“What should we do when we find students giving inconsistent responses on a measurement instrument?” The answer depends on the cognitive models chosen as the basis for the assessment.

The common theoretical basis for many assessment tools assumes that students possess internal models or have general abilities that are responsible for creating the observed measurement results. Under this assumption, those doing assessment would look for consistency between student responses on correlated items in a test in order to infer the existence and quality of those internal abilities. In this case, inconsistent data is often rejected as noise. However, as indicated by research in education, cognitive science, and neuroscience, the cognitive response is a dynamic process that is highly context dependent. Any results observed from a measurement instrument are the results of the interactions between students’ internal characteristics and the specific settings of the instrument. Therefore, inconsistent student behavior on correlated items not only represents a common phenomenon but provides important information about the state of the student’s thinking. If those doing assessment take this perspective, the inconsistency of student behavior becomes part of the signal needed for analyses.

How, then, can those doing educational assessment extract information from such inconsistency and make it useful for education practice? In this chapter, the authors introduce a new quantitative method based on analogies with methods of quantum physics that works with research-based multiple-choice questions to give insight into students’ conceptual states. Both the consistent and inconsistent features of student behaviors are interactively represented under a coherent theoretical basis, which can be used directly to provide direction for instructional interventions.

I. Student Behavior in Learning (Physics) and the Goals for Assessment

The recent decades of education research has brought us a rich collection of student difficulties in learning science and mathematics and many cognitive theories developed to interpret them. We begin the discussion with a brief review of observations and interpretations of common types of student learning behavior. Our interest in assessment has led us to consider four general areas:

- Learning is affected by students’ previous knowledge.
- Students can create new alternative ideas which can be different from both their previous knowledge and the scientific ones.
- Students’ knowledge structures can be different from that of experts in many ways.
- There are commonly observed inconsistencies in students’ use of their knowledge in equivalent contexts.

Effects on learning from students’ previous knowledge

It is well established that students possess a system of knowledge prior to instruction, which develops from personal experience and previous learning. Such knowledge, often called
preconceptions, misconceptions, or alternative conceptions by physics educators, can have significant effects on students’ learning. Many studies in physics education have shown that certain preconceptions can pose strong barriers to understanding physics knowledge correctly (Viennot 1979, 1984, Clement 1982, McDermott 1984, Halloun and Hestenes, 1985a, b). Researchers in other areas of science education have also found similar results where students’ misconceptions were found to be “entrenched” and difficult to change by conventional instruction (Posner et al, 1982; Scott, Asoko & Driver, 1991, NRC 1999).

These studies and others lead us to consider learning in terms of a constructive process that is affected by the information an individual encounters including what they already know and under what conditions they encounter it. Instruction is most effective when it helps students build on (rather than getting rid of) what they already have. (Redish 2002). There is a large body of studies in physics education research documenting common types of students’ prior knowledge and on developing instructional strategies to deal with them (McDermott and Redish, 1999).

Students’ creations of alternative ideas

During instruction, it is widely observed that students can create new alternative conceptual understandings that are different from both the students’ previous knowledge and the scientific knowledge introduced in the instruction. One form of such creation is the development of synthesized (or hybrid) models of a concept during learning. For example, based on the mental model framework, Vosniadou investigated children’s development of the concept of earth and identified a unique situation: during learning children can develop synthesized mental models to reconcile the conflicts between their initial naïve models and the evidences that support the scientific model (Vosniadou 1994). A hybrid mental model incorporates some pieces of the correct understandings while keeping large part of one’s existing knowledge system unchanged. Since it has some part of the correct understandings, a hybrid model can provide explanations consistent with some parts of correct knowledge. This creates a temporary solution for the solving the conflicts created by using one’s prior knowledge and thus can be used as a quick (and easy) fix when cognitive conflicts are encountered in learning.

Vosniadou and Brewer (1992) reported that children often have an initial model of a “rectangular earth” which reflects the common sense of land being flat. Through learning, a child started to recognize that the earth has to be a sphere. To reconcile the conflict between a spherical earth and the experience of land being flat, children can develop a hybrid model of a “hollow sphere” where people can stand inside the sphere on a flat plane.

Not only children develop such hybrid models. In a study of student understanding of quantum mechanics, senior level undergraduate students also develop hybrid models to reconcile quantum phenomena with their classical intuitive beliefs (Bao, 1999). For example, in learning quantum mechanics, two popular naïve models are: (1) the wavefunction of a particle represents the energy of the particle; and (2) an electron will lose energy as it passes through a potential barrier. Most students were able to memorize the correct shape of the wavefunction representing a particle going through a potential barrier, in which the wavefunction shows a decaying shape in the region of the barrier. By putting all three pieces together, students create a tentatively consistent explanation of quantum tunneling (for this particular context): that the particle will lose energy through the barrier causing the wavefunction to show a decaying curve. This explanation is neither prior knowledge nor correct knowledge, but a creation by students through reassembling bits and pieces from both their prior knowledge and the instruction.
The creation of knowledge can also take place at an implicit and abstract level where no apparent pieces of prior knowledge are explicitly involved; but instead, primitives of reasoning patterns developed with students’ prior knowledge can affect students’ creations of new knowledge without students’ explicit awareness. We begin this discussion with a brief review of the work by diSessa and Minstrell.

DiSessa investigated people’s sense of physical mechanism, that is, their understanding of why things work the way they do. 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 can’t 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* (diSessa 1993). 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. Minstrell refers to such statements as *facets* (Minstrell 1991).

Sometimes diSessa’s p-prims are general in character and sometimes they refer to specific physical situations and are similar to Minstrell’s facets. We restrict the term “primitive” to describe the finest-grained cognitive element in our hierarchy, the reasoning primitive — the abstract p-prims in diSessa’s collection and the abstractions of those that refer to more specific physical situations. We use the term facet to refer to the mapping of a reasoning primitive into a physical situation. An example of where we would separate differently from diSessa is in the p-prim “continuous push” — a constant effort is needed to maintain motion. We would refer to this as a facet, which is a mapping of the reasoning primitive: a continued cause is needed to maintain a continued effect.

During instruction, the mapping of a reasoning primitive from students’ prior knowledge system with newly encountered contexts can create new types of facet-like knowledge. Such mapping is often implicit to students, i.e., students are explicit about the facet they use but are much less, if at all, aware of which particular primitive they are using to create the new facet. In other words, we consider that a primitive such as “more cause leads to more effect” represents a very general reasoning pattern of proportionality and is not directly applicable in real contexts before the actual object or entities of the contexts to be associated with the p-prim is determined. A facet based on this primitive in the context of mechanics would be “more mass produces more force”. It is suspected that this particular facet and the underlying primitive are developed through real world experience and are parts of the student prior knowledge framework. (The facet “more mass produces more force” is often considered a one of the very entrenched naïve preconceptions on mechanics.) Later, when students start to learn electricity, in which the contexts are new to most students and therefore there is not much chance for them to develop preconceptions, the same p-prim can be mapped to create new knowledge such as “more resistance indicates more difficulty to conduct the current and thus causes less current”, which can produce correct answers in many situations.

We consider the creation of hybrid models as an explicit creation process in which students are explicitly aware of using pieces of prior and new knowledge. This creation is a reassembling of those pieces of ready-to-use knowledge. A second type of creation is the more implicit process, in which abstract reasoning patterns and beliefs such as reasoning primitives are remapped to the contexts to form new types of applicable knowledge such as facets. In such cases, students are
usually not explicit about which particular abstract reasoning pattern has been used – the use of such abstract reasoning often appears as a spontaneous intuitive feeling which cannot be further explained, i.e., “it is the way it works”. In general, we may summarize that the first creation process is a reassembling of pieces of explicit and concrete type of knowledge, whereas the second creation process is a mapping of implicit and abstract type of knowledge.

**Students’ knowledge structures**

In studies on students’ approaches in solving problems, differences were found between experts and novices, that provide useful information about how knowledge may be structured and organized by the two populations. For example, experts not only have more physics knowledge than novices, but also they have their knowledge organized in a more coherent fashion. As reflected in problem-solving, experts categorize problems based on the principles used in their solutions while novices tend to categorize problems using surface features of the problem situations (Chi, Feltovich, & Glaser, 1981); experts store information in chunks while novices have the information stored individually (Larkin, 1979); experts are seen to solve problems first by determining the underlying principles useful to a solution while novices jump directly to trying to relate the unknowns to those in memorized mathematical equations (Chi, Feltovich, & Glaser, 1981, Larkin 1981, 1983).

It is also observed that after students “learn” certain knowledge in a particular context, it is hard for them to transfer such knowledge to even a slightly different context. For example, mathematics students who have been trained in traditional instruction do very poorly on non-routine problems (Eisenberg & Dreyfus, 1986). These difficulties are suspected to have their origin in the fact that the knowledge of novice students tends to be in fragments tied to specific surface features of the contexts. A large part of the difficulties for students facing non-routine problems in mathematics is the inability to recognize the deep-level mathematical structure of the problems, confusing contextual information with superficial features (Gliner, 1989, 1991). This same confusion of mathematical and surface structure was found to have a large effect on attempts to transfer problem-solving procedures between different disciplines (Bassok, 1990; Bassok & Holyoak, 1989): When problems were given in similar contexts, students were able to transfer their knowledge between different disciplines. On the other hand, when the surface details were different, students were unable to recognize that the new problems were similar to ones they had seen before. Clearly a deeper conceptual understanding is necessary for students to go beyond the surface details to succeed in transferring knowledge.

**Inconsistency of students’ using their knowledge in equivalent contexts**

In research on student learning of science and mathematics, it is well documented that students can use inconsistent reasoning in responding to tasks that are related to a single concept and considered equivalent by experts (Clough and Driver, 1986; Palmer, 1993; Bao and Redish, 2001). When measuring students’ knowledge, we often give several questions that are related to a single concept topic and are considered equivalent by experts. We call a set of such questions **expert-equivalent**. An expert would use a single model to solve them. However, when presented with such a set of expert-equivalent questions, a student who does not see the coherent deep structure, may use different (and often contradictory) types of understandings to solve these questions.

