A Model Analysis of Proof Schemes

Todd CadwalladerOlsker

California State University, Fullerton

Abstract

This paper describes model analysis, a method of analysis used effectively in physics education research, with particular emphasis on the use of model analysis to study the proof schemes held by students. Model analysis is a particularly effective way to study student's proof schemes because it allows for the fact that student proof schemes may not consistently fall into a single category. This paper also describes the results of a pilot study in which model analysis is used to study students proof schemes.

1 Introduction and Background

In recent years, much attention has been paid to students' understanding of mathematical proof, both when reading and writing proofs, at different points in students' eduction (Weber & Alcock, 2004; Solomon, 2006; Ellis, 2007). One framework for studying student understanding of proof is Harel and Sowder's taxonomy of *proof schemes* (Harel & Sowder, 1998, 2007), which categorizes the arguments an individual (or a community) finds to be convincing when *ascertaining* the validity of a theorem, or when *persuading* another of the validity of the theorem. While the taxonomy of proof schemes makes a useful framework, there is some evidence to suggest that students' proof schemes do not always *consistently* fit into a single category (Housman & Porter, 2003). However, there is little current research on these inconsistencies. Furthermore, while there has been a considerable amount of study conducted on the proof schemes held by *individuals*, there has been little research on the proof schemes held by *communities*.

Model analysis, a quantitative analysis technique pioneered by physics education researchers Bao and Redish (1999; 2001, 2006), can be used to correct both of these deficiencies in the literature. As described by Bao and Redish, "Model analysis. . . applies qualitative research to establish a quantitative representation framework." (Bao & Redish, 2006) That is, model analysis allows researchers to use a qualitative framework, such as Harel and Sowder's taxonomy of proof schemes, to establish a quantitative framework that can be used to study a larger group, such as a community of students. Furthermore, model analysis specifically allows for, and analyzes, inconsistencies in students' responses.

The purpose of this paper is to present an adaptation of Bao and Redish's model analysis to the study of proof schemes. The method of model analysis, as applied to proof schemes, will be discussed, and the results of a small pilot study will be presented.

2 Model Analysis and Proof Schemes

Bao and Redish describe model analysis as a way to investigate students' *mental models*, which they define as "robust and coherent knowledge element[s] or strongly associated set[s] of knowledge elements." (Bao & Redish, 2006) In the context of students' understandings of proof, Harel and Sowder's concept of *proof scheme* can be thought of as such a mental model, which can be analyzed using model analysis. This section contains a description of Bao and Redish's method of model analysis, as well as a description of the adaptation of model analysis to the study of proof schemes.

Bao and Redish describe the method of model analysis as consisting of five steps (2006). In each of the following subsections, a single step is described.

2.1 Step 1

> (i) Through systematic research and detailed student interviews, common student models are identified and validated so that these models are reliable for a population of students with a similar background. (Bao & Redish, 2006)

This step describes what Harel and Sowder accomplished when creating their taxonomy of proof schemes (Harel & Sowder, 1998). The taxonomy is very detailed; in order to effectively define usable student models, I will only consider the most basic level of the taxonomy. At this level, proof schemes fall into three categories: *external conviction, empirical*, and *deductive*.

The category of external conviction describes proof schemes held by students who are convinced by something external to the meaning of the theorem or its proof. For example, a student who holds an external conviction proof scheme may be convinced a theorem is correct based on the form or symbols used by the proof, or by an outside authority such as an instructor or textbook. Empirical proof schemes are held by students who are convinced by empirical evidence, such as examples or a diagram. Finally, students who are convinced by logical reasoning are said to have deductive proof schemes.

2.2 Step 2

> (ii) This knowledge is then used in the design of a multiple-choice instrument. The distracters are designed to activate the common student models, and the effectiveness of the questions is validated through research. (Bao & Redish, 2006)

For this step, I created a questionnaire consisting of five theorems, and four purported proofs of each theorem. Each proof was designed to evoke one of the three categories of proof scheme. Some theorems had more than one proof fitting a single category, in all, there were seven proofs evoking the external conviction category of proof scheme, six evoking the empirical category, and seven evoking the deductive category.

