Chapter 4: Model Analysis Algorithms II: Model Estimation

Introduction

Student mental model states are not directly measurable. The only observable data that we can get from an experiment is student responses in a variety of environments including interviews and different problem-solving situations. According to our model of the learning process, each question is considered as an instance of a physics context and should trigger students for associations to certain models. The student responses are considered as the outcome of the students' applying these models into the physics contexts. Using student responses as input, with proper algorithms, we can obtain quantitative information of students using their models.

In PER, the instruments most commonly used to study student understandings include free response questions, interviews, and multiple-choice questions. Free response questions and interviews often provide very rich information on student understandings. But they are also very difficult to analyze and expensive to conduct. Multiple choice tests have the advantage of being easy to analyze and cheap to conduct in large scale. But the traditional way of analyzing scores doesn't provide much useful information especially on student real understandings of physics.

The goal of this chapter is to develop an algorithm/tool to do numerical analysis of student data on multiple choice tests to obtain quantitative evaluations of student models. In order to do so, it is important to understand the general behavior of our students. As indicated by PER, for a same physical concept, a single student can have

- a dominant model (not necessarily a correct one), which is used consistently through out various questions related to a same concept
- a mixed model state where the student uses several models (correct and incorrect ones) in an inconsistent way (slightly different questions on the same concept can trigger different models)
- or no model at all, which often lead the student to give a response based on random guessing (no serious logical reasoning involved).

In general, an average student in our introductory physics class is often found to be in a mixed model state where different models are used inconsistently.

Consider the case where we have a whole class of students and each student has a unique model state (consistent and inconsistent, correct and incorrect). Then the student responses are generated by a large number of different model operations (triggering and applying) with all sorts of models (correct, incorrect, mixed and random). Therefore, the results can be very complicated and a simple analysis using scores alone often fail to provide useful information on the student real understanding of the physics concept. For example, a low score can be caused by a consistent incorrect model, random guessing, or a temporarily triggered incorrect model in a mixed model state. These different situations reflect important information on student understanding of physics. But none of these can be obtained with analysis based solely on scores.

In education research, factor analysis is a very popular tool that researchers often use to study the possible models underlying student answers on a test.[1] It calculates the correlation between student scores on different questions and uses this correlation (which measures the consistency of student scores on different questions) to infer possible "factors" (models) from the data. For more details on factor analysis, please see related references in the end of this chapter. In the later part of this chapter, I will give a more detailed comparison between factor analysis and model analysis.

A major disadvantage of factor analysis is that it only evaluates the consistency of student scores and doesn't look at the systematics of incorrect answers. Underlying this is the assumption that the individual student is consistent with their models. For example, suppose we give students a test containing two groups of questions, one group on Newton III and one group on Force Motion. If all students have consistent models related to the two physics concepts, the correlation (consistency) of students' scores on the questions within same group should be higher than the correlation of scores on questions from different groups. In this case, factor analysis will give two factors, which can be inferred as two "models" that students use consistently to solve problems in the two groups.

As we can see, factor analysis does not consider the possibility that the students may each possess more than one model and may be inconsistent in using these models. With the above example, if the students have more than one model on Newton III and/or on Force Motion and are inconsistent, their scores will show a great deal of randomness and no factors will be obtained. This result has been demonstrated in the literature (a good explanation can be implied from the discussion in this chapter).[2] A more detailed analysis on the factor analysis is discussed in later part of this chapter.

From the above example, we can also see that factor analysis does not provide explicit meanings on the nature of the inferred "factors". In fact each of the two factors only implies a broadly defined "model" that affects the consistency of student scores. This type of "model" does not give the operational details on how a student may use it to generate a response. Therefore, the "models" corresponding to the factors from factor analysis are fundamentally different from the student models we are talking about. According to our understanding of learning, the actual process to generate consistent/inconsistent scores is very complicated and can not be described by the results from a factor analysis.

For example, consider the case that one group of questions on a single physics concept is given to a class of students. Suppose each individual student has more than one model and is inconsistent in using these models (see the random process in model triggering described in chapter 2). Since the individual student may use different models on questions related to a single physics concept, the correlation between scores is low and no dominant factors can be identified. This implies that the students' behaviors are random (sometimes giving correct answers and sometimes giving incorrect ones with no systematic relation. But different types of randomness are involved). Students with inconsistent scores can be consistent on giving incorrect responses of the same type. We may observe student responses switching between correct answers and incorrect ones that can be associated with a single incorrect model. This type of random process happens in model triggering and is fundamentally different from the random behavior when students generate answers by guessing (no model involved), or by human errors (mistakes). With no information on the incorrect responses, factor analysis won't work with such model-based random processes.

Now consider the case where each individual student is consistent in using a model, i.e., in the class, some of the students use one correct model consistently while others use a second (incorrect) model consistently. In this case, student scores will have very high correlation/consistency on all the questions (either all correct or all wrong depending on the models used). Then factor analysis will identify one factor from these data, which reflects the consistency of student scores on this set of questions. But still this factor does not have any information on why students are consistently giving correct or incorrect answers. All we can tell is that there might be one *hidden factor* underlying the student scores on this set of questions so that the answers from the individual students will either be correct or incorrect to all questions.

What makes the students behave like this is not explicitly reflected by the result. In fact, in this case, factor analysis implies a wrong story – only one "factor" in student data rather than that that students are consistently using different models in generating their responses. With this example, even in cases where students all have consistent models, factor analysis fails to give a correct description of student modeling situations. In addition, suppose each question is designed with a substantial number of choices, which makes a small probability on guessing out a correct answer. In this situation, factor analysis will again fail to identify the difference between students who always get the same type of incorrect answers, which indicates a coherent incorrect model, and students who get all different kinds of incorrect answers, which is apparently by random guessing (no coherent model). The information of students using their incorrect models is embedded in the incorrect answers chosen by the students. Consistent incorrect models will result in consistent incorrect answers and we have to look for this information in order to decide the real story about our students. Since factor analysis uses only scores, it can't provide information on how and why the students are doing wrong.

As a short summary for the above discussions, since factor analysis is based on scores, there is no way for it to distinguish the different model operations among students. The results of factor analysis only give the information on the consistency of student correct scores on different questions, and this is not enough to extract the real structure of student models.

More sophisticated versions of factor analysis often include a student ability term to account for possible student "latent" models.[3] But again information on student "models" is implicit and derived solely on the correlation of the scores, which contains no information on student incorrect responses.[4] Therefore, factor analysis, even the modified versions of it, doesn't provide explicit information on what makes students produce high/low scores, what makes a question be easy/difficult, or what makes a student perform consistently/randomly on similar questions related to a single concept.

Therefore, to understand student modeling with multiple-choice data, we need to analyze the complete structural information of responses from each student. R. Thornton has conducted research using students' incorrect responses to study their models where student mixed model state is characterized as a "transitional student view".[5]

In data processing, if the results are based on calculating the total number of students holding different types of models, then the results only contain the proportion of students inferred as having different models by their responses and the structural information for individual student responses is lost during the averaging process. As a result, the possibility that individual student can have multiple models at the same time is not fully represented, and the structural information of the individual student model state is not retained.

In this chapter, I introduce a new method to analyze the structural information of the individual student's responses and to obtain quantitative evaluations of student model states including detailed information on students use of mixed models. Since the development of this method depends heavily on our understanding of student models, and the results also provide explicit information on student model states, this tool is called *Model Analysis*.

The basic approach is to study the information contained in all the student responses rather than just the correct ones. The students' full response patterns are analyzed and stored in a model density matrix that retains part of the structural information of the individual student responses. This model density matrix is then analyzed to extract information on student models.

In the following sections, I will first introduce ways to identify students' models based on their responses on a multiple-choice test. With this modeling process, student raw responses are transformed into model-based responses. Using these model-based responses, a model density matrix (for a class) can be constructed, which preserves important information on individual student model state. Then the density matrix is decomposed to obtain eigenvalues and eigenvectors. These eigenvectors can be used to represent the common student model states and the eigenvalues give the prevalence of the respective models. Different ways to present and analyze the model states will be discussed in detail. In the final part of the chapter, I will use this method to analyze FCI data and discuss the results. This method can be used to investigate a single student model state as well as the overall model states of a class.[6] Due to the availability of test data, the method developed in this chapter is focused on evaluating the model states for a class.

Identifying the Physical Models

Before we can make measurements on student models, we need to know the physical models associated with the topics of study. Then we can design appropriate instruments such as multiple-choice concept tests to probe students' use of these models. Therefore, the first step to do model analysis is always a systematic and detailed investigation of student difficulties on a physics concept. In these researches, we often give students free-response conceptual questions to probe their general understandings of the physical concepts and conduct interviews to obtain more details of their understanding. The interviews can also be used to confirm and elaborate what we have learned from the student responses on short answer questions.

Based on these detailed studies, we can define the set of physical models as those found to be common to the students. Each set of physical models usually consists of one correct model, a few incorrect ones and a null model to account for uncertain student behaviors.[7] Different physics concept domains may have different sets of physical models associated with them and the structure of these models may also be affected by the background of the students. As often observed in research, students models may involve not only misuse of physics principles but also inappropriate understandings of the physics concepts completely different from those traditionally used by physicists. For example, for a physicist, velocity and acceleration represent two distinct concepts. But for some students, these are combined in a single undifferentiated concept of "motion". In other cases, students may differentiate a single physics concept into two parts. For example, a student may decide that some moving objects "want to stop" (a box sliding on a rough floor), while others (a thrown baseball) "want to keep going". A physicist would classify both the same but say that one (the box) is acted on by an outside force (friction). Therefore, when we study student models, we have to be very specific about both the physics concept and the background of the students. We also have to exercise considerable care and effort in determining and selecting common student models.

Once the physical models are determined and confirmed by research, we can start to design multiple-choice questions that can be used to study students' use of these models. Not all types of multiple choice questions are appropriate to use with model analysis. The questions have to be carefully designed so that the choices of the questions represent attractive results corresponding to the common student models. Each physical model needs at least one choice to represent it. The design of these choices should also avoid situations where students with different physical models can legitimately generate the same choice. To reduce the measurement uncertainty, chances of random errors, and incorrect model assignment to students who do not hold the model, it is also necessary to include a significant number of choices in each question. All questions need to be tested and confirmed by research to make sure that the results one is getting from the tests agree with the interview results.

Based on these criteria, questions designed to intentionally eliminate the attractive distracters are not appropriate. On such questions, students unable to generate the correct responses will be forced to make a guess or to use "exam skills". Thus, the results do not contain accurate information on why students get the questions wrong.

The FCI questions are developed based on research and the distracters are designed to represent the incorrect models found in interviews. Usually, there are several concept groups in a test. With the FCI test, I will focus on the questions that probe student understanding on two physical concepts, Newton III (N3) and Force-Motion (FM). For simplicity, the FM group is used as the example in developing the algorithm.

The FCI test has 5 questions in the FM group (5, 9, 18, 22 and 28, see chapter 2 and Appendix A for details). As discussed in chapter 2, student responses to the five FCI questions involve three physical models:

Model 1: It is not necessary to have a force to maintain motion and there is no such thing as a "force in the direction of motion". (Correct)

Model 2: A force is needed to maintain motion. This model also includes the ideas that there is always a force in the direction of motion and that the force is directly related to the velocity of motion. (Incorrect)

Model 3: Other ideas and incomplete answers. (A null model)

These physical models represent different types of reasonings and we can represent them with three orthogonal vectors, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ defined as the physical model vectors, which also span the model space (see chapter 2 for details). For convenience, the graphical representation of a 3-D model space (figure 2-4) is reproduced in figure 4-1.

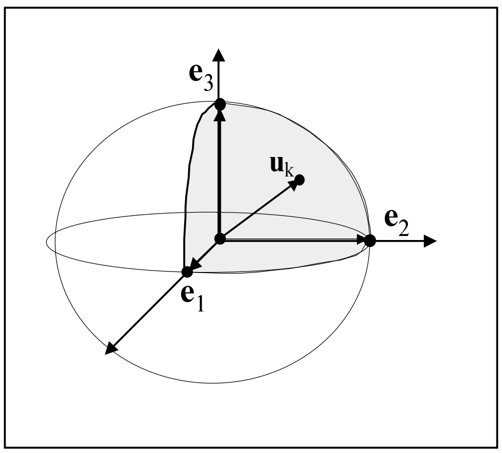

Figure 4-1. A 3-D model space.

