Abstract- Distance education, reaching students through accessible, high quality, individualized instruction is a major challenge for today’s universities. Increased access to computers and the Internet has made this idea more feasible, but researchers are still working on the most effective ways to deliver individualized instruction in inexpensive, expandable ways. This paper explores the evolution of the promising field of fault tolerant teaching (FTT). When fully developed, fault tolerant teaching systems will automatically assess student knowledge and use this assessment to direct remediation of knowledge. FTT methods use statistical techniques to interpret student responses to questions, and are constructed to tolerate the usual errors that occur during testing – such as a student answering a question correctly without knowing how, or accidentally missing a question they understand well. These methods do not require any knowledge about the instructional topic. This paper presents the evolution of fault tolerant teaching and the promising results of an experiment to test FTT methods.

1 Introduction
College and university campuses are now challenged to deliver quality service to students of widely ranging abilities and backgrounds across all distances. To rise to this challenge, distance education must pair individuals with expert instruction that is interactive, adaptive, and accessible at a distance. The promise of distance education is great, but research into methods of data mining in education is still a new, open, and rich field.

Sections 1.1 and 1.2 discuss how data mining, teaching, and assessment interrelate. Section 1.3 gives an overview of the challenges researchers face in developing online educational systems. In Section 2 we discuss relevant work in student modeling and fault tolerant teaching. Section 3 defines the problem to be solved and presents our research results. Section 4 discusses conclusions and future work.

1.1 Data mining, teaching, and assessment
Data mining is the process of finding hidden characteristics or common concepts in a large amount of data. One complex application of data mining is to determine what a student understands during the course of instruction. Human get things wrong for the right reasons, and get things right for the wrong reasons. The complex task of assessment requires tolerance of such errors while still robustly measuring student knowledge. Instructors perform data mining every day, using student behavior, assignments, and tests to assess each student’s level of knowledge and expertise. From this assessment, instructors adapt their teaching to fill the gaps in student knowledge.

Successful teaching, assessment, and remediation usually require a human with real intelligence; our goal is to perform these key elements of data mining with artificial intelligence. Fault tolerant teaching systems will capture that ability: to assess student knowledge in a way that tolerates performance errors and to effectively guide the teaching process. The Q-matrix method of fault tolerant teaching will provide distance education tools with the ability to adapt tutorials for any subject to all types of students, and can also be used to measure and tailor the effectiveness of a variety of teaching strategies.

1.2 Adaptive tutorial systems and student assessment
The Alberta Research Council (1995) reports that, “in order to allow instruction to be individually designed, it is first necessary to capture the student’s understanding of the subject” [10]. Knowledge assessment is the key to tailoring the learning experience, so it is an essential component of any adaptive teaching system. Without assessment, there is no way to measure the results of teaching, or tailor further education.

“The results of a good test or assessment … represent how a student performs on the objective which those items were intended to assess” [4]. A good assessment, in other words, does not just report those questions a student missed, but offers a stronger reflection of the skills and understanding underlying a student’s performance. The results of a good assessment offer reliability and robustness across time and circumstance [4].

A good assessment can be used to make decisions about the teaching process. Not only will a good assessment provide an adaptive tutorial system with a method to measure student understanding and readiness to move on, it will also provide education researchers with a
method for comparing teaching strategies. Since an assessment provides a measure of success in teaching, we can compare assessments before and after applying several teaching strategies to a large group of students to determine which strategies are the most effective.

1.3 Research challenge
Ideally, computer based educational systems can approximate the benefits of private tutoring at a much lower cost. To deliver this performance, researchers must develop a systematic approach to designing, implementing, and evaluating online educational systems.

One of the main challenges of creating tutoring systems online is knowledge assessment [10]. In other words, a computer tutor must be able to diagnose and correct student misconceptions, and distinguish these from careless errors or guesses. Indeed, much of the research in diagnosing misconceptions acknowledges the importance if distinguishing ‘slips’ from true misconceptions [e.g. 3, 11, 1]. Once student knowledge is assessed, a tutor, human or computer, must then determine how to best lead a student to reach her educational goals.