Students were asked to respond to each theorem and its associated proofs in two ways. First, students were asked to mark all of the proofs they found to be convincing. Students were also allowed to mark a response of "none of the above." Second, students were asked to mark the single proof they found to be the most convincing (again, "none of the above" was allowed).

In the next section, the results of a pilot study are presented. Because this study was intended as a pilot study, the effectiveness of the questionnaire was not validated. In order for a model analysis to yield useful results, the validity of the data must be assured. For this reason, the results of this study should not be taken as representative. Instead, the results should be taken as an indication of the kind of information that can be extracted using model analysis. Further discussion is found in the final section of this paper.

2.3 Step 3

> (iii) One then characterizes a student's responses with a vector in a linear "model space" representing the (square roots of the) probabilities that the student will apply the different common models. (Bao & Redish, 2006)

The model space described in this step is represented mathematically by a linear vector space, where each common model is represented by an element of an orthonormal basis. That is, each of the three categories of proof scheme will be assigned a dimension in the vector space. We also assign a fourth dimension to a "null" model, representing other, less common mental models that do not easily fit into any of the three categories of proof scheme.

For each student, we will create a vector inside this model space that represents the student's responses to the questionnaire. Each entry in the vector is meant to represent the (square root of the) probability with which the student uses the associated category of proof scheme to respond to similar types of questions. Of course, these probabilities can only be approximated by the student's responses to the questionnaire. The square roots of the probabilities are used for the sake of making the next step easier, as discussed in the next subsection. There is no real difference in the information contained by a probability and the square root of the probability.

I calculated the probabilities associated with each category of proof scheme in two ways. First, I calculated the probabilities using students' responses to the line in which they marked only the *most* convincing proof. The calculated probability for each category of proof scheme was calculated as the number of responses in the category was divided by the total number of responses given by the student. Responses of "none of the above" were considered to be in the fourth, "null" category.

For example, one student marked one external conviction response, one empirical response, two deductive responses, and one null response. Her model state is represented as the vector

$$\mathbf{u} = \begin{pmatrix} \sqrt{p_{ext}} \\ \sqrt{p_{emp}} \\ \sqrt{p_{ded}} \\ \sqrt{p_{nul}} \end{pmatrix} = \begin{pmatrix} \sqrt{1/5} \\ \sqrt{1/5} \\ \sqrt{2/5} \\ \sqrt{1/5} \end{pmatrix},$$

where p_{ext}, p_{emp}, p_{ded}, and p_{nul} represent the calculated probabilities that the student's response falls into the external conviction category, empirical category, deductive category, or "null" category, respectively.

The above vector calculation was done exactly according to Bao and Redish's description. However, this calculation depends on the assumption that students may only mark one response for each item on the questionnaire. It is reasonable to suppose that students may, simultaneously, find more than one proof to be convincing. For this reason, students were also asked to mark all of the proofs they found to be convincing for each theorem.

For this second method of computing a student's model state vector, each probability was calculated by dividing the number of proofs found to be convincing from each category by the total number of proofs the student marked as convincing. In this way, the probabilities for each student total 1, and each student's vector has equal weight, whether they found a large number of proofs to be convincing or a small number to be convincing. This method of computing probabilities for multiple-response data is most consistent with Bao and Redish's description of model analysis.

As an example, another student marked five external conviction responses, two empirical responses, three deductive responses, and one null response (eleven total). His model state is represented as

$$\mathbf{u} = \begin{pmatrix} \sqrt{5/11} \\ \sqrt{2/11} \\ \sqrt{3/11} \\ \sqrt{1/11} \end{pmatrix}.$$

These model state vectors will be used as a data point for each student. In the next step, these data will be combined in order to analyze the models held by the *community* of students.