The student models can be analyzed in two ways. One is to model the whole class for the class's model state. The other is to model an individual student to find the single student's model state. Since the resolution of the estimated model state depends on the sample size and there are only 5 questions but 3 physical models in the FM group, the uncertainty for an estimated single student model state will be very large. As discussed in chapter 2, this uncertainty is estimated with a solid angle. According to Eq. 2-5, with 5 measurements (5 questions), the uncertainty for the estimated student model vector is $\pi/10$, i.e., an estimated single student model vector can be in an uncertain direction within a solid angle equal to $\pi/10$.

With a class, since there are many students as input, the uncertainty of the estimated class model state will be much smaller. In the following discussion, I focus mostly on analyzing the class model state.

Modeling Responses of a Single Student

For the three physical models in FM group, the associations of the responses corresponding to the FCI questions in FM cluster are listed in table 4-1.

Table 4-1. Modeling of student responses for FM group

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

Then the student raw responses are collapsed into model-based responses. Using table 4-1, each student's response for each question can be modeled by one of the base vectors ($\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$) in the physical model space. For example, if a student answers the five questions with "a", "d", "a", "d" and "b", the five responses are transformed to four vectors as $(010)^T$, $(100)^T$, $(010)^T$, $(100)^T$ and $(001)^T$ respectively. The five vectors are then summed up to get an overall model response vector for that student, which in this case is found to be $(221)^T$. Define the model response vector of the k^{th} student as $\mathbf{r}_k$. Assuming we have a total of N students in a class, $\mathbf{r}_k$ can be represented as:

$$\mathbf{r}_k = \begin{pmatrix} n_{1k} \\ n_{2k} \\ n_{3k} \end{pmatrix} \qquad k = 1, 2, \ldots, N \qquad (4\text{-}1)$$

where $n_{\eta k}$ is the number of questions in the concept group that the k^{th} student answers using the η^{th} physical model. Obviously,

$$n_{1k} + n_{2k} + n_{3k} = m \qquad (4\text{-}2)$$

where m is the total number of questions in the group (5 in this case).

Error Analysis on Item-Based Modeling (IbM)

The above modeling procedure is an item-based modeling (IbM) process. That is, when we detect a response from a student, the corresponding physical model is assigned as the student's model state in that instance. But there are many random process involved. A student can generate a response by random guessing or by incorrectly applying a model unrelated to the response. Therefore, we have to analyze the uncertainty of this modeling process. The following is a discussion on the mathematical analysis of the uncertainty using item-based modeling method. We will see that in most situations, such uncertainty does not significantly degrade the results. Therefore, although I show the mathematics as how one may handle the problem, it is not applied in later analysis.

To simplify the discussion, I use a 3-model example similar to the FM case discussed previously. Assume we have a group of "m" questions with 3 physical models denoted as M_1, M_2, and M_3. M_3 is defined as the random null model. Let each question have "L" number of different choices. Define Q_1, Q_2, and Q_3 as the three types of responses corresponding to the three physical models. These represent the types of responses not the real response itself. For example, a question can be designed with multiple choices corresponding to one model and all these responses will be characterized with the same Q_η for that model. Define q_1, q_2, and q_3 as the probability for a student to be triggered into the corresponding physical model states. Obviously, we can write

$$q_1 + q_2 + q_3 = 1$$

That is, a student has to be triggered into one of the physical model states (since one of the categories is "none of the above models"). When a student is triggered into a certain physical model state, it is still not guaranteed that this student will generate a response corresponding to this physical model. We can define a set of *ability variables*, α_η (η = 1, 2, 3), as the probability for a student using model η to correctly generate the answers corresponding to that model. This probability can be measured with interviews where students solve a problem aloud. Since students often have more experience with their old models than with the new ones, the abilities for applying the different physical models may be different. For the random model, the ability to guess a response is considered to be 1.

In the context of the N3 and FM examples, the logical reasoning and mathematical operations involved in both the correct and incorrect physical models are similar. Therefore, for simplicity, a single ability variable, α, is assumed for all the non-random physical models.

Suppose we detect a response of Q_1. What are the probabilities for the student to be in any of the different physical model states? For simplicity, in this example, it is assumed that Q1 and Q2 each consists of 1 choice and the random response Q3 consists of "h" random choices, where L = h + 2.

Define $p_{\eta\mu}$ as the probability of a student in η^{th} physical model state (M_η) to generate a μ^{th} type of response (Q_μ). Then the $p_{\eta 1}$'s can be found as in table 4-2, where

$$\beta_1 = \alpha + (1-\alpha)\frac{1}{L}, \quad \beta_2 = (1-\alpha)\frac{1}{L}, \quad \beta_3 = \frac{1}{L}.$$

Table 4-2. The probability for a student to generate Q_1 with different M_η's

$p_{11} = q_1\alpha + q_1(1-\alpha)\frac{1}{L} = q_1\beta_1$	The first term describes the probability of a student triggered into M_1 generates a Q_1. The second term represents that a student with M_1 generates a response other than Q_1 but still picked Q_1.
$p_{21} = q_2(1-\alpha)\frac{1}{L} = q_2\beta_2$	The student triggered into M2 but unable to generate Q2 and picked Q1.
$p_{31} = q_3\frac{1}{L} = q_3\beta_3$	The student triggered into a random model and guessed Q1.

A detailed schematic for p_{11} is shown in figure 4-2.[8]

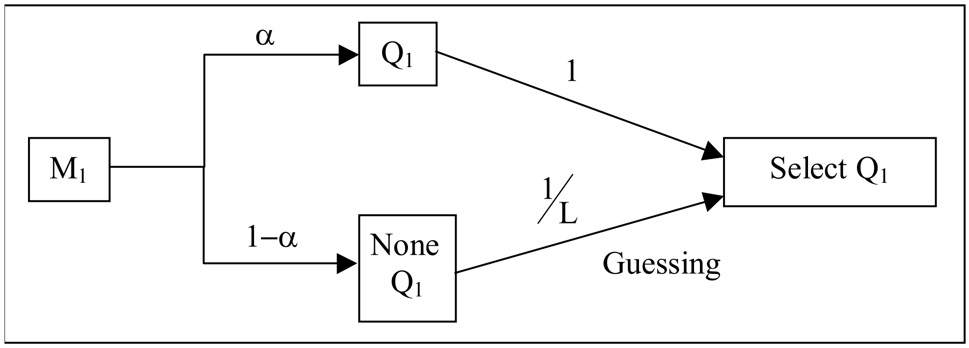

Figure 4-2. Schematic for p_{11}

As we can see, it is possible for a student in different physical model states to generate a same response but with different probabilities. Once a definite response is detected, the inferential probabilities for a student to be in either of the physical model states are directly related to the $p_{\eta\mu}$'s. For example, if p_{11} is much larger than the other $p_{\eta 1}$'s, when a Q_1 is detected, the student will have a much larger probability to be in a M_1 state.

Then define $P_{\eta\mu}$ as the inferential probabilities of a student that have generated a μ^{th} type of response but is in η^{th} physical model state. The $P_{\eta\mu}$ can be found as

$$P_{\eta\mu} = \frac{p_{\eta\mu}}{\sum_{\eta=1}^{w} p_{\eta\mu}}$$

where w is the total number of physical models. With this example, if a Q1 is detected, the probabilities for the student in different physical model states are

$$P_{11} = \frac{p_{11}}{p_{11} + p_{21} + p_{31}}, \quad P_{21} = \frac{p_{21}}{p_{11} + p_{21} + p_{31}}, \quad P_{31} = \frac{p_{31}}{p_{11} + p_{21} + p_{31}}$$

Obviously $\sum_{\eta=1}^{w} P_{\eta\mu} = 1$

That is, a student has to be in one of the physical model states (which explains why a random model is always needed). Similarly, we can find the $p_{\eta\mu}$'s for other responses (μ= 2, 3) as in table 4-3

Table 4-3. The probability of a student to generate Q2 and Q3 with different M_η's

Q2 is detected	Q3 is detected*
$p_{12} = q_1(1-\alpha)\dfrac{1}{L} = q_1\beta_2$	$p_{13} = q_1(1-\alpha)\dfrac{h}{L} = q_1\beta_2 h$
$p_{22} = q_2\alpha + q_2(1-\alpha)\dfrac{1}{L} = q_2\beta_1$	$p_{23} = q_2(1-\alpha)\dfrac{h}{L} = q_2\beta_2 h$
$p_{32} = q_3\dfrac{1}{L} = q_2\beta_3$	$p_{33} = q_3\dfrac{h}{L} = q_3\beta_3 h$

*That the h is included in $p_{\eta 3}$ is because the probability to guess one of the h random responses from the total L choices is h/L.

The process of a student generating a response can be represented as in figure 4-3.

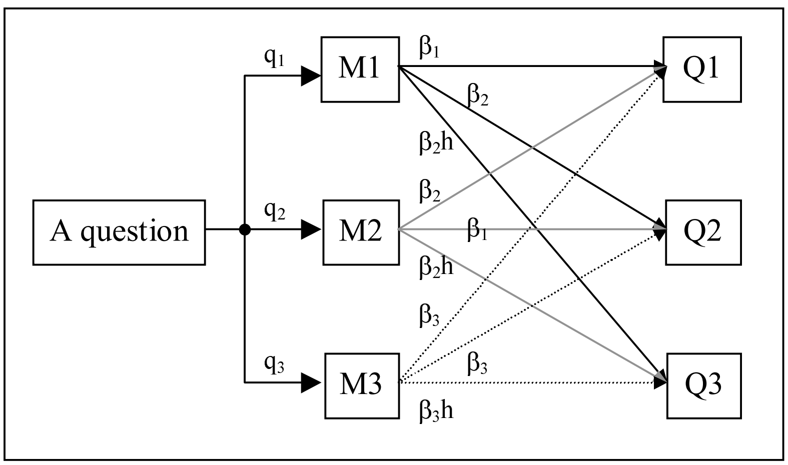

Figure 4-3. A three-model example of possible paths for a student to generate a response

Based on the $p_{\eta\mu}$'s, we can calculate the $P_{\eta\mu}$'s for all the different situations. The final results are put in a matrix form as:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

This **P** is defined as the *model-reliability matrix* of a test. As an example for an average student, suppose we have

$\alpha = 0.8$, $q_1 = 0.5$, $q_2 = 0.4$, $q_3 = 0.1$, $L = 5$, and $h = 3$

The reliability matrix for this example $\mathbf{P}_A$ is calculated to be

$$\mathbf{P}_A = \begin{bmatrix} 0.92 & 0.05 & 0.36 \\ 0.04 & 0.90 & 0.28 \\ 0.04 & 0.05 & 0.36 \end{bmatrix}$$

The diagonal elements represent the probabilities of correct matches between responses and student model states while the off-diagonal elements represent the "cross-talk" from mismatched model states and responses. Therefore, larger diagonal elements indicate higher reliability for the corresponding item-based modeling schemes. With this example, the modeling for M_1 and M_2 is comparatively reliable with uncertainties around 10%. The modeling for the random model M_3 is quite unreliable with a total uncertainty equal to 64%. Since M_3 is a random model and often affects less than 10% of the students (see the analysis with real data in later part of this chapter), the low reliability of this model measurement will not cause large errors to the overall results.

From a signal processing point of view, one often uses signal-to-noise ratio (SNR) to evaluate how much the "cross-talk" can degrade the accuracy of a measurement. For a η^{th} response, define SNR_η as the SNR between the signal from a correct match and those from the "cross-talks". SNR_η can be calculated as (in dB)

$$SNR_\eta = 10\log\frac{P_{\eta\eta}}{1-P_{\eta\eta}}$$

For the example here, the different SNR_η's are found to be

$SNR_1 = 10.6$ dB, $SNR_2 = 9.3$ dB, $SNR_3 = -2.5$ dB

The negative value of SNR_3 means that the noise is larger than the signal. If a value of zero is obtained, it represents the case where the signal and the noise are at the same level and the item-based modeling is not usable (50% uncertainty).

The Implications for Question Design

To design a reliable test, we want the reliability matrix to have large diagonal elements. Table 4-4 lists some results for the diagonal elements of **P** under different conditions.

Table 4-4. Diagonal elements of **P** at different settings

$\alpha = 0.6$ $q_1 = 0.5, q_2 = 0.4, q_1 = 0.1$	$\alpha = 0.8$ $q_1 = 0.5, q_2 = 0.4, q_1 = 0.1$	$\alpha = 0.8$ $q_1 = 0.7, q_2 = 0.2, q_1 = 0.1$
$P_{11} = 0.867$	$P_{11} = 0.921$	$P_{11} = 0.955$
$P_{22} = 0.819$	$P_{22} = 0.894$	$P_{22} = 0.778$

As we can see, a 30% variation of the student ability and model triggering probabilities only result in 10% changes of the diagonal elements. Therefore, if α and q_η's are not too ill conditioned (close to zero), the reliability of the test should be acceptable ($P_{\eta\eta} \sim 0.8$).