The Q-matrix method described in this paper examines the inputs of many students to automatically extract relationships between questions and underlying concepts, and then uses those relationships in diagnosing and correcting student misconceptions. The Q-matrix method builds in fault tolerance and robustness, optimizes for student performance, and can be used to optimize teaching strategies for effectiveness. This combination of assessment with the ability to optimize teaching strategies will help make it possible to deliver individualized, high quality education at a distance – near or far.

2 Related work
In this section, we will discuss the work that led to the development of fault tolerant teaching. We first examine rule-based systems used to predict and diagnose student errors in mathematics, including work by Tatsuoka, Brown, VanLehn, and others [e.g. 1, 3, 11, 12]. Inspired by this work, Robert Hubal compared the actual performance of thousands of students to Tatsuoka’s predictions and randomly generated rules [6]. Taking this work to another level of abstraction, Patrick Brewer tested the theory that rules can be derived strictly from student responses without knowledge of the subject area [2]. Susan Jones and Jennifer Sellers wrote interactive lessons using question generators and answer judging, and collected data for small empirical tests of Q-matrix theory [7, 9]. In addition, Sellers compared Q-matrices derived from student data with those determined by subject instructors, testing the interpretability of the extracted “rules” or concepts [9].

2.1 Procedural models of student knowledge
Knowledge assessment is imperative in the design of individualized instruction. One way to assess student knowledge is to create a model of the subject being taught and a corresponding model of the knowledge a student has attained so far. These two models can be used together to evaluate and guide a student’s learning until his knowledge model closely corresponds to the subject area model.

Researchers investigating learning processes have typically built models such as these in areas such as mathematics and the sciences, where the structure of knowledge is assumed to be highly procedural. Brown and his associates have built a procedural model for solving arithmetic problems, and have designed similar procedural models of how students solve arithmetic problems as their understanding is growing [1, 3]. Van Lehn extended these ideas to build Bayesian networks to model student procedures while learning basic physics [12]. These and other systems offered researchers important insights into human learning and cognition.

Like Brown, VanLehn, and others, Tatsuoka et al. began investigating the procedures that students use in solving algebra problems [1,11]. Since Tatsuoka’s goal was to diagnose and correct misconceptions, rather than to understand and model human cognition, her research took a practical turn that led to the development of Q-matrix theory and fault tolerant teaching.

Procedural models of knowledge have been very important in building understanding of human cognition and learning, and can help direct remediation, but have several drawbacks in adaptive tutorial systems. These models require extensive time and effort to construct – and are not extendable to other specific subject areas. After they are constructed, research has shown that even extensive libraries of errors cannot predict all the procedural errors that students commit, and not all student errors can be explained by erroneous procedures. Even in cases where erroneous procedures can explain student errors, it is not certain that these procedures are those being used by students. In addition, these systems have no ability to explain, predict, or remedy non-procedural errors that students may commit. Student errors are not consistent over time, so these diagnoses may not consist of accurate assessments of student knowledge.

Because of these limitations, procedural models of knowledge may not be the best choice for assessment in adaptive tutorial systems. Nevertheless, these models have formed the basis fault tolerant teaching theory. The following section describes how a procedural model evolved into Q-matrix theory.

2.1.1 Rule space theory and Q-matrices
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 [1, 11]. 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 [1].
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. This space allowed instructors to map and understand 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 [1]. 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. In Tatsuoka’s work, a Q-matrix, also called an attribute-by-item incidence matrix, contains a one if a question is related to the concept, and a zero if not. 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 [2].

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>e₁</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>e₂</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Tatsuoka’s rule space research showed that it is possible to automate the determination 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.

### 2.2 Q-matrices and student behavior

Tatsuoka’s Q-matrix methods offered researchers an automated way to diagnose student misconceptions, partially bridging the gap between procedural models of knowledge and student behavior. Q-matrices were originally designed to encompass procedures for solving problems, but also included non-procedural problem “attributes”, e.g. having a negative number in a problem. This is a stronger model of student behavior, encompassing both rules and attributes in a way that has transcended rules, assuming that students with similar responses have similar understandings and misconceptions.

