The Q-matrix Method:
Mining Student Response Data for Knowledge

Tiffany Barnes

Department of Computer Science
University of North Carolina at Charlotte
Charlotte, NC 28223
tbarnes2@uncc.edu

Abstract

Although many talented researchers have created excellent tools for computer-assisted instruction and intelligent tutoring systems, creating high-quality, effective, scalable but individualized tools for learning at a low cost is still an open research challenge. Many learning tools create complex models of student behavior that require extensive time on the part of subject experts, as well as cognitive science researchers, to create effective help and feedback strategies. In this research, we propose a different approach, called the q-matrix method, where data from student behavior is "mined" to create concept models of the material being taught. These models are then used to both understand student behavior and direct learning paths for future students. We describe the q-matrix method and present preliminary results that imply that the method can effectively predict which concepts need further review.

Introduction

Computer-aided instruction has great promise in expanding the accessibility of high quality education. In particular, it is the capability to adapt to individual students that makes intelligent tutoring systems such as those in (Conati, et al., 2002; Heffernan & Koedinger, 2002; Van Lehn & Martin 1998) attractive. However, the majority of these systems require the construction of complex models that are applicable only to a specific tutorial in a specific field, requiring the time of experts to create and then test these models on students. In fact, these are only a few of the tradeoffs ITS system developers face (Murray 1999). One system, REDEEM, was built to ameliorate the time needed to create an ITS, and allow teachers to apply their own teaching strategies in an existing computer-based training (CBT) system, and has been shown to be more effective than a non-expert human tutor in improving student test scores (Ainsworth, et al., 2003). Another system, ASSERT, was built to replace the need for expert model design, using theory refinement to learn student models from behavior, and generate feedback for remediation (Baffes & Mooney, 1996).

The q-matrix method, has the same goal of being able to quickly modify an existing CBT tool to be a simple ITS. The method employs knowledge discovery and data mining techniques to model student knowledge and direct knowledge remediation. In this work, we briefly discuss the history of the q-matrix method, describe the q-matrix algorithm and evaluations, and present two experiments that demonstrate the application of the method in understanding student behavior and directing learning.

Background and related work

The original inspiration for the q-matrix method came from Tatsuoka et al., who explored student misconceptions in basic math concepts, such as adding fractions (Birenbaum, et al. 1993; Tatsuoka, 1983). The main goal of this research was diagnosis of students' misconceptions, which could be used to guide remediation, assess group performance as a measure of teaching effectiveness, and discover difficult topics (Birenbaum, et al. 1993). Tatsuoka developed a rule space, based on a relatively small set of rules and ideas, in which hypothesized expert rules and actual student errors in fraction addition could be mapped and compared. For example, for -1 + -7, one "rule" is that the student might add the two absolute values. This answer, 8, would them be compared with student answers. This space allowed instructors to map student errors without having to catalog every possible mistake. The expert point in rule space closest to the student response corresponds to the rule the student is assumed to be using. This method improves on other procedural models, by creating a space where all student responses can be compared to expert predictions.

This idea of determining a student's knowledge state from her test question responses inspired the creation of a q-matrix, a binary matrix showing the relationship between test items and latent or underlying attributes, or concepts (Birenbaum, et al., 1993). Students were assigned knowledge states based on their test answers and the constructed q-matrix. An example of a binary q-matrix is given in Table 1. A q-matrix, or "attribute-by item incidence matrix", contains a one if a question is related to the concept, and a zero if not. For example, in this q-matrix, questions q1 and q6 are both related by concept con1, while q1 is also related to q2 and q4 by concept con2. Brewer extended these to values ranging from zero to one, representing a probability that a student will answer a question incorrectly if he does not understand the concept (1996).

