this dominant group.

In the case when students are different but not “so” different (with a distribution of different but nonorthogonal model states), an eigenanalysis will not give appropriate model states. Rather, it will provide a set of orthogonal model vectors representing unique features of all the average students’ model states. In the case that the eigenvalues are small, a scatter plot of the individual students’ eigenvectors can suggest whether it might be useful to perform a cluster analysis, separating the class into distinct populations and determining the characteristics of those populations. Based on computer simulations and the analysis of a large-scale data set (which will be discussed in future papers), it is suggested that when the eigenvalue of a primary eigenvector is less than 0.65 and the student model states are mixed, the students in the class will have a somewhat “flat” distribution of nonorthogonal model states. In such cases, plotting the angular distribution of the individual students’ model states and/or conducting cluster analysis may provide more details on the population. Still, the eigenanalysis can provide a simple indicative evaluation of the population when we combine to consider both eigenvalues and eigenvectors in our data analysis. In the example reported below, the primary eigenvalue is close to 0.8, which indicates that most students have similar model states.

C. Representing the class model state: The model plot

In many situations we have encountered, students often have two dominant models: a correct one and a common misconception. To conveniently represent and study the states and changes of student models in this situation, we construct a two-dimensional graph or *model plot* to represent the class use of the two models. For example, suppose we study the first two models in a three-model situation. A class model state (an eigenvector of the class model density matrix) $\mathbf{v}_\mu = (v_{1\mu}, v_{2\mu}, v_{3\mu})^T$ can be represented as a point in a two-dimensional space in which the two axes represent the probabilities that a representative student in the class will use the corresponding models over the whole set of expert-equivalent questions of the probe instrument. The state is represented by a point (point B in Fig. 4) that we refer to as the *class model point* on a plot with $P_1 = \sigma_\mu^2 v_{1\mu}^2$ as the vertical component and $P_2 = \sigma_\mu^2 v_{2\mu}^2$ as the horizontal component.

When the eigenvalue of a class model state is small, the class model point will be close to the origin. On the other hand, a state with a large eigenvalue will be close to the line going through (0,1) and (1,0), which is the upper boundary of the allowed region of the model plot. [Since the two coordinates represent probabilities and the sum of the probabilities must be less than or equal to 1, a class point must lie below the line $P_1 + P_2 = 1$. In addition, since each probability must be positive, the class point must lie within the triangle bounded by the points (0,0), (1,0), and (0,1).] In the case when a class model state vector has small elements on model dimensions that are not considered ($v_{3\mu}$ in this case), which often occurs, we can make an approximation letting $\sigma_\mu^2(1$

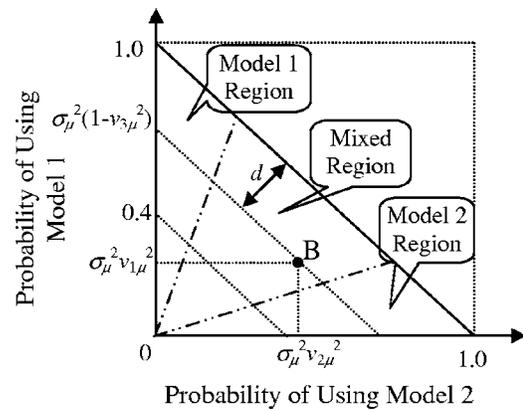


FIG. 4. Model regions on model plot. The model 1 (model 2) region represents comparatively consistent model states with dominant model 1 (model 2) components. The mixed model region represents mixed model states.

$-v_{3\mu}^2) \cong \sigma_\mu^2$. Then the distance between a model point and the upper boundary can be used to estimate the eigenvalue of the corresponding model state. Defining d as the distance between a model point and the upper boundary, this estimation can be calculated with