Though Q-matrices seemed to allow the diagnosis of non-procedural errors, research was still needed to compare actual student responses with Q-matrix predictions. In 1992, Robert Hubal investigated how well student data in Tatsuoka’s experiments supported the Q-matrices she created [6]. Hubal found that, although students did seem to be using a small set of possible rules, none of these rules was predicted by Tatsuoka’s Q-matrix.

In using Q-matrices to understand student behavior, we can choose to either: 1) select a Q-matrix and try to create or find questions whose relationships to underlying questions are supported by student data, or 2) use questions we already have and try to construct a Q-matrix based on actual student data. In his research, Hubal experimented with the first option, by generating questions, testing them on students, and deleting those questions that did not exhibit the relationships predicted by Tatsuoka’s Q-matrix.

Limiting questions to those that exhibit behavior predicted by a hand-constructed Q-matrix may not be the best choice. First, the questions have been designed by an instructor to test student performance, and a Q-matrix is a much more abstract measure of the relationships among questions. We might assume that the questions designed to test students are a more accurate reflection of the teaching objectives than an abstract construct which relates questions to underlying concepts. Second, the selection of questions requires many students to answer questions that will later be ignored, and data from these questions is then lost. It is quite possible that questions that do not exhibit the relationships predicted by a Q-matrix can reveal unknown aspects of student behavior and understanding.

An alternative strategy is to extract a Q-matrix which explains student behavior, and reveals the underlying relationships between questions. Using such a strategy, student answers to existing questions would be analyzed to create a Q-matrix to explain the observed data. Experts can then examine the Q-matrix to ensure that the extracted relationships seem to be valid, and then use that Q-matrix to guide the generation of new problems.

### 2.3 Q-matrix extraction methods and fault tolerance

Inspired by Hubal and Tatsuoka’s work, Brewer [2] investigated methods of extracting the Q-matrix directly from student data. Since it is impossible to truly know what a student knows, Brewer created ‘ideal’ students, to design an experiment to study the feasibility of extracting the Q-matrix when student knowledge states are known.

Starting with a pre-determined Q-matrix, Brewer generated all possible concept states to create ‘ideal’ students. He then generated their predicted response vectors based on the Q-matrix. Random bit flips in their 0/1 answers were added at rates from no flips to as many as fifty percent changes. For each noise rate, Brewer extracted the Q-matrix using two techniques: the “Q-matrix method”, an iterative hill-climbing technique, and common factor analysis. Comparing the two techniques, Brewer found that...
the Q-matrix method was able to recover the original Q-matrix for as few as 25 students with error rates as high as fifteen to twenty percent, whereas factor analysis required hundreds or thousands of students for equal recovery. Brewer’s research demonstrated that Q-matrix extraction could succeed in spite of noise, introducing the possibility of fault tolerance in assessing student knowledge.

The goal of Q-matrix construction is to extract underlying, or latent, variables which account for students’ differential performance on questions. This extraction may employ any method which will successfully account for student response patterns. Brewer’s research, along with this work, have shown that traditional iterative hill-climbing techniques augmented with several random starting positions are as effective, if not more effective, than factor analysis. In the next two sections, we will give a brief overview of factor analysis and the Q-matrix method.

2.3.1 Factor analysis
Factor analysis is a statistical process that attempts to extract latent (i.e. unknown and unobservable) variables from observed data [8]. It starts using a correlation matrix, and then iteratively extracts eigenvalues and eigenvectors (called factors) from the correlation matrix until a certain percentage of the variance in that matrix is explained by the extracted factors. Once the factors are extracted, they are ‘rotated’, i.e. mathematically modified, so that people may better understand their meaning.