Table 1: Example q-matrix

	Questions					
	q1	**q2**	**q3**	**q4**	**q5**	**q6**
con1	1	0	0	0	0	1
con2	1	1	0	1	0	0
con3	1	1	1	0	0	0
con4	1	0	1	0	0	0

Tatsuoka's rule space research showed that it is possible to automate the diagnosis of student knowledge states, based solely on student item-response patterns and the relationship between questions and their concepts. Through promising, the rule space method is very time consuming and topic-specific, and requires expert analysis of questions. The rule space method provides no way to measure or validate that the relationships derived by experts are in fact those used by students, or that different experts will create the same rules.

In 1992, Hubal studied the correspondence between expert-derived q-matrices and student data, and found that these two did not necessarily coincide. In 1996, Brewer created a method to extract a q-matrix from student data, and found that the method could be used to recover knowledge states of simulated students. Sellers applied the q-matrix method to small groups of students (1998). In (Barnes & Bitzer, 2002), we applied the method to larger groups of students, and in (Barnes, et al., 2005) found the method comparable to standard knowledge discovery techniques for grouping student data. In particular, the method outperformed factor analysis in modeling student data and resulted in much more understandable q-matrices, but had higher error than k-means cluster analysis on the data. However, cluster analysis is not as suitable for automated direction of student learning as the q-matrix method, because human intervention would usually be required to create behaviors to associate with each cluster.

Q-matrix algorithm

The q-matrix algorithm, as devised by Brewer in 1996, is a simple hill-climbing algorithm that creates a matrix representing relationships between concepts and questions directly from student response data. The algorithm varies c, the number of concepts, and the values in the q-matrix, minimizing the total error for all students for a given set of n questions. To avoid of local minima, each hill-climbing search is seeded with different random q-matrices and the best of these is kept.

First, c, the number of concepts, is set to one, and a random q-matrix of concepts versus questions is generated with values ranging from zero to one. We then cluster student response data according to "concept states", and compute the total error associated with assigning students to concept states, over all students.

After the error has been computed for a q-matrix each value in the q-matrix is changed by a small amount, and if the overall q-matrix error is improved, the change is saved. This process is repeated for all the values in the q-matrix several times, until the error in the q-matrix is not changing significantly.

After a q-matrix is computed in this fashion, the algorithm is run again with a new random starting point several times, and the q-matrix with minimum error is saved, to avoid falling into a local minimum. It is not guaranteed to be the absolute minimum, but provides and acceptable q-matrix for a given number of concepts.

To determine the best number of concepts to use in the q-matrix, this algorithm is repeated for increasing values of c. The final q-matrix is selected when adding an additional concept does not decrease the overall q-matrix error significantly, and the number of concepts is significantly smaller than the number of questions. This is comparable to the "elbow" criterion in choosing the number of factors for a factor analysis. For this study, q-matrices with an error rate of less than 1 per student were selected. Other built-in criteria could also be used to protect from over-fitting the data.

Q-matrix evaluation

In the q-matrix method, student responses are grouped into clusters by concept states. Each cluster in the q-matrix method is represented by its concept state, a vector of bits where the ith bit is 0 if the students do not understand concept i, and a 1 if they do. Each concept state also has associated with it an ideal response vector (IDR). We use the concept state with the q-matrix to determine the IDR. For each question q in the q-matrix we examine the concepts needed to answer that question. If the concept state contains all those needed for q, we set bit q in the IDR to 1, and otherwise to 0. When the q-matrix contains only binary values (not probabilities between 0 and 1), this can be calculated for a concept state c and the q-matrix Q by the following procedure, composing $\neg c$ with Q:

$$IDR = \neg ((\neg c)Q)$$

For example, given concept state c = 0110 and the q-matrix Q given in Table 1, $\neg c$ = 1001, $(\neg c)Q$ = 101001. Therefore, IDR = $\neg((\neg c)Q)$ = 010110. This can be explained by viewing $(\neg c)Q$ as all the questions that require knowledge in the concepts that are unknown for a student in concept state c. Thus, the IDR for c is exactly the remaining questions, since none of these require concepts that are unknown for a student in concept state c.