The change of the q_η's reflects a restructuring of student model states, which is part of the data and is dependent on the individual students. Smaller q_η's lower the reliability of the detection of less favorable models; however, it also increases the reliability of the detection of models that are popular. On the other hand, the decrease of α reduces the reliability on the detection of all the models. For this case, one must carefully design the questions so that the physics and mathematics involved are well matched with the level of the students, to make sure that the majority of students won't have operational difficulties on the content of the questions. The student ability also depends heavily on correct interpretation of the questions. Therefore, the wording and structure of a question should also be straightforward to the students to minimize any possible confusions. A misinterpretation of a question can also result in a triggering of an inappropriate model. This must be evaluated through interviews.

Student abilities of applying different physical models in various contexts can often be evaluated with problem-solving interviews. Based on research, we can also develop multiple-choice questions to assess the ability. For example, on a test we can design a cluster of questions to measure a particular model starting with several simple questions to identify the triggering of certain models. Our data analysis indicates that within a cluster of questions with similar context settings, usually only one physical model is triggered (see chapter 5 for details). Then with similar context settings, we can include a few questions in which students need to apply their models. By analyzing the responses on the questions in the cluster, we can extract information about student ability in applying the corresponding physical model. Again, all the design of the questions and the interpretations of the results should be based on rigorous qualitative researches.

From table 4-2 and table 4-3, it is easy to see that in order to increase the SNR, a larger L is helpful. With L = 5 in the example, the probability of a student guessing any non-random response is about 20%.

In the previous discussion, I have assumed that each non-random physical model only has one corresponding response. In some cases, students can generate responses with some variation on certain insignificant features. For example, on the FM concept, the major incorrect model emphasizes whether there is always a force in the direction of motion. However, in solving different problems, students may come up with a variety of responses on the behavior of such a force (changing or constant, what causes it, etc.). All these different responses still reflect the common incorrect physical model. Therefore, in these cases, having more than one choice to include the different possible student responses can help the detection of student models.

Finally, the most important criteria for a test to be usable with item-based modeling is that a choice in a question cannot be related to more than one physical model. A example is question 33 in the FMCE test (see Appendix A), where students with four different models can legitimately generate the same response (more details are discussed in chapter 5).[9] In such cases the uncertainty for the item-based modeling is very high (>75% when there are four possible models). As discussed earlier, when the uncertainty is greater than 50%, the noise is larger than the signal. Therefore, the item-based method is not usable in such cases and we have to use other methods to extract the information.

Student Model Density Matrix

In the next few sections, I will discuss the details of the mathematical formulation of the model evaluation algorithm. Before going into all the mathematics, I would like to define the variables and indices as in table 4-5.

Table 4-5. Definition of variables and indices

Items	Description
k	indices for students
N	total number of students
m	total number of questions in a concept group
w	total number of physical models in a concept group
η, μ	indices for different physical models and student class model states
$\mathcal{D}$	student model density matrix
$\rho_{\eta\mu}$	an element of $\mathcal{D}$
$\mathcal{D}_k$	student model density matrix for the k^{th} student
$\rho_{\eta\mu}^{k}$	an element of $\mathbf{R}_k$
$\mathbf{V}$	student model vector matrix – eigenvector matrix of $\mathcal{D}$
$\mathbf{v}_\mu$	the μ^{th} eigenvector of $\mathcal{D}$
$v_{\mu\eta}$	an element of $\mathbf{V}$
$\mathbf{u}_k$	the k^{th} student model vector obtained with student responses
$\mathbf{r}_k$	the k^{th} student model response vector
σ_μ^2	the eigenvalue corresponding to the μ^{th} eigenvector of $\mathcal{D}$
$\mathbf{e}_\eta$	the base vector representing the η^{th} physical model

In this section, I introduce a crucial element in this algorithm – the student model density matrix denoted as $\mathcal{D}$. Student responses will be represented with this $\mathcal{D}$ in the model space. This matrix retains the structural information on individual student responses with respect to different physical models. For the three-model example discussed earlier, the density matrix is a 3×3 matrix.

Creating the Single Student Model Density Matrix

To create $\mathcal{D}$, the first step is to find the individual student model states. The student model response vector $\mathbf{r}_k$, when normalized, gives the estimation on the probabilities for a single student to be triggered into the different physical models. As defined in chapter 2, the student model vector, $\mathbf{u}_k$, is obtain by taking the square root of the elements of the normalized mode response vector:

$$\mathbf{u}_k = \begin{pmatrix} u_{1k} \\ u_{2k} \\ u_{3k} \end{pmatrix} = \frac{1}{\sqrt{m}} \begin{pmatrix} \sqrt{n_{1k}} \\ \sqrt{n_{2k}} \\ \sqrt{n_{3k}} \end{pmatrix} = \left| \mathbf{u}_k \right\rangle \qquad (4\text{-}3)$$

As we can see, $\mathbf{u}_k$ represents the probability amplitude.

Then we can construct the single student *model density matrix* of the k^{th} student as:

$$\mathcal{D}_k = \left|\mathbf{u}_k\right\rangle\left\langle\mathbf{u}_k\right| = \left\{\rho^k_{\eta\mu}\right\} = \frac{1}{m}\begin{bmatrix} n_{1k} & \sqrt{n_{1k}n_{2k}} & \sqrt{n_{1k}n_{3k}} \\ \sqrt{n_{2k}n_{1k}} & n_{2k} & \sqrt{n_{2k}n_{3k}} \\ \sqrt{n_{3k}n_{1k}} & \sqrt{n_{3k}n_{2k}} & n_{3k} \end{bmatrix} \qquad (4\text{-}4)$$

Eq. (4-4) indicates that $\rho^k_{\eta\mu} = \rho^k_{\mu\eta}$ and all elements are positive. Then we can write:

$$\mathcal{D}_k = \mathcal{D}_k^T \qquad (4\text{-}5)$$

Table 4-6 is a list of a few typical single student model density matrices for different student model states. The calculation is based measurements with a group of five questions on a concept domain with three physical models (m = 5, w = 3).

Table 4-6. Samples of single student model density matrix

Student Model Responses	Student Model Vector	Student Density Matrix
(500)	$\begin{pmatrix}1\\0\\0\end{pmatrix}$	$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}$
(410)	$\frac{1}{\sqrt{5}}\begin{pmatrix}2\\1\\0\end{pmatrix}$	$\frac{1}{5}\begin{bmatrix}4&2&0\\2&1&0\\0&0&0\end{bmatrix}$
(320)	$\frac{1}{\sqrt{5}}\begin{pmatrix}\sqrt{3}\\\sqrt{2}\\0\end{pmatrix}$	$\frac{1}{5}\begin{bmatrix}3&\sqrt{6}&0\\\sqrt{6}&2&0\\0&0&0\end{bmatrix}$
(221)	$\frac{1}{\sqrt{5}}\begin{pmatrix}\sqrt{2}\\\sqrt{2}\\1\end{pmatrix}$	$\frac{1}{5}\begin{bmatrix}2&2&\sqrt{2}\\2&2&\sqrt{2}\\\sqrt{2}&\sqrt{2}&1\end{bmatrix}$
…	…	…

Creating the Class Model Density Matrix

As discussed earlier, the uncertainty of our determination of individual student models depends on the number of the questions in the group, i.e., we need m >> 3 for the example here, which is not satisfied. But if we have a lot of students in a class, we can construct the density matrix for the whole class and evaluate the student model states of the class. The density matrix for the class is obtained with Eq. (4-6).

$$\mathcal{D} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} = \frac{1}{N}\sum_{k=1}^{N}\mathcal{D}_k = \frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix} \rho_{11}^{k} & \rho_{12}^{k} & \rho_{13}^{k} \\ \rho_{21}^{k} & \rho_{22}^{k} & \rho_{23}^{k} \\ \rho_{31}^{k} & \rho_{32}^{k} & \rho_{33}^{k} \end{bmatrix} \tag{4-6}$$

where

$$\rho_{\eta\mu} = \frac{1}{N}\sum_{k=1}^{N}\rho_{\eta\mu}^{k}$$

From this definition of the class density matrix, we can see that the diagonal elements of the density matrix equals the number of student model-based responses (in percentage) corresponding to each of the physical models, which can be written as

$$\rho_{\eta\eta} = \frac{1}{N \cdot m}\sum_{k=1}^{N} n_{\eta k}$$

Since we have (see Eq. 4-2)

$$\sum_{\eta=1}^{m}\sum_{k=1}^{N} n_{\eta k} = \sum_{k=1}^{N}\sum_{\eta=1}^{m} n_{\eta k} = \sum_{k=1}^{N} m = N \cdot m\,,$$

it is easy to see that the sum of the diagonal elements equals 1:

$$\sum_{\eta=1}^{w}\rho_{\eta\eta} = \sum_{\eta=1}^{w}\frac{1}{N \cdot m}\sum_{k=1}^{N} n_{k} = 1$$

Notice that according to the discussion on measurement resolution, for good results, we need N>>w, which is usually satisfied.

Since $\mathcal{D}$ is symmetric and all data are real, it is a Hermitian matrix. Therefore, we can perform eigenvalue decomposition on it to get its unique set of eigenvectors and eigenvalues. Since the matrix is nonnegative definite (all data are ≥ 0), the eigenvalues are all real nonnegative numbers, which are denoted by $\sigma^2{}_1$, $\sigma^2{}_2$, … $\sigma^2{}_w$. Define the eigenvectors of $\mathcal{D}$ as $\mathbf{v}_\mu$ (a column vector) with $\mu = 1, \ldots, w$ as indices for different class model states. Then the matrix of eigenvectors can be written as

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

This matrix transforms $\mathcal{D}$ into a diagonal form. For a 3-D model space we can write

$$\mathbf{V}^{\mathrm{T}}\mathcal{D}\mathbf{V} = \left[\Sigma^2\right] \overset{w=3}{=} \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

We can reconstruct the class model density matrix from the eigenvectors and eigenvalues. Suppose we are given the student class model states represented by $\mathbf{v}_\mu$ with eigenvalues σ_μ^2. We know for each $\mathbf{v}_\mu$, defined similarly as in equation (4-3), the component corresponding to the η^{th} physical model is $v_{\eta\mu}$. Therefore, this model state will have a contribution equal to $v_{\eta\mu}^2$ from the η^{th} physical model. In addition, from the eigenvalues, we know that σ_μ^2 represents the weight on $\mathbf{v}_\mu$. Then the frequency of the student responses reflecting the η^{th} physical model, $R_{\eta\eta}$, will have contributions from all the class model states and can be calculated as

$$\rho_{\eta\eta} = \sum_{\mu=1}^{w} \sigma_\mu^2 \cdot (\mathbf{v}_\mu \cdot \mathbf{e}_\eta)^2 = \sum_{\mu=1}^{w} \sigma_\mu^2 \cdot v_{\eta\mu}^2 \qquad (4\text{-}10)$$

This result can be verified with (w = 3)

$$\mathcal{D} = \mathbf{V}\left[\Sigma^2\right]\mathbf{V}^{\mathrm{H}} = \begin{bmatrix} \sigma_1^2 v_{11} & \sigma_2^2 v_{12} & \sigma_3^2 v_{13} \\ \sigma_1^2 v_{21} & \sigma_2^2 v_{22} & \sigma_3^2 v_{23} \\ \sigma_1^2 v_{31} & \sigma_2^2 v_{32} & \sigma_3^2 v_{33} \end{bmatrix} \cdot \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}$$

It is easy to see that the diagonal elements are the same as equation (4-10), and the off-diagonal elements can also be obtained with

$$\rho_{\mu\nu} = \sum_{\eta=1}^{3} \sigma_\eta^2 (\mathbf{v}_\mu \cdot \mathbf{e}_\eta)(\mathbf{v}_\nu \cdot \mathbf{e}_\eta) = \sum_{\eta=1}^{3} \sigma_\eta^2 v_{\mu\eta} v_{\nu\eta}$$

which follows directly from the spectral decomposition of $\mathcal{D}$,

$$\mathcal{D} = \sum_{\mu=1}^{3} \sigma_\mu^2 \left|\mathbf{v}_\mu\right\rangle\left\langle\mathbf{v}_\mu\right|$$

The eigenvalues and eigenvectors of $\mathcal{D}$ will be used to study the student class model states. These eigenvectors represent the model states of the class of students. Before going into more details of how to use these eigenvalues and eigenvectors, it is helpful to know the meanings of all these new elements such as the density matrix, the eigenvalues, and the eigenvectors. In the following discussion, I will only focus on the student class model states.