$$\sigma^2 \cong 1 - d\sqrt{2}. \tag{9}$$

D. Describing model-mixing features

When analyzing student model structures in a case where the model space is dominated by two models we can represent the student model states on a two-dimensional model plot as shown in Fig. 4. In order to describe the different regions of the plot, we separate the plot by drawing two straight lines from the origin with slopes equal to 1/3 and 3, respectively (see Fig. 4). We also draw the line corresponding to the condition $P_1 + P_2 = 0.4$. With these lines, we partition the model plot into four regions: the *model 1 region*, *model 2 region*, *mixed region*, and *secondary model region* (model states with eigenvalues smaller than 0.4) as shown in Fig. 4. When a class has a primary model point in model 1 region (or model 2 region), it suggests that statistically the students in the class have similar model states which have a dominant component on model 1 (or model 2). When a class has a primary model point in the mixed region, the students in the class often have predominantly mixed model states; i.e., most of the students are inconsistent in using the different common models. The secondary model region represents model states with small eigenvalues, which reflect less popular features of the class behavior. In most cases we have studied, there is one primary model state with an eigenvalue 3–4 times larger than the second largest eigenvalue. In these cases, the primary model state alone provides a good overview of the class’s model state.

The model plot can visually present much information about the student model states on the same graph (e.g., the consistency of the class population, the consistency of the individual students in the class, and the types of models used). We can also put the pre- and post-model states from

TABLE II. Results of class model density matrices and class model states on the force-motion concept with data from UMD students.

Tutorial	Pre			Post		
Density matrix	$\begin{bmatrix} 0.27 & 0.23 & 0.02 \\ 0.23 & 0.69 & 0.07 \\ 0.02 & 0.07 & 0.04 \end{bmatrix}$			$\begin{bmatrix} 0.66 & 0.28 & 0.03 \\ 0.28 & 0.31 & 0.02 \\ 0.03 & 0.02 & 0.03 \end{bmatrix}$		
Eigenvalues	0.80	0.17	0.03	0.82	0.15	0.03
Eigenvectors	$\begin{pmatrix} 0.40 \\ 0.91 \\ 0.09 \end{pmatrix}$	$\begin{pmatrix} -0.92 \\ 0.39 \\ 0.07 \end{pmatrix}$	$\begin{pmatrix} 0.03 \\ -0.12 \\ 0.99 \end{pmatrix}$	$\begin{pmatrix} 0.87 \\ 0.48 \\ 0.05 \end{pmatrix}$	$\begin{pmatrix} -0.49 \\ 0.87 \\ 0.02 \end{pmatrix}$	$\begin{pmatrix} 0.03 \\ 0.04 \\ -0.99 \end{pmatrix}$
Traditional	Pre			Post		
Density matrix	$\begin{bmatrix} 0.27 & 0.22 & 0.03 \\ 0.22 & 0.68 & 0.08 \\ 0.03 & 0.08 & 0.05 \end{bmatrix}$			$\begin{bmatrix} 0.46 & 0.25 & 0.03 \\ 0.25 & 0.50 & 0.05 \\ 0.03 & 0.05 & 0.04 \end{bmatrix}$		
Eigenvalues	0.79	0.17	0.04	0.74	0.23	0.03
Eigenvectors	$\begin{pmatrix} 0.40 \\ 0.91 \\ 0.12 \end{pmatrix}$	$\begin{pmatrix} -0.92 \\ 0.39 \\ 0.03 \end{pmatrix}$	$\begin{pmatrix} 0.02 \\ 0.12 \\ -0.99 \end{pmatrix}$	$\begin{pmatrix} 0.67 \\ 0.73 \\ 0.08 \end{pmatrix}$	$\begin{pmatrix} -0.74 \\ 0.67 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} 0.01 \\ 0.10 \\ -0.99 \end{pmatrix}$

different classes together on the same plot, making it much easier to see the patterns and shifts of the different class model states.

V. MODEL ANALYSIS OF FCI DATA

Using the model estimation method, we analyzed FCI data from the pre-post testing of 14 introductory mechanics classes (Physics 161) at the University of Maryland [data collected by J. M. Saul at the University of Maryland (UMd)]. The students were mostly engineering majors. All the classes had traditional lectures three hours per week and were assigned weekly readings and homework consisting of traditional textbook problems. All of the students also had one hour per week of small-group ($N \sim 30$) teaching-assistant- (TA-) led recitations. In half of the classes recitations were traditional TA-led problem-solving sessions (students asking questions and the TA modeling solutions on the board). The other half received recitations taught with *tutorials* (McDermott & Shaffer, 1998). These sessions consisted of students working together in groups of three to five on research-based guided-discovery worksheets. The worksheets often used a cognitive conflict model and helped students develop qualitative reasoning about fundamental physics concepts. In the following analysis, we use the five FCI questions on the force-motion concept as an example to demonstrate the model estimation algorithm.