2.4 Step 4

> (iv) The individual student model states are used to create a "density matrix," which is then summed over the class. The off-diagonal elements of this matrix retain information about the confusions (probabilities of using different models) of individual students. (Bao & Redish, 2006)

For each model state vector, a *density matrix* is created by taking the outer product of the model state vector with itself. That is, for a model state vector, **u**,

$$\mathbf{u} = \begin{pmatrix} \sqrt{p_{ext}} \\ \sqrt{p_{emp}} \\ \sqrt{p_{ded}} \\ \sqrt{p_{nul}} \end{pmatrix},$$

the density matrix **D** is given by

$$\mathbf{D} = \mathbf{u} \otimes \mathbf{u}^T$$

$$\mathbf{D} = \begin{bmatrix} p_{ext} & \sqrt{p_{ext} \cdot p_{emp}} & \sqrt{p_{ext} \cdot p_{ded}} & \sqrt{p_{ext} \cdot p_{nul}} \\ \sqrt{p_{emp} \cdot p_{ext}} & p_{emp} & \sqrt{p_{emp} \cdot p_{ded}} & \sqrt{p_{emp} \cdot p_{nul}} \\ \sqrt{p_{ded} \cdot p_{ext}} & \sqrt{p_{ded} \cdot p_{emp}} & p_{ded} & \sqrt{p_{ded} \cdot p_{nul}} \\ \sqrt{p_{nul} \cdot p_{ext}} & \sqrt{p_{nul} \cdot p_{emp}} & \sqrt{p_{nul} \cdot p_{ded}} & p_{nul} \end{bmatrix}.$$

Notice that the diagonal entries of the density matrix are simply the probabilities calculated in step (iii). These entries are the squares of the entries in the original vector; this was the reason for using the square roots of the probabilities to create the model state vector in the previous step. Also notice that the off-diagonal entries are non-zero only when a student has responses from more than one category of proof scheme. Thus, when a student is inconsistent, those inconsistencies are preserved by the off-diagonal entries of the matrix.

To study a large number of data points, the density matrices will be averaged together: that is, the entries in each position are added together and divided by the total number of data points. Bao and Redish refer to the resulting matrix as the *class density matrix*. The diagonal elements of the class density matrix give the average of the probabilities that a student response falls into each category of proof scheme.

The class density matrix yields valuable information regarding the inconsistencies of students. As noted earlier, any time a student uses more than one category of proof scheme, than inconsistency is preserved by the off-diagonal elements. When the class density matrix is created, non-zero off-diagonal elements indicate that some of the students have responded using more than one category of proof scheme. When the off-diagonal elements are large, the class density matrix indicates a high degree of inconsistency in the students. Bao and Redish suggest that an off-diagonal element is sufficiently "large" when the off-diagonal element exceeds half of the product of the square roots of the two corresponding diagonal elements.

2.5 Step 5

> (v) The eigenvalues and eigenvectors of the class density matrix give information not only how many students got correct answers, but about the level of confusion in the state of the class's knowledge. (Bao & Redish, 2006)

The class density matrix contains information on the students' responses to the questionnaire. An eigenvalue decomposition allows for trends in the data to be identified. In particular, Bao and Redish identify two situations in which an eigenvalue decomposition is particularly useful. First, if a large majority of the students have similar model states, the eigenvalue decomposition will yield one large eigenvalue, and the eigenvector associated with this value will be indicative of the model state vectors held by the majority. In this context, an eigenvalue is considered "large" if it is greater than 0.65. Second, if there are two subgroups of students whose collective model state vectors are close to orthogonal, the eigenvalue decomposition will yield two dominant eigenvalues, and the associated eigenvectors will be indicative of the model state vectors held by each subgroup of students. In this way, the eigenvalue decomposition allows for information about the class as a whole to be extracted from the data.

3 Results of a Pilot Study

In order to study the utility of model analysis in analyzing proof schemes, a small pilot study was conducted. This study was conducted in two sections of an introductory proof writing course at a large public university in California. The questionnaire described in Step 2, above, was distributed to students in both sections at the beginning and again at the end of the semester. A total of thirty-four students completed the questionnaire at both times.