Many studies have been conducted to model this situation. Maloney and Siegler (1993) proposed a framework of *conceptual competition*. They suggested that the student might both enter and leave the course with several different understandings of a concept that would coexist and compete with, rather replace, the previous understandings. The understanding that wins the
competition on a given problem will be used to represent that problem. The winning of the
competition of a particular understanding is dependent on the nature of the understanding and the
features of the problem.

In a similar study, Thornton (1994) used the term student views to represent the different
understandings of a physics concept. He developed a phenomenological framework to identify
student views of the physical world and to explore the dynamic process by which these views are
transformed in learning. In his research, students were found to have different views coexisting at
the same time during instruction. Such mixing of student views was defined as a transitional state.
His research also suggests that in the learning process, many students often move from a somewhat
consistent but incorrect view, through a transitional state, towards a consistent and correct view.

These research results imply that students may simultaneously hold more than one type of
understandings in their minds. Which type they choose to apply is likely to depend on the presented
context and their experiences with similar contexts. While these mixed learning states have not been
the subject of much research, they should not come as a surprise. Ample research shows us that
students develop naïve frameworks or models as a result of everyday, concrete experiences. These
models prove sufficient to explain most of the students’ observations until they encounter formal
instruction in physics. At the time of that instruction, we should not expect the students to change
suddenly to a new, more generalized model. Instead, the change is likely to be gradual. Students
will apply the newly learned model to some situations while retaining the previous model for others.
The mixed learning state will occur as the students make the transition, which can be an important
indicator for students’ achievement in learning.

Expectations on assessment

The ultimate goal of education research is to help students learn better. To achieve this goal, we
need not only to understand the learning process but also to equip the instructors with effective
assessment instruments so that they can obtain in-time feedback on specific issues and difficulties
that are emerging during the course of instruction. Due to the complexity and the large variety of
possible student behaviors during learning, it is necessary for assessment instruments to be able to
provide specific types of information. Along the four categories of student behaviors summarized
above, we expect the assessment instruments to:

- measure the existence and “popularity” of specific pieces of students’ prior knowledge,
- measure specific types of created knowledge, e.g., hybrid models,
- measure and distinguish structural information about students’ knowledge such as deep-
  level understanding vs. superficial understanding, and
- properly handle the “inconsistency” of students’ using their knowledge in equivalent
  contexts.

To achieve the goals listed above, we need a model of the cognitive process so that we can
identify the needed information and understand how to extract such information from student
behaviors in learning. Further, based on the cognitive model we need to identify what variables we
can control and how to systematically control them to obtain robust measurement results.

II. Understanding Student Behaviors in Terms of Context Dependence

We have already seen the importance of contexts in learning. To achieve the goals we set for
assessment, it is necessary to systematically study the effects of contexts in learning and in
measurement. In the literature, there are multiple perspectives of what can be defined (or included) as a context. In our research, we consider four kinds of factors:

- **Content-based context factors**: These are the actual scenarios and specific features of context scenarios employed in or related to the learning of a particular piece of knowledge. For example, in teaching physics, the instructor can use a demonstration in a lecture. The specific objects, the features of such objects, and how these objects are put together to operate are considered as the content-based context feature of this learning activity. One can also use virtually posed scenarios such as a sample problem on the board or a question to the students. All the physical features of such examples and questions are considered as the content-based context factors.

- **Learning environment context factors**: These are specific educational settings and features of such settings used in teaching and learning, which may include the class formats, teaching and learning styles, specific approaches used in learning activities, etc.

- **Student-teacher internal cognitive status**: These include the students’ and instructors’ general views and attitudes on the learning and teaching of a particular content of knowledge as well as the knowledge background on the content area. Specific items may include motivations, goals, attitudes, epistemological/philosophical views, affective status, conceptual understanding status, etc.

- **Specific student cognitive activation**: As a result of a combination of the above factors and specific activities carried out by the student just before or during the measurement, particularly knowledge items may be activated leading to a higher probability of relating specific associated elements.

The content and learning-environment-based context factors are often physically determinable and can be manipulated during the instruction. One can use a particular content-based context scenario (e.g. a physics demo) in very different learning settings (e.g., a lecture, an interactive lab, a tutorial, etc.). Similarly, one can also teach different content topics through identical or different learning settings. Distinguishing between the two types of contexts allows a clearer analysis of how the learning process may be affected by the variety of context issues (both content based and environment based) during the instruction.

Students and instructors’ internal mental states are not directly determinable and controllable. However, these internal states may be activated or formed in relation with the content and learning environment contexts. As a result, these internal states may not be independent variables but often are not easily changed over a short period of time (e.g., in a class session period). Students will often maintain, either explicitly or implicitly, such views and attitudes throughout a class session unless they are significantly manipulated (e.g. the class explicitly emphasizes a particular view and or attitude that is not normally activated by the students). Either way, the time constant for general views to change are often larger than that for students to change understandings of one piece of content knowledge. Therefore, during the learning of a particular piece of content knowledge, one can consider the general views and attitudes as a slow-changing background of the learning process, which can also have significant impact to the learning.

On the other hand, when students have mixed, possibly contradictory elements of knowledge that may arise do to partial or incomplete learning, local cues can strongly affect which elements are activated. This adds a short term, rapidly-changing fluctuation to the student response.
To fully understand the context dependence of learning, we may need to look into the entire context issues simultaneously, which can be very difficult to realize in research. Therefore, it is a practical approach to first isolate certain variables and focus on a narrow slice of the problem. The research discussed in this paper emphasizes the content-based contexts and their effects on learning.

The effects of content-based contexts on learning

The contextual features of the content knowledge are often directly involved in the learning as part of the knowledge and as the cues for activations of knowledge. Here, we try to carefully inspect the four categories of student behaviors discussed in part I by using the context dependence as a theme for a more generalized understanding of the learning process.

Inconsistency of student behaviors

Contextual features of physics scenarios can cause students to activate different types of knowledge. For example, in our study on Newton’s Third Law, after instruction we provide students a set of expert-equivalent questions in which the objects, the mass, and the velocity of the objects would vary. The results show that both the objects (e.g., a football or a cart) and the features of the objects (e.g. mass and or velocity) can cause a single student to use different types of knowledge, which usually exit in two dominant forms, an expert type and an Aristotelian type (Bao, Hogg & Zollman 2002). Other research also suggested similar results (Palmer 1997, Schecker & Gerdes, 1999).

When presented with a physics scenario, students who have not yet developed a principle-level understanding of the concept, often make associations to their prior knowledge developed from real world experience based on the surface features of the contexts. It was observed in our interviews that when presented with a scenario of a collision between two carts similar to the demonstration shown in class, most students were able to use the correct expert knowledge (Bao 2002b). In the same interview with the same student, when presented with scenarios of collisions between real world objects such as footballs, cars, etc., the student would often use her Aristotelian type of knowledge. A plausible reason that students use different types of knowledge is that they associate the similarities of the contextual features to the contexts in which the different types of knowledge are developed. That is, contexts similar to real world situations activate students to use the knowledge they developed in the real world; whereas the contexts similar to the ones used in class often activate the knowledge discussed in instruction. We also found that the students were not actively looking for consistency among the set of given scenarios; however, when requested, some of them were able to identify the similarity and developed a more generalized understanding that was consistent to all scenarios. In such cases, students usually started to realize that their prior knowledge was inappropriate.

These studies suggest that the physical features of the context scenarios related to the teaching and learning of a concept are an important factor causing the inconsistency in students’ uses of their knowledge. Based on the same idea, we can interpret the phenomenon of entrenched misconceptions and the fragmentations (and transfer) of knowledge in terms of context dependence. Before we do so, it is useful to further inspect the deeper level causes of the context dependence.

The underlying mechanisms of context dependence

The origin of the context dependence may be traced directly to the very fundamental level neural and cognitive operations, especially the brain’s ability to develop associations between sensory input and stored information. In neural science, the learning process and the knowledge are
usually modeled with forms of associational networks such as pattern association networks and autoassociation networks (Rolls & Treves, 1998). In psychology, association is used to represent and interpret the relations between conditioning and learned responses (Anderson, 2000). Although many approaches exist to interpreting learning at cognitive levels, it is in general accepted that associations are the basic type of functions underlying all levels of cognitive processes.

One way to model learning at the cognitive level is to use the constructive process of knowledge development in terms of associative networks (Rumelhart & McClelland 1986). Extending this idea to the macroscopic cognitive level, we can then consider mental associations between hypothetical cognitive constructs such as episodic memories (images) of context scenarios, mental models, facets and p-prims. Then a particular type of student understanding of a concept can be represented in terms of networks of associations connecting the different elements (Bao 2002a,b). The phenomenon of context dependence can then be interpreted as the results of patterns of association connecting episodic images of context scenarios and more abstract level constructs such as mental models.

![Figure 1. Patterns of associations between contexts and student models.](image)

Figure 1 shows a simple example to illustrate how we can use association patterns to represent features of context dependence. In this example, C1 … C4 represent four expert-equivalent context scenarios related to a single concept. M1 and M2 are two dominant types of understandings of the concept: here we may refer to M1 as the correct expert mental model and M2 as the naïve model. The lines connecting Ci and Mj represent the associations a single student or a population may develop between context scenarios and conceptual understandings. In practice, such an association represents a probability that a context scenario would activate a student or a population to use a specific type of knowledge, which can be measured using conceptual surveys and interviews (Bao 2002b). Part (a) of the diagram shows an association pattern reflecting inconsistent use of different types of knowledge, i.e., some of the scenarios activate a student with the correct model and others activate the naïve model. On the other hand, part (b) of the diagram shows a consistent use of the correct knowledge in all given contexts.