Using the item-based modeling scheme in Table I and following the procedures in Eqs. (2)–(8), we calculated the average student initial model state on force and motion by combining all classes (778 students). The results are shown in Table II. As we can see from this table, the eigenvalues for the class states corresponding to the null models are very small. This indicates that most students use either the correct expert model or the incorrect naive model and the model space defined from the qualitative research matches well with this population. In addition, the primary class model states (states with the largest eigenvalue) of all classes have eigenvalues around 0.8. Therefore, the primary state alone can give a fairly good description of the class. Using the results in Table II the class model states on the force-motion concept are displayed on a model plot spanned by model 1 (expert model) and model 2 (naive model) (see Fig. 5). For each type of class, we plot the class primary model state. The initial states of both types of classes are nearly the same and can be interpreted as that before instruction most students in the two classes consistently use the incorrect model on all the questions related to force and motion.

After instruction, the model state of the tutorial classes indicates that most students use the correct model rather consistently. On the other hand, the primary model state of the traditional classes indicates a mixed model state, which shows that most students in the class are inconsistent in using their models. Since the model state is nearly a perfect

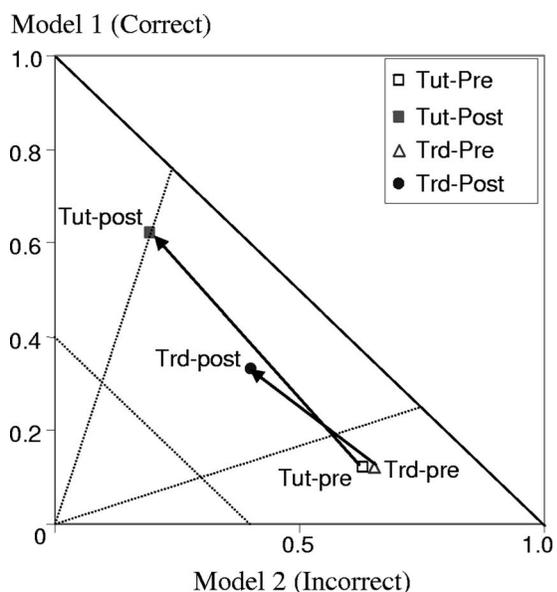


FIG. 5. Model plot of student class model states on force and motion with FCI data from the University of Maryland. For each type of class, we plotted, for pre- and post-results, the first two class model states (states with the first and second largest eigenvalues). The two arrows represent the shifts of the first model states the for pre- and post-results of tutorial (Tut) and traditional (Trd) classes.

mix (half and half), a particular student is likely to use the correct model on half of the questions and use the incorrect model on the other half of the questions. This result provides a piece of evidence that validates our treatment of context dependence with the representation of mixed model states in a population and also indicates that the five FCI questions on force and motion are well designed and are appropriate for the assessment of students' conceptual knowledge concerning context dependence.

VI. COMPARING MODEL ANALYSIS AND FACTOR ANALYSIS

Factor analysis is one of the statistical methods widely used in educational and psychological research. To compare model analysis and factor analysis, we should understand to which extent factor analysis is applicable. In other words, what should one use factor analysis for?

Basically, factor analysis extracts information from a correlation matrix usually built from students' scores on different test items. The factors (eigenvectors) extracted from a correlation matrix provide a measure of how different test items may be related in terms of consistencies among student responses. Factor analysis is not designed to provide the reasons for such relations.

In a test instrument, researchers usually design several equivalent questions on a single concept with varying contextual features. This represents the experts' view on how test items are clustered. However, due to the context dependence of learning, the different contextual features of the equivalent questions may cause the students to respond differently. In such cases, there will be low-consistency among

students' scores to the cluster of questions that the experts would consider equivalent.