When the q-matrix consists of continuous probabilities, we compute the IDR as explained above, but the negation symbol is interpreted as the probability of the opposite outcome, so in each case where a not appears, we interchange any following values x with 1-x.

A q-matrix is evaluated on its fit to a set of student responses, and is measured as error per student. We now describe the error computation method. First, we create an array whose indices are answer vectors, from 0 up to 2^q-1, where q is the number of questions in the tutorial. We then tally the number of student responses with each answer vector. Then, for each response with at least one student, we compare the response with all IDRs and choose the one closest in Hamming distance. This distance is called the "error" for the student response. We sum all errors over all students to determine the overall error for the q-matrix.

Method

In this research, we apply the q-matrix method to a large group of students and analyze the results in several ways. Our hypotheses are: 1) expert and extracted q-matrices will differ, but extracted q-matrices will be useful in interpreting student data, and 2) the q-matrix method can be effective in guiding student remediation (at least as effective as a student directing his own learning).

The Binary Relations tutorial was administered to over 100 students in the Fall 2002 Discrete Mathematics course, CSC 226, at NC State University. The tutorial was required for credit in the class, but only a completion grade was assigned to each student. Students were randomly assigned to guided or self-guided groups. In the guided group, after taking a tutorial quiz, students were automatically directed to review a question it was determined that the student least understood. In the self-guided group, students were asked to choose which question on the quiz they least understood, but were not redirected to review any particular items. Students in both groups were allowed to review any topics they wished after either a q-matrix redirection or listing their least-understood question. To receive credit for taking the tutorial, students were also required to complete a survey, which asked students whether they felt the tutorial knew which questions they least understood.

The q-matrices used for directing guided students were initialized to those found by Sellers in 1998. After students took the tutorials, their new data was used to extract a new set of q-matrices, which are compared with expert models, to check our hypothesis that student concepts do not always correspond to expert analysis. These extracted q-matrices are also examined to investigate our hypothesis that they can be used to understand student behavior. Finally, the choices that students made for remediation are compared to those made by the q-matrix method, and survey results relevant to remediation are presented.

Results and Discussion

Comparison of model with expert analysis

Jennifer Sellers (1998) designed the binary relations tutorial and conducted an experiment to determine if the q-matrix extracted from student responses corresponded well with expert-created q-matrices. With her small sample of students, she found that Sections 1 and 2 corresponded well with expert analysis but that section 3 did not. We further tested the hypotheses that the q-matrices formed through analyzing student responses will be interpretable and perhaps also comparable to an expert analysis of the subject. In this section we compare the three q-matrices extracted for the Binary Relations tutorial in Fall 2002 with the expert q-matrices generated in Sellers' experiment.

Expert v. extracted q-matrices for the binary relations tutorial section 1

Section 1 of the binary relations tutorial covers Cartesian products and binary relations in general. The topics in this section are listed in Table 2. In this section we will discuss how experts related these topics (and their corresponding questions) by concepts, and compare this to the q-matrix extracted from the fall 2002 discrete math class.

Table 2: Binary Relations Tutorial Topics, Section 1

Section 1	**Cartesian Products & Binary Relations**
Question	**Topic**
q1.1	Cartesian product (one point)
q1.2	Cartesian product (all points)
q1.3	Relations as subsets of Cartesian products
q1.4	Matrix representations of relations
q1.5	Composite of two relations

Table 3 lists the Section 1 matrix extracted by the q-matrix method both in Sellers' work and in our fall 2002 experiment. Table 4 lists the Section 1 matrix created by 3 instructors in Sellers' experiment. Concept 1 in Table 3, the extracted q-matrix, corresponds to concept 3 in Table 4Table 4, the expert q-matrix. The extracted q-matrix tells us several things: most students answered questions 1-4 correctly, because their predicted answers to those questions would be correct answers. Students found question 1.5, that on composite relations, the most difficult. This corresponds with the expert analysis of this tutorial. It is no surprise that the q-matrix method did not find the expert concept 1, since this concept consists of a relationship to all the tutorial questions. In other words, this concept was the over-arching concept that relates all questions in this tutorial, so most students would understand the main concept by the time they took the quiz, and their answers would not reflect this relationship among all the questions. Concept 2 in the expert matrix relates those questions dealing with relations, which are subsets of Cartesian products. In this tutorial, there was not enough difference in responses to extract this concept.