During the learning, the brain receives inputs from virtually all association areas in the neocortex (Van Hoesen, 1982; Squire, Shimamura and Amaral, 1989), and has available highly elaborated multimodal information from different sensory pathways. The brain integrates these inputs to form multimodal representations of snapshots of scenarios, which are further processed to create episodic memories. Therefore, all the physical features of the context scenarios can be part of the cues that are integrated in the formation of the conceptual knowledge and that can later cause the activation of the previously formed knowledge.

**Interpreting preconceptions, fragmentation, and transfer**

Students’ preconceptions are formed through real world experiences, which establish extensive association networks based on a wide variety real life scenarios. These scenarios and their features
can then act as cueing agents to activate a student’s knowledge to deal with given situations. It is then plausible to expect that when presented scenarios are similar to those from the real world, students will have large probabilities to activate their preconceptions. Since the students’ preconceptions can provide practical explanations that are validated through large amount of experience, the association networks developed with the preconceptions can be very robust or, in other words, entrenched and very difficult to modify. (A conceptual level modification might require a restructuring of the network that connects many different pieces of the knowledge system developed based on considerable personal experience.)

During the instruction, physics knowledge is often introduced with abstract scientific representations and non-real overly-simplified examples. It is then likely for students to develop an association network that only connects the correct physics knowledge with the in-class context scenarios. As a result, students can develop a fragmented view treating physics as non-real and only applicable to the situations presented in class.

Since learning is a gradual constructive process, students may first develop somewhat incomplete association networks, which can function successfully within small localized domains of contexts. When the context domains are crossed, problems can happen in the forms of failures to activate the related knowledge or the activations of inappropriate knowledge. This may explain the phenomenon of students’ developing fragmented conceptual knowledge and the difficulties in transfer of knowledge.

*To create or to activate*

The phenomenon of students’ creating knowledge is also dependent on the context. Here we distinguish two mental processes: activation and creation. Suppose a student is given a context scenario. If the settings of the context are similar to the ones in the student’s experience from which the student has developed prior knowledge, such a context can often activate that knowledge which is then applied in the given context to create a response. We call this an *activation* process that retrieves previously developed knowledge. If the given context has no good matches from the ones in the student’s experience, it is then more likely for the student to create-on-the-spot a new type of understanding based on certain features of the given context. This represents a *creation* process that creates new types of knowledge in responding to a given context. Depending on the familiarity of the content knowledge to the student, one process can be more “dominant” than the other. For example, in learning many topics in classical mechanics, the dominant process is the activation of previously developed knowledge. On the other hand, in learning quantum mechanics, students often create new types of interpretations of the physical settings on the spot using both classical ideas and quantum mechanical ones (Bao 1999, Bao 2002c). It is also observed that in the creation process, the pieces (or features) of a context scenario become the primary cues that establish the associations to many different pieces of knowledge (both old and new) which are then integrated (reassembled) to create an understanding for the presented context.

The p-prim and resource framework (diSessa 1993, Hammer 1996, 2000) emphasizes the phenomena that occur in the creation process. However, this framework doesn’t provide an explicit explanation on the mechanisms of how different resources may be activated, created and applied in contexts, and how such process can be affected with both external and internal factors. In addition, it also pay less attention to the process of a student’s activating previously formed knowledge and applying it directly in context, which is also an important part of the learning process.
The arguments of this section lead to two major conclusions: (1) the creation process and the activation process are both important in learning; and (2) context dependence is a significant part of the mechanism underlying the activation and the creation processes. Using the context dependence framework, both the activation and the creation can be interpreted coherently as indispensable parts of a complete learning process.

Since contexts play a crucial role in all four categories of student learning behaviors, we consider contexts as the primary group of variables that we need to study, control and utilize.

III. Developing a Measurement Variable and a Measurement Representation

Our goal is to develop assessment methods and instruments that give us information about student learning that can be interpreted within our cognitive model. We started our study by taking our first look at the inconsistency issue, which led us to investigate the effects of context in learning. Our second look at the inconsistency issue revealed ways to explain the inconsistency as the results of the learning being context dependent. Now we will take a third look at the inconsistency issue to treat it as “signals” that contain important information about the status of students’ knowledge and the learning process, and we will develop assessment methods that measure and utilize such information. One of the results of this is that what used to be treated as sources of random variations now becomes valuable and interpretable signal.

Developing a measurement variable

We begin by making a list of requirements on a measurement variable. Firstly, it should contain quantitative information on specific types of knowledge that are possible for a student to activate or to create during learning. These often include various types of preconceptions, the correct expert knowledge and possible forms of students’ creations. Secondly, the measurement variable should contain quantitative information on the ways that students use their knowledge – i.e., the information about the consistency of student behaviors. Finally the contextual information of the measurement instruments needs to be integrated in the analysis as an important part of the signal.

Many existing quantitative assessment tools focus on a binary question: Have the students applied the knowledge correctly or have they not? The results of such a measurement are the percentage of students who get the correct answer. The simultaneous possession of alternate frameworks or misconceptions are not reflected in the measurement. In addition, the assessment tools seldom address the context issue or whether students may apply the correct knowledge in some situations and alternative types in others.

We can code richer information about student knowledge in our model by taking our measurement variable beyond the one dimensional type (e.g. a score) and to a multi-dimensional form (e.g. a vector). To construct such a vector, we need first identify the types of knowledge that are to be represented by the vector, which requires a detailed cognitive model of learning and a representation for the knowledge.

A cognitive representation and the inferential measurement

In previous research, we developed a theory-based cognitive representation in the form of associative networks (Bao 1999, Bao & Redish 2001, Bao 2002a). This framework explicitly emphasizes the nature of cognitive measurement being inferential. That is, that the cognitive constructs are all hypothetical in nature and are defined by researchers to interpret student behaviors. These constructs are not directly measurable and can only be inferred from observations.
of student behaviors. Depending on the chosen cognitive theory, the hypothetical constructs and the representation will have different forms. The representation framework we use has three general layers of elements based on their relations with the contexts (see Figure 2).

<table>
<thead>
<tr>
<th>Abstract Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Layer</td>
</tr>
<tr>
<td>Context Layer</td>
</tr>
</tbody>
</table>

Figure 2. Layers of mental constructs in the order of context dependence and measurement.

The context layer consists of the actual physically determinable (controllable) entities such as a student’s raw responses to given tasks and concrete features of the contexts which may include the specific scenarios, tasks, settings of a scenario and settings of the learning environment.

The functional layer contains a category of functional mental constructs that can be activated by or be created on the spot for the contexts. The applications of these functional constructs in the related contexts are directly responsible for the specific responses students would create in responding to the given tasks. Students are explicitly aware of their use of such mental constructs. Examples of these elements include mental models, facets (Minstrell 1991), or simply propositional or statement-like ideas that can be directly applied in the context. The feature of “directly applicable” is crucial, since it establishes the causal relation between such a mental construct and the student’s responses. Based on this casual relation, one can perform inferential analysis on the measured student responses to make inference about the existence and characteristics of those hypothetical mental constructs. Notice that a functional construct is not just a recalled fact. It has certain generality over a domain of contexts (it can often be applied to more than one context setting).

Finally, we have an abstract layer – mental constructs that are general over a large domain of contexts. For example, ontology, general beliefs, and reasoning primitive-like mathematical and logical reasoning patterns are less tied to specific contexts. These elements are often not directly applicable to particular contexts, but rather they contribute to the formation of functional mental constructs in the functional layer. In a way, functional constructs are the context-based “expressions” of the abstract layer elements. For instance, in the example discussed earlier, a reasoning primitive such as “more cause indicates more effect” is simply a very general reasoning pattern of proportionality, which can be related to virtually all situations, but is not directly applicable in any contexts before the actual objects or entities to be associated with the “more” is determined. A facet based on this primitive in the context of mechanics would be “more mass indicates more force”, which is direct applicable in many contexts of mechanics to produce responses. The use of a functional construct such as this facet is explicit to the student. On the other hand, the involvement of the general elements such as a reasoning primitive can be implicit to the student who is explicit about using the corresponding facet. Therefore, the inference about the existence and characteristics of a general element cannot be obtained directly from analyzing the student responses. Such an inference needs to be made based on analysis of a range of functional constructs that involve the same general element.
Based on the orders of the context dependence, we can start to identify variables from the three layers of elements and to develop a cognitive representation for research and assessment. The research process starts with systematic qualitative studies emphasizing the context issues. From the results, we locate a set of context variables. We can then change the context variables to form measurement settings and use them to observe student behaviors in the different settings. These results are analyzed to make inference about the possible functional constructs and the probabilities for the different functional constructs to be used by the students in those contexts. Then based on the results of multiple concept domains, we can make inference about the possible higher order abstract constructs and their relations with the lower order variables. Shown in figure 3 is a flow chart of the general process of inferential measurement for the different layers of cognitive constructs.

![Figure 3. Inferential measurement of mental constructs](image)

**Developing a measurement representation: the basic ideas of Model Analysis:**

In the following sections, we apply this framework to build a measurement representation. As discussed earlier, for a particular physics concept, questions designed with different context settings can cause a student to use different types of knowledge. In our research, we are particularly interested in the functional layer of cognitive constructs. This layer forms functional packages of knowledge interpolating between the abstract and the particular contexts. Elements of this layer can be created, stored and retrieved, and once activated they can be applied by the students to create explanatory understandings of the contexts. For simplicity, we call such a package of knowledge a *model*, which is, in many ways, similar to what is call a facet (Minstrell 1991) and a mental model (Vosniadou 1994). There are a number of features we like to explicitly emphasize toward the definition of a model.

- **Models are context dependent:** The formation and the activation of a model is context dependent. Context features are explicit in the model and can be used as cues to activate previously developed models.

- **Models are productive and directly causal:** A model is a functional package that can be directly applied in contexts to create responses. As a result, we treat the responses produced by a student in a given context as the outcome of the student’s applying the model – i.e., there is a direct causal relation between the model and the student’s responses. This is an important feature for measurement, since in a measurement one can only obtain the student’s responses and then one must make inferences about the models and other abstract mental constructs.