The interpretation of this result depends on the researchers' model about student learning. If one considers that the consistency among students' scores reflects their understandings of the test items, the result can lead researchers to think that from the students' point of view, those items which should be grouped together in the experts' view are actually not. However, it can be argued that such an interpretation is valid only when students are in pure model states so that they are consistent in using their knowledge in different but equivalent contexts. When students have mixed model states, the low consistency among students' scores on different equivalent items is primarily caused by the context dependence of their knowledge. Therefore, in such cases, the analysis of the correlation matrix will not be able to identify a strong factor, which is evident from the results in the study by Huffman and Heller.⁴⁰ Let us consider an idealized example to demonstrate how the two methods deal with the issue of context dependence.

Suppose we give four multiple-choice questions to a class of 100 students ($m=4$, $N=100$) and that all four questions probe the understanding of a single physics concept that might activate one of two models: model A and model B ($w=2$). Consider two situations.

Case 1. All students in the class are self-consistent. Half of them use model A on all four questions, and the other half use model B on all four questions.

Case 2. All students are equally mixed between model A and model B: They use model A and model B equally, so each student applies model A to two questions and model B to the other two, but the choices of which questions correspond to which model is random.

In case 1, the results from both methods are calculated in Table III. As we can see, the results from model analysis show two states with equal weights, indicating that the class has two groups, each of which consistently use one of the models. The results from factor analysis give a single factor, which shows that all the students are consistent. The result from factor analysis does not tell in which way the students are being consistent. (This can, however, be supplemented by the information about the class scores.)

The results for case 2 are displayed in Table IV. For both methods, eigenvectors with eigenvalues equal to zero are omitted. Since the students are assumed to be equally mixed with model A and model B, the probability for a single student to use either model A or model B is equal for all questions. As we can see, the results from model analysis indicate a single perfectly mixed class model state with 100% occupancy (the eigenvalue equals 1). On the other hand, since the students are inconsistent in answering the questions, factor analysis gives a randomlike correlation between different questions and shows no dominant factors. Such a situation is often interpreted as if there is no factor in the data. In terms of consistency among students' scores, both methods produce the correct results—students' scores are not consistent. However, in the second case, it is obvious that students in the class are behaving similarly in terms of their model states; they all have an identical mixed model state. This information can only be retrieved from model analysis.

TABLE III. Results from model analysis and factor analysis for a class having two equal populations each with a consistent model.

Model analysis		Factor analysis	
Density matrix	$\frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	Correlation matrix	$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$
Eigenvalues	$\sigma_1^2 = \frac{1}{2}, \sigma_2^2 = \frac{1}{2}$	Eigenvalues	$\sigma_1^2 = 4, \sigma_j^2 = 0 (j=2,3,4)$
Class model states	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$	Factors	$\frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

In the two hypothetical situations, we can see that both methods respond well to the consistency of students' scores. The method of model analysis goes beyond the score consistency and represents the student's knowledge state in a model space. The advantage of using model analysis benefits from the fact that it uses a multidimensional representation for the student's knowledge state and that it is based on the fundamental assumption of learning being context dependent. Therefore, model analysis may better address issues such as context dependence in assessing conceptual knowledge.

Besides factor analysis, there are many quantitative modeling methods developed for education research including structural equation modeling and item response theory.^{41,42}

These methods assume certain latent constructs responsible for students' responses on different test items and rely on the measure of consistency among student responses to extract such latent constructs. This does not respond well to the context dependence of student knowledge and works only when students have pure model states. When students have mixed model states, the low consistency among students' responses often causes failure in detecting any latent constructs. In such cases, the results do not yield much insight into the students' cognitive states.

Model analysis method is fundamentally different. Unlike correlation-based analysis methods, which usually attempt to draw the dimensions of student understanding from test data, model analysis puts information from qualitative research into the analysis to determine the mental space. This space is then used to measure and represent the states of student learning in terms of probabilities for the students to apply different knowledge in a range of contexts.

From an information-processing point of view, what researchers categorize as signal or noise depends on the underlying model of cognition. The context dependence of student knowledge will behave as a random process in the observers' viewpoint with respect to the types of knowledge used by students in changing contexts. In correlation-based data analysis, this randomness is often regarded as a source of uncertainty causing low-consistency in responses. In model analysis, such randomness is treated as signal from the data that represents important features of the student knowledge states—the mixed model states.