The Meaning of Model Density Matrix and Student Model States

The purpose of introducing the model density matrix is that this matrix can store important structural information about the individual student models. In general, there are three typical model conditions for a class of students. A class can be in either one of these types or in a mixed situation formed by a combination of these types.

1. Most students have the same physical model (not necessarily a correct one) and they are always consistent about it.

2. Students have several different physical models but each student has only one model and is consistent about it. Thus the class of students can be partitioned into several groups each with a different but consistent physical model.

3. Many students in the class can each have multiple physical models and they are not consistent in using these models.

The model density matrix has very different forms corresponding to these different types of situations. The diagonal elements of the $\mathcal{D}$ reflect population of students with model based responses corresponding to the different physical models and the off-diagonal elements indicate the mixing of students using the different physical models in generating their responses. Large off-diagonal elements indicate low consistency in using the physical models (very mixed student model states).

For the above three situations, if most of the students are in favor of one clear and consistent model, one of the diagonal elements will be much larger than the other diagonal elements and the off-diagonal ones will be almost 0. If the individual students are still consistent in using their models but different students may have different types of models, the off-diagonal elements are still zero. If a significant number of students are inconsistent in using their models, the off-diagonal elements of $\mathcal{D}$ will be comparatively large.

In general, the diagonal elements give the distribution of the probability of students' using the different physical models, while the off-diagonal elements indicate consistency of the individual students' using their models. Figure 4-4 shows three examples with different student modeling situations.

When students have consistent models, the density matrix will only have diagonal elements and the base vectors for the physical model space, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, are the eigenvectors of the density matrix (see figure 4-4 (a) and (b)). Then the class model states are the eigenstates of the physical model space. The eigenvalues are the diagonal elements of $\mathcal{D}$, representing the proportions of the students with the corresponding models. However, in most cases students often have inconsistent models, so the density matrix will have both diagonal and off-diagonal elements and the class mode states will not be the basis vectors of $\mathcal{D}$.

From case (b) and (c) in figure 4-4, it is easy to see that the students with consistent models and inconsistent mixed models can generate the same model-based responses corresponding to different physical models (the diagonal elements). Therefore, even if we model the student responses with the physical models but only count the numbers, which can produce the diagonal elements, there is still not enough information to tell if the models of the individual students are mixed or consistent. With a model density matrix, information for such differences is stored in the off-diagonal elements and can be extracted for further analysis.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.2 \end{bmatrix} \qquad \begin{bmatrix} 0.5 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.2 \end{bmatrix}$$

(a) Consistent one-model

(b) Consistent three-model

(c) Inconsistent three-model

Figure 4-4. Samples of student class model density matrix: (a) an extreme case corresponding to the first type of class model condition where everyone has the same physical model (model 1); (b) the second type of class model condition where the class consists of three different groups of students each with a consistent physical model; (c) the third type of class model condition where many students have multiple physical models and are inconsistent in using these models.

Now let us see how exactly the information of the individual student model states is stored in $\mathcal{D}$ and what the eigenvalues and eigenvectors represent. Consider a class with N students. The k^{th} student model vector is represented with $|\mathbf{u}_k\rangle$, where k = 1, … N. Then the class model density matrix is

$$\mathcal{D} = \frac{1}{N}\sum_{k=1}^{N}\mathcal{D}_k = \frac{1}{N}\sum_{k=1}^{N}\left|\mathbf{u}_k\right\rangle\left\langle\mathbf{u}_k\right| \tag{4-11}$$

Define the eigenvectors of $\mathcal{D}$ as $\mathbf{v}_1, \ldots, \mathbf{v}_w$ with eigenvalues as $\sigma^2_1, \sigma^2_2, \ldots \sigma^2_w$. Then $\mathcal{D}$ can also be written as

$$\mathcal{D} = \sum_{\mu=1}^{w}\sigma_\mu^2 \cdot \left|\mathbf{v}_\mu\right\rangle\left\langle\mathbf{v}_\mu\right| \tag{4-12}$$

Using Eq. (4-11), we can write

$$\mathcal{D}\left|\mathbf{v}_\mu\right\rangle = \frac{1}{N}\sum_{k=1}^{N}\left|\mathbf{u}_k\right\rangle\left\langle\mathbf{u}_k\middle|\mathbf{v}_\mu\right\rangle = \sigma_\mu^2 \cdot \left|\mathbf{v}_\mu\right\rangle \tag{4-13}$$

Define $a_{\mu k}$ as the agreement between the k^{th} student model vector $\mathbf{u}_k$ and the μ^{th} eigenvector:

$$a_{\mu k} = \left\langle\mathbf{u}_k\middle|\mathbf{v}_\mu\right\rangle = \left\langle\mathbf{v}_\mu\middle|\mathbf{u}_k\right\rangle \tag{4-14}$$

Then Eq. (4-12) can be rewritten as

$$\mathcal{D}\left|\mathbf{v}_\mu\right\rangle = \frac{1}{N}\sum_{k=1}^{N}\left|\mathbf{u}_k\right\rangle\left\langle\mathbf{u}_k\middle|\mathbf{v}_\mu\right\rangle = \frac{1}{N}\sum_{k=1}^{N}a_{\mu k}\cdot\left|\mathbf{u}_k\right\rangle = \sigma_\mu^2\cdot\left|\mathbf{v}_\mu\right\rangle$$

Therefore, we have

$$\sigma_\mu^2\cdot\left|\mathbf{v}_\mu\right\rangle = \frac{1}{N}\sum_{k=1}^{N}a_{\mu k}\cdot\left|\mathbf{u}_k\right\rangle$$

$$\therefore\quad \left|\mathbf{v}_\mu\right\rangle = \frac{1}{\sigma_\mu^2\cdot N}\sum_{k=1}^{N}a_{\mu k}\cdot\left|\mathbf{u}_k\right\rangle \qquad (4\text{-}15)$$

Thus an eigenvector of $\mathcal{D}$ is a weighted average of all the individual student model vectors with weights equal to the agreements between the eigenvector and the single student model vectors. Therefore, the class model states represented by these eigenvectors are the set of states that most resemble the salient features of all the individual student model vectors.

Eq. (4-15) also indicates that the structure of $|\mathbf{v}_\mu\rangle$ will have more contributions from student model vectors that are similar to $|\mathbf{v}_\mu\rangle$. Therefore, if there exist a group of $|\mathbf{u}_k\rangle$'s that are very similar to each other but different from the rest, this group of $|\mathbf{u}_k\rangle$'s will have a common effect to make one of the eigenvectors ($|\mathbf{v}_\mu\rangle$'s) similar to them.

If we left multiply Eq. (4-15) with the same eigenvector, we have

$$\left\langle\mathbf{v}_\mu\middle|\mathbf{v}_\mu\right\rangle = \frac{1}{\sigma_\mu^2\cdot N}\sum_{k=1}^{N}a_{\mu k}\cdot\left\langle\mathbf{v}_\mu\middle|\mathbf{u}_k\right\rangle = \frac{1}{\sigma_\mu^2\cdot N}\sum_{k=1}^{N}a_{\mu k}^2 = 1$$

$$\therefore\quad \sigma_\mu^2 = \frac{1}{N}\sum_{k=1}^{N}a_{\mu k}^2 \qquad (4\text{-}16)$$

This result indicates that the μ^{th} eigenvalue is the average of the squares of the agreements between the μ^{th} eigenvector and the individual student model vectors. Since the agreement $a_{\mu k}$ is a dot product of two vectors, it is possible for $a_{\mu k}$ to have negative values. The outcome of the dot product gives the agreement of two model vectors in probability amplitudes. On the other hand, $a_{\mu k}^2$ is always positive and gives the agreement based on response output (the model vector is amplitude based and the squares of the elements of a model vector produce the normalized model responses). Therefore, an eigenvalue represents the average response-based agreement between the corresponding eigenvector and all the individual student model vectors.

It can be inferred from Eq. (4-16) that the eigenvalue is affected by both the similarity of the individual student model vectors and the number of students with similar model vectors. In order to have a large eigenvalue, it is needed to have not only large individual

$a_{\mu k}^2$, but also a good number of them, which implies that more students in the class have more similar single student model vectors. That is, the students in the class behave more similarly to each other (the consistency between students is high).

In cases when the performances of students are very similar to each other, an eigenvalue will be mostly determined by the number of students with model vectors similar to the eigenvector corresponding to that eigenvalue.

As an example, suppose there are N students in a class and they can be partitioned into several groups where students in each group are similar to each other and have a same single student model state. Students belong to different groups will have different single student model states. These single student model states can be mixed states, but we assume them to be orthogonal to each other so that the similarity between different student model vectors are 0. (This assumption also determines that the total number of such student groups obtained with this method will always be less than the dimension of the model space – this method will collapse all the students into three groups that best resemble the actual situation.). Define N_μ as the number of the students in a group with a same student model vector $\mathbf{v}_\mu$, which can be a mixed state of the physical models. According to Eq. (4-15), each group of student model vectors will produce an eigenvector exactly the same as themselves because in this case, $a_{\mu k}$ will be either 1 or 0. Thus, the eigenvectors of $\mathcal{D}$ will be the same as these student model vectors. The eigenvalues can be calculated with

$$\sigma_\mu^2 = \frac{1}{N}\sum_{k=1}^{N} a_{\mu k}^2 = \frac{N_\mu}{N} = \rho_{\mu\mu}$$

Therefore in this case, σ_μ^2 reflects the population of the students with the model vector represented by $\mathbf{v}_\mu$.

If the students are not consistent to each other and have different model states, the eigenvalue will not have a direct association with a group of students. However, it still partially reflects the size of a more consistent group of students. In general, when $a_{\mu k}^2$ has values close to a 0/1 binary mode, the eigenvalue will contain more information on the student population distribution for the different model states.

In the above discussion, there are two types of consistency involved that will affect different aspects of the results. The consistency of individual students using different physical models is reflected by the off-diagonal elements of $\mathcal{D}$ and determines the structures of the class model states (mixed models or pure physical models). The consistency among different students determines the eigenvalues.

Student Class Model States

When students have inconsistent models, as in case (c) of figure 4-4, the class model density matrix will not be a diagonal matrix. To find the eigenvalues and eigenvectors of

$\mathcal{D}$ in such situations, we can perform eigenvalue decomposition. For the three-model example here, the eigenvector matrix, **V**, can be written as

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \tag{4-17}$$

and the three eigenvectors are

$$\mathbf{v}_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ v_{31} \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} v_{12} \\ v_{22} \\ v_{32} \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} v_{13} \\ v_{23} \\ v_{33} \end{pmatrix} \tag{4-18}$$

Each of the eigenvectors can also be represented with the basis vectors as

$$\mathbf{v}_\mu = \sum_{\eta=1}^{w} v_{\eta\mu}\mathbf{e}_\eta = v_{1\mu}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_{2\mu}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_{3\mu}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad \mu = 1, 2, 3$$

Then we can represent $\mathbf{v}_i$ by a vector in the physical model space with coefficient, $v_{\eta\mu}$, as component on the η^{th} base. For the example in figure 4-4(c), the eigenvalues and eigenvectors are found to be:

$$\left[\Sigma^2\right] = \begin{bmatrix} 0.66 & 0 & 0 \\ 0 & 0.20 & 0 \\ 0 & 0 & 0.14 \end{bmatrix}, \quad V = \begin{bmatrix} 0.81 & -0.58 & 0.14 \\ 0.52 & 0.58 & -0.63 \\ 0.29 & 0.58 & 0.77 \end{bmatrix} \tag{4-19}$$

The eigenvectors for student model states can be represented as in figure 4-5. In the figure, only the components of $\mathbf{v}_1$ are shown. The eigenvalues are represented as points on the correspondent eigenvectors. In general, the new eigenvectors have components on all basis vectors.

As discussed in the previous section, the class model vectors are the weighted average of all the individual student model vectors. As a result, a class model state contains all the signature information from all the individual student model states with contributions controlled by the similarity between them. Therefore, the structure of a class model state often represents an outstanding salient feature embedded in a large number of the individual student model states. The prevalence of such features is described by the eigenvalues.