- **Models are domain oriented:** For a particular concept, there are usually a set of models developed by students and experts that each represents a different type of understanding of the concept. We can also identify a domain of contexts related to the concept, which are
involved in the formation and activation of the models. When a model is activated, it is a construct that can be applied to a subset of context instances belong to the context domain related to the concept topic.

**Common models**

Now that we have decided to interpret student behaviors in terms of the formation and application of various models, the measurement variable needs to represent the information about the existence of specific models and the ways that students use the models in contexts. We take an approach that combines both qualitative and quantitative methods to develop the measurement: we use qualitative instruments such as interviews and open-ended surveys to identify the possible types of models that are typically activated or created by students in our particular population. Based on these qualitative results, we then develop quantitative instruments to probe how individual students or a population may use the different models in changing contexts.

Suppose for a particular concept topic, we prepare a sequence of expert-equivalent questions or scenarios in which an expert would use a single, coherent model. During learning, the particular pieces of knowledge or models that can be activated by one of such expert-equivalent questions are highly dependent on the settings of the specific question. Research also suggests that when studying a finite size group of students, the number of different models with which the students can activate by different context settings is also finite and often a small number (Halloun & Hestenes 1985, Marton 1986). Then based on systematic qualitative studies with a student population, for a particular concept topic one can often identify a small set of models that have non-trivial probabilities to be activated or created by students. As a practical treatment, we often lump all the insignificant models or models that yet to be identified and that have a total probability smaller than 5% into a single *null model*. With the null model included, the set of models becomes a complete set, i.e., any student response can be categorized.¹ We then use the term *common models* to represent this complete set of models related to a single concept topic. These models include the null model, the models held by the students with non-trivial probabilities and an expert model (if not already included in the models held by the students).

**Model space and student model state**

With the common models identified, we can quantitatively measure and analyze how students use their models in contexts. We have observed three ways in which a single student may use his/her models when presented with a set of expert equivalent questions (Bao 1999, Bao & Redish 2001, Bao et al. 2002). Each of the situations is defined as a *model state*:

1. The student consistently uses one of the common models in answering all questions. When this is measured, we call it a *pure model state* (or a consistent state).

2. Each question activates one model for the student, but different questions may activate different common models. We call this an *implicit mixed model state* representing inconsistent use of models.

¹ Of course, in addition to collecting random and incoherent student responses, models that have not yet been understood as a stable construct 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 becomes an indicator suggesting that a better understanding of the range of student responses needs to be developed through qualitative research.
3. Each question can activate multiple models for the student. In this case, we call it an explicit mixed model state. (Bao, 2002b).

If a set of questions has been carefully designed to probe a particular concept, we can estimate the probabilities for a single student to activate the different common models in response to these questions and we can use these probabilities to represent the student’s model state. Thus, a student’s model state can be represented by a specific configuration of the probabilities for the student to use different common models in a given set of situations related to a particular concept.

Figure 4 shows a schematic of the process of cueing and activating a student’s model, where \( M_0 \) \ldots \( M_{w-1} \) represent the different common models (assuming a total of \( w \) common models including the null model), and \( q_0 \) \ldots \( q_{w-1} \) 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. For convenience, we consistently define \( M_1 \) to be the expert model and \( M_0 \) to be the null model. The possible incorrect models are represented with \( M_2 \) \ldots \( M_{w-1} \).

![Figure 4. Using a set of questions designed for a particular physics concept, we can measure the probability for a single student to use different common models in solving these problems. In the figure, \( M_0 \) \ldots \( M_{w-1} \) represent the different common models (there are a total of \( w \) common models including a null model), and \( q_0 \) \ldots \( q_{w-1} \) represent the probabilities for a student being activated with the corresponding models.](image)

By putting this set of probabilities in a \( w \)-dimension vector, we can develop a mathematical representation for the model state. This vector is in a linear space spanned by a set of basis vectors that each represents a unique common model. We call such a space the model space (Bao 1999, Bao & Redish 2001).

### IV. Model Analysis: Mathematical Representations and Analysis Methods

The vector representing a single student’s model state can be expressed as a linear combination of the basis vectors of the model space. For example, suppose we give a student \( m \) multiple-choice single-response (MCSR) questions on a single concept for which there exists \( w \) common models. First, let’s consider the probability vector. Define \( \tilde{Q}_k \) as the \( k \)th student’s probability distribution vector measured with the \( m \) questions. Then we can write
where \( q^k_\eta \) represents the probability for the \( k^{\text{th}} \) student to use the \( \eta^{\text{th}} \) model in solving these questions and \( n^k_\eta \) represents the number of questions in which the \( k^{\text{th}} \) student applied the \( \eta^{\text{th}} \) common model. We also have

\[
\sum_{\eta=0}^{w-1} n^k_\eta = m .
\]

In eq. (1) we have taken the probability that the \( k^{\text{th}} \) student is in the \( \eta^{\text{th}} \) model state to be \( q^k_\eta = n^k_\eta / m \). Note that \( q^k_\eta \) is affected by the specific question set chosen.

We do not choose to use \( \tilde{Q}_k \) to represent the model state of the \( k^{\text{th}} \) student. Instead, we choose to associate the student state with a vector consisting of the square root of the probabilities, which has unit length in the model space, \( u_k \):

\[
\begin{pmatrix}
\sqrt{q^k_0} \\
\sqrt{q^k_1} \\
\vdots \\
\sqrt{q^k_{w-1}}
\end{pmatrix} = \frac{1}{\sqrt{m}} \begin{pmatrix}
\sqrt{n^k_0} \\
\sqrt{n^k_1} \\
\vdots \\
\sqrt{n^k_{w-1}}
\end{pmatrix}
\]

where

\[
\langle u_k | u_k \rangle = \sum_{\eta=0}^{w-1} q^k_\eta = 1
\]

We choose to represent the vector using the “bra-ket” notation where the ket state \( | u_k \rangle \) represents the column vector defined as the model state vector of the \( k^{\text{th}} \) student. The bra state \( \langle u_k | \) is a row vector which is the transpose of the ket state. By the standard rules of matrix multiplication, the two put together in “bra-ket” form gives the dot product of the two vectors, which is 1 in this case, i.e., the model state vector is unitary.

The student model state represents an interaction between the student and the particular instrument chosen. 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 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 comparative evaluation on situations of student models for different populations.
When using models and model states, it is important to distinguish the difference between a hybrid model and a mixed model state (Bao 2002b): a hybrid model is treated as any other type of model and is given a unique dimension in the model space. The mixed use of different models describes a type of student behavior and is represented as a mixed model state. That is, a model is a hypothetical cognitive construct defined by researchers. A model state is the measured results of students’ using their knowledge in given contexts which are interpreted against the model space defined by the researchers. Therefore, a hybrid model is considered in our research as an independent mental construct, which represents a completely different entity comparing to a mixed model state: a hybrid model is a model defined by researchers to represent unique types of knowledge whereas a model state represents the ways that students’ use the different types of knowledge rather than the types of knowledge used by students.

**Measuring Students’ Model States with Multiple-Choice Instruments**

We use multiple-choice instruments as an example to show how to apply the mathematical representation to measure and analyze student learning behavior. The development of an effective instrument should always begin with systematic investigations on 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 are designed to probe the different common models. 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 popularly available instruments on concepts in Newtonian mechanics are the Force Concept Inventory (FCI) and Force-Motion Concept Evaluation (FMCE) (Hestenes, Wells, & Swackhammer, 1992; Thornton, 1994). The questions are designed to probe critical conceptual knowledge and their distracters are chosen to activate common naïve 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 data from the FCI test taken by 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 (Bao 1999).

**The force–motion concept**

Student understanding of the concept of 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 (Viennot, 1979; Watts, 1983; Champagne, Klopfer & Anderson, 1980; Clement, 1982; Galili & Bar, 1992; Halloun & Hestenes 1985). A commonly observed student naïve model is the “motion indicates force” – i.e., 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 force–motion concept. The correct expert model would state that an unbalanced force is associated with a change in the velocity – an acceleration. Therefore, for this concept, we can define three common models:

- **Model 0:** Null model.
- **Model 1:** An object can move with or without a net force in the direction of motion. (expert model)
Model 2: There is always a force in the direction of motion. (incorrect student model)

In the FCI, five questions activate models associated with the force-motion concept (questions 5, 9, 18, 22, 28). As an example, consider question 5. (See figure 5.) 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. 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%). Such probabilities can often be evaluated with interviews. With this method, if a student answers “d” on this question, we assume that it is very likely that the student has a correct model. A method to estimate the probability for such inference can be found in the reference (Bao 1999). Choice “e” is rarely held by students in our introductory physics class and thus categorized as reflecting a null model.

**Figure 5. Question 5 of FCI test**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Model 0</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>e</td>
<td>d</td>
<td>a, b, c</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
<td>a, d</td>
<td>b, c</td>
</tr>
<tr>
<td>18</td>
<td>c, d</td>
<td>b</td>
<td>a, e</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>a, d</td>
<td>b, c, e</td>
</tr>
<tr>
<td>28</td>
<td>b</td>
<td>c</td>
<td>a, d, e</td>
</tr>
</tbody>
</table>

The associations between the three models and the responses corresponding to the five FCI questions are listed in table 1. Notice that the mappings between model and item do not have to be one-to-one. However, to ensure that meaningful inference can be made, it is appropriate to map several choices to one model but not one choice to multiple models. Using table1, we can obtain a quantitative estimation of individual students’ model states from the students’ responses. For
example, if a student answers the five questions with “d”, “a”, “b”, “d” and “c”. From table 1, we may infer that the student have used the expert model (Model 1) on all five questions. Thus, the student probability vector is

$$
\hat{Q}_k = \frac{1}{5} \begin{pmatrix}
0 \\
5 \\
0
\end{pmatrix} = \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix},
$$

which shows that the student has the probability of 100% to use the expert model and 0% to use either the null model (Model 0) and the naïve model (Model 2). The corresponding model state vector is

$$
|u_k\rangle = \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix}.
$$

This state can be described as a pure expert state – the student consistently applies the expert model on all five questions related to the concept being probed.