VII. SUMMARY AND DISCUSSION

In this paper, we have introduced model analysis, a method to analyze student's knowledge states in large classes with multiple-choice questions. It begins with the cognitive observation that students are often inconsistent in their use of mental models in situations that an expert would consider equivalent. We suggest that the best way to treat this situa-

TABLE IV. Results from model analysis and factor analysis for a class having a single population with an equally mixed model state.

Model analysis		Factor analysis	
Density matrix	$\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$	Correlation matrix	$\begin{bmatrix} 1 & -0.33 & -0.33 & -0.33 \\ -0.33 & 1 & -0.33 & -0.33 \\ -0.33 & -0.33 & 1 & -0.33 \\ -0.33 & -0.33 & -0.33 & 1 \end{bmatrix}$
Eigenvalues	$\sigma_1^2 = 1, \sigma_2^2 = 0$	Eigenvalues	$\sigma_1^2 = 1.33, \sigma_2^2 = 1.33, \sigma_3^2 = 1.33, \sigma_4^2 = 0$
Class model states	$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	Factors	$\frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \frac{-1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

tion is by considering the student as being able to simultaneously possess multiple models with a distribution of probabilities for the activation of the different models. Model analysis allows the assessment on the probabilities of students' use of these alternative models. The results can be used to analyze student understanding and/or the features of the measurement instruments.

Model analysis presents a way to integrate the qualitative knowledge gained from student interviews with the quantitative analysis of multiple-choice instruments. The complete process of using this method is recapped below.

(i) Through systematic research and detailed student interviews, common student models and the contextual features of questions that can activate those models are identified and validated so that these models are reliable for a large population of students with a similar background.

(ii) This knowledge is then used in the design of a multiple-choice instrument. For each concept topic, we need multiple (usually three to five) equivalent questions designed with different contextual features so that we can measure if a student's model state is mixed or pure. In each question, the distracters are designed to capture the common student models and the validity of the questions and the distracters is validated through research.

(iii) With the measurement data, one then classifies a student's responses by corresponding common models and creates a state in the model space representing the student probabilities in applying the different common models. The individual student model states are used to create a density matrix, which is then summed over the class. The eigenvalues and eigenvectors of the class density matrix give information about the state of the class's knowledge.

In constructing a measurement of student conceptual understanding, there is often a "communication" problem; students can use the same terminology (or a statement) as used by an expert but with a different understanding. A simple word or a statement often fails to extract the actual underlying reasoning, which usually can only be obtained by analyzing how students apply their knowledge. Model analysis, although a quantitative tool, relies heavily on qualitative methods. By conducting systematic qualitative research, including careful validation of the test instruments, it is expected that the identified student models reflect the majority of different types of student understandings and that the multiple-choice instruments do not contain significant communication problems—the distracters are designed to reflect not a simple use of a word or statement but rather the results of students' application of their models. That is, we use interviews to identify the students' actual reasoning common to a large population and use research-based multiple-choice instruments, with the algorithms in model analysis, to measure the students' use of these popular types of reasoning in learning.

The combination of the two methods can partially solve the communication problem and yet provide an effective and reliable tool to probe large classes. Once a reliable package is developed, it can be applied in instruction to obtain feedback from students with comparatively rich information on the students' actual understandings.

It is often argued that by putting in researchers' knowledge of student learning in constructing the representation

framework, we limit the framework for students' possible models. In model analysis, we always include a null model space to include possibilities that may be missing when the test is designed. In early stages of research, model analysis could be used with open-ended questions and the results classified by common models using phenomenography.³² If a large null model element is identified in the analysis, it immediately alerts researchers that the population being tested may have possible models that are not understood and suggests the need for further research. Therefore, model analysis can also be used to evaluate the features of the instruments as part of a cyclic process of research, modeling, and development.

The results from model analysis provide more explicit information on improving instruction than score-based analysis. With the knowledge of students' model states and changes of such states with specific contextual features in different equivalent questions, instructors can see more directly the possible causes of the student difficulties and develop better instructional strategies to help students.

Most analyses of the results from the FCI and FMCE compare the pre- and post-test scores of a class and measure an overall "efficiency of instruction" by calculating the fraction of the possible gain g attained by the class.⁴³ While giving a global overview of teaching effectiveness, such a result blends together a variety of distinct learning issues and makes it difficult for an instructor to draw any detailed conclusions about what in his or her instruction was effective or ineffective. This limits the utility of such tests for providing specific guidance to a teacher or researcher for the reform of instruction.