Table 3: Binary Relations Section 1 Extracted q-matrix

	q1.1	q1.2	q1.3	q1.4	q1.5
con 1	0	0	0	0	1

Table 4: Binary Relations Section 1 Expert q-matrix

	q1.1	q1.2	q1.3	q1.4	q1.5	Description
con 1	1	1	1	1	1	Cartesian Products
con 2	0	0	1	1	1	Relations
con 3	0	0	0	0	1	Composites

Concept 1 represents question 5, which is composition of relations. This is a difficult concept for most students to understand. This q-matrix agrees with that determined for Section 1 in Seller's thesis, both by q-matrix analysis and by expert analysis.

Expert v. extracted q-matrices for the binary relations tutorial section 2

The topics taught in section 2 of the Binary Relations tutorial are given in Table 5. This section gives definitions and examples of the properties of binary relations. In this section we discuss how experts related these topics by concepts, and compare this to the q-matrix extracted from the fall 2002 discrete math class.

Table 5: Binary Relations Tutorial Topics, Section 2

Section 2	Properties of Binary Relations
Question	Topic
q2.1	Reflexive
q2.2	Symmetric
q2.3	Irreflexive
q2.4	Antisymmetric
q2.5	Asymmetric
q2.6	Transitive
q2.7	Equivalence relation
q2.8	Partially-ordered set (poset)

Jennifer Sellers compared her extracted q-matrices to those constructed by instructors. For two of her q-matrices, expert and extracted q-matrices corresponded in all but one matrix value. In Table 6, element (2,6) shows the only difference between the expert and extracted (auto) q-matrices for one set of questions. In this case, instructors indicated that concept 2 indicated the application of more than one definition, and had a zero in position (con2, q2.6) since it corresponds to applying the definition of a transitive relation. However, the extracted q-matrix, with a one in that position, reflects that students find applying this definition much more complex than applying any combination of two other properties of binary relations. Upon a second look, experts agreed that this extracted q-matrix was appropriate.

Table 6 Binary Relations Section 2 Sellers' Expert v. Extracted Q-matrices

	q2.1	q2.2	q2.3	q2.4	q2.5	q2.6	q2.7	q2.8	Descrip.
con 1	1	1	1	1	1	1	1	1	Special properties of relations
con 2	0	0	0	0	1	exp 0/ auto 1	1	1	Multiple properties

In the analysis of fall 2002 data, we extracted a q-matrix with 4 concepts. In Seller's work, with only 17 students, it was more appropriate to extract a q-matrix with fewer concepts. With 2 concepts, we have 4 concept states available and expect about 4 students to be assigned to each concept state. With so few students, even a 1-concept q-matrix might have been more appropriate. With 142 students, the fall 2002 data can support a more complex model. With 4 concepts, there are 16 concept states available, so we'd expect an average concept state to contain about 8-9 students each. If we had used 3 concepts, we'd expect about 18 students in each of 8 states, perhaps a bit high for our purposes.

The 4-concept q-matrix for fall 2002 is given in Table 7. This q-matrix is quite different from those given by both experts and Seller's previous analysis. In this matrix, questions on antisymmetric (q2.4), equivalence relations (q2.7), and partially ordered sets (q2.8) were all partitioned into separate concepts. These questions were the biggest separators among students, since 74 out of 142 students missed only one question. Out of the remaining students, 39 missed only two questions. This skew towards missing only one question caused the q-matrix analysis to work harder to separate these large groups of students into groups with the lowest error. This resulted in the simple separation by the questions causing the most single errors.