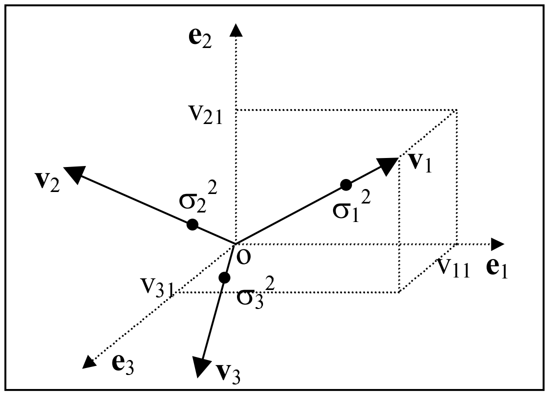

Figure 4-5. Student class mode state vectors

A large dominant eigenvalue indicates that many students are similar to each other and the single student model vectors for different students are similar. On the other hand, if students are all different from one another, the individual student model vectors will have different structures. As a result, it is difficult to find a vector that agrees well with a large number of different single student model vectors. In such cases, there will be no dominant eigenvalues. Therefore, the eigenvalues can be used as a measure to tell if the students are similar or different from one another. The obtained class model states usually have components from all the basic physical models. To see how the class model states agree with the physical models, we can calculate the agreement with a dot product of the $\mathbf{v}_\mu$'s and $\mathbf{e}_\eta$'s. Since the class model states represent the probability amplitude, the probability-based agreement between a physical model and a student class model state is defined as the square of the dot product between the two model vectors

$$\left\langle \mathbf{v}_\mu \middle| \mathrm{e}_\eta \right\rangle^2 = \mathrm{v}_{\eta\mu}^2 \qquad (4\text{-}20)$$

The higher the agreement the better the student models agree with the physical models. For a more complete evaluation, the eigenvalues also need to be considered. Since larger eigenvalue indicates more students with single student model states similar to the corresponding class model state, we would like to have a large eigenvalue for the model state that agrees best with the favorable physical model.

Evaluating and Presenting Student Model States

Class model states derived from the class density matrix can be used as an evaluation to assess class performance. In the following sections, I will introduce a few tools developed to present these student model states and to obtain numerical evaluations on various features of student models.

Model Plane Plot

In many cases, students often have model states structured with dominant components on two physical models, a correct one and a major "misconception". In this case the null model element is often very small (for a three-model situation). For many examples in our data analysis, the measured null elements are around 5%. Then, we can construct a two-dimensional *model plane* by using the two physical models as axes. Each class model state obtained from the class model density matrix is represented with a point on this two-dimensional model plane. As shown in figure 4-6, a class model state, $\mathbf{v}_\mu$, is represented by point B on the model plane with $\sigma_\mu^2 v_{2\mu}^2$ as the horizontal component and $\sigma_\mu^2 v_{1\mu}^2$ as the vertical component. Obviously, when the eigenvalue for a specific class model state is small, the model point will be close to the origin. As discussed earlier, small eigenvalues represent insignificant model states (less popular). If there exists a dominant state, with an eigenvalue lager than 0.8, this state alone is often enough to represent the whole class. In such cases, usually only the dominant state is plotted and it is considered as the *primary model state*.

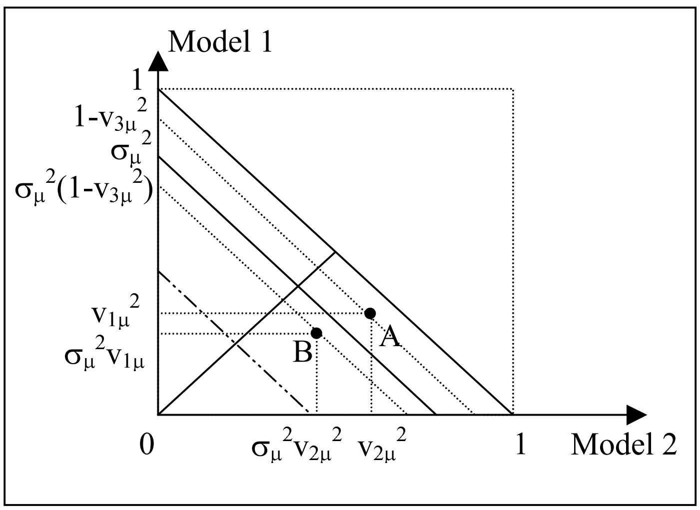

Figure 4-6. Sample student model plot

Since the eigenvector is unitary, the sum of squares of the three components should be equal to the eigenvalue, σ_μ^2. Under such constraint, if the third model (the null model) component, $v_{3\mu}$, is zero, the point representing a model state will be on a line going through point $(0, \sigma_\mu^2)$ and $(\sigma_\mu^2, 0)$. This line is called as the *ideal model line*. If $v_{3\mu}$ is not zero, then the model line will go through the points $(0, \sigma_\mu^2(1-v_{3\mu}^2))$ and $(\sigma_\mu^2(1-v_{3\mu}^2), 0)$. Thus by showing the two model lines together, the third component of the eigenvector can be represented with the shift between the two model lines.

In figure 4-6, all the model lines are represented with dashed lines and the ideal model lines are in solid lines. For the extreme case of a 100% model concentration, i.e., σ_i^2 equals 1, the ideal model line will go through the points (0, 1) and (1, 0), which is the upper boundary of the model region. No model points can exist above this line. This also gives a way to represent the eigenvalues – the distance between the ideal model lines and the upper model boundary where larger distance corresponds to a smaller eigenvalue. According to our experience with student data, the square of the third components (represents the null model) of the primary model vector (eigenvector with largest eigenvalue) is often small (~ 0.02). Therefore, we can make an approximation to write

$$\sigma_\mu^2(1-v_{3\mu}^2) \cong \sigma_\mu^2$$

This indicates that the model points are often very close to their ideal model line. Thus the distance between a model point and the upper boundary is approximately equal to the distance between its ideal model line and the upper boundary, which is directly associated with the eigenvalue. Therefore, on the model plot the distance between a model point and the upper boundary line can be used to get an estimation of the eigenvalue of the corresponding model state.

As we can see from figure 4-6, model points on the boundary line (100% concentration on one eigenvector) will have an eigenvalue equal to 1. The model point at (0,0) has the largest distance (= 0.707) and the eigenvalue is equal to 0 (under the assumption of $v_{3\mu}^2$ being negligible). Thus the eigenvalues of model points at other places on the model-plane can be approximated with

$$\sigma^2 \cong \left(1-\frac{d}{\frac{1}{\sqrt{2}}}\right) = \left(1-d\cdot\sqrt{2}\right) \qquad (4\text{-}21)$$

where "d" is the distance between the model point and the upper boundary line.

Also in figure 4-6, a centerline, going through point (0, 0) and (0.5, 0.5), separates the model plot into two regions. The lower right part is the region for model states with preference on the incorrect model (model 2) and the upper left region is for model states with more preference on the expert model (model 1). When model points are close to the centerline, their model components on model 1 and model 2 will have similar values, indicating more mixed model states. When the model points are closer to the two corners at (1, 0) and (0, 1), one of the model components will be much larger than the other, which implies that the student model states are more consistent and close to the physical models. The point of (1, 0) and (0, 1) represent pure physical models of "Model 2" and "Model 1". In such situations, $\mathcal{D}$ is diagonal and student model states are identical to the physical models.

Angular Presentations of Model States and Model Mixing Features

When analyzing student model structures, we can project the student model states on the plane spanned by the two dominant physical models. Then we can use the angles between the physical models and the projections of the student model states to represent the mixing feature of a model state. Using $\phi_{\eta\mu}$ to represent the model projection angle on the plane spanned by the η^{th} and the μ^{th} physical models, we can write (see Eq. 2-3):

$$\phi_{\eta\mu} = \mathrm{arctg}\left(\frac{\sqrt{q_\mu}}{\sqrt{q_\eta}}\right)$$

Notice that $\phi_{\eta\mu}$ is generally defined so that it can be used for both the measured single student model states and the calculated class model states. Then we can represent the student model states on an angular distribution plot shown in figure 4-7b. The angular distribution plot is evenly partitioned for an easy estimation on the mixing features of student model states.

When analyzing class model state, the model plot shown in figure 4-8 is a better tool since it gives the eigenvalues and is plotted in terms of probability. Define the model angle on the model plot with $\Phi_{\eta\mu}$, then we can write:

$$tg(\Phi_{\eta\mu}) = \frac{q_\mu}{q_\eta} = tg^2(\phi_{\eta\mu})$$

Then the boundaries separating different model regions in the model states angular distribution plot can be translated into the model plot as two straight lines from the origin with a slop equal to 1/3 and 3 respectively (see figure 4-8, and $tg^2(\pi/6)=1/3$).

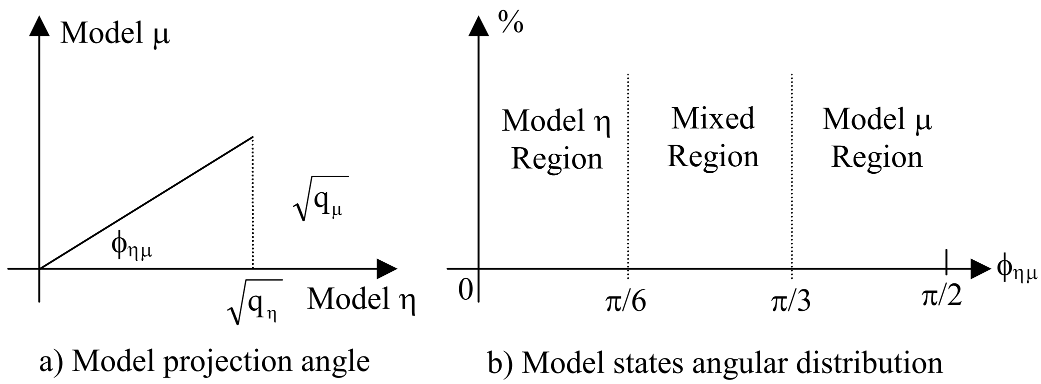

Figure 4-7. Schematics of model projection angle and model states angular distribution.

In the three-model example, the model plot is partitioned into four regions, the "model 1 region", the "model 2 region", the "mixed region" and the "secondary model region" as shown in figure 4-8. The "model 1 region" (and "model 2 region") contains models with dominant model 1 (and model 2) components. For the example described in table 4-1, models in the "model 2 region" will then indicate a strong misconception on FM and models in the "model 1 region" will imply good understandings on FM.

Model points in the "mixed region" can represent a mixed model state, where no physical model is in domination and the individual students are inconsistent in using the different physical models. The "secondary model region" represents model states with small eigenvalues and therefore they are usually considered secondary with much smaller effects on the overall class performance. In most cases from our data, there is often one primary model state with an eigenvalue 3 to 4 times larger than the second largest eigenvalue. Therefore in these cases, showing the primary model state alone can provide a good overview of the student model situations.

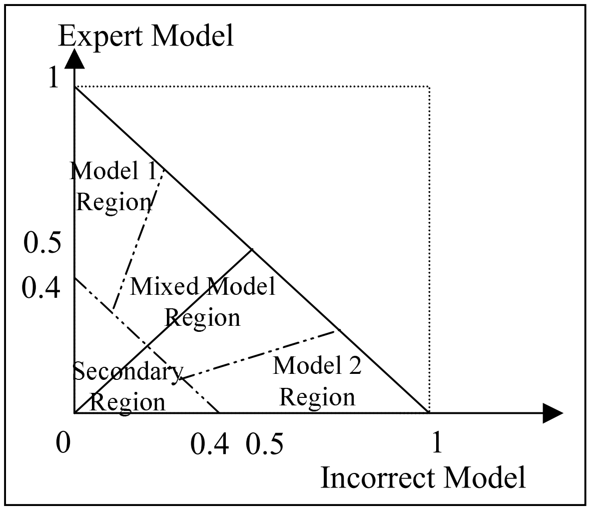

Figure 4-8. Model regions on model plot

The model plane allows us to visually present most information about the student model states on the same graph. We can also put the pre and post model states from different classes together on the same plot, making it much clearer to see the patterns and shifts of the different student model states. Furthermore, it allows us to do quantitative analysis of the student model changes.

Model Improvement

Based on the model plane plot, we can construct a numerical evaluation, the fraction of possible model improvement, denoted as $\mathcal{M}$, to evaluate the shift of the student models on pre and post tests. Figure 4-9 shows a typical model plot with point A as the initial state and point B as the final state.