In educational statistics, researchers employ advanced methods such as factor analysis to extract possible latent model-like traits (factors) that underlie students' responses. However, most of these methods assume consistency in the students' activation and application of conceptual knowledge and rely on such consistency to extract latent cognitive factors. In addition, many of these methods rely solely on score-based data. These limitations can lead to difficulty in extracting explicit information on student conceptual models. For example, a factor analysis of FCI results leads to the conclusion that there are no distinct factors, other than the obvious cluster that refers to the conceptually distinct Newton's third law.⁴⁰

From a more general methodological perspective, attempts to extract possible latent cognitive factors from purely test data have fundamental difficulties. In a test situation, there are many hidden processes that can lead to a student's giving a particular type of response. When analyzing test data, researchers need to consider many potential causal pathways for the inferential analysis as well as the issue of context dependence. One way or the other, assumptions have to be made to reduce the complexity of the system. Therefore, results of qualitative research have to be used as the basis for the theoretical assumptions to be employed in the data analysis.

Our approach combines both qualitative and quantitative methods. It assumes that the most commonly used mental models are identified through extensive qualitative research. These known factors can then be mapped onto the choices of

a multiple-choice test design based on results from qualitative research. The mental states of the individual students tend to be mixed, especially when they are making a transition from an initial state dominated by a naive incorrect model to an expert state. Model analysis allows us to take a measure of the degree of confusion in the student's state.

ACKNOWLEDGMENTS

We thank all members of the PER group at UMD and OSU for their help in conducting this research. We also thank many members of the PER community who have given us valuable insights. These include but not limited to Rebecca Lindel, Lillian McDermott, David Meltzer, Ron Thornton, Alan Van Heuvelen, Dean Zollman, and many more. This work is supported in part by NSF Grant Nos. REC-0087788 and REC-0126070.

APPENDIX: INTERPRETATIONS OF THE EIGENSTATE ANALYSIS OF \mathbf{D}

Since \mathbf{D} is symmetric and all data are real and non-negative definite, the eigenvalues are all real non-negative numbers, which we denoted by $\sigma_1^2, \sigma_2^2, \dots, \sigma_w^2$. Define the eigenvectors of \mathbf{D} as \mathbf{v}_μ (a column vector) where $\mu = 1, \dots, w$ are the indices for different class model states. Then the matrix of eigenvectors can be written as

$$\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_\mu, \dots, \mathbf{v}_w].$$

This matrix transforms \mathbf{D} into a diagonal form. For a three-dimensional model space we can write

$$\mathbf{VDV}^T = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}.$$

Now let us see how the information of the individual students' model states are stored in \mathbf{D} and what can be learned from the eigenvalues and eigenvectors. Consider a class with N students. The class model density matrix can be written as

$$\mathbf{D} = \frac{1}{N} \sum_{k=1}^N \mathbf{D}_k = \frac{1}{N} \sum_{k=1}^N \mathbf{u}_k \otimes \mathbf{u}_k^T. \quad (\text{A1})$$

Using the eigenvectors and eigenvalues, \mathbf{D} can also be written as

$$\mathbf{D} = \sum_{\mu=1}^w \sigma_\mu^2 \mathbf{v}_\mu \otimes \mathbf{v}_\mu^T. \quad (\text{A2})$$

Apply this to an eigenvector \mathbf{v}_μ , we can write

$$\mathbf{D}\mathbf{v}_\mu = \frac{1}{N} \sum_{k=1}^N \mathbf{u}_k \otimes \mathbf{u}_k^T \mathbf{v}_\mu = \sigma_\mu^2 \mathbf{v}_\mu. \quad (\text{A3})$$

Define $a_{\mu k}$ as the agreement between the k th student's model vector \mathbf{u}_k and \mathbf{v}_μ , the μ th eigenvector of \mathbf{D} . We have

$$a_{\mu k} = \mathbf{u}_k^T \mathbf{v}_\mu = \mathbf{v}_\mu^T \mathbf{u}_k. \quad (\text{A4})$$