2.3.2 The Q-matrix method
The Q-matrix method is a simple hill-climbing algorithm which creates a matrix representing relationships between concepts and questions directly. The algorithm varies $c$, the number of concepts, and the values in the Q-matrix optimizing for minimum total error for each student for a given set of $n$ questions. To address the problem 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. Also, $2^c$ concept states are generated. A binary string of length $c$ represents each state where a zero in position $k$ represents that a student does not understand concept $k$, and a one represents that he does. For each concept state $q$, its corresponding ideal response vector is generated based on the Q-matrix.

For example, for the state 01 (not knowing concept 1, knowing concept 2) and the Q-matrix in Table 1, the ideal response vector would be 11000. Since the Q-matrix tells us that questions 1-2 depend only on knowing concept 2, the student should answer questions 1-2 correctly. Since questions 3, 4, and 5 all depend on understanding concept 1, this student should miss questions 3, 4, and 5, giving us the predicted ‘ideal’ answer vector of 11000.

For Q-matrix values between 0 and 1, the value $Q(x, y)$ represents the probability that a student will answer question $y$ incorrectly given that he does not understand concept $x$ [2]. Using this value, the most probable ideal response vector is calculated for each concept state.

The next step is to compute the total error for the Q-matrix computed over all students. For efficiency, we create an array of size $2^c$ of all possible response vectors, and the $i$th element of the array contains the number of students with response vector $i$. Each observed response vector is compared with each ideal response vector, and is assigned to the one that is closest to it in Hamming distance. This distance is the error for that response vector. To compute the total error for the Q-matrix, the individual errors for each observed response vector are multiplied by the total number of students with that response, and summed over all observed response vectors.

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 the entire Q-matrix, until the error in the Q-matrix is not changing significantly.

To avoid local minima, the algorithm is run with several new random starting points, and the Q-matrix with minimum error is saved. This Q-matrix is not guaranteed to be the absolute minimum, but provides an acceptable Q-matrix for a given number of $c$ 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.

Brewer found this hill-climbing method to be effective in extracting Q-matrices for small groups of ideal students with random noise added to their responses. This method was much more effective for small groups of students than factor analysis. This difference in performance is most probably a result of pre-processing the student response data before implementing factor analysis. When forming a correlation matrix, we lose individual student data in favor of calculating average relationships between questions. The Q-matrix method is optimized to assign each student the most appropriate knowledge state, using all available response data for each student.

2.4 Question generation and Q-matrix extraction
Once Brewer verified the Q-matrix extraction is feasible even in the presence of noise, it was important to test this idea with real questions and students. Soon after Brewer completed his research, Jones [7] and Sellers [9] both wrote interactive lessons on NovaNET, a computer-based educational network, developing techniques for question generation and automatic answer judging. They each also collected student data and created Q-matrices from these data. In both experiments, and in recent work, students reported feeling positive about the online lessons. Also in both experiments, the extracted Q-matrices appeared...
reasonable, even with small numbers of students (7 and 17). These results suggest that it is feasible to extract Q-matrices from larger groups of student responses.

Sellers also compared the extracted Q-matrices to those constructed by instructors. For two of her Q-matrices, expert and extracted Q-matrices corresponded in all but one Q-matrix value. Upon a second look, experts agreed that the extracted Q-matrix was appropriate. For more complex problems, expert and extracted Q-matrices differed more greatly. For these complex problems, computer extraction of concepts based on real student data may be preferable to expert analysis, since experts tend to explain complex problems in different ways, resulting in conflicting models.

Some of Sellers' recommendations for future research are: 1) use few concepts for many questions and many students, 2) compute Q-matrices for small chunks of questions, 3) design a thorough experiment to test the correlation of Q-matrix information and instructor assessment and question analysis, 4) further investigate the meaning of Q-matrix values, and 5) to test the validity of using the Q-matrix to guide teaching for remediation.