In Table 7, concept 1 relates questions on irreflexive (q2.3), asymmetric (q2.5), and transitive (q2.6) properties of binary relations. Out of the properties given in this section, asymmetric and irreflexive are the two properties whose definitions contain negations. These are notorious for causing student difficulty. However, these two properties are relatively easy to check in a matrix. Transitivity, on the other hand, has a more complex definition, and is also difficult to check for in a matrix.

It makes sense for the irreflexive and asymmetric properties to be paired, since an asymmetric relation must be both antisymmetric and irreflexive. Antisymmetric (q2.4) was not paired with these since it was grouped on its own to explain students who missed this as the only question they answered incorrectly. The transitive property (q2.5) was probably grouped with irreflexive and asymmetric since, if a student missed more than one question, it was probably in this group of three questions.

Concept 2 is only for the antisymmetric property, which causes the most difficulty for students in remembering and applying its definition. Concept 3 represents only equivalence relations, and concept 4 represents only partially ordered sets. Answering these questions on a matrix is much more complex than the other properties since each of these requires the student to check for three properties. Equivalence relations must be reflexive, symmetric, and transitive. Partially ordered sets must be reflexive, antisymmetric, and transitive. The difficulty in checking each of these usually lies in checking for transitivity. In addition, it is easy to make a mistake in checking for so many properties in one question.

Table 7: Binary Relations Section 2 Fall 2002 Extracted q-matrix, 4 concepts, Err/stud: 0.74

	q2.1	q2.2	q2.3	q2.4	q2.5	q2.6	q2.7	q2.8
con 1	0	0	1	0	1	1	0	0
con 2	0	0	0	1	0	0	0	0
con 3	0	0	0	0	0	0	1	0
con 4	0	0	0	0	0	0	0	1

We have also listed the q-matrices for Section 2 fall 2002 data with 3 concepts in Table 8, and for 2 concepts in Table 9. As we stated before, the 3-concept model for this problem would place about 18 students in each concept state, and we'd expect about 36 students per concept state for the 2-concept model. As we see below, the errors for each of these are more than 1 per student. The two-concept q-matrix does not correspond to the 2-concept expert q-matrix given in Table 6.

Although each run of the q-matrix method starts with random values, it is often the case that q-matrices with increasing numbers of concepts have similar or even identical concepts. As we can see by comparing Table 9 with Table 6, concepts 1 and 4 in the 4-concept q-matrix correspond exactly with the concepts in the 2-concept q-matrix. Similarly, concept 3 in the 3-concept q-matrix (Table 8) appears again in the 4-concept model (Table 7) as concept 2.

Table 8: Binary Relations Section 2 Fall 2002 Extracted q-matrix, 3 concepts, Err/stud: 1.11

	q2.1	q2.2	q2.3	q2.4	q2.5	q2.6	q2.7	q2.8
con 1	0	0	0	0	1	1	0	1
con 2	0	0	1	0	0	1	1	0
con 3	0	0	0	1	0	0	0	0

Table 9: Binary Relations Section 2 Fall 2002 Extracted q-matrix, 2 concepts, Err/stud: 1.27

	q2.1	q2.2	q2.3	q2.4	q2.5	q2.6	q2.7	q2.8
con 1	0	0	0	0	0	0	0	1
con 2	0	0	1	0	1	1	0	0

The recurring nature of concepts in increasing q-matrix models indicates that the q-matrix models created are fairly robust. In addition, although the q-matrix method is a heuristic hill-climbing method, running the q-matrix analysis again has almost always returned the same q-matrix, even though the q-matrices are started with random values each time. This is another indication that the q-matrix method may be a robust algorithm.

Expert v. extracted q-matrices for the binary relations tutorial section 3

Section 3 of the binary relations tutorial covers Hasse diagrams and properties of these diagrams. The topics in this section are listed in Table 10. In this section we will discuss how experts related these topics (and their corresponding questions) by concepts, and compare this to the q-matrix extracted from the fall 2002 discrete math class.