Then Eq. (A4) can be rewritten as

$$\mathbf{D}\mathbf{v}_\mu = \frac{1}{N} \sum_{k=1}^N \mathbf{u}_k (\mathbf{u}_k^T \mathbf{v}_\mu) = \frac{1}{N} \sum_{k=1}^N a_{\mu k} \mathbf{u}_k = \sigma_\mu^2 \mathbf{v}_\mu,$$

which yields

$$\mathbf{v}_\mu = \frac{1}{\sigma_\mu^2 N} \sum_{k=1}^N a_{\mu k} \mathbf{u}_k. \quad (\text{A5})$$

Thus an eigenvector of \mathbf{D} is a weighted average of all the individual student model vectors with weights equal to the agreements between the eigenvector and the single student model vectors.

Therefore, the class model states represented by these eigenvectors are the set of states that reflect the salient features of all the individual student model vectors.

If we left multiply Eq. (A5) by its conjugate, we then have

$$\mathbf{v}_\mu^T \mathbf{v}_\mu = \frac{1}{\sigma_\mu^2 N} \sum_{k=1}^N a_{\mu k} \mathbf{v}_\mu^T \mathbf{u}_k = \frac{1}{\sigma_\mu^2 N} \sum_{k=1}^N a_{\mu k}^2 = 1$$

and

$$\sigma_\mu^2 = \frac{1}{N} \sum_{k=1}^N a_{\mu k}^2. \quad (\text{A6})$$

This result indicates that the μ th eigenvalue is the average of the squares of the agreements between the μ th eigenvector and the individual students' model vectors. Consequently, the eigenvalue is affected by both the similarity of the individual students' model vectors and the number of students with similar model state vectors. Thus, if we obtain a large eigenvalue from a class model density matrix, it implies that many students in the class have similar single student model state vectors (the class has a consistent population). On the other hand, if we obtain several small eigenvalues, it indicates that students in the class behave rather differently from one another. Therefore, we can use the magnitude of the eigenvalues to evaluate the consistency of a class' population.

Further details and developments of model analysis methods can be found at www.modelanalysis.net

*Electronic address: lbao@mps.ohio-state.edu

†Electronic address: redish@physics.umd.edu

¹L. C. McDermott and E. F. Redish, Resource Letter PER-01: Physics Education Research, Am. J. Phys. **67**, 755 (1999).

²A. diSessa, Towards an epistemology of physics, Cogn. Instruct. **10**, 105 (1993).

³J. Minstrell, in *Physics Learning: Theoretical Issues and Empirical Studies, Proceedings of an International Workshop, Bremen*,