The model analysis yields class density matrices and eigenvalue decompositions. Recall that in Step 3 of the model analysis, two data sets were created: the first data set uses responses in which only the most convincing proof was marked by the student, and the second uses responses in which the student marked all of the proof he or she found to be convincing. I will refer to the first data set as the "single-response" data set, and the second as the "multiple-response" data set.

The results of the model analysis using the single-response data set are below:

Pre-test class density matrix:

$$\begin{bmatrix} 0.2176 & 0.0982 & 0.2499 & 0.0485 \\ 0.0982 & 0.1353 & 0.1224 & 0.0420 \\ 0.2499 & 0.1224 & 0.5647 & 0.0774 \\ 0.0485 & 0.0420 & 0.0774 & 0.0824 \end{bmatrix}$$

Post-test class density matrix:

$$\begin{bmatrix} 0.2059 & 0.0176 & 0.2851 & 0.0083 \\ 0.0176 & 0.0176 & 0.0185 & 0 \\ 0.2851 & 0.0185 & 0.7118 & 0.0751 \\ 0.0083 & 0 & 0.0751 & 0.0588 \end{bmatrix}$$

Pre-test eigenvalue decomposition:

0.7497	0.1262	0.0690	0.0552
0.4550	0.4209	0.6655	0.4158
0.2505	0.7097	−0.2361	−0.6147
0.8419	−0.4882	−0.1699	−0.1552
0.1465	0.2845	−0.6874	0.6520

Post-test eigenvalue decomposition:

0.8472	0.0923	0.0389	0.0157
0.4060	0.8006	0.4153	0.1472
0.0289	0.1121	0.1056	−0.9876
0.9089	−0.3113	−0.2749	−0.0381
0.0908	−0.4995	0.8607	0.0380

The model analysis provides a large amount of information, so some interpretation is necessary. The diagonal entries of the pre-test class density matrix show that even at the beginning of the course, the student's model space vectors were largely in the direction of the deductive category of proof scheme, but that there was a significant element of external conviction proof scheme in the model space vectors as well. Furthermore, the off-diagonal entry "0.2499" shows that a number of students were inconsistent; these students marked both external conviction responses as well as deductive responses. The off-diagonal entry "0.0982" also exceeds half the product of the square roots of the corresponding diagonal entries, indicating that a portion of students marked both external conviction and empirical responses to the questionnaire.

However, by the end of the semester, very few students marked empirical responses at all, as indicated by the diagonal entry corresponding to the empirical category. The strength of the deductive category increases, but there was very little change in the level of external conviction responses. In fact, there was a slight increase in the entry showing student inconsistencies between the external conviction and deductive categories.

The eigenvalue decompositions show similar information about the class. Both the pre-test and post-test show that the class largely had a single dominant eigenvector, indicating that a majority of students had similar model state vectors, with a large entry in the deductive category and a smaller entry in the external conviction category.

The results from the multiple-response data are similar, except for a higher rate of empirical responses in both the pre-test and post-test. The class density matrices and eigenvalue decompositions are below:

Pre-test class density matrix:

$$\begin{bmatrix} 0.2146 & 0.1946 & 0.2905 & 0.0528 \\ 0.1946 & 0.2221 & 0.2985 & 0.0489 \\ 0.2905 & 0.2985 & 0.5182 & 0.0608 \\ 0.0528 & 0.0489 & 0.0608 & 0.0450 \end{bmatrix}$$

Post-test class density matrix:

$$\begin{bmatrix} 0.2254 & 0.1204 & 0.3240 & 0.0254 \\ 0.1204 & 0.1168 & 0.1932 & 0.0152 \\ 0.3240 & 0.1932 & 0.6140 & 0.0640 \\ 0.0254 & 0.0152 & 0.0640 & 0.0438 \end{bmatrix}$$

Pre-test eigenvalue decomposition:

0.8956	0.0553	0.0271	0.0220
0.4606	0.4680	0.0772	0.7503
0.4710	0.3450	−0.6864	−0.4337
0.7445	−0.5897	0.2896	−0.1190
0.1088	0.5606	0.6626	−0.4846