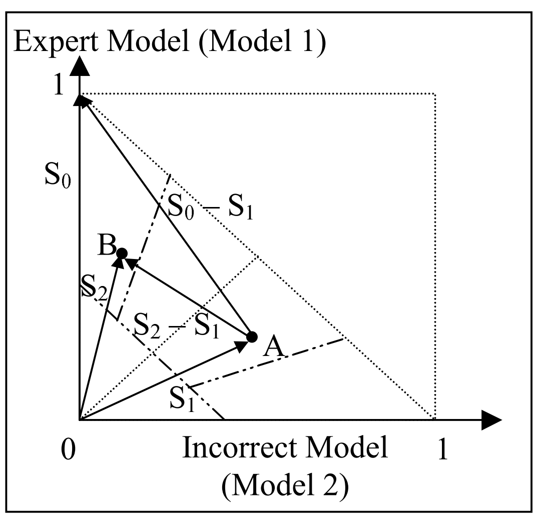

Figure 4-9. Shifts of model states

Define $\mathbf{S}_1$ and $\mathbf{S}_2$ as the two vectors to represent the initial and final student model states. The third vector, $\mathbf{S}_0$, represents the most favorable final state, assuming model 1 is the favorable physical model. For each initial $\mathbf{S}_1$, the best shift it can get is $\mathbf{S}_0 - \mathbf{S}_1$. The real shift, however, is $\mathbf{S}_2 - \mathbf{S}_1$. The fraction of possible model improvement is then calculated with

$$\mathcal{M} = \frac{(\mathbf{S}_0 - \mathbf{S}_1)\cdot(\mathbf{S}_2 - \mathbf{S}_1)}{\left|(\mathbf{S}_0 - \mathbf{S}_1)\right|^2} \tag{4-22}$$

Using Eq. (4-22), $\mathcal{M}$ has a value within the range $(-\infty, 1]$ (1 for the most favorable shift). The negative values indicate a shift towards the unfavorable model. We can also re-scale the unfavorable shift to be within the range [-1, 0]. To do so, we need to locate an unfavorable model state and replace the vector $\mathbf{S}_0$ with the vector starting from the initial state to the unfavorable state. Thus Eq. (4-22) will have two forms depending on the direction of the student models changes and can give a scaled evaluation from –1 to1. Although this calculation uses a two-model example, the general formulation is the same for multi-dimension model spaces.

Relative Model Mixing Rate

As discussed in previous sections, the diagonal elements of the class model density matrix give the probabilities of the class using different physical models. With same diagonal elements, the off-diagonal elements can have different values corresponding to different structures of the measured single student model states. Therefore, with a given distribution of the diagonal elements, the largest model mixing happens when all students have the same model state that gives the probability distribution equal to the diagonal elements. In a three-model example, under this maximum model mixing situation we can write the class model density matrix as:

$$\mathcal{D} = \begin{bmatrix} \rho_{11} & \sqrt{\rho_{11}\rho_{22}} & \sqrt{\rho_{11}\rho_{33}} \\ \sqrt{\rho_{22}\rho_{11}} & \rho_{22} & \sqrt{\rho_{22}\rho_{33}} \\ \sqrt{\rho_{33}\rho_{11}} & \sqrt{\rho_{33}\rho_{22}} & \rho_{33} \end{bmatrix}$$

Define $\gamma_{\eta\mu}$ as the relative model mixing rate between physical model η and physical mode μ. Then we can write

$$\gamma_{\eta\mu} = \frac{\rho_{\eta\mu}}{\sqrt{\rho_{\eta\eta}\rho_{\mu\mu}}} \tag{4-23}$$

where $\rho_{\eta\mu}$ is the measured off-diagonal elements. The relative model mixing rate gives the similarity between a measured class and the possible maximum model mixing situation with same diagonal elements. Therefore, if there exits several groups of students each with a quite consistent model state (similar to case b in figure 4-4), $\gamma_{\eta\mu}$ will give a value close to 0. If all students are having a similar model state, $\gamma_{\eta\mu}$ will give a value close to 1. In this case, the common student model state is close to the one reflected by the diagonal elements of the class model density matrix. Thus $\gamma_{\eta\mu}$ can give us an estimation on whether the students in a class are having similar mixed model states or they may have different types of model states (consistent and mixed). Such information is also reflected in the eigenvalues, however $\gamma_{\eta\mu}$ can give a comparatively direct estimation on the relative model mixing situation between any two physical models of interests. To use $\gamma_{\eta\mu}$, it is necessary to show that Eq. (4-23) always generates a value between 0 and 1. The mathematical proof is discussed in Appendix C.

Measurement Concern on Model State

The model angle of a measured student model state has a non-linear dependence with respect to the measurement in terms of numbers of questions. For example, suppose there are two physical models and the total number of questions used in the measurement is m. The measured model angle can be calculated with:

$$\phi_{\eta\mu} = \operatorname{arctg}\left(\frac{\sqrt{n_1}}{\sqrt{n_2}}\right) = \operatorname{arctg}\left(\frac{\sqrt{n_1}}{\sqrt{m - n_1}}\right)$$

where n_1 (n_2) represents the number of questions that the student uses physical model 1 (physical model 2). Obvious, $n_1 + n_2 = m$. Then the angular dependence of the measured student model state can be graphed as shown in figure 4-10.

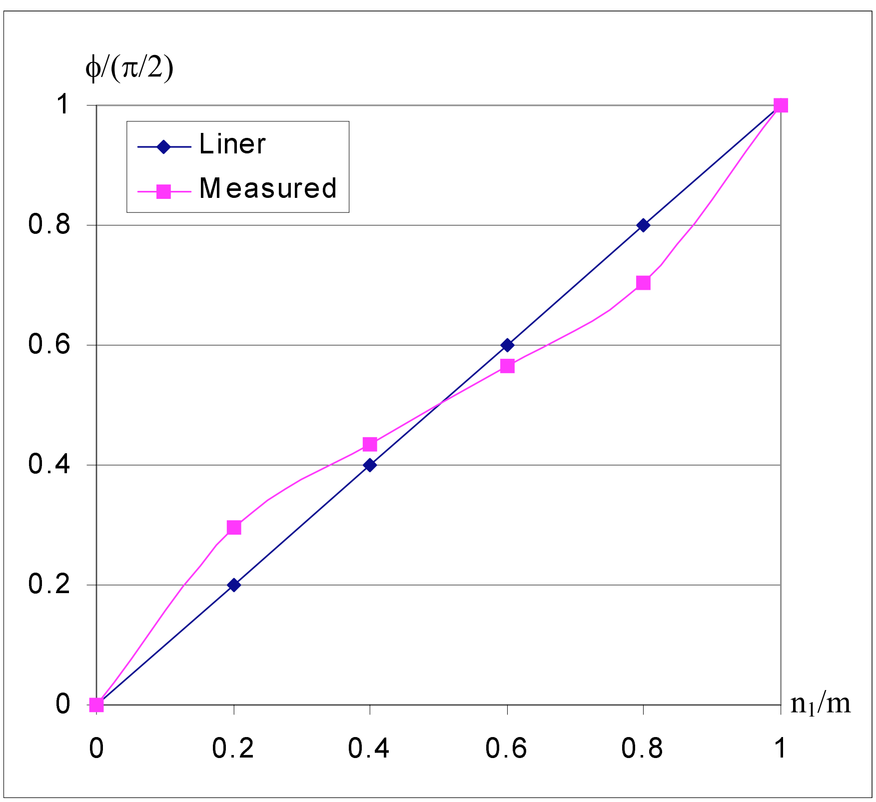

Figure 4-10. Angular dependence of measured student model state

As we can see, the uncertainty of the measured model angle is not liner and it gets larger when the model angle is closer to 0 or $\pi/2$ (model state close to physical model 1 or 2). In the center (perfectly mixed model state), the uncertainty is reduced. Therefore, when setting the resolution for measuring single student model states, we need to take into account this non-linear effect. If most students' model states are close to the physical models, we need to increase the number of questions used in measurement.

Application Examples

Student Model States from FCI data

In this section, I discuss some application examples with FCI data. The main goal of this exercise is to see how this tool performs in practical situations. The data set used in the analysis includes pre and post FCI results from 14 Physics 161 classes at UMd (calculus-based physics for engineers). Our study will focus on two groups of questions of the FCI test, the N3 group and the FM group. In the literature, it has already been shown that on FCI test the performances of the students from tutorial classes and traditional classes are very different.[10] Here I will also compare the results from the two types of classes. The purpose is not to rediscover any of the known differences but to look for new ways of studying the problem and extracting information, which is unavailable with traditional analysis methods.

- **Modeling the Test**

The three physical models of the FM group have already been discussed in the previous sections. Now we need to model the N3 group consisting of 3 questions (2, 11, and 13 see chapter 2 for details). For FCI questions, the N3 group can also be summarized into three physical models (see chapter 2 for detailed discussion on the physical models with Newton's Third Law).

Model 1: The force has the same magnitude and opposite direction (correct).

Model 2: The force is related to the dominant agent (mass, velocity, acceleration, etc.)[11] (incorrect).

Model 3: Other irrelevant ideas (null model).

The model scheme for the corresponding student responses is listed in table 4-7.

Table 4-7. Modeling the Responses of the N3 Group (FCI)

Question	Model 1	Model 2	Model 3
2	e	a	b, c, d
11	e	d, b	a, c
13	a	c, b	d, e

As we can see, the modeling of the N3 group is a little more complicated than the FM group described in table 4-1. The N3 concept domain involves multiple physical features such as mass, velocity, acceleration, etc. Ideally, we would like to have for each physical agent a few questions with different contexts so that we can isolate the student understandings on the different physical features and have more independent information. Unfortunately, the FCI questions are not designed to isolated these physical features, therefore, we have to collapse all the incorrect student models into one generally defined "dominant agent" model.

- **Student Class Model States**

To find the averaged student model state, I analyzed the data from seven tutorial classes with about 500 students and seven traditional classes with about 300 students. The average results are obtained by putting all the students from the same group of classes together. Table 4-8 lists the calculated results of the averaged density matrices, eigenvalues and eigenvectors for the two types of classes.

From the diagonal elements of the class model density matrix, we can roughly see the distribution of student responses on the physical models. In table 4-8, it appears that, with similar initial situations, the tutorial classes have larger first diagonal elements, r_{11}, for their density matrices on the post test, showing larger shifts towards the favorable model. The primary model vectors of the post density matrices for the tutorial classes also have larger elements on the favorable physical models, indicating the individual students are more consistent in using the favorable physical models. In the following sections, more details of the student models are discussed using the new tools developed in the previous sections.

Table 4-8. Results form student FCI data (UMd)

		Tutorial			Traditional		
		Density Matrix	Eigen value	Eigen vector v1 v2 v3	Density Matrix	Eigen value	Eigen vector v1 v2 v3
N3	Pre	0.41 0.27 0.05	0.18	-0.61 0.76 -0.22	0.38 0.24 0.06	0.20	-0.57 0.79 -0.25
		0.27 0.50 0.09	0.75	0.78 0.63 0.03	0.24 0.51 0.11	0.72	0.81 0.59 0.02
		0.05 0.09 0.09	0.07	-0.16 0.16 0.97	0.06 0.11 0.11	0.08	-0.17 0.19 0.97
	Post	0.75 0.21 0.04	0.82	0.32 0.94 -0.11	0.55 0.35 0.16	0.88	0.59 0.62 -0.52
		0.21 0.21 0.03	0.14	0.95 -0.33 -0.03	0.35 0.35 0.16	0.10	0.76 -0.65 0.09
		0.04 0.03 0.04	0.04	0.06 0.10 0.99	0.16 0.16 0.12	0.02	0.28 0.45 0.85
FM	Pre	0.27 0.23 0.02	0.17	-0.40 0.91 -0.11	0.27 0.22 0.03	0.18	-0.39 0.91 -0.12
		0.23 0.69 0.07	0.79	0.92 0.40 0.01	0.22 0.68 0.08	0.78	0.92 0.39 0.01
		0.02 0.07 0.04	0.04	-0.26 0.10 0.99	0.03 0.08 0.05	0.04	-0.06 0.11 0.99
	Post	0.66 0.28 0.03	0.82	0.49 0.87 -0.05	0.46 0.25 0.03	0.23	-0.67 0.74 -0.10
		0.28 0.31 0.02	0.16	0.87 -0.49 -0.03	0.25 0.50 0.05	0.74	0.74 0.67 -0.00
		0.03 0.02 0.03	0.02	0.05 0.03 0.99	0.03 0.05 0.04	0.04	-0.07 0.08 0.99

1. Student model states for the N3 group.

Based on the calculation in table 4-8, the averaged results of student model states in N3 group are plotted in figure 4-11. The results of the individual classes are also plotted in Appendix C. As discussed earlier, model 2 (the horizontal axis) represents the major misconception and the model 1 (the vertical axis) is the favorable physical model. Since the largest eigenvalues for all the classes are around 0.8 and are 3~5 times larger than the second largest one, the primary model can well represent about 90% of the students and should give a pretty good description of the class performance. Therefore, only the primary models (models with largest eigenvalues) are plotted in the graph.