- Germany, edited by R. Duit, F. Goldberg, and H. Niedderer (IPN, Kiel, Germany, 1992) p. 110.
- ⁴L. Viennot, Spontaneous reasoning in elementary dynamics, *Eur. J. Sci. Educ.* **1**, 205 (1979); I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985); Common sense concepts about motion, **53**, 1056 (1985); J. Clement, Students' preconceptions in introductory mechanics, *ibid.* **50**, 66 (1982); and many others.
 - ⁵D. Hestenes, M. Wells, and G. Swackhammer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
 - ⁶R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula, *Am. J. Phys.* **66**, 338 (1998).
 - ⁷E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, New York, 1997).
 - ⁸P. Thagard, *Mind: Introduction to Cognitive Science* (MIT Press, Cambridge, MA 1996).
 - ⁹J. M. Fuster, *Memory in the Cerebral Cortex* (MIT Press, Cambridge, MA, 1999).
 - ¹⁰J. Bransford, A. Brown, and R. Cocking, *How People Learn* (National Academy Press, Washington, DC, 1999).
 - ¹¹J. R. Anderson and C. Lebiere, *The Atomic Components of Thought* (Lawrence Erlbaum Associates, Mahwah, NJ, 1998).
 - ¹²A. Baddeley, *Human Memory: Theory and Practice* (Allyn & Bacon, Needham Heights, MA, 1997).
 - ¹³E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 4th ed. (McGraw-Hill, New York, 2000); P. S. Churchland and T. J. Sejnowski, *The Computational Brain* (MIT Press, Cambridge, MA, 1993).
 - ¹⁴E. F. Redish, in *Proceedings of the Varenna Summer School in Physics, 'Enrico Fermi,' Course CLVI, Varenna, Italy*, edited by E. F. Redish and M. Vicentini (IOS Press, Amsterdam, 2004), p. 1.
 - ¹⁵D. Hammer, A. Elby, R. Scherr, and E. Redish, in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2004), Chap. 3.
 - ¹⁶M. Sabella and E. F. Redish (unpublished).
 - ¹⁷R. Steinberg and M. Sabella, Performance on multiple-choice diagnostics and complementary exam problems, *Phys. Teach.* **35**, 150 (1997).
 - ¹⁸*Handbook of Psycholinguistics*, edited by M. A. Gernsbacher (Academic Press, San Diego, CA, 1994).
 - ¹⁹G. Morgan, *Images of Organization* (Sage Publications, Newbury Park, CA, 1986).
 - ²⁰A. H. Schoenfeld, *Mathematical Problem Solving* (Academic Press, Orlando, FL, 1985).
 - ²¹A. diSessa and B. L. Sherin, What changes in conceptual change?, *Int. J. Sci. Educ.* **20**, 1155 (1998).
 - ²²A. Caramazza, M. McCloskey, and B. Green, Naive beliefs in "sophisticated" subjects: misconceptions about trajectories of objects, *Cognition* **9**, 117 (1981).
 - ²³S. Vosniadou, Capturing and modeling the process of conceptual change, *Learn. Instr.* **4**, 45 (1994).
 - ²⁴V. Otero, in Ref. 14, p. 409.
 - ²⁵D. Norman, in *Mental Models*, edited by D. Gentner and A. L. Stevens (Lea Publishing, Hillsdale, NJ 1983), p. 1.
 - ²⁶F. Reif and S. Allen, Cognition for interpreting scientific concepts: A study of acceleration, *Cogn. Instruct.* **9**, 1 (1992).
 - ²⁷D. P. Maloney and R. S. Siegler, Conceptual competition in physics learning, *Int. J. Sci. Educ.* **15**, 283 (1993).
 - ²⁸T. Shallice, *From Neuropsychology to Mental Structure* (Cambridge University Press, Cambridge, UK, 1998).
 - ²⁹C. Spearman, "General intelligence," objectively determined and measured, *Am. J. Psychol.*, **15**, 201 (1904).
 - ³⁰L. Bao, Ph.D. thesis, University of Maryland, 1999.
 - ³¹L. Bao, K. Hogg, and D. Zollman, Model analysis of fine structures of student models: An example with Newton's third law, *Am. J. Phys.* **70**, S766 (2002).
 - ³²F. Marton, Phenomenography—a research approach to investigating different understandings of reality, *J. Thought* **21** (3), 28 (1986).
 - ³³E. T. Rolls and A. Treves, *Neural Networks and Brain Function* (Oxford University Press, New York, 1998).
 - ³⁴L. Bao and E. F. Redish, Concentration analysis: A quantitative assessment of student states, *Am. J. Phys.* **69**, S45 (2001).
 - ³⁵A. Champagne, L. Klopfer, and J. Anderson, Factors influencing the learning of classical mechanics, *Am. J. Phys.* **48**, 1074 (1980).
 - ³⁶J. Clement, Students' preconceptions in introductory mechanics, *Am. J. Phys.* **50**, 66 (1982).
 - ³⁷I. Galili and V. Bar, Motion implies force: where to expect vestiges of the misconception?, *Int. J. Sci. Educ.* **14**, 63 (1992).
 - ³⁸I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985); Common sense concepts about motion, **53**, 1056 (1985).
 - ³⁹L. Bao (unpublished).
 - ⁴⁰D. Huffman and P. Heller, What does the force concept inventory actually measure?, *Phys. Teach.* **33**, 138 (1995).
 - ⁴¹K. G. Jöreskog, in *Structural Equation Models in the Social Sciences*, edited by A. S. Goldberger and O. D. Duncan (Seminar Press/Harcourt Brace, New York, 1973).
 - ⁴²R. K. Hambleton and H. Swaminathan, *Item Response Theory: Principles and Applications* (Kluwer, Nijhoff, 1984).
 - ⁴³R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).