Similarly, if the student answers the questions with “b”, “b”, “a”, “c” and “d”, we may infer that the student have used the naïve model (Model 2) on all five questions, which will produce a pure naïve state

$$
|u_k\rangle = \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix},
$$

which describes that the student consistently applies the naïve model on all five questions.

More interestingly, if the student answers the five questions with “a”, “d”, “a”, “d” and “b”, we will obtain a probability vector

$$
\hat{Q}_k = \frac{1}{5} \begin{pmatrix}
1 \\
2 \\
2
\end{pmatrix},
$$

which gives a model state vector

$$
|u_k\rangle = \frac{1}{\sqrt{5}} \begin{pmatrix}
1 \\
\sqrt{2} \\
\sqrt{2}
\end{pmatrix}.
$$

Compared to the two model states discuss above, this state vector has multiple non-zero components, which reflects that the students is using multiple models on this set of expert-equivalent questions. This is what we call a mixed model state. As discussed earlier, there are two types of mixed states, implicit and explicit mixed states. The method shown here gives an implicit mixed state, i.e., the student is using multiple models on the set of questions but only one model for each question. Methods on measuring explicit mixed state are discussed in the reference (Bao
2002a). In this paper, we only show examples for measuring implicit mixed states, which will be simply called as mixed state.

As we can see, using the model state vector, we can represent quantitatively two types of information: (1) which models (and their probabilities) are inferred as being used by a student in responding to a set of questions; (2) which ways the student uses the models in terms that if the student uses these models consistently or inconsistently (and how inconsistent in terms of probabilities).

Notice again that the state is a measured state using a finite number of questions. Therefore, it is the outcome of the student interacting with the given questions, which reflects certain internal properties of the student’s mind but not the mind itself. This is why we call it a “state” meaning a specific instance of interactive situation of the student’s mind and the contexts. If a different set of questions were used in a measurement with the same student, one should generally expect the result to yield a different model state. However, through multiple measurements with diverse contexts, one may approach statistically an expectation value of the model state for a student or a population. The measurement of the state also has uncertainties due to a variety of factors, some of which are discussed in details in the reference (Bao 1999).

**Analyzing a Population with the Model Density Matrix**

As discussed above, we can use model state vectors to represent how a single student is responding to a set of questions. Now we want to analyze the same types of information for a population. When assessing a class, each student is measured with an individual model state. We may simply add all the individuals’ states together and obtain an average state for the class. However, this will “erase” the information on the consistency of individual students’ uses of models. For example, suppose a class has two groups of students of the same size. Students in one group all have a pure expert model state (eq. 6) and the students in the other group all have a pure naïve state (eq. 7), then the average state for the class obtained by summation of all the individuals’ states are

$$|u_{\text{avg}}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} .$$  \hspace{1cm} (10)

Now consider another situation in which all the students in the class have the identical mixed state

$$|u_k\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} .$$  \hspace{1cm} (11)

The above two situations will produce the same average state for the class. Therefore, the summation of the individuals’ states will eliminate the information about the mixing carried by the individuals’ states.

One can also perform statistical analysis or cluster analysis of the individuals’ states (Bao 2002b). These analyses can provide very detailed information about the class and the individual students. However, as the dimension of the model space increases, the orders of the calculation and the total number of possible combinations of states will increase exponentially (in the form of $2^w N m$ where $w$ is the dimension of the model space, $N$ is the number of students, and $m$ is the number of
questions used). Here we review a method, the model density matrix, which can retain a significant portion of the information on the individuals’ states without much increase of the complexity in calculation. (The increase of calculation is in the order of $w^2 Nm/2$.)

With the individual students’ model states measured, we can calculate a model density matrix for the class denoted $\mathcal{D}$ (eq. 12 is based on the same example discussed above and $N$ is the total number of the students in a class):

$$\mathcal{D} = \frac{1}{N} \sum_{k=1}^{N} |\mathbf{u}_k \times \mathbf{u}_k| = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{m} \begin{bmatrix} n_0^k & \sqrt{n_0^k n_1^k} & \sqrt{n_0^k n_2^k} \\ \sqrt{n_1^k n_0^k} & n_1^k & \sqrt{n_1^k n_2^k} \\ \sqrt{n_2^k n_0^k} & \sqrt{n_2^k n_1^k} & n_2^k \end{bmatrix}. \quad (12)$$

The class model density matrix is a symmetric matrix. The diagonal elements of the matrix add to 1 and give the probabilities of how the class of students altogether uses the different models, which is equivalent to the summation of the model probability vectors. The off-diagonal elements retain the information about whether the individual students use their models consistently. For example, consider a class of students with diverse background. In responding to the set of questions, the class behavior can be interpreted as a combination of three types of situations:

1. All students in a class use the same model (not necessarily a correct one) and are individually consistent in using it. The students all have the same pure model state.

2. The class population holds several different models but each student only uses one model consistently. Thus the class of students can be partitioned into several groups that each consistently uses a different model. All students have pure model state but there can be different types of pure states.

3. Students in the class can each have multiple models and use these models inconsistently. The individual students have mixed model states.

Corresponding to these different situations, the class model density matrix will show different structures (see figure 6).

![Figure 6. Examples of student class model density matrix: (a) an extreme case corresponding to the first type of class condition where everyone has the same model (model 1); (b) the second type of class condition where the class consists of three different groups of students that each uses a consistent model; (c) the third type of class condition where individual students have multiple models and are inconsistent in using their models.](image-url)
As shown in this example, the diagonal elements of the $\mathbf{D}$ give the inferred probabilities for the class of students to use the different models and the off-diagonal elements reflect the consistency of the individual students’ uses of their models. Notice that in the latter two of the three situations shown in figure 6 the diagonal elements are the same but the off-diagonal elements are different. Large off-diagonal elements indicate low consistency (more mixing) for individual students in their model uses. Therefore, with this matrix, we can represent information on students’ using their models for both the class and the individual students. Since the matrix is symmetric, the effective number of elements that need to be calculated is $w(w+1)/2$. This is larger than what is needed for the model state vector, but much less than the total possible combinations of the model states needed in performing cluster analysis.

Extracting Information from the Model Density Matrix

There are multiple ways to extract information from the model density matrix (Bao 1999, 2002b). Here, we will show examples using the method of eigenvector analysis.

Since a class model density matrix is symmetric, we can obtain a complete set of eigenvalues and eigenvectors (Bao & Redish 2001). The magnitude of an eigenvalues is affected by the similarity between the individual students’ model vectors and the number of students that have similar model state vectors. Thus if we obtain a large eigenvalue ($> 0.8$) from a class model density matrix, it implies that many students in the class have similar individual model state vectors. On the other hand, if we obtain several small eigenvalues, it indicates that students in the class behave differently from one another, i.e., the population is diverse. Therefore, we can use the magnitudes of the eigenvalues to evaluate the diversity of a class’ population.

Since the total number of eigenvectors equals the number of dimensions of the model space, they cannot retain all the detailed information about the individual model states. Such information can be extracted through statistical methods or through cluster analysis methods which allow the clusters to be non-orthogonal and unlimited in number – the only constraint is the measurement precision which determines the resolution of a measured model state which further determines the number of total distinguishable state vectors. In general, this resolution is inversely proportional to the number of questions used in a measurement. (See Bao 1999 for details on the measurement resolution.)

The information contained in the eigenvectors reflects a set of unique and sometimes dominant features of the individual model states, which needs to be interpreted in combination with the eigenvalues. For example, if there is an eigenvector with a large eigenvalue ($>0.8$), it represent a dominant state vector that is similar to most individual states. In this case, we refer to this as a primary eigenvector. The additional eigenvectors act as corrections to the primary state reflecting less popular features that are not represented by the primary state.

When considering the class as a single unit, a primary eigenvector gives good evaluation of the overall structure of the class. However, if we regard the class as a composition of individual students, there can exists interesting details that cannot be extracted with a simple eigenvalue decomposition due to the reasons discussed above. For instance, 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 provide good results for two cases:
1. When the model states from different groups are nearly orthogonal, the eigenvalue
decomposition will produce eigenvectors that are similar to these model states. In this case,
the eigenvalues reflect the sizes of the different groups.

2. When one of these student groups has a dominant population, the eigenvalue decomposition
will produce a primary 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 in that they have a range of
different but non-orthogonal model states, an eigenanalysis will not give appropriate model states. If
the eigenvalues are small, we can do a scatter plot of the individual students’ model state vectors,
which can further suggest whether it might be useful to perform a cluster analysis, separating the
class into distinct subgroups of populations and determining the characteristics of those populations.
In general, when the eigenvalue of a primary eigenvector is less than 0.65 and the student model
states are mixed, it often indicates that the students in the class have a less concentrated distribution
of non-orthogonal model states.

**Representing the Class Model State – The Model Plot**

In many situations we have encountered, students often have two dominant models: a correct
expert model and a common misconception. To conveniently represent and study the student model
states and the changes of the states in this situation, we construct a two-dimensional graph called a
model plot to represent the student usage of the two models (Bao 1999, Bao & Redish 2001). For
example, suppose we study model 1 and model 2 in a 3-model situation (model 0 is a null model). A
class model state, $v_\mu = (v_{0\mu}, v_{1\mu}, v_{2\mu})^T$ with an eigenvalue $\lambda_\mu$, can be represented as a point in a two-
dimensional space in which the two axes represent the probabilities that a student in the class will
use the corresponding models on one of the items of the probe instrument. The state is represented
by a point (point B in figure 7) that we refer to as the $I$ on a plot with $P_1 = \lambda_\mu v_{1\mu}^2$ as the vertical
component and $P_2 = \lambda_\mu v_{2\mu}^2$ 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. In the case
when a class model state vector has small elements on model dimensions that are not considered
($v_{0\mu}$ in this case), we can make an approximation letting $\lambda_\mu (1-v_{0\mu}^2) \approx \lambda_\mu$. Then the distance between
a model point and the upper boundary can be used to estimate the eigenvalue of the corresponding
model state.