2.5 Data mining and Q-matrix methods

Q-matrix methods are used to extract underlying relationships among observed variables, which is also one of the main goals of data mining, “the computer automated exploratory data analysis of (usually) large complex data sets” [5]. Data mining researchers today often use principal components analysis, one form of factor analysis, as a preliminary step in processing large data sets. Since factor analysis and the Q-matrix method described herein both offer methods of understanding larger data sets in terms of a much smaller set of ‘concepts,’ it is quite feasible that Q-matrix methods can be used as data mining techniques. The advantage of applying Q-matrix methods to data mining is the potential improvement in interpretability that Q-matrix methods may offer. As Sellers found in her research, the results obtained through Q-matrix analysis seem to describe relationships among variables in interpretable ways. Factor analysis and principal components analysis, in contrast, do not readily offer interpretable results. As they are developed, new methods of pattern recognition will be extremely important in gaining advances in data mining.

3 Current Research

The research reported in this paper is an ongoing experiment to test FTT methods across multiple topics and difficulty levels. Our experiment involves implementing FTT methods in three NovaNET lessons, testing them with 50 students each, and evaluating the extracted Q-matrices for each lesson. The three lessons include: Sellers' binary relations lesson, an existing propositional proof tutorial, and a new lesson in combinatorics. These three tutorials involve problems at three levels of thinking. The proofs program involves synthesis of proofs, a very high level of thinking. The binary relations lesson tests memory and application of definitions, while the counting program provides a wider range of problem-solving skill.

We have completed one trial run of Sellers’ NovaNET lesson on binary relations with a group of 70 students taking the Spring 2001 Discrete Mathematics course at North Carolina State University. The binary relations lesson consists of text explanations of definitions, examples, and small quizzes for each of the three sections of the lesson. The lesson also links to audio lectures and slides created with the Web Lecture System [14].

During the course of the lesson, after each quiz, the lesson used a Q-matrix to determine which concept the student least understood, and directed the student to repeat the lesson related to that concept. In this program, all students were automatically directed to review material, and only one concept was reviewed after each quiz. For future lessons, it will be important to develop a measure of competence, or ‘concept closure,’ based on the Q-matrix, and use this measure to determine when a student is ready to proceed to the next section of a lesson. Another issue to note is that, during this lesson, the Q-matrix is such that, when a student misunderstands several concepts, the program directs students to review the one that is most difficult. This may not be the most effective teaching strategy [see 13 for a discussion of teaching strategies]. The Q-matrix method appears to be an effective assessment tool, but research is still needed to determine the most appropriate teaching strategies to apply.

In this trial, about 60 of 70 students completed the entire lesson and the accompanying attitudinal survey. The survey results show high satisfaction among the students, with all but one student indicating that they would recommend the lesson to other students. Many students also found the program to be more effective than classroom lectures and working alone with a textbook. More than one third of the students requested additional lessons of this type, and several returned to use the program a second time to study. According to the students, the best features of the program were the ability to work as many examples as you care to ask for, and the optional audio lectures and slides. One student even emailed his thanks for the program, stating that he answered all the questions on binary relations on his in-class test after taking the program.

Only one student reported having a negative experience with the lesson. This student was using a Macintosh terminal, which requires that the student use the web browser directly to access the audio lectures. Since the NovaNET lesson prompted him to press a key to hear the audio lectures, he became very frustrated, and after a short while, completed his survey and left. Although we assured him the trouble was our mistake, his frustration was so high that he no longer wished to use the program. This anecdote is an effective reminder of the importance of quality of service and help materials in delivering instruction online.
For the most part, during the experiment, students used the NovaNET lesson independently after a very brief introduction to the system. When glitches did occur, such as a student accidentally ending the lesson, someone was always nearby to restart the lesson and bring the student back to where they left off, which usually took less than thirty seconds. Occasionally, students asked questions about the material, indicating areas that could use more explanation in the binary relations tutorial.

Students became frustrated when there were glitches or errors anywhere in the tutorial. Nevertheless, most students were patient with the few errors that occurred. As in any system, when difficulties arise, it is best to have someone nearby to reassure students that they have not made a mistake and that the lesson can be recovered quickly and restarted where they left off. These are important findings to note for future lessons - each lesson should be fully tested, irregularities noted, and resolved or reported to students, who are remarkably adaptive when informed.