Post-test eigenvalue decomposition:

0.8724	0.0592	0.0386	0.0298
0.4720	0.4397	0.4927	0.5840
0.2888	0.6547	−0.6820	−0.1510
0.8287	−0.4347	0.0116	−0.3524
0.0838	−0.4348	−0.5403	0.7156

These results show that at the beginning of the semester, students marked empirical responses with almost equal probability to external conviction responses. Furthermore, there was a high level of interaction between the empirical category and deductive category, between the external conviction and empirical categories, and between the external conviction and deductive categories. This shows an extremely high level of inconsistency in student responses.

A comparison of these results to those of the single-response data shows that the empirical category is stronger in the multiple-response results. This indicates that students were willing to mark empirical responses as convincing, but not as the *most* convincing argument. The eigenvalue decomposition shows a similar increase in the empirical direction. Also, the largest eigenvalue is quite large in both the pre- and post-test results.

4 Discussion

While the results of the pilot study are interesting, caution should be taken when interpreting the results. Because the questionnaire has not been validated through student interviews, the results should not be considered to be an accurate description of the proof schemes used by the students. In particular, the high level of inconsistency indicated by the multiple-response results suggests that the questionnaire may not accurately reflect student proof schemes. Instead, the results should be thought of as a sort of "proof of concept" for model analysis. That is, the results show that model analysis can be a valuable way to examine the proof schemes held by students.

The students' responses to a questionnaire like the one used in this study can be analyzed in many other, more traditional ways. For instance, the pre- and post- data can be analyzed to see if there are any significant changes in the rate of responses in any of the categories of proof scheme. What model analysis adds is the ability to analyze students *inconsistencies*. The results of this pilot study show a high level of inconsistency in students, particularly between the external conviction and deductive categories of proof scheme, which is not captured by simpler quantitative analysis.

The eigenvalue decompositions make a significant contribution as well. In the results of this pilot study, the dominant eigenvectors indicate that the class had a single dominant understanding of proof schemes, rather than one group holding deductive proof schemes and a second group holding external conviction proof schemes. The ability to distinguish between a uniform community of students and a split community is very valuable, and is not captured by traditional statistical methods.

In order to conduct a proper model analysis of proof schemes, a questionnaire needs be developed and validated. Once we can be sure that the questionnaire accurately reflects student thought, model analysis will allow for a new way of capturing important information about the proof schemes of students.

It should be noted that the method of model analysis, as described here, can only measure students responses to proofs written by others, rather than the proofs that students write themselves. That is, it is possible to use model analysis to measure the proof schemes that students use when ascertaining the validity of a proof, but not when trying to persuade another of the truth of a theorem.

Even with this limitation, model analysis appears to be a valuable method for analyzing student proof schemes. When a reliable instrument for gathering data is available, model analysis will give valuable insights into the proof schemes held by students.

References

Bao, L. (1999). *Dynamics of student modeling: A theory, algorithms, and application to quantum mechanics*. University of Maryland, College Park: Doctoral dissertation.

Bao, L., & Redish, E. F. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics*, *69*(S1), S45-S53.

Bao, L., & Redish, E. F. (2006). Model analysis: Representing and assessing the dynamics of student learning. *Physical Review Special Topics – Physics Education Research*, *2*(1), 010103.

Ellis, A. B. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education*, *38*(3), 194–229.

Harel, G., & Sowder, L. (1998). Students proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–282). American Mathematical Society: Washington, D.C.

Harel, G., & Sowder, L. (2007). Toward a comprehensive perspective on proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* National Council of Teachers of Mathematics.

Housman, D., & Porter, M. (2003). Proof schemes and learning strategies of above-average mathematics students. *Educational Studies in Mathematics*, *53*, 139–158.

Solomon, Y. (2006). Deficit or difference? the role of students epistemologies of mathematics in their interactions with proof. *Educational Studies in Mathematics*, *61*(3), 373–393.

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics*, *56*(2/3), 209–234.