We divide the model plot into four regions: the consistent model 1 region, the consistent model
2 region, the mixed region, and the small eigenvalue region. When a class has a primary model state
plotted in the model 1 region (or model 2 region), it indicates that most students have comparatively
consistent model states with a dominant probability on model 1 (or model 2). When a class has a
primary model state plotted in the mixed region, most of the individual students have mixed model
states, i.e., most of the individual students are inconsistent in using their models. The small
eigenvalue region represents model states with small eigenvalues (<0.4), which reflect less popular
features of the class behavior. In the example reported below, the primary eigenvalue is close to 0.8,
which indicates that most students have similar model states and the primary eigenvector can
represent most of the information of the class.
Figure 7. Model regions on model plot. Model 1 (Model 2) region represents comparatively consistent model states with dominant model 1 (model 2) components. Mixed model region represents mixed model states.

The model plot is a useful tool that can visually present a variety of information which includes the diversity (or similarity) of the population, the consistency of individual students in using their models, the types of models used by individual students, and the probabilities for individual students and the class population to use the different models. We can also put the pre and post model states from different classes together on the same plot to study the patterns and shifts of the different classes’ model states.

V. Examples Using Model Analysis Methods

Model Analysis of FCI Data

The results shown below are based on FCI data from the pre and post testing of 14 introductory mechanics classes at the University of Maryland (UMD). The students are mostly engineering majors. All of 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~30) recitations led by teaching assistants (TAs). 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 3-5 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 methods of class model density matrix and eigenvector analysis, which are summarized in figure 8.

Details of the calculations are discussed in the references (Bao 1999, Bao & Redish 2001). We show in table 2 the eigenvectors of the pre and post student model density matrices obtained by combining all classes together (N ~ 800).
A student’s responses on 5 questions related to a single concept

<table>
<thead>
<tr>
<th>Q</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer</td>
<td>a</td>
<td>d</td>
<td>a</td>
<td>d</td>
<td>b</td>
</tr>
</tbody>
</table>

Relations between the 3 models and the responses to each question

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>e</td>
<td>d</td>
<td>abc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td>ad</td>
<td>bc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td>b</td>
<td>ae</td>
<td>c</td>
</tr>
</tbody>
</table>

$$V^T D V = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$$

$$|u_k\rangle = \frac{1}{\sqrt{m}} \begin{bmatrix} \sqrt{n_0^k} \\ \sqrt{n_1^k} \\ \sqrt{n_2^k} \end{bmatrix}$$

Single student model state

$$\mathcal{D} = \begin{bmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{10} & \rho_{11} & \rho_{12} \\ \rho_{20} & \rho_{21} & \rho_{22} \end{bmatrix} = \sum_{k=1}^{N} |u_k\rangle \langle u_k|$$

Figure 8. A schematic representation of calculating the class model density matrix and extracting the class’ model states.

Table 2. Results of class model density matrices and class model states on Force-Motion concept with data from UMD students.

<table>
<thead>
<tr>
<th>Force – Motion</th>
<th>Tutorial</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Density Matrix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04 0.02 0.07</td>
<td>0.03 0.03 0.02</td>
<td></td>
</tr>
<tr>
<td>0.02 0.27 0.23</td>
<td>0.03 0.66 0.28</td>
<td></td>
</tr>
<tr>
<td>0.07 0.23 0.69</td>
<td>0.02 0.28 0.31</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03 0.80 0.17</td>
<td>0.03 0.82 0.15</td>
<td></td>
</tr>
<tr>
<td>Eigenvectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.03) (0.40) (-0.92)</td>
<td>(0.03) (0.87) (-0.49)</td>
<td></td>
</tr>
<tr>
<td>(-0.12) (0.91) (0.39)</td>
<td>(0.04) (0.48) 0.87</td>
<td></td>
</tr>
<tr>
<td>(0.99) (0.09) (0.07)</td>
<td>(-0.99) (0.05) (0.02)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Force – Motion</th>
<th>Traditional</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Density Matrix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05 0.03 0.08</td>
<td>0.04 0.03 0.05</td>
<td></td>
</tr>
<tr>
<td>0.03 0.27 0.22</td>
<td>0.03 0.46 0.25</td>
<td></td>
</tr>
<tr>
<td>0.08 0.22 0.68</td>
<td>0.05 0.25 0.50</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04 0.79 0.17</td>
<td>0.03 0.74 0.23</td>
<td></td>
</tr>
<tr>
<td>Eigenvectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.02) (0.40) (-0.92)</td>
<td>(0.01) (0.67) (-0.74)</td>
<td></td>
</tr>
<tr>
<td>(0.12) (0.91) (0.39)</td>
<td>(0.10) (0.73) 0.67</td>
<td></td>
</tr>
<tr>
<td>(-0.99) (0.12) (0.03)</td>
<td>(-0.99) (0.08) (0.06)</td>
<td></td>
</tr>
</tbody>
</table>

From table 2, we can see that the eigenvalues for the eigenvectors corresponding to the null models are very small. This indicates that most students use either the correct expert model or the incorrect naïve model (or both) and the model space defined from the qualitative research matches well with this population. In addition, the primary class model states (state 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 2, the student class model states on the force-motion concept are displayed on a model plot in figure 9. For each type of class, we plot the class' primary model states for both pre and post tests. The initial states of both types of classes are nearly the same and in the consistent model 2 region indicating that before instruction most students in the two classes consistently use the incorrect model on all the questions related to force-motion.

Figure 9. Model plot of student class model states on Force-Motion with FCI data from University of Maryland. For each type of class, we plotted, for pre and post results, the first class model states (states with f largest eigenvalues). The two arrows represent the shifts of the first model states for pre and post results of Tutorial and traditional classes.

After instruction, the model state of the tutorial class moves to the consistent model 1 region indicating that most students use the correct model rather consistently. On the other hand, the primary model state of the traditional class is mixed, which suggests that most students in the class are inconsistent in using their models. Since the model state is nearly a perfect mix (half to half), a student randomly selected from the class would have equal probability to use either the expert model or the naïve model on a question arbitrarily selected from this set of five questions on force-motion.

Assessing the Fine Details of Context Dependence: Effects of Specific Context Features

To further investigate the finer details of context dependence, we conducted an experiment at Kansas State University, in which we implemented a new measurement instrument that can isolate the effects of different contextual features of expert equivalent scenarios (Bao et al. 2002). This research demonstrates that context scenarios can involve multiple contextual aspects, which we refer to as physical features. These features of the contexts can be integrated in students’ reasoning as part of the knowledge or as cues to activate their knowledge. It has been observed that students can develop different types of models with different physical features. For example, in learning
Newton’s Third Law, for a given context scenario, students are usually attend to four types of physical features including the mass, the velocity, the object that is actively pushing, and the acceleration. Students can activate a naïve model when cued with one physical feature and an expert model when cued with another physical feature as if the physical features are independent variables. The changes of students’ models also show different processes with the different physical features.

To probe this phenomenon, we developed a multiple-choice instrument that isolates the different physical features in measuring the student model states. Existing instruments such as FCI often mix several physical features in a single question. For example, the two situations shown in figure 10 can be used to design questions to measure student understandings on Newton’s Third Laws. Students are often asked to compare the forces exerted by the two persons on each other. There are usually two naïve models involved: (1) the person who pushes exerts a larger force; (2) the person with larger mass exerts a larger force. The situation shown in figure 10 (a) is one from FCI test. It mixes the two physical features, mass and pushing, together. If a student answers item a) saying that Bob exerts a larger force, no further evidence can be obtained to infer if the incorrect response is generated based on consideration on mass, on pushing or on both.

![Figure 10. Comparison of question design on isolating context features.](image)

In our new multiple-choice instrument, each question only measures students’ reasoning related to a single physical feature of Newton’s Third Law. Figure 10 (b) is a modified version of the two persons pushing each other, in which both now push at the same time. If a student responds that Bob exerts a larger force, we then have evidence to infer that this student is using an incorrect model based on the physical feature of mass. Through this design, the question can isolate the physical feature of mass and allow the probing of the students’ models involving only mass. In order to measure the possible mixing of students’ use of their models, we designed three questions using different context settings for each of the four physical features, (Bao et al. 2002).

With the new instrument, we studied five introductory physics courses at Kansas State University (all taught with traditional lecture-recitation type instruction). These courses are:

- Physical World (PW), an conceptual physics course for non-science majors with no math pre-requisites;
- General Physics 1/2 (GP1/2), the first/second semester of a two-semester algebra-based physics course;
- Engineering Physics 1/2 (EP1/2), the first/second semester of a two-semester calculus-based course for physics and engineering majors.
Among these classes, only GP2 and EP2 had previously received instruction on mechanics, therefore, we treat the results of GP2 and EP2 as post-instructional data and the results from PW, GP1 and EP1 as pre-instructional data. In addition, the level of students’ preparation/background on physics knowledge can be described in an increasing order from PW \(\rightarrow\) GP1 \(\rightarrow\) EP1 for pre-instruction classes and from GP2 \(\rightarrow\) EP2 for post-instruction classes.

The results are shown in four model plots each representing the classes’ primary model states for one physical feature (see Figure 11). Each point on one model plot represents a class’s primary state (there are five points each plot representing five classes). The arrow in the diagram shows how the model states of different classes change as the level of the class increases.

![Figure 11. Student class model states on Newton III with the four physical features: Velocity (V), Mass (M), Pushing (P), and Acceleration (A). The data is taken from 5 introductory physics course at Kansas State University.](image)

From Figure 11, we can see that for the physical features of mass and velocity, the model states of all the classes stay in the consistent model 2 region. This indicates that most students have a dominant consistent incorrect model – “object with larger mass or velocity exerts a larger force.
during an interaction". The general direction of the changes of the model states is horizontal and towards 0. This suggests that the popularity of the incorrect model decreases with higher-level courses – from 90% (GP1) to 70% (EP2), but the model states stay in the model 2 region showing that most students in these five classes apply their models consistently, i.e., there is little mixed use of different models. The process can be interpreted as that a small fraction of students totally understand the concept and became very consistent in using an expert model. The remaining students, whose sizes would reduce as the level of the class increases, didn’t get it at all and were very consistent in using the naïve model.