Table 10: Topics in Binary Relations Tutorial Section 3

Section 3	Posets and Hasse Diagrams
Question	**Topic**
q3.1	Hasse Diagrams
q3.2	Maximal elements
q3.3	Minimal elements
q3.4	Upper bounds
q3.5	Lower bounds
q3.6	Least upper bound
q3.7	Greatest lower bound

The expert-generated q-matrix for the binary relations tutorial, Section 3, from Sellers' work is given in Table 11. Instructors broke this section of 7 questions down into 3 concepts. The first concept was named "Hasse diagrams" and contained all questions, since this is the general topic of this section. The second concept grouped together questions that examined subsets of partially ordered sets for upper and lower bounds. The third concept grouped together questions that combined the ideas of maximal and minimal elements with upper and lower bounds.

Using the q-matrix method, we would be very unlikely to extract concept 1 from this tutorial. If these questions were grouped with a significant group of tutorial questions, not all of which were based on Hasse diagrams, we might expect the q-matrix method to extract this concept because of the relative relationships among these questions in comparison to other quiz topics. When comparing this matrix to that extracted for fall 2002, we notice that q3.6 is singled out both in the expert and the extracted q-matrices. This agreement between instructors and the tutorial indicates that, as perceived by experts and by students, questions relating to lower bounds are likely to give students trouble. We also see q3.2 and q3.4 grouped together in concepts in both models. This means that more students are missing questions on maximal elements and upper bounds together, suggesting that students may have a hard time interpreting the meaning of upward movement in a Hasse diagram.

Table 11: Binary Relations Section 3 Expert q-matrix

	q3.1	q3.2	q3.3	q3.4	q3.5	q3.6	q3.7	Expert Descrip.
con 1	1	1	1	1	1	1	1	Hasse diagrams
con 2	0	1	1	1	1	0	0	Groups of elements
con 3	0	0	0	0	0	1	1	Max & min

194 students completed the Binary Relations Section 3 quiz, and we extracted a 3 concept q-matrix for this data. We have 8 concept states, so we expect an average of 25 students per state. This is quite a large number for our error to be less than one per student on this section, indicating that there was quite a bit of overlap in student responses. In fact, there were only 78 distinct answers on this tutorial. The q-matrix extracted from fall 2002 data for the binary relations tutorial, section 3 is given in Table 12. When we compare this q-matrix with the expert q-matrix in Table 11, we find that concept 3 is similar in both of these – in both, q3.6 and q3.7 are related, but in the fall 2002 q-matrix, these are also related to q3.4.

In the fall 2002 q-matrix, concept 3 combines questions q3.4, q3.6, and q3.7. These questions are on upper bounds, least upper bounds, and greatest lower bounds. Concepts 2 and 3 both refer to q3.4 and q3.6, implying that these two questions, by being related to 2 concepts each, are more complex than other questions. This concept grouping would indicate to instructors that upper bounds and least upper bounds were not as well understood as the other questions. This might encourage instructors to add more practice working with these ideas, or improving these two sections of the tutorial.

In Table 12, concepts 1 and 4 both select one question each: concept 1 represents only q3.1 on Hasse diagrams. and concept 4 contains only q3.5 on lower bounds. This suggests that some significant groups of students had difficulty with these two questions, or that these questions were missed alone in the tutorial (i.e. no other questions were missed when a student missed one of these).

Concept 2 combines questions q3.2, q3.4, and q3.6 into one concept. These questions are on maximal elements, upper bounds, and least upper bounds. This combination is significant, since least upper bounds depend on understanding upper bounds. We might also suppose that, due to the arrangement of Hasse diagrams, with edges leading upward, that students have more difficulty interpreting behavior in the poset moving up the diagram, as in finding a maximum or an upper bound.