It is important to note that most students began to appear restless during the third and final section of the lesson, after taking the lesson for about 2 hours. The difficulty of the material increases as the lesson progresses, and students are not given an indication on the lesson of how many topics are left to complete. The results of this trial suggest that lessons should be limited to a maximum of 2 hours; students should have more control over navigating through the lesson; and students should be given a list of topics to get a sense of their progress and the remaining time in the lesson. We have implemented these changes in the binary relations lesson, including keys to go back in the lesson sequence, and menus of the lesson topics.

In addition to these qualitative findings, the trial run results were analyzed and compared to the Q-matrices extracted by using Jennifer Sellers’ data and those determined by experts. For the first two sections of the tutorial, the extracted Q-matrices corresponded to those Sellers found with only 17 students. For the final section, the extracted Q-matrix was quite different from that found by Sellers, and was again different from the expert-created Q-matrix. These results are not surprising, since the material in the third section was much more complex, and there is only one question per topic in the lesson. For fault tolerance and robustness, it will be necessary to ask more questions per topic, to build more reliable Q-matrices.

In the binary relations lesson, only answers from quizzes were recorded for Q-matrix analysis, but in future experiments, all example answers can be collected for analysis. It is probable that student understanding changes as they work problems, and this additional data may allow us to track the evolution of a student’s knowledge.

A preliminary version of the counting tutorial, including lessons in permutations and combinations with and without replacement, was completed and tested with students in Fall 2001. Its interface was built to include the functionality students desired in the binary relations program, including menus for direct access to topics, and keys to move forward and back in the program. Each problem was programmed to randomize numbers in the problems given, and to interpret answers automatically. More in-depth help was also developed for this program, increasing student satisfaction. We are currently developing more challenging problems for this tutorial.

The propositional proofs program has offered us the greatest challenge in applying Q-matrix techniques. Since writing proofs is very open-ended, it is difficult to extract relationships between problems. We have designed a solution to this problem. Instead of simply marking whether a student has successfully completed a proof or not, we will also consider each student use of a particular axiom as a question. In this way, we can compare work among students, and create a model that can extract the relationships among proof axioms. This will provide another good test of the Q-matrix method, since axioms can be readily compared to one another. We realize that this comparison of student answers falls short of a full analysis of student responses, and plan to apply several different analyses to student responses to find the best application of FTT methods.

4 Conclusions
Fault tolerant teaching methods appear to be promising in improving computer based education. Previous research results have suggested that these methods can be used effectively to assess student knowledge and find relationships among tutorial questions. Recent work also demonstrates positive student satisfaction and performance using three FTT-based lessons. Though these results are promising, it is still necessary to further validate the FTT methods described herein. During this spring, we are applying FTT methods in these three NovaNET tutorials with larger numbers of students to develop guidelines for the general application of these methods.

The most important aspect of this research is student knowledge assessment. During the course of instruction, the tutorials developed for this research will collect data about student answers and estimate the appropriate knowledge state for each student, without knowing the subject area or specific questions. This estimate of student knowledge will be made based on statistical analysis of a body of student responses to the same questions. Once this analysis relates questions by their underlying concepts, student responses can be diagnosed for conceptual understanding. Using this diagnosis, the tutorial can then tailor the lesson to students’ individual needs. In addition, group assessments can reveal which concepts are being taught more effectively, to direct further development of teaching materials.

There are two aspects to evaluating system knowledge assessment. The first aspect is to compare automatic
knowledge assessments for each student to those made by the student’s instructor. The second aspect is to evaluate the use of this knowledge assessment in directing learning. One group of students will be allowed to review material freely, while another will be directed to review concepts found to be lacking by the diagnostic engine. Self-guided student learning paths will be compared with paths chosen by the diagnostic engine for each student. Students learning with guidance from the diagnostic engine will be compared with self-guided students for performance, time to completion, and satisfaction.