Student model states with the physical feature of acceleration appear to be in the opposite situation where most students hold a consistent “correct” model considering acceleration irrelevant. In this case, the pattern of the changes in model states suggests that higher level students started to have less confidence on this model. The cause of this phenomenon is still being studied and our current results from interviews implies a possible explanation which suggests that the lower level students often respond to the questions with their dominant attentions on mass and velocity, and therefore treat acceleration as irrelevant. As students gain more knowledge from instruction on mechanics, they can get confused on the correct relations between acceleration, force and velocity and thus develop the incorrect model that treats acceleration as a factor to affect the magnitudes of forces during the interaction.

With the physical feature of pushing, student model states show a different structure. The low level classes still dominantly have a consistent incorrect model. As the level of class becomes more advanced, student model states become more mixed. The most advanced class (EP2) has nearly a perfectly mixed model state. This is very different from the situations with the other physical features and implies a different process in conceptual development. As clearly shown in the model plot, the three classes without instruction all stay in the region representing a consistent incorrect model. The two classes with instruction are in the mixed region. The shifts of the class model states suggests the process in which students start with consistent incorrect model and go through stages of mixed model states towards building a consistent correct model (e.g., see the previous example with tutorial instruction shown in Figure 9). This is a different process of conceptual development compared to the students’ models on other three physical features.

As recognized by many researchers, the stage of mixed model state is often an important intermediate step on the way to a complete favorable conceptual change. Further investigation into the causes of this phenomenon suggests that most students remembered the sentence “when you push something, you get pushed back” introduced by the instructors, and believed it since they can easily relate this to their personal experience in real world. When students see a situation involving pushing, they often get “reminded” to apply this piece of knowledge. It is also found that some students still have the tendency to think that the one who pushes exerts a larger force. So students can sometimes give contradictory answers on similar questions with pushing resulting in a mixed model state. With questions that do not explicitly involve the issue of pushing, students would look for mass or velocity immediately in their reasoning without even bother to think about the pushing. On those questions, many students are very consistent in using their naïve models (Bao 2002b).

From the research results using both qualitative and quantitative methods, we can infer a possible explanation for why student model states are different with the physical feature of pushing. It appears that “pushing” is often the first and the most common issue in examples used to introduce Newton’s Third Law. More importantly, most students all have the experience of being pushed back when they are pushing an object. Integrating this piece of student experience as examples in
instruction can make this side of the concept directly linked to students’ life experience and presumably more meaningful for them to make sense. Therefore, students can have significant changes of their models on this physical feature even with traditional instruction. On the other hand, students’ strong naïve models associated with mass and velocity often receive inadequate treatment (or ineffective treatment) through traditional instruction and students’ changes on their models with these physical features are small.

This example shows clearly the context dependence of learning with respects to the formation and activation of students’ prior knowledge and the formation of conceptual knowledge during the instruction. It is particularly interesting that for the same concept students can develop very different models for subsets of the context features and can also go through a completely different conceptual change process on different contextual features. This is strong evidence suggesting that we have to explicitly address the context issues in assessment and in instruction with extreme care and completeness.

VI. Cognitive Models and Treatments on Information from Measurement

The results we obtained using model analysis methods are often difficult to get from any 1-dimensional measurement variables such as a score. Further, our studies on context dependence suggest that the mixed state of students’ use of their knowledge is fundamental to the learning process and is part of the information needs to be assessed. Such information cannot be obtained through analysis that assumes consistency of student behaviors. Below, we show an example that compares model analysis with other methods such as factor analysis.

Factor analysis uses a correlation matrix to analyze the consistency of student behaviors in terms of covariance and correlations of student responses on different test items. Usually, the measurement variable is the score on an item. By analyzing the covariance or correlation matrix of students responses on test items, a set of “factors” can be obtained, which are often interpreted as representing certain latent abilities or conceptual constructs possessed by the students. There are many concerns about this method. For example, the factors are non-unique and can be rotated, which can cause arguments on what the factors actually represent. Here we emphasize the connection to the underlying cognitive model. If one uses factor analysis or similar methods that attempt to extract hidden variables from analyzing correlations of student behaviors, it is implicitly assumed that hidden variables should lead to a consistency in student behaviors in given contexts, which produces the correlation.

To show how factor analysis may fail to produce meaningful results, we reference a study conducted by Huffman and Heller (1995) who used factor analysis to explore whether the results of the FCI can reveal any factors and patterns along the six conceptual dimensions based on which the FCI was designed (Hestenes, Wells and Swackhamer 1992). After performing a factor analysis on a data set from nearly 1000 students, no factors could be extracted to match the specific conceptual dimensions. They concluded that “the items appear to be only loosely related to each other” and that the FCI “actually measures bits and pieces of students’ knowledge that do not necessarily form a coherent force concept.” In our view, this research result is strong evidence showing that the activation of conceptual knowledge is context dependent and thus causing inconsistency of students’ responses on expert-equivalent items. Therefore, the correlations between students’ responses on a cluster of questions designed to measure a single concept dimension can be insignificant unless the students are all in pure model states.
With the same data used in the above examples, we calculated the correlation matrices for four questions on the force-motion concept (see table 3). We used only the data from the classes under traditional instruction. The correlations on the pre-test are somewhat higher than on the post-test. This makes sense since model analysis shows (see Fig. 9) that most students appear to have a pure naïve model state on pre-test and mixed states on the post-test. With this kind of correlations, it is very hard to extract any meaningful factors.

Table 3. Correlation matrix of student scores on four questions related to the force-motion concept (traditional instruction only)

<table>
<thead>
<tr>
<th></th>
<th>Pre-instruction</th>
<th>Post-Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R  Q5  Q9  Q18  Q28</td>
<td>R  Q5  Q9  Q18  Q28</td>
</tr>
<tr>
<td>R</td>
<td>1    0.64  0.47  0.42</td>
<td>1    0.51  0.31  0.33</td>
</tr>
<tr>
<td>Q5</td>
<td>0.64  1    0.54  0.55</td>
<td>Q5    0.51  1    0.38  0.34</td>
</tr>
<tr>
<td>Q9</td>
<td>0.47  0.54  1    0.51</td>
<td>Q18   0.31  0.38  1    0.44</td>
</tr>
<tr>
<td>Q18</td>
<td>0.42  0.55  0.51  1</td>
<td>Q28   0.33  0.34  0.44  1</td>
</tr>
<tr>
<td>Q28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Then we do the analog of a model analysis: we would like to study how the students use both the expert model and the naïve model, since the two models both describe rational and stable reasoning patterns about the concept dimension. Here the null model is treated as pieces and bits of rather random ideas that are less rational and less stable. In the calculation, if we find a student giving a response that reflects either the naïve model or the expert model, we assign a model score of 1 to the student. Similarly, if a student uses the null model, we assign 0 to the students. Then we calculate the correlations between students' model scores on the four questions. The correlation matrices are shown in table 4.

Table 4. Correlation matrix of student model scores on four questions related to the force-motion concept (traditional instruction only)

<table>
<thead>
<tr>
<th></th>
<th>Pre-instruction</th>
<th>Post-Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R  Q5  Q9  Q18  Q28</td>
<td>R  Q5  Q9  Q18  Q28</td>
</tr>
<tr>
<td>R</td>
<td>1    0.97  0.86  0.74</td>
<td>1    0.96  0.86  0.9</td>
</tr>
<tr>
<td>Q5</td>
<td>0.97  1    0.82  0.7</td>
<td>Q5    0.96  1    0.83  0.89</td>
</tr>
<tr>
<td>Q9</td>
<td>0.86  0.82  1    0.67</td>
<td>Q18   0.86  0.83  1    0.78</td>
</tr>
<tr>
<td>Q18</td>
<td>0.74  0.7   0.67  1</td>
<td>Q28   0.9    0.89  0.78  1</td>
</tr>
<tr>
<td>Q28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This shows much stronger correlations between the model scores on the four questions, which indicates that most students used either the naïve model or the expert model on the set of questions. Based on such correlations, one would get one strong factor showing commonalities between the four questions (see table 5).

From many examples we have studied, we find that the use of null models on the FCI questions is less than 5%. Therefore, we can conclude that the majority of students we studied would use either the expert model or the naïve model on the two clusters of FCI questions we studied. When many individual students are inconsistent in using the different models on the set of questions related to a single concept as a result of context dependence, low correlations between student
scores on the questions will result. In this case, using factor analysis will not produce meaningful results.

Table 5. Factor loading obtained from student model scores on four questions related to the force-motion concept (traditional instruction only)

<table>
<thead>
<tr>
<th></th>
<th>Pre-instruction</th>
<th>Post-Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loading</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Q5</td>
<td>0.97</td>
<td>-0.14</td>
</tr>
<tr>
<td>Q9</td>
<td>0.95</td>
<td>-0.17</td>
</tr>
<tr>
<td>Q18</td>
<td>0.91</td>
<td>-0.18</td>
</tr>
<tr>
<td>Q28</td>
<td>0.84</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Another popular method in education assessment is the Item Response Theory (IRT), which, in contrast to correlation analysis, analyzes students’ full responses on specific items. This method assumes that students possess certain abilities that are latent and can affect the probabilities of students’ responding to test items. Thus, one can establish causal relations between the assumed latent abilities and the features of the test items (often in terms of difficulty and discrimination parameters). Estimations of the latent variables and the adjustment of the parameters are performed from fitting the measurement data to the assumed causal relation between the latent variables and the students’ responses.

Evidently, both factor analysis and IRT methods need to assume a stable causal relation between certain latent variable of the students and the students’ responses to items. Further, it is assumed that the variance of this relation is only caused by random errors which can be treated by ensemble measurements. In addition, the IRT ignores specific interactions between student knowledge states (both long and short term) and specific items, such as cueing.