Table 12: Binary Relations Section 3 Fall 2002 Extracted q-matrix, 4 concepts, Err/stud: 0.72

	q3.1	q3.2	q3.3	q3.4	q3.5	q3.6	q3.7
con 1	1	0	0	0	0	0	0
con 2	0	1	0	1	0	1	0
con 3	0	0	0	1	0	1	1
con 4	0	0	0	0	1	0	0

Summary of comparison of expert and extracted q-matrices

In this section, we have compared the extracted q-matrices with expert q-matrices created for the Binary Relations tutorial. We have found that there is sometimes overlap in the expert and extracted q-matrices, but often these q-matrices do not correspond. Since the design of the q-matrix method of extraction was based on the inaccuracy of expert q-matrices in explaining student data, we would not expect expert q-matrices to correspond particularly well with those extracted from student data. In some cases, we did find a correspondence, and this usually occurred on the most difficult or complex questions.

In addition to comparing the expert and extracted q-matrices, we also examined the extracted q-matrices to understand the relationships in the questions in this tutorial. In both sections 2 and 3, the extracted q-matrices separated out several questions into one concept each. In this tutorial, we would expect something like this since the questions here are very much the application of definitions. Many students taking this section missed only 1 or 2 questions at a time, making a q-matrix with several 1-question concepts a good model. As we hypothesized, understanding the relationships shown in the q-matrices was not difficult, and our interpretations can be used to both understand student data and determine which questions were most difficult.

Comparison of predicted direction vs. student for remediation

In this experiment, the q-matrix method was effective in automatically choosing questions that students least understood. When compared with the choices that students made on their own, the q-matrix method chose the same question as students did for more than half the students in all three sections of the binary relations tutorial, as demonstrated in Table 13. Our intention was to have half of the students choose their own progress, and half to be self-guided, concurrently-running students were often assigned the same value for this choice, so fewer students guided their own remediation process.

Table 13 Student v. q-matrix Selection of Least-Understood Question

	Section 1	Section 2	Section 3
Total Students	255	251	246
Auto-guided	106	204	199
Self-guided	149	47	47
# self-guided who chose other than predicted least understood concept	11	14	20
Percentage	7.38%	29.78%	42.55%

In this experiment, random students were chosen to select the question they least understood upon completion of each section of the binary relations tutorial. Most of these students chose a question related to the q-matrix predicted "least understood concept". In Table 13, we see that on section 1, only 7% of self-guided students chose to review a different question than selected by the q-matrix. Since our q-matrix was a one-concept q-matrix relating only to question 5, the most difficult on the quiz, this is not a surprising result. On section 2, 30% of self-guided students chose a different question to review than the q-matrix would have. This means that 70% of the time the q-matrix method predicted the same concept to review as students did. On section 3, however, 43% of students chose differently than the q-matrix method would have. Since this number is high, we cannot conclude that the q-matrix choice corresponds to student choices often. This result is not discouraging, though, since on a more complex topic such as this one, students may need to review more than one concept. In order to better measure the choices made by the q-matrix method, for students who chose to review questions that the system would not have chosen for them, we examine student performance on these two questions on the final exam.

Table 14 lists the performance on the final exam of those students who chose differently than predicted. (Since the final exam was not required, this data set is much smaller than the number of students who took the tutorial). On the final exam, these students performed equally or worse on the q-matrix predicted questions than they did on the questions they selected as those they least understood. This suggests that sometimes, a student may not realize when he should review a particular topic.

In this small sample, the q-matrix method did at least as well as the students did in choosing their "least understood question". Since we had such a small sample size, we cannot conclude more from these data. We also note that, since the final exam was optional, and there may be several confounding effects that determine student scores, so these results should be interpreted as suggestive and not conclusive. Further experiments will be needed to determine if the "least-understood question" method is truly a good educational choice.