The graph shows that the initial states of both types of classes are very close to each other and these states are at the boundary of the model 2 region. The student initial model can be interpreted as a quite consistent incorrect model with a strong belief that "an object with dominant agent will exert larger force during the interaction". After the instruction, the tutorial classes move to the model 1 region showing that most students ($\sigma_1^2 \approx 0.8$) are having a more consistent model in favor of the correct concept. The traditional classes only make to the mixed model region on the model 1 side, indicating that their model states are still very mixed and inconsistent with significant influence from the misconception. From the model plot, we can see that the tutorial classes have a much larger shift (about twice as much) towards the favorable physical model.

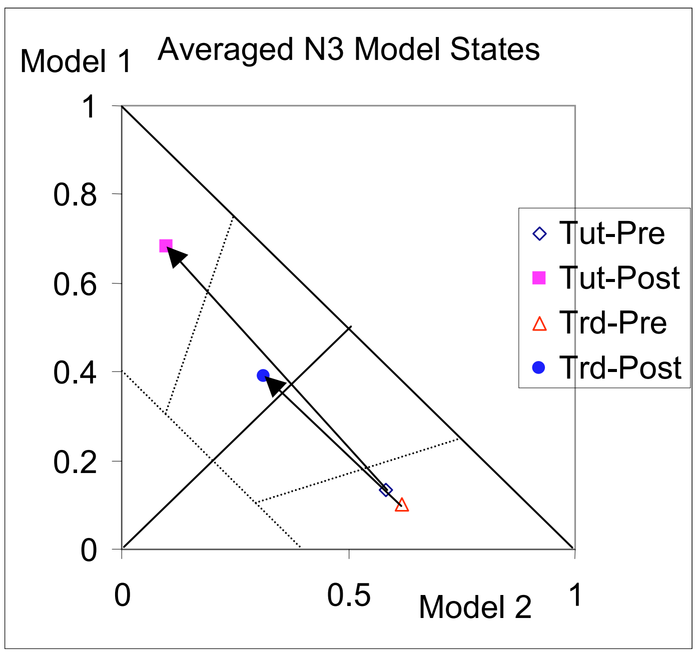

Figure 4-11. Student models on Newton III (FCI, UMd)

2. Student model states for the FM group.

Similarly the averaged results of the student model states for the FM group are plotted in figure 4-12. (The results of the individual classes are shown in Appendix C.) As we can see in this case, the initial states of both types of classes fall into the model 2 region, which indicates that the students are having a consistent but incorrect model (strong misconception) – "there is always a force in the direction of motion". After the instruction, the tutorial classes show some promising improvement towards the favorable model.

Although still in the mixed model region, the tutorial classes make a quite large favorable shift comparing to its problematic initial state. On the other hand, the situation for the traditional classes is not so optimistic. Its final state doesn't even cross the centerline, i.e., the primary model fails to make the transition to the favorable side. This situation indicates that many students are still in favor of their initial misconception. Since the final state is also very close to the centerline, the student model can be interpreted as very mixed and inconsistent under equal influences from the correct physical model and the initial misconception.

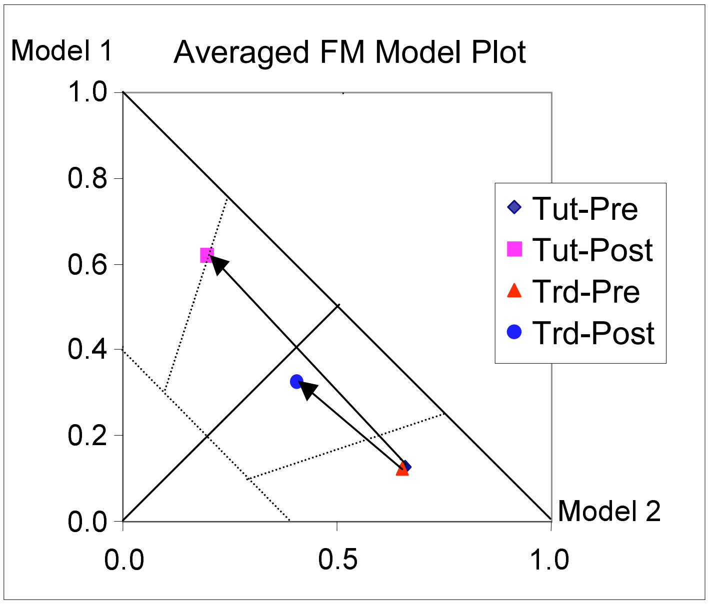

Figure 4-12. Student models on Force-Motion (FCI, UMd)

3. Numerical evaluations

The numerical evaluations reveal similar results. In table 4-9, the quantitative evaluations are calculated.

Table 4-9. Improvement of student class model states (FCI, UMd)

Context Domain	Classes	Possible Model Improvement ($\mathcal{M}$)
N3 Group	Tutorial	0.70
	Traditional	0.38
FM Group	Tutorial	0.62
	Traditional	0.29

The results also show that the students of the tutorial classes have about twice as much improvement towards the favorable model. It appears that the FM group is more difficult for the students compared to the N3 group, but the tutorial classes still have improvement about twice larger than the traditional classes.

Comparison with Factor Analysis

In the beginning of this chapter, I have discussed certain problems with factor analysis. After the introduction of the basic algorithms in model analysis, I would like to make more detailed comparison on the two methods.

Different assumptions and data models

The fundamental assumptions of factor and model analysis are different. The model analysis is constructed on our understandings of student models. It has physical models first, i.e., the basis of the physical model space for a physics concept domain is already determined and model analysis is used to study all different possibilities of students using their models. On the other hand, factor analysis is to factor out a possible model. When a factor is found out, the physical interpretation for such a factor is still uncertain and not unique. With model analysis, the physical meaning is always clear and represents the dynamical process of student modeling in learning. Based on the student model states, we can even reconstruct the model density matrix and make predictions on other tests with similar physical contexts.

Density matrix and correlation matrix

In factor analysis, a correlation matrix is used where the diagonal elements are always equal to 1 and the off-diagonal elements indicate the correlation between the student scores on different questions. With model analysis, we have a model density matrix where the diagonal elements represent the probability of students using the different physical models and the off-diagonal elements represents the consistency of the individual students using these models.

Different models of random process

With model analysis, there are three types of random processes (see chapter 2 for details). The first one is the random process happens in the model triggering where the context can trigger a variety of student models. The second one is that a student can produce a response by random guessing. The third one is the random error due to human mistakes.

In factor analysis, there is only one random variable corresponding to human errors and guessing. The data model is often written as

$$S = T + e$$

where S is the student score, T is the "true" score, and "e" is the error.[12] Since only the score is evaluated in this model, there is no direct information that can be obtained to study the student underlying models. In addition, for some questions, students with different models can sometimes produce the same score. This can create very confusing results if only the score is considered (examples are discussed in chapter 5).

An example on differences between model analysis and factor analysis

Suppose we give four multiple-choice questions to a class of 100 students (m=4, N=100). All four questions are based on one physics concept domain, which has two physical models, model A and model B (w=2). Consider two cases:

Case 1: All students in the class are consistent and half of them use model A and the other half use model B.

Case 2: All students are equally mixed between model A and model B.

The following calculations are performed on computer simulated data generated based on the definition of the example. For case 1, the results from both methods are calculated in table 4-10. As we can see, with model analysis, the results represent that the whole class has two consistent models with equal weightings. The result from factor analysis gives a single factor. This result indicates that all the students are consistent – the students give either all correct answers or all incorrect answers. However, it doesn't tell in which way the students are being consistent. In fact, when the students are consistent the same factor can be obtained regardless how many of the students get the questions correct or wrong. Therefore, the results from factor analysis can only provide limited and indirect information on student behavior.

Table 4-10. Results from model analysis and factor analysis for students with consistent models

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

When the individual students are inconsistent as in case 2, factor analysis often fail to provide useful information. The results for case 2 are calculated in table 4-11. (In both cases, eigenvectors with eigenvalues equal to zero are omitted.)

The students are assumed to be equally mixed with model A and model B. Therefore, in doing the calculation, the probability for a single student to use either model A or model B is set equal for all the questions. As we can see, the results from model analysis indicate a perfectly mixed class model state with 100% dominance (σ^2=1). The results from factor analysis show no dominant factors. Since the students are inconsistent in answering the questions, factor analysis only gives insignificant random-like correlation between the different questions. Consequently, such situations are often interpreted as if there is no factors in the test data.

Table 4-11. Results from model analysis and factor analysis for students with a equally mixed model

Model Analysis		Factor Analysis	
Density Matrix	$\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$	Correlation Matrix	$\begin{bmatrix}1 & -0.34 & -0.34 & -0.32\\ -0.34 & 1 & -0.32 & -0.34\\ -0.34 & -0.32 & 1 & -0.34\\ -0.32 & -0.34 & -0.34 & 1\end{bmatrix}$
Eigenvalues	$\sigma_1^2 = 1, \sigma_2^2 = 0$	Eigenvalues	$\sigma_1^2 = 1.36, \sigma_2^2 = 1.32,$ $\sigma_3^2 = 1.32, \sigma_4^2 = 0$
Class Model States	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix}$	Factors	$\frac{1}{2}\begin{pmatrix}1\\ -1\\ -1\\ 1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ 1\\ -1\\ 0\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\ 0\\ 0\\ 1\end{pmatrix}, \frac{-1}{2}\begin{pmatrix}1\\ 1\\ 1\\ 1\end{pmatrix}$

This example shows that factor analysis cannot deal with the random process in model triggering.

The Special Features of Model Analysis

To justify the using of Model Analysis, I have to clarify two frequently raised questions.

1. Is it worth doing the additional math? What is it that the model analysis can provide and a simple evaluation of scores can not?

2. Can we only use the model response vectors rather than the model density matrix?

The score doesn't tell the whole story

It is obvious that the major disadvantage of score based evaluations is that they don't provide any information on why students are doing wrong. Especially when the score is low, the information on the behavior of the majority of students is completely lost. With model analysis, the complete set of data is used and the results reflect the behavior of all the students including those giving correct, incorrect, and random types of responses. This evaluation provides comprehensive information with clear and straightforward meanings. On the other hand, what a score can really tell is often quite an argument and depends heavily on the background of the students as well as the design of the specific questions (how students fail to give correct answers is implicit with the score based evaluations).

According to our model of learning, the student data represents a multi-dimensional output based on multiple underlying models. There are also different random processes involved. The score-based evaluation only provides a one-dimensional data set, therefore it is very difficult, if not impossible, to extract the multi-dimensional model based information with one-dimensional data. Even when fancy statistical tools are used (such as Latent Class/Model analysis, which is intrinsically similar to factor analysis), the result is still constrained by its one-dimensional nature. Therefore, there can be many uncertainties because different configurations of any multi-dimensional structure can often create a same 1−D projection. The results are also very much dependent on the selection of ways to carry out the 1-D projection. Therefore, it is likely that sometime a score-based analysis gives reasonable results and sometime it doesn't work at all.

Score-based methods are also vulnerable to any model-based random processes in model triggering (these appear to be colored noises compared to the white random noise). For example, the modified versions of factor analysis often characterizes student involvement or their latent models with an ability factor. It represents the probability of students with a certain background to get a question correct. But the ability is a very fuzzily defined concept. It still doesn't say exactly why a student makes mistakes. From our research as well as many others, it is observed that in many situations, students have the "ability" if measured with math, language comprehension and logical reasoning. But they often possess an incorrect mental model on the topic of the question. Therefore, their failure to give a correct answer is due to that they are using their "ability" to generate an incorrect answer, which is reasonable to them based on an incorrect model. Their "ability" in using such models, including the correct ones if they have them, is not bad. However, in these situations, the results coming out of the score-based analysis will often characterize the students with a low "ability", which implies a completely different story – students with poor "ability" on basic mathematical/logical operations and students with good "ability" but incorrect models can generate the same low score. The score-based evaluation methods have no way to distinguish these two different cases.

Let's see another example. Consider if we have a class of 100 students with average score of 50% on questions in a concept group. From the score, what we know is that an average student will most likely answer half of the questions correctly. But there is still a lot more we would like to know. For example, what makes the students go wrong? (Are they picking incorrect responses randomly or with some preference on certain incorrect choices?) Are these students consistent with themselves (are they always using the same model or using different ones at different times)? Are these students consistent with each other (with regards to not only the scores but also the incorrect responses)? All these questions are not likely to be answered by doing score-based evaluations.