If FTT methods can successfully assess student knowledge, this method will provide a powerful tool for investigating teaching strategies: once a student’s knowledge state is known, the effects of any teaching strategy on changes in student knowledge can be measured. This research will be the first that we know of to implement a fully automated, general diagnostic tool and use it to compare two different teaching strategies.

The concept structure used for knowledge assessment is also an important outcome of this research. For each tutorial, the extracted concept structure can be examined for interpretability — i.e., for whether the structure makes sense to a human expert. A hypothesis of this research is that this structure will be understandable and usable.

In addition to the analysis of FTT methods, this experiment will also yield knowledge about differences in performance and satisfaction of online education. Students taking the online lessons will be assessed and surveyed and compared with a control group of students enrolled in the same class but not taking the online lessons. It is hoped that students will find the tutorials more effective than self-study, and perhaps at least as effective as in-class lectures.

This research contributes to automating two very important aspects of computer-based education and fault tolerant teaching research: 1) comparing teaching strategies for effectiveness, and 2) determining and correcting student misconceptions. These two results can provide a very strong, scientific basis to the continued development of computer based education and fault tolerant teaching.

4.1 Future work

This research will lay the groundwork for future work in fault tolerant teaching and computer-based education. Any online lesson with the capability to assess student answers can be augmented with the FTT methods. As FTT methods are applied to more lessons and more students, they can be further validated and improved.

Fault tolerant teaching methods can be applied to other topics, not limited to mathematics, since they rely only on the correctness or incorrectness of student answers. They may be applied to multiple-choice tests and much more open-ended questions. These results are limited only by the automation of judging responses, and not by the type of student response. Future work could extend the model to include partial credit for student responses.

Future experiments can compare teaching strategies by applying different teaching strategies to different groups of students and comparing the time students take to master a set of concepts. Future research also might test the effectiveness of different strategies on individual students, comparing the change in student knowledge before and after each strategy is applied.

In a similar extension to this work, Q-matrix methods might be used to extract other student characteristics, such as student learning style. Such an experiment could determine students’ learning styles using a test such as the Felder-Silverman Index of Learning Styles independently, and administer a lesson with two styles, such as visual and verbal examples, and attempt to extract students’ learning styles using the Q-matrix extraction method. This type of experiment would suppose that a student’s learning style has an effect on question responses similar to the effect of “concepts” supposed in the current research.

Automated student knowledge assessment can be applied in research to understand the changes in student knowledge as they learn. In this case, researchers would map student knowledge states after each question or group of questions to track changes. This could be used in several different ways. First, it could be used to determine the most effective areas of a tutorial, and identify those that are not causing changes in student knowledge. It could also be used to further understand the process of human learning, assessing whether students are learning in great leaps or with small, gradual changes. Alternatively, this could also be used to redirect a student’s learning. Tracking changes in student knowledge can alert an automated system to change its strategy.

Another major future application of this work may be in data mining and latent variable analysis. If the Q-matrix extraction techniques in the research yield interpretable results, this could be applied in other areas, such as data mining, knowledge discovery, and latent variable analysis. To date, these areas use many different artificial intelligence techniques to detect underlying relationships in data. As an example, a search engine on the web could record the occurrence of all keywords in a document, and use Q-matrix analysis to find concepts that relate keywords, and related documents to these concepts. Then, when a user entered a keyword search, the search engine can look for concepts and not just word occurrences.

In conclusion, this research will provide an experimental analysis of the effectiveness of fault tolerant teaching methods across three topics and difficulty levels, with three groups of students. Regardless of the outcome, this analysis will provide valuable data about the problems of automatic knowledge assessment, and make a significant contribution to the fields of computer based education and fault tolerant teaching.
Acknowledgements
This research was supported in part by a grant from the National Science Foundation (HRD-9813902) and an IBM Fellowship.

References
[14] Web Lecture System (14), a system for creating and transmitting online presentations on the Internet, was developed by NC State University Regional Training Center for Parallel Processing. [http://renoir.csc.ncsu.edu/14]