Table 14: Final Exam for Self-Directed Students

	Section 1	Section 2	Section 3
# students with diff choices	11	14	20
# of these with finals	7	10	13
# who missed the self-directed question on the final exam	0	2	0
# who missed the q-matrix predicted question	2	2	1
# who performed the same on both	5	6	12

In the tutorial survey, students were asked to agree or disagree with the statement, "I felt like the program knew which concepts I did not understand, and directed me back to lessons on concepts I understood the least." Forty-six percent of students agreed that the Binary Relations tutorial seemed to know what they did not understand and directed them to study the concepts they least understood. Only 13% of students disagreed, feeling that the tutorial did not know which concepts they least understood.

Six percent felt that the best aspect of the tutorial was its ability to take students back to review material they did not understand. Some of these pointed out that it would be more useful to review these sections a different way, or that they wanted to review more than just one topic after taking a section test. Future work will address these concerns.

Final Remarks

This research represents an initial study of the effectiveness of using one data mining technique, the q-matrix method, in understanding and directing student learning. As predicted, expert and extracted q-matrices did not often coincide, but we were still able to understand student responses based on extracted q-matrices. We also compared the questions that self-guided students chose to review with the questions that the q-matrix method would have chosen for them. We found that the q-matrix method often chose the same questions for review as the self-guided students chose for themselves. We also found that students who chose differently than the q-matrix method could have benefited from reviewing a q-matrix selected concept. However, due to the small sample size of self-guided students, choosing to review different questions than would have been suggested, who also took the final exam, these findings are preliminary and not conclusive.

Future work will address several important questions about the q-matrix method. Although the method has been validated using simulated students (Brewer, 1996), a comparison of q-matrix results on varying class sizes would yield a measure of the robustness of the method. We also plan to compare error for expert models on explaining student performance with q-matrix models.

References

Ainsworth, S.E., Major, N., Grimshaw, S.K., Hayes, M., Underwood, J.D., Williams, B. & Wood, D.J. 2003. REDEEM: Simple Intelligent Tutoring Systems From Usable Tools, in T. Murray, S. Blessing & S.E. Ainsworth (eds). *Advanced Tools for Advanced Technology Learning Environments*. pp. 205-232. Amsterdam: Kluwer Academic Publishers.

Baffes, P. & R.J. Mooney. 1996. A Novel Application of Theory Refinement to Student Modeling. *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (AAAI-96), pp. 403-408, Portland, OR, August 1996.

Barnes, T. & Bitzer, D. 2002. Fault Tolerant Teaching and Automated Knowledge Assessment. *Proceedings of the ACM Southeast Conference (ACMSE 2002)*, Raleigh NC, April 2002.

Barnes, T., D. Bitzer, & M. Vouk. 2005. Experimental analysis of the q-matrix method in knowledge discovery. To appear in *International Symposium on Methodologies for Intelligent Systems*, Saratoga Springs, MY, May 2005.

Birenbaum, M., Kelly, A., & Tatsuoka, K. 1993. Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education*, 24(5), 442-459.

Brown, J.S. & Burton, R. 1978. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-191.

Conati, C., A.S. Gertner, & K. VanLehn. 2002. Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Model. User-Adapt. Interact*. 12(4): 371-417.

Heffernan, N.T. & K.R. Koedinger. 2002. An Intelligent Tutoring System Incorporating a Model of an Experienced Human Tutor. *Intelligent Tutoring Systems,* p. 596-608.

Hubal, R. 1992. Retro-adaptive testing and evaluation system. Computer Science Masters Thesis, North Carolina State University.

Murray, Tom. 1999. Authoring Intelligent Tutoring Systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education* (1999), 10: 98-129.

NovaNET educational network. Online. http://www.pearsondigital.com/novanet/

Sellers, J. 1998. An empirical evaluation of a fault-tolerant approach to computer-assisted teaching of binary relations. Computer Science Masters Thesis, North Carolina State University.

Tatsuoka, K. 1983. Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.

VanLehn, K., & Martin, J. 1998. Evaluation of an assessment system based on Bayesian student modeling. *International Journal of Artificial Intelligence and Education*, 8(2), 179-221.