With model analysis, these questions can be answered and most of the information is included in the model plot. Figure 4-13 is a scatter plot of some possible model states with the same score equal to 50% (the data is from computer simulations for different possible configurations of student responses). Due to the complexity of the calculation, there are only a handful of simulated data points. But these should give the basic idea.

As we can see from the diagram, with the same score, it is possible for model states to be in different places in a comparatively large region (marked with the dashed lines). I will not go into the details of defining the boundaries of this region. Even with the limited number of data points here, we can see very different physical meanings reflected by these states.

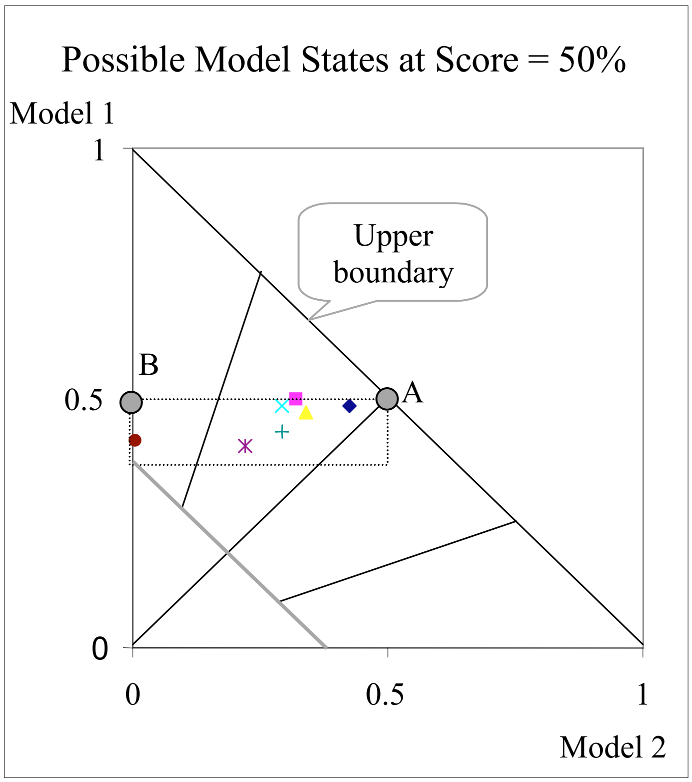

Figure 4-13. Scatter plot of possible model states at S = 50% from simulated data. Point "A" represents the case where all students have a perfectly (equally) mixed model state. Point "B" represents the case where half the students have a pure model 2 state (the remaining half can have any type of model state configurations between model 2 and model 3 but with no model 1 element). All other points represent other simulated cases where the model states are somewhere between case A and B.

As discussed earlier, the closer a point is to the upper model boundary, the larger the eigenvalue will be. Therefore, for the example shown in figure 4-13, point “A” describes a situation where the eigenvalue of the primary model is 1 indicating that the behaviors of all the students are identical (which also implies that they all have the same single student model state). Since the model state is also in the center of the mixed model region, the student model is also a perfectly mixed state, which can be represented as

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

Point “B” represents another situation where the model state is a pure “Model 1” and the eigenvalue of the corresponding state is 0.5 (it is on a model line half way between the origin and the upper model boundary). This indicates that exactly half of the students have perfect and consistent “Model 1”. The rest of the students, depending on the model states of the second and third class model states (not shown in figure 4-13), can be either with pure “Model 2”, pure “Model 3” or any mixed states of both (what is certain here is that the second and third model states will be points on the horizontal axis with zero components on Model 1 axis due to the constraint of S=50%).

All other model points represent some intermediate situations between these two extreme cases. These variations on the student models are all under the constraint of the score being 50%. Suppose we only have the information on the score, all these subtle details of the student model states will then be unavailable. Therefore, Model Analysis can provide a much more complete description of the student understanding than what scores can tell.

Density matrix vs. model response vector

Working with the model response vector is much simpler than working with model density matrix. For the case of a single student, the model response vector provides the same information on student model structures as the density matrix does (under the assumption that only one response vector is calculated for a single student). If we simply add all the model responses from different students together and find an averaged model vector, this model vector can only maintain the information on the distribution of the student model-based responses generated with different physical models. All other information concerning the consistency of students (described with eigenvalues) and the structure of individual student models (described by the different model states) will no longer exist.

As an example with 3 physical models, suppose we have 60 students answering a group of 4 questions. If we sum up all the model responses, we can obtain an average model response vector. For a same averaged response vector, there can be a variety of different configurations of the individual model responses. Table 4-12 lists two typical cases that will both generate an average response vector equal to $2/3\cdot(1,1,1)^T$.

As we can see from table 4-12, the two types of responses reveal very different student model situations. For type 1, there are three groups of students and within each group all students have a same consistent pure physical model. For type 2, there are also three groups of students and within each group all students have the same mixed model state. Since both cases produce the same average model response vector. It will be impossible to distinguish these two cases if we only keep the average vector.

Table 4-12. Different model states with same average model response

	Type 1			Type 2		
Students $\Rightarrow$	n1	n2	n3	n1	n2	n3
Average	20	20	20	20	20	20
$\frac{2}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}$	$\begin{pmatrix}2\\0\\0\end{pmatrix}$	$\begin{pmatrix}0\\2\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\\2\end{pmatrix}$	$\begin{pmatrix}1\\1\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\\1\end{pmatrix}$	$\begin{pmatrix}1\\0\\1\end{pmatrix}$
Density $\Rightarrow$ Matrix	$\frac{1}{3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$			$\frac{1}{3}\begin{pmatrix}1&0.5&0.5\\0.5&1&0.5\\0.5&0.5&1\end{pmatrix}$		

These two cases are very simple examples of possible forms of student responses. In reality, the student responses can be very complicated. Therefore, in order to retain the detailed information on individual student mental model structures, it is necessary to use model density matrix.

From these calculations, it can also be implied that in order to identify the mixed model states, the number of questions in a concept group has to be larger than 1 (the larger the better). According to the uncertainty relation between model states and measurement instances, if we only have one question, the uncertainty for possible mixed model states is the whole model space. That is the students can be in a definite physical model state at one instance and get triggered in a different state at the next instance. And the probability of such model triggering can not be obtained with only one question. In other words, with only one question, the resolution is so low that it is impossible to detect any mixed model states.

Limitations of Eigenvectors

As indicated from Eq. (4-15) and Eq. (4-16), the primary eigenvector contains the most dominant features of all single student model vectors. The additional eigenvectors act as corrections which reflect less popular features. When considering the class a single unit, the primary eigenvector gives good evaluation on the overall model structure of the class. However, if we regard the class as a composition of individual students, there can exists colorful details that are often unable to be extracted with simple eigenvalue decomposition.

For example, a class can contain several groups of students, where students in each group all have very similar model states and students from different groups have different model states. For the following two situations, eigenvalue decomposition can give good results. 1. When the model states from different groups are very different (orthogonal), eigenvalue decomposition will produce eigenvectors that represent these model states. 2. When one of these groups has a dominant popularity, eigenvalue decomposition will produce a primary vector very close to this dominant state. That is, eigenvalue decomposition can give good results when most students are similar to each other or when students are very different (with orthogonal model states). In the case when students are different but not "that" different (with different but non-orthogonal model states), and if we want to identify the different groups with such different but non-orthogonal model states, it is not helpful to use eigenvalue decomposition. In such cases, we can use model-based cluster analysis to identify these clusters of student model states.[13] To identify clusters of individual student model states, it is necessary to have a relatively high resolution in measuring the single student model states. Otherwise, the identified cluster model states can have high uncertainty and complicate the interpretation. As an empirical recommendation, when the eigenvalue of a primary eigenvector is less than 0.65 and the student model states are not orthogonal, it is suggested to use cluster analysis to study the details of the student model structures. In our data, we often get primary eigenvalues close to 0.8, which indicates that most students have similar model states. In these cases, the primary eigenvector can give good evaluation of the class and the individual students.

Advantages of Using Model Analysis

Using model analysis tool, we can obtain useful results about the student models in understanding physics, which are unavailable with score-based analysis. There are certain advantages to use this method.

A major advantage is that with model analysis, the data is transferred into model space. It is often a problem that the research in PER is very much dependent on the background and context of the experiment. For example, with a similar topic researches by different researchers are often affected by many factors including student background, instructions, and research instruments. Due to the "human study" nature of the research, it is also very difficult to repeat any experiment with the same exact condition. Just like in quantum mechanics, once a measurement is performed, the original student states will also be changed. Therefore, variations of the backgrounds of students and the different probing instruments used in different studies all make it very difficult to directly compare the results from different experiments. Using model analysis, the instruments are research based and validated. This can significantly reduce the variations of the results due to the use of individually designed probes. In addition, results from model analysis are represented and analyzed in model space, which can reduce possible misinterpretations of student real understandings due to the weakness of score-based evaluations.

Since model analysis is constructed based on student models, the physical meanings in all the processes are straightforward. The model density matrix also retains a lot of useful information of the student data. Therefore, it is a more useful format to store these data. In addition, the results of student model states provide information on student real understanding of the related topics, which is very important in designing new instructions to help students.

Summary

In this chapter, I have introduced a new algorithm to do quantitative evaluations of student mental models. It makes more use of the data than score-based analysis, (traditionally if we only calculate the correct answers, a lot of information is wasted) and allows us to study the student models in a quantitative way. It can serve as a more comprehensive quantitative evaluation for the student performance, especially with respect to the student models.

With this method, student raw data are transferred into states in the model space. The results can be used to analyze student understandings and/or the features of the instruments.

Model analysis can also provide more explicit information on how to improve instructions. Since it gives more detailed knowledge of the models the students have, it allows us to see more directly about the possible causes of the student difficulties rather than the just the difficulty itself. Therefore we can develop more appropriate instruction strategies right upon the weak points and help the student more effectively.

The results from model analysis match very well with the recognized expectations from other researches using different evaluation methods. These can be evidences for the plausibility of the results from model analysis. In addition, model analysis also provides more detailed quantitative information on the student models, which is otherwise unavailable from traditional score-based evaluations. The density matrix can be used as a new way to store the student responses, where a lot of information is retained and can be easily extracted for different purposes. The various graphical representations of the data also make the results much easier to understand.

References and Endnotes:

[1] E. E. Cureton and R. B. D'Agostino, *Factor Analysis An Applied Approach*, Lawrence Erlbaum, 1983.

[2] D. Huffman and P. Heller, "What does the force concept inventory actually measure?" *Phys. Teach.* **33**, 138-143 (1995)

[3] W. J. van der Linden, R. K. Hambleton, Eds, *Handbook of Modern Item Response Theory*, Springer, 1997.

[4] Theoretically, we can construct a test that each question probes one model at a time. In such cases, we need to design questions to use incorrect responses as scores which seems awkward. Say if we want student responses on Newton III, we need a set of questions that only provide incorrect responses such as "big boy exerts more force" and no correct ones, so that students with incorrect models will pick this one and student with correct model will pick "none of above". But these are unnecessary with model analysis where the full response is used.

[5] R. K. Thornton, "Conceptual Dynamics: Changing Student Views of Force and Motion," Proceedings of the Int. Conf. on Thinking Science for Teaching: The Case of Physics. Rome, Sept. 1994.

[6] For analysis of single student model state, with multiple-choice single response questions, we can measure the probabilities directly (see chapter 2). With multiple-choice multiple-response questions, depending on the design of the questions, usually each individual question doesn't give the measurement on the probability and if multiple questions are used, we often need to use model density matrix again (see later part of this chapter and examples in chapter 5).

[7] See chapter 2 for more detailed definitions on physical models.

[8] It is assumed that when a student is unsuccessful in applying a physical model, there won't be a choice coincide with the result the student will produce.

[9] It is believed that certain questions in FMCE test are used for special purposes by the designer and should not be treated with traditional evaluation. Here in the discussion, we only study the general method to design a question to be used with IbM and do not intend to analyze the features of a particular test.

[10] R. R. Hake, "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," Am. J. Phys. **66**, 64-74 (1998). This paper presents pre and post FCI data from a large number of classes at many high schools, colleges, and universities. Most active-engagement classes showed much greater improvement than traditional classes.

[11] I. A. Halloun and D. Hestenes, "Common sense concepts about motion," Am. J. Phys. **53** (11), Nov. 1985.

[12] E. J. Pedhazur and L. P. Schmelkin, *Measurement, design, and analysis: an integrated approach*, Hillsdale, N.J.: LEA, 1991.

[13] A set of model-based cluster analysis algorithms have been developed to deal with these situations. Due to time limitation, the details are not included in this dissertation. We will report this method and results in a paper in preparation. Pre-prints and details are available upon requests.