However, neither assumption may be appropriate, depending on the types of knowledge being measured. As evident from the above discussions on context issues, there is sometimes a strong context dependence in students’ use of their conceptual knowledge, which can cause inconsistency in their responses to expert equivalent questions. This inconsistency, which is measured in terms of mixed model states, is related to a student’s general abilities but such relations contain large uncertainties caused by unknown context issues. That is, at a given value of ability, there is a range of probabilities rather than a unique value for a student to respond with a particular type of answer to a question, depending on how the question interacts with specific items in the student’s knowledge structure. In addition, the probability in an IRT model is interpreted as the likelihood of ability or inability to produce a particular type of response, which is of an unknown (or implicitly expressed) random origin. Based on the context dependence, one can expect that such probabilities in producing different types of responses are in fact supported by rational reasoning that is activated by specific features of the contexts. Therefore, the response function used in an IRT fitting should produce, at a given value of ability, multiple values of probabilities rather than a single value which is the form being currently used. Since the variance of the response probability is not caused by random errors, it cannot be removed by ensemble measurements. In fact, this variance represents important information on students’ mixed model states (or the inconsistency of their behaviors), which needs to be extracted and studied. The IRT can therefore only provide broad averaged information about a student’s knowledge. The kind of detailed match between problem cues (such
as the physical features discussed above) and student knowledge discussed above, would be missed by an IRT analysis using one ability variable.

Further discussions on IRT are beyond the scope of this paper. In general, the IRT model works better in measuring knowledge and abilities that are more context general so that different items designed to measure such abilities will not be affected significantly by the specific context features of these items and can provide good fittings for the assumed causal relations. For examples, the definition of “difficulty” level of an item is often meaningless when the context features of the item can significantly affect an individual in activating a set of coexisting conceptual knowledge. Therefore, one may consider two general types of measurement and methods when context dependence is considered: 1) Use model analysis to identify the probability distribution of the context dependent activations of students’ knowledge; 2) Then further identify any possible context general knowledge which is less affected by the specific context settings and use IRT type of methods to analyze such knowledge.

VII. Values of Measuring Mixed Model States

Empirical Observations of Students’ Conceptual Development Process

Based on the literature and our own investigations on student models of physics (Thornton 1994, Maloney & Siegler 1993, Bao & Redish 2001, Bao et al. 2002), we can put together a picture to characterize the different stages of conceptual development in terms of the situations of students’ use of models. It is suggested that in learning a particular concept topic, a student’s learning can be described with five typical stages shown in Figure 12 (Bao 2002b). The process represents a general progressive learning path and a student can begin and end with any one of the stages.

![Figure 12. Stages of conceptual development](image)

The first is the pre-naïve stage, which represents the situation where students by themselves won’t even be able to develop/posses a logically consistent model of the concept topic. Many examples of this situation have been observed in our interviews, where students fail to provide any logically-sound reasoning for a physical phenomenon. Many students just can’t provide reasoning of any kind. Some students attempt to put together reasoning but they usually fail to recognize the relevant variables from the context and don’t have appropriate causal relations in their reasoning –

Bao & Redish

Page 32
e.g. a result can also be the cause of itself. Basically, these students have a very limited logical manipulation ability. Students in this stage can also be measured with model-based multiple-choice instrument, which will show a high probability on the null model dimension (an example of this is shown in Bao et al. 2002).

The second is the naïve model stage. In this case students consistently hold one or several models of the concept topic. These models are often developed by students based on their real-life experience and previous school learning. While these models may have a similar level of coherence and robustness as an expert’s model, they can produce correct results in a limited set of cases and may not interface well with scientific models.

The third and forth stages correspond to the two mixed model states – the implicit and explicit mixed states. The fifth stage is the expert model state in which case the students are able to apply the expert models of the concept topic consistently in a wide variety of appropriate context settings. Starting from the third stage, it is possible to bring students into explicit awareness of conflicts of the different models with carefully prepared context settings. Also starting from the third stage, students can develop hybrid models as a result of instruction.

From many examples in the literature (McDermott & Redish, 1999), it appears that an average undergraduate population in an introductory physics course often starts with the naïve model state and ends up somewhere in the mixed stages. Our previous research indicates that with traditional instruction, students often move to the middle of the mixed stages, and with research-based instructions (e.g. Tutorials), students can move very close to the expert state.

**Possible Applications in Instruction**

As indicated by many studies on student conceptual change process, when students hold naïve models, effective teaching strategies often require bringing students into explicit awareness of the conflicts of their existing naïve models (Posner, Strike, Hewson & Gertzog 1982, McDermott & Shaffer 1998). Therefore, the measurement of what types of context settings can cue multiple models and the explicit awareness of conflicts can play a valuable role in instruction. It provides useful information on both the students’ states of understanding and on if (or which ones of) the context scenarios of the questions are appropriate to be incorporated in teaching to help students. The possible instructional applications of the results of such measurements are summarized below (Bao 2002b)

1. When students are measured as in the pre-naïve stage, the first task of the instruction is often to help students recognize the relevant variables of the contexts and understand the basic causal relations that the students will need to manipulate the variables and to put together some reasoning, i.e., to help them make observations of the basic characteristics of the physical phenomenon. At this stage, since many students’ reasoning capabilities are quite “fragile”, encouragement and facilitation on students’ development of the ability to do logical reasoning (not necessarily the correct types) is often the major theme of the instruction. Activities to confront students’ incorrect ideas should be carefully controlled to avoid causing frustrations among students.

2. If pure naïve model is detected, the instructor can begin with helping students develop some basic understanding of the expert model and or confronting students’ naïve models. In this stage, strong confrontation of students’ existing models can cue students’ desire to change their current understandings; however, since the students don’t have much of the correct model yet, it
is often difficult for them to quickly get into the correct direction and thus can cause significant frustrations. Therefore, it may work better if the instruction first helps students learn some new ideas in reasoning without being troubled by any contradiction. Later, the instruction can guide the students to re-interpret their existing examples (including the ones used in the questions) and see the conflicts of using their old models.

(3) If implicit mixed model state is detected, it shows that the students have started to understand the expert model but such understandings are often strongly tied to limited contexts and students may treat different expert-equivalent contexts as unrelated situations (context dependent fragmentation of knowledge). It also implies that the context scenarios of the questions won’t spontaneously activate explicit awareness of multiple models and the conflicts between these models. To use these context scenarios in instruction, it is often necessary to use several questions that each may cue a different model and guide the students to realize the conflicts by comparing their reasoning on the different questions.

(4) If explicit mixed model state is measured (Bao 2002b), it reveals that the students have further developed understanding of the concept topic and start to recognize the relevance and applicability of the expert models in a wider range context settings (usually they start to recognize the commonalities among the set of expert-equivalent context scenarios and may actively seeking consistency over a range of contexts); however, they still keep their naïve models at the same time. These students are the most readily prepared ones to explicitly work out the conflicts between the different models. Since the context scenarios used in the measurement can simultaneously cue multiple models, they can be used directly as examples in teaching. For instructional purposes, these context scenarios may need some modification so that they not only bring up multiple models but also activate students into explicit awareness of the conflicts among those models.

(5) As a related issue, we may found students developing hybrid models during instruction. In this situation, it is important to identify the specific contexts in which the hybrid models are developed. Since a hybrid model is often locally consistent within the contexts in which it is developed, to create cognitive conflicts it is necessary to prepare different context settings that the students can recognize as equivalent and that the application of the hybrid model will produce contradictory results.

In the 3rd, 4th and 5th cases, since the students already know the correct model (or parts of the correct model), the instruction can more directly confront the students’ naïve models and guide them to apply the correct model under different contexts. Regardless what states the students are in, the instruction should always involve as many real-life examples as possible (since most of the students have rich experience and also incorrect interpretations with such examples), and guide the students to re-interpret these examples using the correct models developed in learning. Without a rigorous re-interpretation process, it is often difficult for students to develop a robust understanding of the concept (rather students can develop fragmented knowledge strongly tied to specific contexts). That is, the features of the contexts that may cue a student into using naïve models need to be systematically addressed so that they would consistently cue a student into using the correct model. Similarly, the instruments used in the measurement should also involve diverse context settings to ensure a more complete evaluation of students’ understanding.
VIII. Summary

In this paper, we take a theoretical perspective built on context dependence of learning to interpret the student learning behaviors. Based on this framework, we find explanations for a range of phenomena observed in research, which include students’ misconceptions and preconceptions, students’ creations of alternative conceptions, fragmentation of students’ knowledge, difficulties in transfer of students’ knowledge, and inconsistency in students’ use of their knowledge in equivalent contexts. These are important issues need to be explicitly addressed in assessment and instruction. Built on the theoretical understanding of the context dependence issue, we developed model analysis, which combines both qualitative and quantitative methods to form a measurement representation and numerical methods for conducting quantitative assessment on the probabilities for students to apply specific pieces of knowledge and the ways that students use the knowledge. In particular, we are now able to quantitatively measure and evaluate the inconsistency of student behaviors in terms of mixed model states. Such inconsistency often causes difficulties for methods such as factor analysis that relies on the existence of correlations in student behaviors. As shown by research, the mixed states represent crucial information about the conceptual development process, which needs to be included in the assessment as part of the signal rather than random errors. The assessment of the mixed states can also provide direct guidance for instruction.

Acknowledgement:

The authors would thank the members of the research groups at The Ohio State University and at the University of Maryland. The related research is supported in part by the NSF grants REC-0087788, REC-0126070, and REC-0087519.

This paper is one chapter of the book to be published by the Indiana University.

The Scholarship Of Teaching And Learning In Post-Secondary Education: The Contributions Of Research Universities

Research Universities and the Scholarship of Teaching and Learning

William E. Becker
Professor of Economics, and Editor
The Journal of Economic Education
Indiana University

Moya L. Andrews
Professor of Speech and Hearing Sciences, and Vice Chancellor and Dean of Faculty
Indiana University
Reference:


