IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

by

JOHN DAVID MOODY

(Under the Direction of Charles H. Atwood)

ABSTRACT

Item Response Theory (IRT) analysis on General Chemistry exams has been used successfully at the University of Georgia (UGA) to identify key topics that students find difficult. Our preliminary analysis of the Online Web-Based Learning (OWL) homework questions in Question mode has shown some overlap between UGA difficult topics and nationwide difficult topics. In addition three more national topics have been identified as difficult based on the preliminary analysis. Unfortunately, question mode questions only involved 15% of the total OWL responses. Many of the questions in the identified difficult topics had fewer than 300 responses. Therefore, we also conducted IRT analysis of the OWL homework database in Mastery (Question Pool) mode. Since most of the data we analyzed involved questions where students were allowed 15 or more attempts (and in many cases, unlimited attempts) to answer a question, we also had to determine which attempt had the question's most normally distributed Total Information Curve (TIC). The parameters of the two modes were also compared to each other and to parameters from JExam homework. In addition, the difficult topics identified in each mode will be examined.

We will also look at small differences between two questions that results in a large difficulty difference between the two questions, such as an equilibrium questions where there are no solids present in the equilibrium vs. a question where a solid is present in the equilibrium.

Lastly, we will summarize the results and make suggestions for an instructor who wants to build homework and testing database where the downloading of response data is easy.

INDEX WORDS: Item Response Theory, Difficult topics, Chemistry, Online homework, Problematic subtopics

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

by

JOHN DAVID MOODY

B.S., Juniata College, 1991

M.A., University of Delaware, 1994

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

by

JOHN DAVID MOODY

Major Professor: Charles H. Atwood

Committee: Richard Morrison

Charles Kutal

Electronic Version Approved:

Maureen Grasso

Dean of the Graduate School

The University of Georgia

August 2013

DEDICATION

This Ph. D. dissertation is dedicated to the following people who I had the pleasure of knowing but who unfortunately are no longer with me, but in a better place:

Edna U. Moody (1900 – 1981)

Reno Snyder (1903 – 1983)

John Daniel Moody (1935 – 1995)

Cora Snyder (1914 – 2010)

ACKNOWLEDGEMENTS

Conducting research in today's world is not an individual effort. In my quest to receive a Ph. D., there are several people I would like to thank.

First, my committee: Dr. Atwood, Dr. Morrison, Dr. Stickney, and Dr. Kutal. They have all been very helpful in giving me advice and keeping me on track. I wish to thank my group members (none of whom are here) Kimberly Schurmeier, Derek Behmke, and Haley Nyguen. They were excellent not only for bouncing ideas, but also keeping me focused. I wish to also acknowledge the people at Cengage Learning and the University of Massachusetts, especially Lisa Lockwood, Dave Hart, Steve Battisti, and Gordon Anderson (especially Gordon Anderson). They were the people responsible for getting me the data that enabled me to conduct my research. I also wish to thank Bill Vining. Talking with him gave me some insight into question difficulty; especially those questions where entering the answer to a question can be cumbersome. Another thank you goes out to the UGA chemistry department. Not only did they support me, they were especially supportive when Dr. Atwood left for the University of Utah. This enabled me to remain at UGA so I could finish my Ph. D

I have to give another acknowledgement to Dr. Charles Atwood. He understood me, and I hold him very responsible for my success, not only professionally, but also personally.

TABLE OF CONTENTS

CHAPTER

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

Section 1.1 – Purpose of This Research

At the University of Georgia (UGA), Item Response Theory (IRT) analysis has been used on the response patterns of examinations given at UGA from 2001 – 2011. The analysis has resulted in the identification of key topics that are difficult for UGA students (1). These topics are:

1. The Particulate Nature of Matter. This topic addresses chemical events at the atomic, molecular, or ionic level. Typical exam and homework questions ask about the number of ions, calculating the number and concentrations of ions, distinguishing between covalent and ionic compounds (including vapor pressure), and which substances exist as atoms, ions, formula units, or molecules.

2. Molecular Polarity and Intermolecular Forces. This topic addresses many subtopics, such as polar bonds vs. polar molecules, intermolecular forces and how those forces affect physical properties. In addition polarity and miscibility are also addressed.

3. Understanding Quantum Numbers. This topic addresses the number of electrons that can have a particular set of one to three quantum numbers and the identification of orbitals based on quantum numbers.

4. Use of the Terms Strong, Weak, Concentrated, and Dilute in Relation to Acidic or Basic Solutions. Questions on this topic ask students to choose a strong or weak acid from a list. In addition, students choose an image which represents a concentrated strong acid. Lastly, students choose from a set of compounds those which agree with the descriptive statement (i.e. conducts electricity well in dilute aqueous solutions).

5. Molecular Image Problems. Problems involving images are more difficult than problems where images are not involved, such as physical and chemical changes.

6. The Mole Concept. Questions about the mole concept contain an image, and ask which image contains the greater number of moles. Alternatively, students are asked "What mass of a substance contains the same number of molecules as 45.3 grams of another substance.

7. Solution Calorimetry. This is a simple calorimeter question, similar to what the students do in the laboratory

8. Inorganic Nomenclature. Nomenclature questions increase in difficulty as the naming sequence becomes more complex. Covalent compounds are the simplest for the students to name. Most difficult are ternary oxyacids and acidic salts of ternary oxyacids.

The goal of our current research is to examine the responses of students nationwide from the OWL (Online Web-Based Learning) database. The questions our research will answer are:

1. Are there any nationwide difficult topics that are the same or similar to the difficult topics identified at UGA?

2. Are there any nationwide difficult topics that are NOT difficult for students at UGA?

3. Are there certain topics where a minor change to a question makes a huge difference in the difficulty of the question?

4. Are there any nationwide easy topics identified by the OWL database?

Identifying the difficult and the easy topics has many implications. Instructors can devote more class time to emphasizing the difficult topics. Homework questions can be written to give students more opportunities to practice questions that address these difficult topics.

Identifying easy topics is also important. Topics that are easy, as identified by the OWL database, indicate topics that do not need as much class time. The class time that is saved can be devoted to more difficult topics, or can be used to introduce other innovative teaching methods, such as Problem Based Integrated Instruction (PBI2). As will be mentioned in Chapter 2, one of the issues with PBI2 was the time that was involved. More class time can be devoted to PBI2 instead of the easy topics (2). The feedback from the OWL system can be used to teach students these easy topics.

Section 1.2 – Dissertation Outline

Chapter 2 of the dissertation will discuss PBI2 that was implemented during the 2010 – 2011 academic year. It was only implemented for one year since PBI2 did NOT show an improvement in the students' performances. Chapter 3 will discuss Item Response Theory (IRT) in detail since the subsequent chapters use IRT to analyze the OWL responses.

Chapters 4 and 5 will discuss the results of the analysis of the OWL questions. Chapter 4 involves questions that were presented in question mode and chapter 5 involves questions that were presented in mastery mode.

In question mode, each question within a module is presented in a linear fashion. Students have to answer all of the questions in the module in order to receive credit for the module. Once a question is correctly answered, it is no longer presented. In mastery mode, a module consists of a given number of questions (usually 3) that are selected from a pool of questions. In order for the student to receive credit for successfully completing the module, the student must answer a given number of questions (usually 2). If the student is not successful, the entire module needs to be redone.

Chapter 6 will compare the two modes and the JExam (an in-house electronic, on-line homework and exam system used by students at UGA) homework not only on a topic basis but also on a question by question basis. Which mode gives the more reliable data? Are the two modes equivalent? Chapter 7 will consider situations where a very minor change in a question dramatically changes the difficulty of a question, such as drawing Lewis structures containing only single bonds vs. those Lewis structures that contain both single and double bonds. Chapter 7 will also look at how the method of entering the answer (i.e. number entry vs. formula entry) affects the difficulty of the question. Chapter 8 summarizes the work, suggests future directions for the work, and also makes recommendations for building an online homework database.

Section 1.3 – Differences Between IRT Analysis of OWL Homework and JExam Tests

JExam is an in-house testing and homework system used by students at UGA. JExam allows students to enter numbers, text, and formulas. JExam can also ask questions where students can view molecular models and even click on individual atoms in that model. However, students cannot enter complete reactions in JExam. JExam also allows for multiple choice (where there is only one correct answer) and multiple answer questions (where there is more than one possible answer). In OWL, students can also enter numbers, formulas, and text. However, in OWL, students can enter the reactants and products of a chemical reaction since the OWL will accept the reactants and products in any order, whereas JExam is not as flexible.

Another major difference is the three way parameterization of questions. Each time the same OWL question is presented to a student, the numbers, chemical reactions, answers (for a multiple choice or answer question), substances, and/or what is asked may change. In JExam, the numbers are different for different students, but if a student gets a question incorrect, they are presented with the exact same question (including the exact same numbers) on their second attempt. In JExam multiple-answer questions, if a student answers the question incorrectly, they are given clues on which answers are correct. Therefore, in homework where students get full credit for getting the answer correct, even on their second or third attempt, there is no incentive to read the question and attempt to think through the question (this is discussed in more detail in chapter 6). The JExam multiple answer issue is not a problem in OWL due to the parameterization of the questions since students will often not see the same choices in their future attempts.

The OWL homework system also has the flexibility of allowing for a different number of attempts, including unlimited attempts. JExam can be set up to allow different numbers of attempts, but not unlimited attempts.

The OWL homework system also has two methods for presenting questions (question mode and mastery mode) whereas JExam only presents questions in question mode and does not parameterize the question after each attempt. In JExam, the numbers, chemical reactions, and compounds are the same for each of the attempts whereas in OWL, the numbers, chemical reactions, answer choices, and compounds change from attempt to attempt.

CHAPTER 2

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

PROBLEM BASED INSTRUCTIONAL INITIATIVE (PBI2)

Section 2.1 – Introduction

One challenge in teaching chemistry is helping students achieve a deep level of understanding of not only mathematical problems, but also conceptual problems. Many approaches have been tried at various schools and colleges around the country. Examples of those approaches include SCALE-UP, POGIL, and PBL (3, 4). These approaches are different but all of the approaches have similar themes: get the students actively involved in their learning, and shift the burden of learning from the instructor to the student.

Additionally, students often perform poorer on the final exam when compared to in-semester exams due to the fact that the students “compartmentalize” the various topics, and do not make connections between seemingly unrelated topics. Students see the trees but not the forest.

In order to accomplish these goals and have students make connections between seemingly unrelated topics, it was proposed that integrated problem based learning be implemented in CHEM 1211 and CHEM 1212. CHEM 1211 and 1212 is the two semester general chemistry course sequence taken by science majors such as biology, pre-med, and pre-pharmacy. The name of this project was Problem Based Instructional Initiative (PBI2). The goals of PBI2 were

1. Decrease the percentage of W's, F's, D's, and C–'s in general chemistry;

2. Improve students' abilities to integrate material from different chemistry topics;

3. Improve students' performances on the ACS (American Chemical Society) Standardized final exam.

Even though problem based learning has become widespread over the past 5 to 10 years, there still are impediments to implementation. SCALE-UP requires that a special classroom be developed. In addition, SCALE-UP integrates the lecture and the lab into one course (3, 4).

Most problems are designed for specific topics, such as the University of Delaware's problem "Saving for a Rainy Day" (5) which only examines heat transfer involving phase changes (equivalent to Whitten et al. chapter 13, sections 9 and 11.

PBI2 was an attempt to utilize problems in which students to integrate material from different chapters. An example of material that can be integrated is Thermodynamics and Equilibrium. These concepts are discussed in the Whitten text (9th edition) in Chapter 14 (enthalpy of solution), Chapter 16 (relationship between ΔG^o_{rxn} and ΔH^o_{rxn}, and Chapter 17 (LeChatelier's Principle). Students may not realize that all of these topics are interrelated, especially since each of these topics is often covered on a different exam.

Section 2.2 – Experimental

In both Fall 2010 (CHEM 1211) and Spring 2011 (CHEM 1212), Dr. Atwood, taught two sections of general chemistry, one from 9:05 am to 9:55 am, and the other section from 12:20 pm to 1:10 pm. For both semesters, the 9:05 am section (morning section) was the traditional lecture and the 12:20 pm (afternoon section) was the PBI2 section (chosen at random by coin flip).

The PBI2 class was 3/4 traditional lecture and 1/4 problem solving groups. During the problem solving sessions (conducted in class) students worked in their groups to prepare a report to solve the PBI2 problem. These problem solving groups were also used to address topics that had not yet been covered in class.

The PBI2 class was separated into problem solving groups of 4 or 5 using CAT-ME software (6). The groups were formed based on schedule, gender, and math prerequisite scores. Students were grouped based on common availability (7). If a student was free on Thursday evening (for example), the software looked at other students who were free on Thursday evening and put those students in a group. The software did not assign groups where the males outnumbered the females. This was done to ensure that the males would not "dominate" the group (8). The software also balanced students based on math prerequisite scores. Each group contained one outstanding math student, two average math students, and one poor match student. In CHEM 1212, the CHEM 1211 grade and whether or not the students were in PBI2 first semester were used as additional criteria for forming the groups. The software ensured that each group was a mix of students who were in PBI2 in Fall 2010 and students who were not in PBI2 in Fall 2010 (9).

PBI2 students solved two problems for each semester. For each problem, students had to write a report on how to solve the problem, had to take a quiz (report defense) showing that they understood the calculations, and evaluated their group members on their contribution to the project. Students were allowed to use their written reports during the report defense. Each problem was worth 10% of the overall course grade (20% of the overall course grade for both problems). Students in the PBI2 section did these two problems in lieu of the second hour exam (also 20% of the overall course grade).

The problems presented in CHEM 1211 were:

1. Examine the change in CO_2 levels based on the amount of CO_2 that is emitted vs. the amount of CO_2 absorbed by forests. Determine if using CO_2 in chemical production can make an appreciable difference in reducing CO_2 levels.

2. Determine the feasibility of separating Plastics #1 - #6 in a recycling center. Find a method to test the plastics after separation to ensure no cross contamination. Also, if one cannot recycle the plastics, propose a method for “disposal”.

The problems that were presented in CHEM 1212 are:

1. Examination of thermodynamic and environmental feasibility of an ethanol based fuel economy, including why some scientists claim that an ethanol based economy results in a net decrease in energy.

2. There was an article in the Daily Item (a Middleburg, PA newspaper) about a device you can install in your car that improves your gas mileage. The device consists of a 1-quart water container with some dissolved NaOH plus electrodes designed to utilize the excess electricity

produced by the alternator to decompose the water into hydrogen and oxygen gas via electrolysis. Discuss the feasibility of this device and whether it can work or not. Assume that we are talking about a car that gets 30.0 miles / gallon when travelling at a speed of 60.0 miles / hour (i.e. the car burns 2.00 gallons of gasoline per hour).

Section 2.3 – Results

The grade distribution for the two semesters is summarized in Figure 2.1

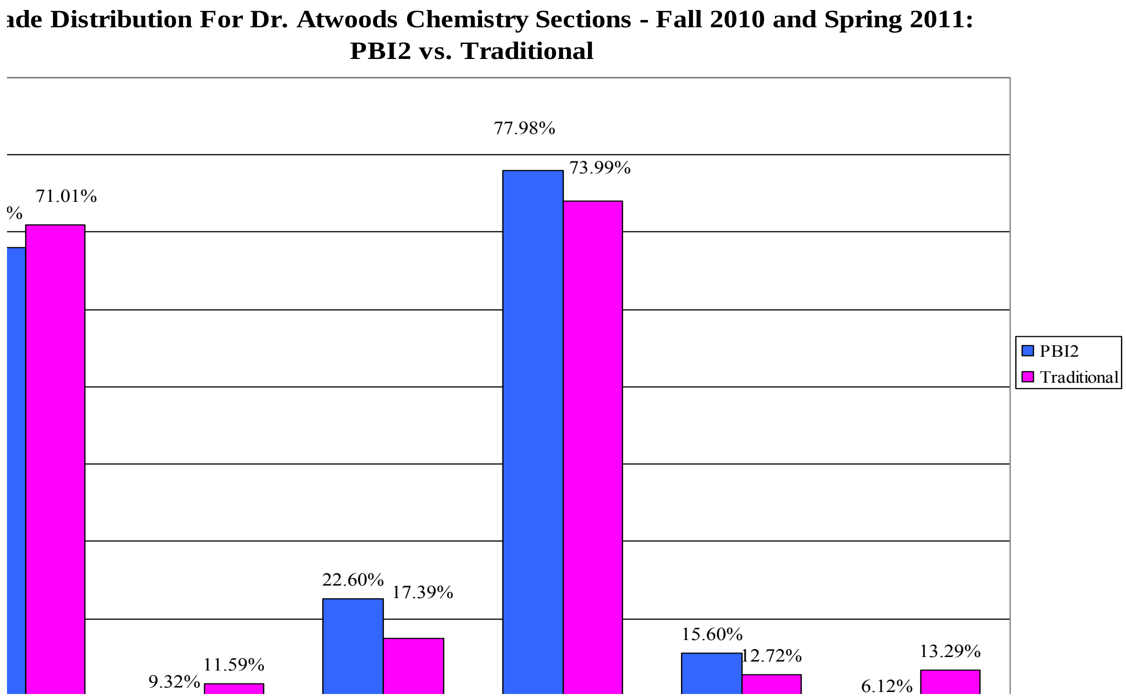


Figure 2.1 – Grade distribution for Dr. Atwood's chemistry sections – Fall 2010 and Spring 2011 for the PBI2 section and the traditional section. The blue bars are Dr. Atwood's PBI2 group and the pink bars are Dr. Atwood's traditional group

The graph shows that PBI2 was not effective during the first semester. The percentage of students getting grades of C–, D, F or W was 31.92% for the PBI2 section vs. 28.99% for the traditional section. Even though the percentage of low grades was smaller, the percentage of students withdrawing was greater. Reasons for the higher percentage of students withdrawing will be addressed in the discussion section.

During the second semester, the trends were reversed. The PBI2 section had a C–, D, F or W percentage of 22.02% vs. 26.01% for the traditional section. The percentage of withdrawals was significantly reduced in the second semester PBI2 class. This will be addressed in the discussion section.

However, the analysis of overall grades is flawed, especially since 20% of each student's grade was determined differently (i.e. hour exam vs. reports). To address the issue, the performances on the hour exams that were taken by Dr. Atwood's students in BOTH sections were examined. The results are shown in figure 2.2.

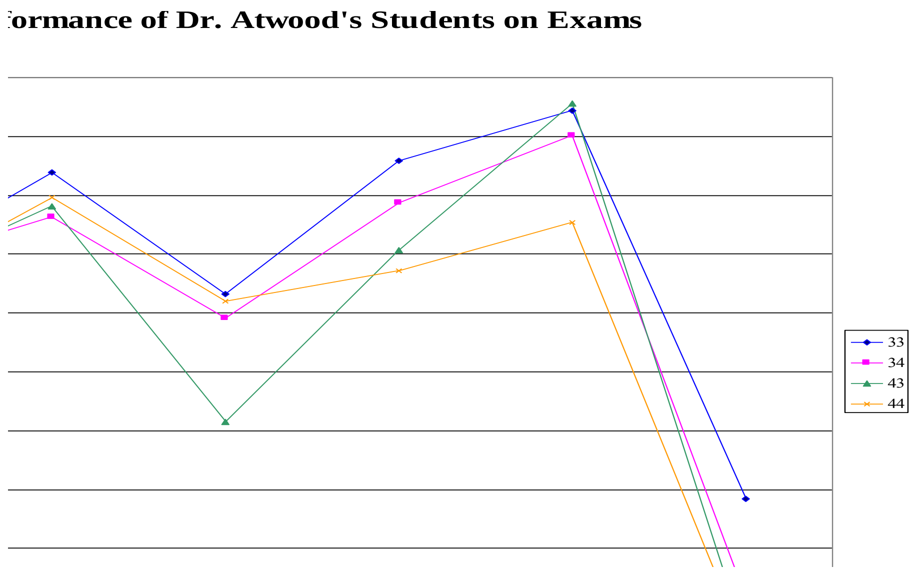


Figure 2.2 – Performance of Dr. Atwood's Students on Exams – The legend 33, 34, 43, and 44 mean the following:

33: Students were in traditional lecture for Fall and Spring

34: Students were in traditional lecture for Fall and PBI2 for Spring

43: Students were in PBI2 for Fall and traditional lecture for Spring

44: Students were in PBI2 for Fall and Spring

There are some noticeable trends. For the Fall 2010 final exam, the group of students that performed the worst were the students who did PBI2 in the Fall, but not the Spring. This means that PBI2 may have negatively affected their grade on the Fall 2010 final exam. In addition, the group of students that performed the best, in general, was not exposed to PBI2 in the fall or the spring. For the Spring 2011 final exam, the students who did not have PBI2 in the Fall, but PBI2 in the spring, preformed better than the students that had PBI2 the first semester, but not the second semester.

With respect to goal number 2, two JExam questions were written for the first exam (multiple answer format). These questions were designed to see how well students integrated material from the first two chapters (unit conversions plus chemical formulas and composition). Unfortunately, the two variations of the question was so difficult that only 4 students out of the 1840 got that question correct. Figures 2.3 and 2.4 show the two variations of the difficult question.

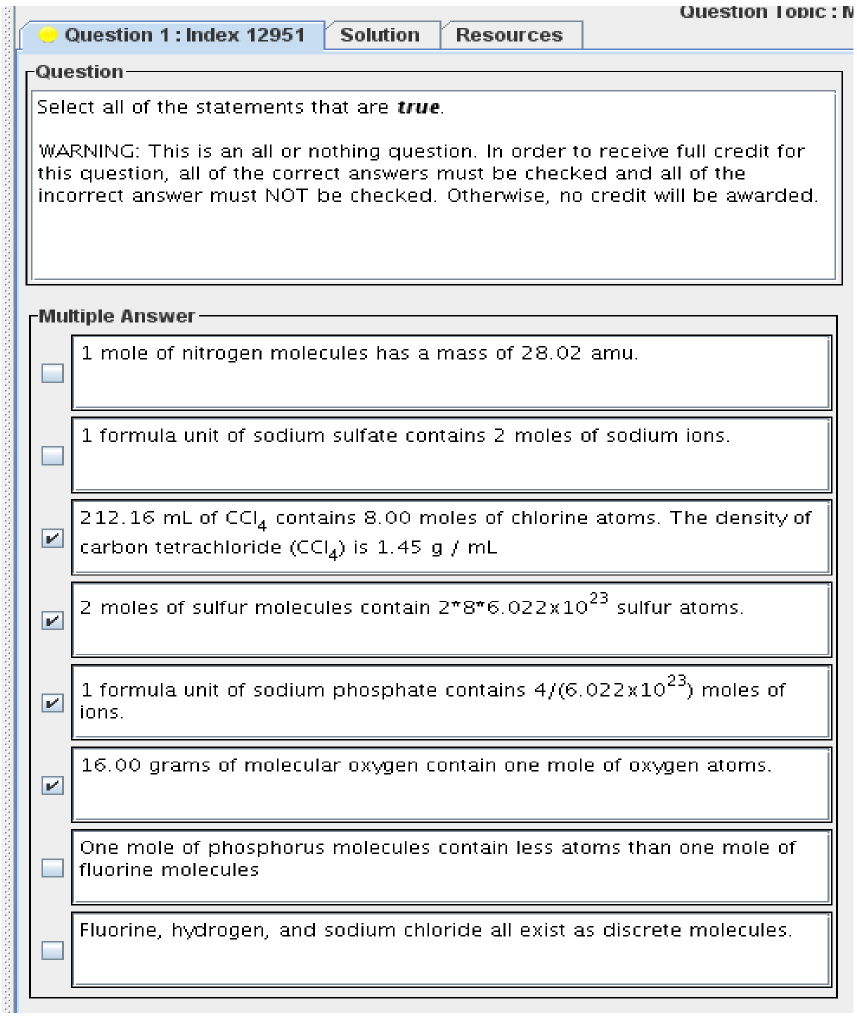


Figure 2.3 – Variation 1 of the difficult question

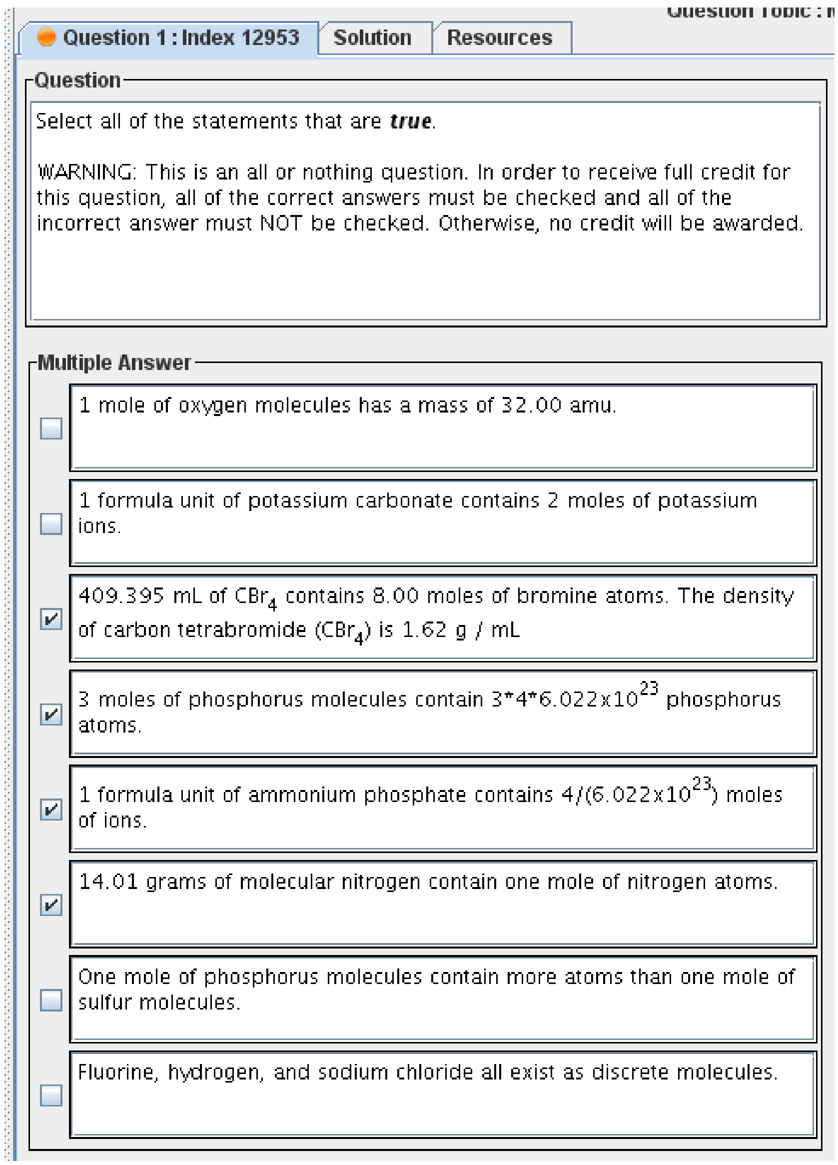


Figure 2.4 – Variation 2 of the difficult question

Goal 3 involved students' performances on the ACS standardized final exam. Even though most of the questions did not involve the integration of material, it was hoped that exposure to different topics and the fact that the students could review some of those topics would help their performance on the final exam.

Since PBI2 students did not take exam #2, the questions from the ACS final examination were categorized based on whether they involved material from the second exam or they involved material from the first and third exams. Table 2.1 summarizes this information. A group performed better if the difference in the IRT abilities were statistically significant at the 90% confidence interval. IRT abilities are further discussed in chapter 3. The IRT ability determines the difficulty of a question for a particular population. Easier questions have negative abilities and more difficult questions have positive abilities.

Table 2.1 – Comparison of Dr. Atwood's traditional students vs. Dr. Atwood's PBI2 students.

Total number of questions on the ACS Exam = 70	Fall 2010	Spring 2011
The no. of questions which PBI2 students did better on Exams 1 and 3	0 / 50	1 / 63
The no. of questions which traditional students did better on Exams 1 and 3	1 / 50	4 / 63
The no. of questions which PBI2 students did better on Exam 2	1 / 20*	0 / 7
The no. of questions which traditional students did better on Exam 2	7 / 20	0 / 7
The no. of questions which PBI2 students did better that were PBI2 related	1 / 12*	1 / 7
The no. of questions which traditional students did better that were PBI2 related	1 / 12	0 / 7

*The question where PBI2 students did better on exam 2 is the same question that is PBI2 related.

Both in Fall 2010 and Spring 2011, traditional students overall did better on final exam questions. However, when it came to the PBI2 questions, the PBI2 students' and the traditional students' performances were essentially the same.

In addition, the attitudes of the students in the PBI2 class were examined. A five point Likert scale was used for both Fall 2010 and Spring 2011 for the first four questions. The answers for the fifth question are provided. The five questions that were asked were:

1. The PBI2 (Problem Based Integrated Instruction) class helped me understand chemistry concepts better than the traditional lecture instruction.

2. The PBI2 (Problem Based Integrated Instruction) class will help me prepare for the final exam better.

3. The workload in the PBI2 (Problem Based Integrated Instruction) class is approximately the same as a traditional lecture class.

4. I enjoyed interacting and working with my group.

5. I would prefer the lecture format to be.

*A. Entirely PBI2 with no lecturing.

*B. More PBI2 than we experienced, but still some traditional lecturing.

*C. About the same as this PBI2 class.

D. Less PBI2, where groups meet less often in class.

E. Entirely traditional lecture, with no PBI2.

*Denotes a neutral or positive response

The results are summarized in Figure 2.5

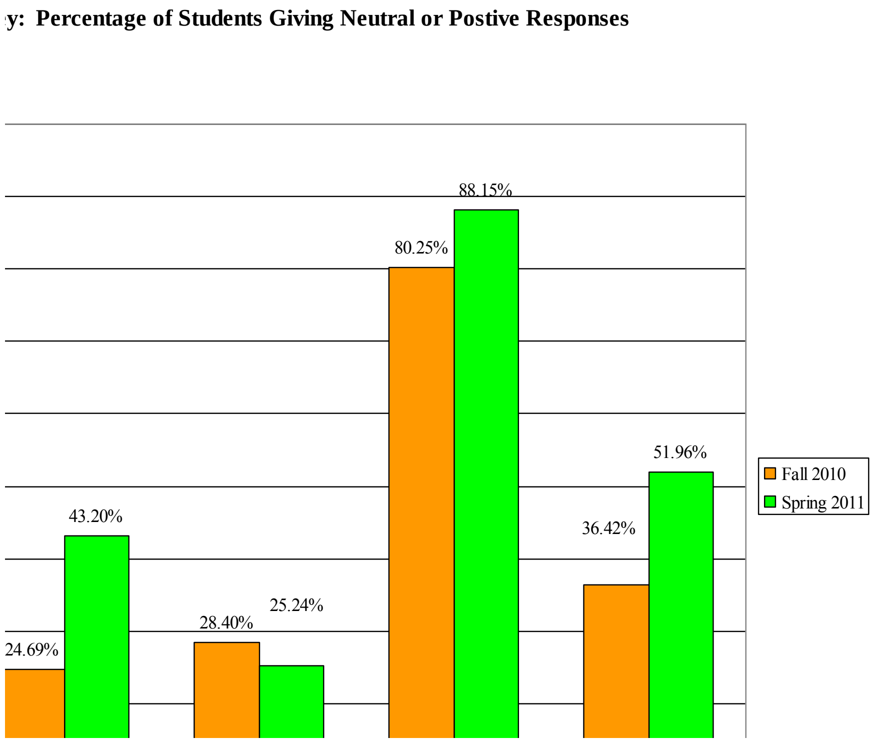


Figure 2.5 – A summary of the surveys. The analysis is based on the percentage of students that gave a neutral or positive response. The orange bars represent the Fall 2010 semester and the green bars represent the Spring 2011 semester

As can be seen, there was improvement in all areas from Fall 2010 to Spring 2011 except for the workload. The reasons for the improvement will be discussed in a later section.

Section 2.4 – Conclusions

PBI2 was not successful in reducing the number of C–, D, F, and W grade for Fall 2010, but was successful in reducing the number of C–, D, F, and W grades for Spring 2011.

PBI2 students' performances on the final exam were worse for PBI2 students in the Fall vs. the traditional lecture, but the differences were not statistically significant in the Spring vs. the traditional lecture.

PBI2 students were more positive in their opinions about the benefits of PBI2 in the Spring as opposed to the Fall.

On final exam questions that involved topics used to solve PBI2 problems, students tended to perform about the same as the traditional students. However, this is in contrast to non PBI2 topics, where PBI2 students tended to perform worse.

Even though PBI2 students did not perform well overall, the difference was greater for the second exam when compared to the first and third exams, especially for Fall 2010. There was less of a difference in the traditional and PBI2 class on ACS final exam questions whose topics were covered on the second hour exam.

Section 2.5 – Discussion

Overall, the PBI2 seemed to hurt students' performances in Fall 2010, but seemed to have an overall neutral effect in Spring 2011. Here are some possible reasons for the discrepancies.

The Fall 2010 semester was the first "draft" of the PBI2. One mistake was the difficulty of the first report defense. This caused many students to withdraw from the course. The other problem with the report defense was the student's perception of the relative stakes. Even though a report defense for a problem was only worth 4% of the overall grade, students really took it seriously, since it was done on JExam. Perhaps a clicker quiz, a quiz on paper or some other method which the students would associate with "low stakes" would have been a better option. This perception could have led to the fact that students thought the time involved was greater for PBI2 than for a traditional class.

Another part of PBI2 that negatively affected students in Fall 2010 was the lack of the second exam. Since students did not need to prepare for the second exam, students did not put effort into understanding the second exam topics and concepts, unless those topics were used in solving any of the PBI2 problems.

The Spring 2011 semester ran smoother. There were some things that were done during the first semester that caused not only more student stress, but also involved more work for the instructor. These were eliminated in the second semester. In addition, there were approximately 35 students who dropped the PBI2 class and added another class, either Dr. Atwood's 9:05 am class, or another CHEM 1212 class. This was observed when the enrollment went from about 360 students to 325 students. This self-selection may have removed students who felt that PBI2 was not beneficial.

Despite the changes made for the second semester, there was still the perception that the workload was not the same as for a traditional lecture session. Again, replacing the JExam report defense with an assessment perceived as "low stakes" may have helped with the workload perception issue.

After Spring 2011 semester, it was decided the PBI2 would not be implemented for a second year. However, a major reason for the suspension of the PBI2 was the fact that Dr. Atwood was leaving the University of Georgia for another position. It was unlikely that another instructor would be willing to have me take over one of their sections for an experimental instructional method which has not been shown to work (or perhaps be detrimental) in the past.

Another change that would help is to identify topics that are very easy. Kimberly Schurmeier (1) did research into general chemistry topics that students found difficult on the exams. Her research also identified topics that students found easy. Spending less time on these easier topics to allow more time for PBI2 may increase the chance of success in PBI2.

One final suggestion would be to build the course around PBI2, and not try to fit PBI2 into the course. PBI2 had to be fit into the existing course to compare two sections of chemistry (taught by the same instructor). Fitting the PBI2 part of the course into the existing course probably caused some duplication of topics, in addition to an increased workload for the students.

CHAPTER 3

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

ITEM RESPONSE THEORY (IRT)

Section 3.1 – Introduction

Item Response Theory (IRT) is a psychometric tool that analyzes individual responses to questions (called items). The responses can be dichotomous or polytomous . Dichotomous data only has 2 values, 0 or 1. Polytomous data has more than two values. One example of polytomous data is a 5 point Likert scale.

IRT is better than Classical Test Theory analysis because IRT indicates the relative difficulty of a question with greater precision (the difficulty parameter, b), and tells us how well a question discriminates between students of different abilities with greater precision (the discrimination parameter, a). Furthermore, the ability of a question is also determined based upon the abilities of students that answered the question correctly. For example, if two questions both have a 25% correct response rate, one of the two questions can have a higher difficulty if students of higher ability answered that question correctly. In addition, IRT also indicates the probability of a student with very low ability getting a question correct by guessing the answer (the asymptote, c). IRT also is better than classical test theory for data sets with more than 200 students (10).

To conduct IRT analysis there are several commercially available programs. Which program is used depends on the type of data that is analyzed. Since our data is dichotomous (which means each item is scored as right, wrong, or not-presented), the BILOG-MG 3.0 program is used (11). The BILOG-MG 3.0 program is designed to analyze dichotomous data. For polytomous data, other programs, such as PARASCALE 4 or IRTPRO 2.1 must be used (11).

In IRT the student ability and question difficulty are be plotted on the same scale. This scale varies from $-\infty$ to $+\infty$ but is more often depicted as varying from –4 to +4 or –3 to +3, depending on the software. Student abilities are normalized so the mean student ability is 0 and the standard deviation is 1. From this information we can determine if a set of questions overall is difficult or easy. The more positive the mean difficulty parameter is for a set of questions, the more difficult the question set.

Different fitting models can be used to analyze dichotomous data. Differences between the models are based upon the number of parameters (difficulty, discrimination, and asymptote) allowed to vary to fit the data.

The one-parameter (Rasch) model (10) varies only the difficulty parameter to fit the data, while fixing the discrimination parameter to 1.000 and the asymptote to 0. Because part of our research involves analyzing how well questions discriminate, in addition to the restrictive nature of the one-parameter Rasch model, the one parameter Rasch model was not used in our IRT analysis.

The two parameter (Birnbaum and Lord) (10) model varies the discrimination parameter and the difficulty parameter to fit the data, while fixing the asymptote to 0. The two parameter model is used for questions where there is no chance of a low ability student answering the question correctly due to random guessing, such as a question where you perform a calculation and enter a number.

The three parameter (guessing) model allows the discrimination parameter, difficulty parameter, and the asymptote to vary to fit the data (10). The three parameter model is used when there is a finite probability of a low ability student randomly guessing the answer to a particular item. This model is commonly used for multiple-choice questions, multiple answer questions, or questions where there are a finite number of logical answers (i.e. how many significant figures are in the final answer).

BILOG MG 3.0 can apply both the 2 parameter model and the 3 parameter model in the analysis of a single assessment. This can be done by setting the initial values of α and β for each item in BILOG MG 3.0. In BILOG MG 3.0 an item with initial values of $\alpha=2$ and $\beta=1000$ fixes the asymptote value to 0.001 (2 parameter model). An item given initial values of $\alpha=5$ and $\beta=17$ in BILOG MG 3.0 allows the asymptote to vary (3 parameter model).

Section 3.2 – Item Characteristic Curves (ICC) and Probability

One graph produced in IRT analysis is an Item Characteristic Curve (ICC). An ICC plots the probability of answering a question correctly vs. the student ability. The probability of a student with an ability of theta (θ) answering a question correctly is given Equation 3.1 (10).

Equation 3.1

Where: θ is the ability level of the student

a is the discrimination parameter of the question

b is the difficulty parameter of the question

c is the asymptote of the question

An example of an item characteristic curve (ICC) using the 3-parameter model is given in Figure 3.1. The difficulty parameter (b) is 1.283, which means the question is relatively difficult. The discrimination parameter (a) is 2.006, which means the question does an excellent job of discriminating between students of different abilities. The asymptote (c) is 0.131, which means there is a 13.1% probability of a very low ability student correctly answering the question by guessing.

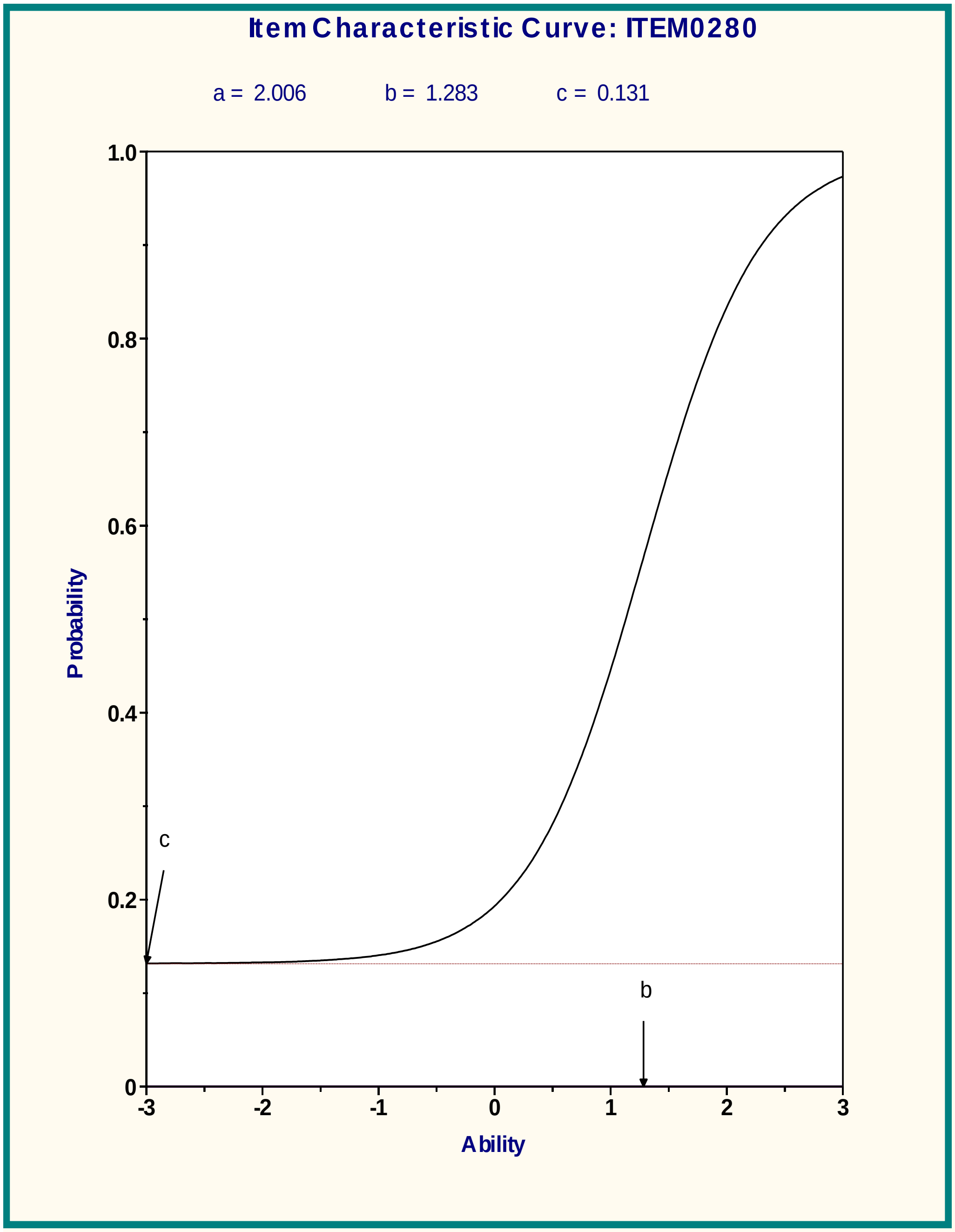


Figure 3.1 – Item Characteristic Curve for Item #280

The inflection point in the ICC corresponds to the question's difficulty parameter. Easier questions have more negative abilities. More difficult questions have more positive difficulty parameters. Figures 3.2 and 3.3 give examples of ICC curves for a difficult and an easy question respectively.

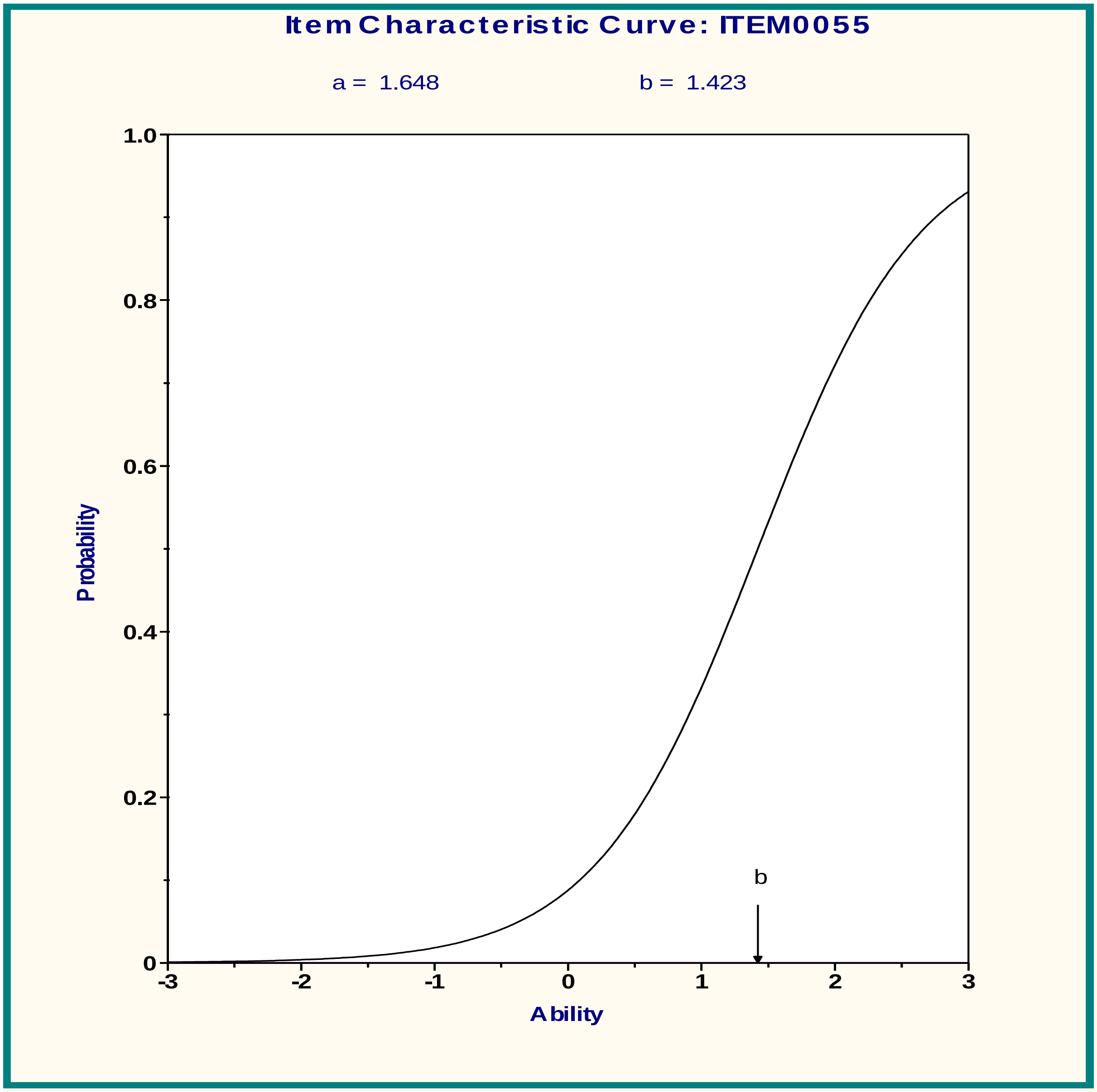


Figure 3.2 – Item Characteristic Curve for a Relatively Difficult Question (Item #55)

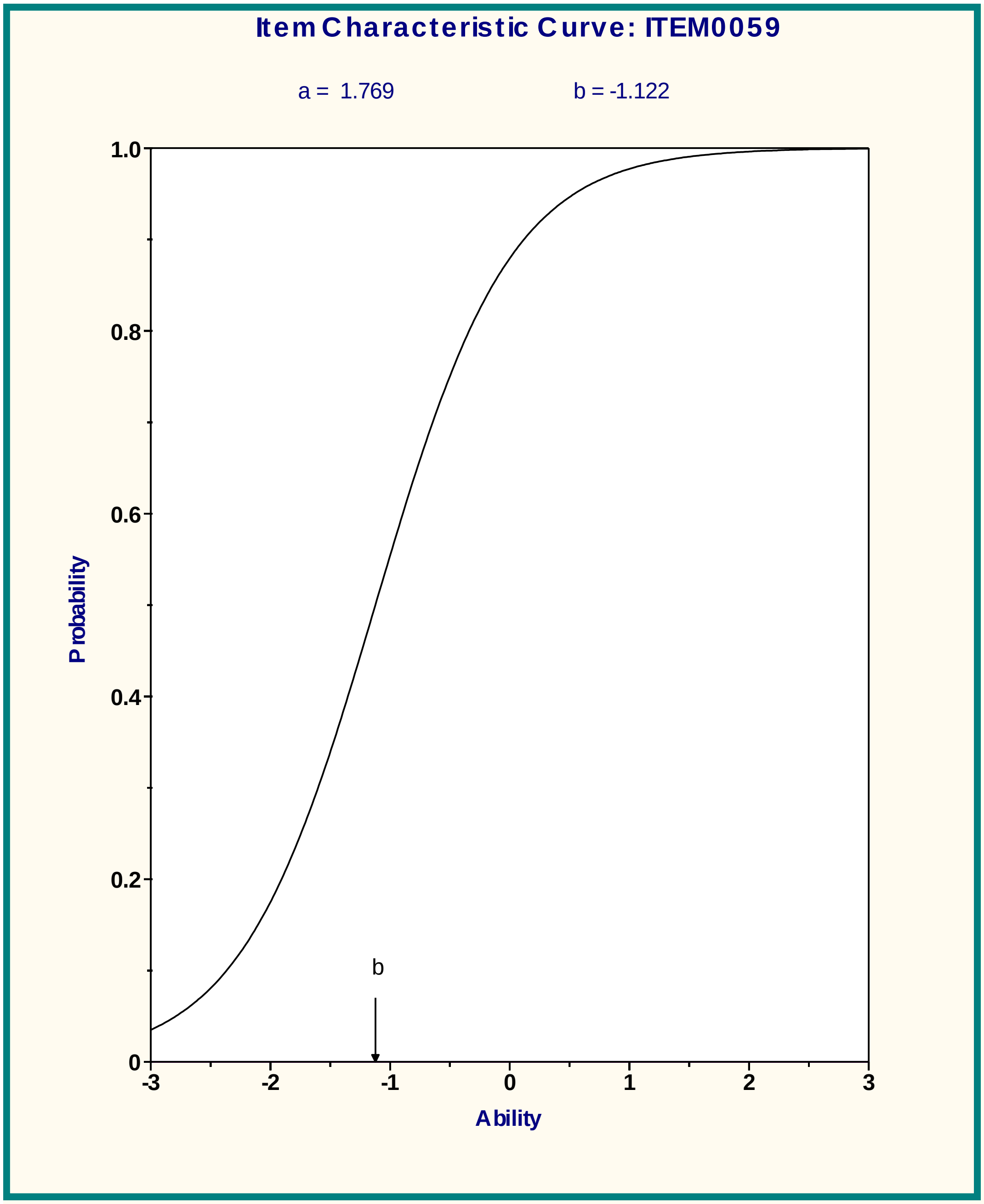


Figure 3.3 – ICC Curve for a Relatively Easy Question (Item #59)

Question discrimination on ICC's is reflected in the slope of the probability curve. Questions that discriminate well have steeper slopes with discrimination parameters greater than 1.000. Questions that discriminate poorly have less steep slopes with discrimination parameters less than 0.500. Figures 3.4 and 3.5 are examples of ICC's for an excellent discriminating question and a poor discriminating question, respectively.

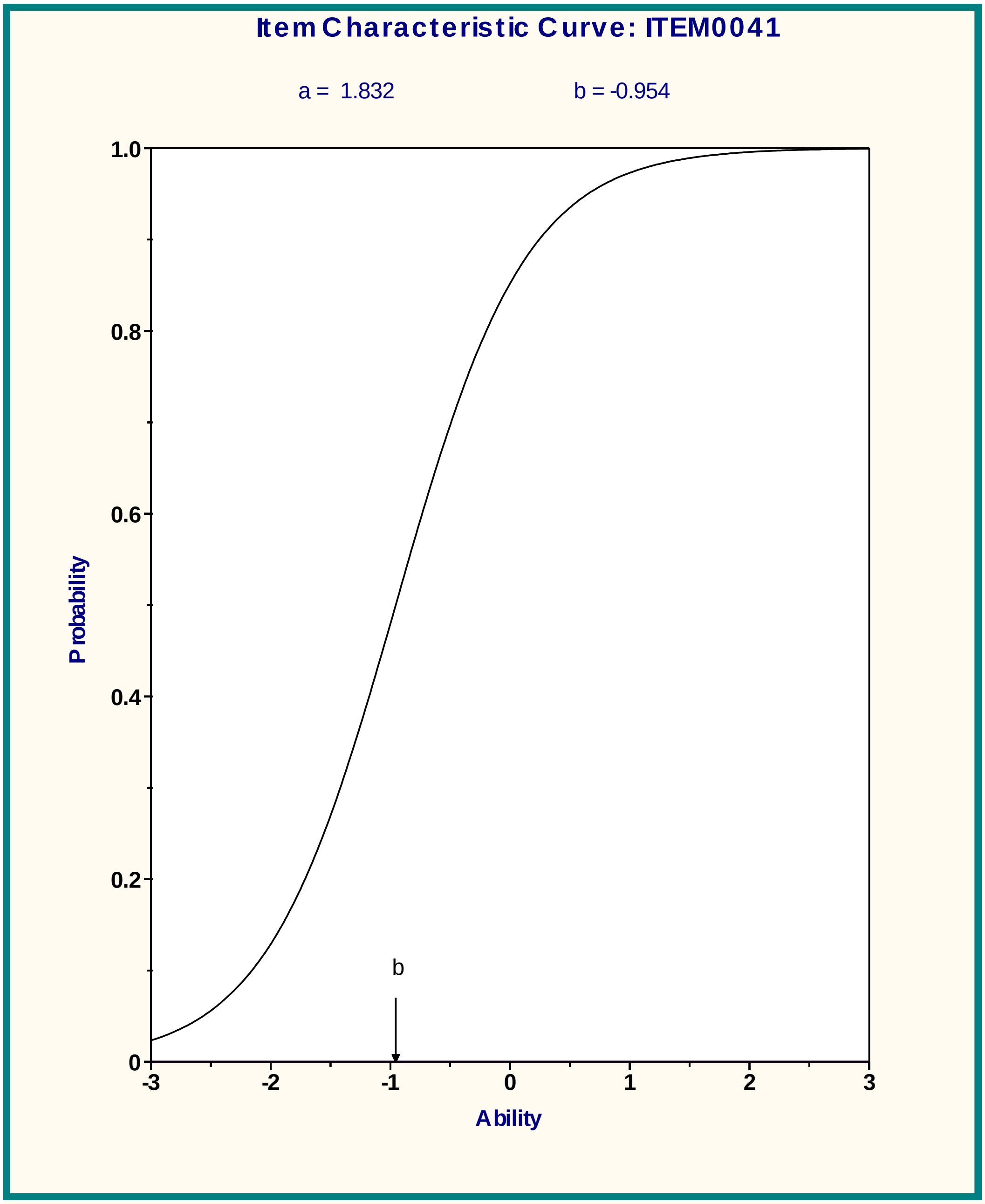


Figure 3.4 – ICC Curve for an Excellent Discriminating Question

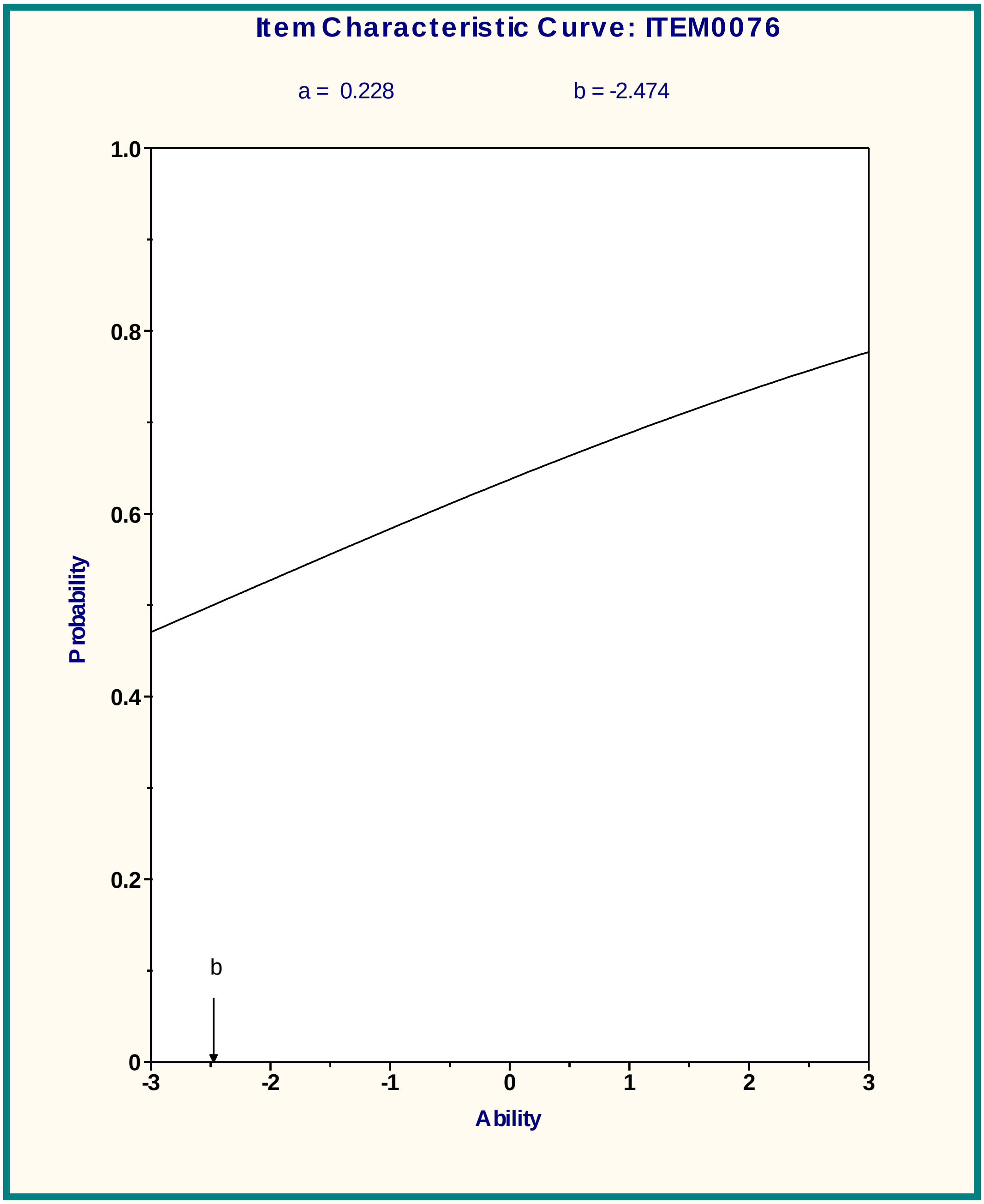


Figure 3.5 – ICC Curve for a Poorly Discriminating Question

Section 3.3 – IRT Information Relationships – Information Curves

In IRT it is possible to determine what range of student abilities each question most effectively assesses. More difficult questions discriminate between students of high and very high ability but do not discriminate between lower ability students primarily because lower ability students always incorrectly answer the question. Easier questions discriminate between students of low ability but not high ability students. This is because the high ability students always answer the question correctly. IRT displays this property using item information curves (IIC). An IIC is a plot of ability range versus information. In IRT vernacular, information is to be interpreted as “this question allows us to effectively assign a student ability to students in this ability range.” The information obtained from a particular item in the 3 parameter model is calculated using equation 3.2 (10).

Equation 3.2

Where: I_i = the amount of information for the item i

D = scaling factor (1 in our case)

a_i = the discrimination parameter for item i

c_i = the asymptote for item i

p_i = the probability of answering question i correctly at an ability of θ

θ = the student ability

For the 2 parameter model, where c = 0, the equation reduces to Equation 3.3 (10).

Equation 3.3

An IIC for a highly discriminating question, item # 280, is given in Figure 3.6. The ICC for item #280 was shown in Figure 3.1.

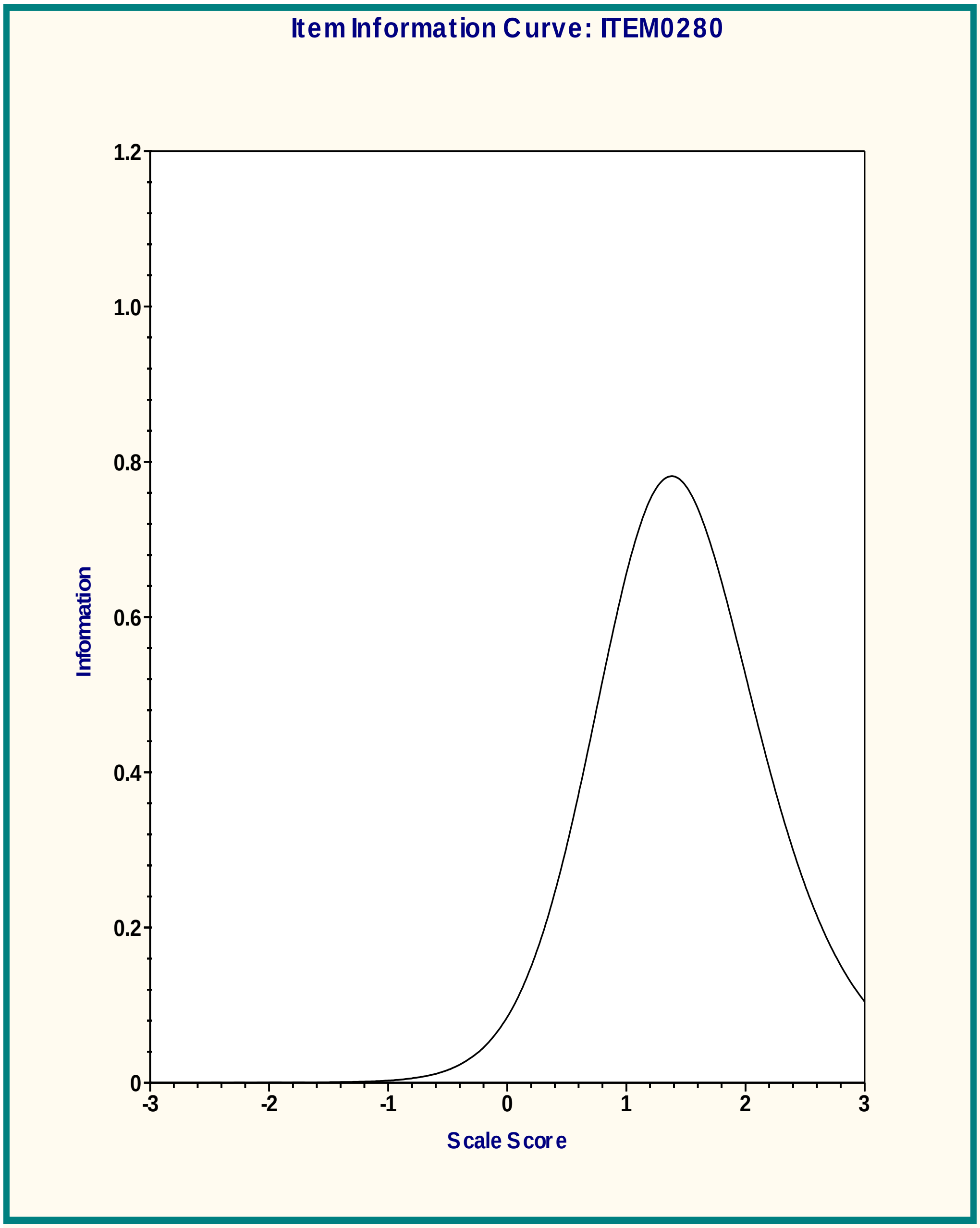


Figure 3.6 – Item Information Curve for Item #280

Note that between abilities of –3 and –1, virtually no information about student abilities was obtained. This is due to the fact that all students with an ability score between –3 and –1 have the same probability (13.1%) of answering the question correctly (see Figure 3.1). The ICC peak is centered at the question difficulty parameter, 1.283 in this case. A well written assessment has items with a variety of difficulty levels. This permits an examiner to obtain information about students of all abilities.

To determine IRT information about students of all abilities on an assessment (such as an exam), it is possible to sum item information curves for all of the questions in the assessment yielding a total information curve (TIC). The TIC, $I(\theta)$ is calculated using Equation 3.4 (11).

Equation 3.4

Where $Q_i(\theta) = 1 - P_i(\theta)$

$P_i(\theta)$ = The probability of an individual item

The standard error of the ability estimates for a student at an ability level of θ is given by (11)

Ideally, the total information curve for an assessment is centered at zero ability and is normally distributed which allows us to acquire maximum information about students of all abilities. (Remember that student abilities are normalized to a mean of 0 and a standard deviation of 1). An example of a total information curve is given in Figure 3.7.

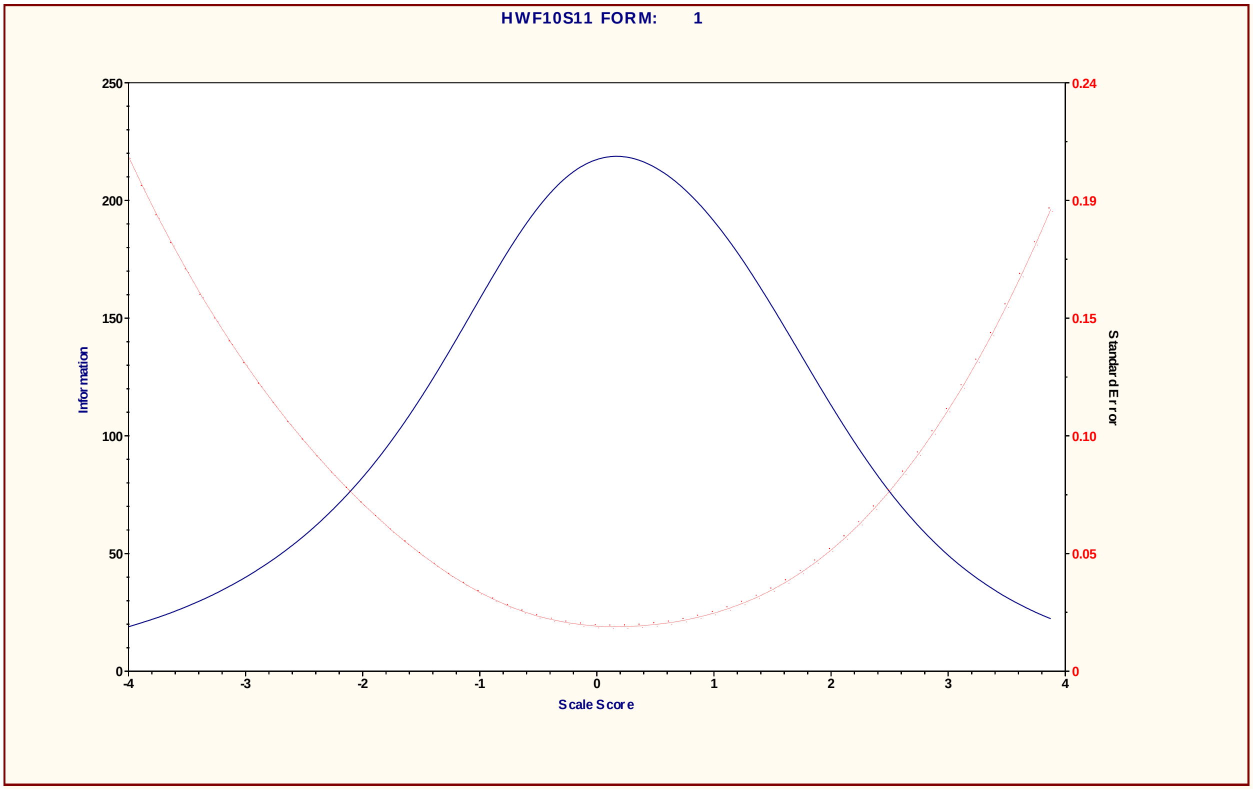


Figure 3.7 – Total Information Curve. Note that the total information curve is centered at an ability of 0.2 (blue solid line). The solid line is the total information vs. the ability. The dashed line is the standard error for estimating the student ability.

Section 3.4 – Student Abilities and Relationship to Classical Statistics

One advantage of IRT over classical statistics is the ability of the student does not depend on the difficulty of the assessment since the student abilities are normalized. A histogram of the distribution of student abilities is shown in figure 3.8.

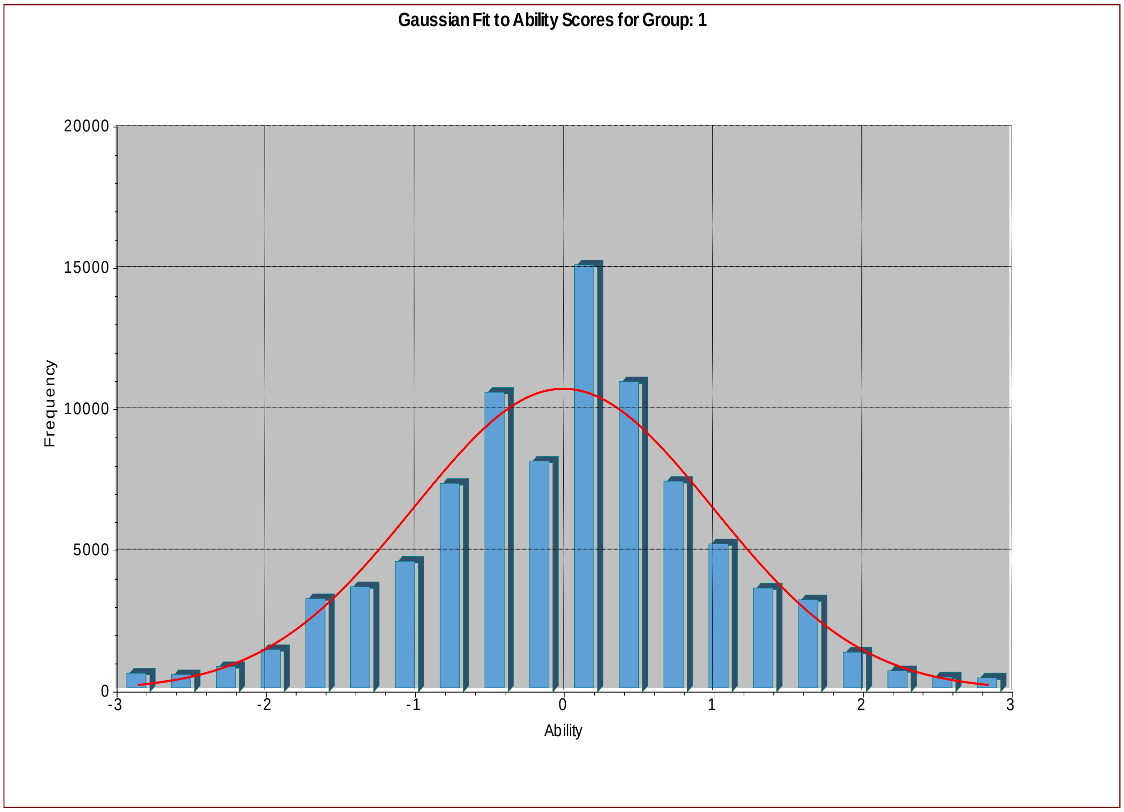


Figure 3.8 – Histogram of Student Abilities

BILOG MG 3.0 can produce a bivariate plot that correlates student ability with the percentage of questions the student correctly answered. In general, an A student correctly answers 90% or more of the questions correctly. B students lie in the range of 80 to 89% correct responses and so forth for the other grades, the bivariate plot provides a method to associate letter grades (A, B, C, etc.) with IRT abilities. An example of a bivariate plot is shown in figure 3.9

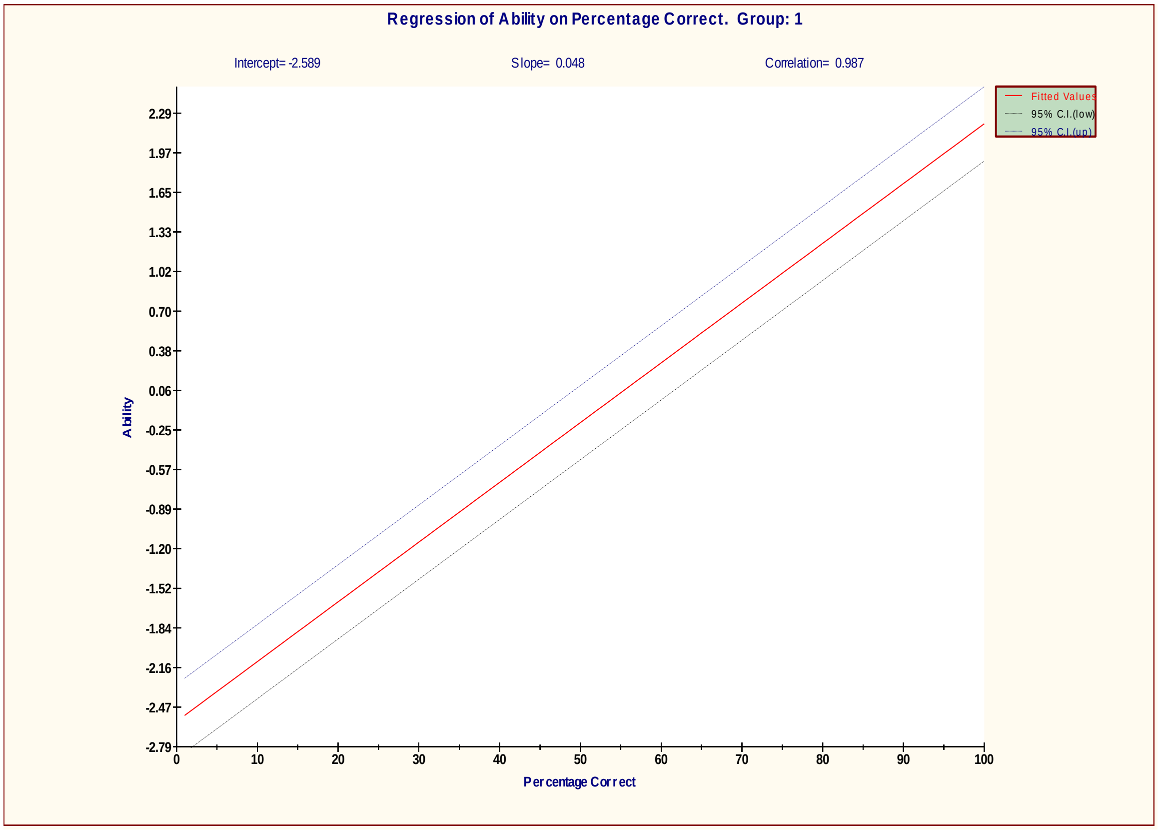


Figure 3.9 – Bivariate Plot of Student Ability vs. Percentage Correct. The red solid line is the best fit line and the dashed lines are the lower and upper limits of the 95% confidence interval

In figure 3.9, the slope of the line is 0.04789 (calculated from the raw data) and the intercept is –2.589. This information can be used to calculate the percent correct for a student having an IRT ability of +1.346.

$1.346 = (0.04789)\ x - 2.589 \qquad\qquad x = 82.17\%$

Even though this student is above average, their IRT ability corresponds to a weak "B" student, assuming a traditional grading scale (90.0% and above = A, 80.0% – 89.9% = B, 70.0% – 79.9% = C, 60.0 – 69.9% = D, 59.9% and below = F).

Section 3.5 – Data Analysis and Output

Before data analysis begins the question response data must be arranged into an input file format that is readable by BILOG MG 3.0. Each line of the data corresponds to the response pattern for a single student. The first 9 characters are a unique identification number for the student. The identification number can be up to 30 alphanumeric characters. A single space follows the identification number, followed by the responses for each question for that student.

A score of "0" means the student responded incorrectly to any part of the question. A score of "1" means the student responded correctly to all parts of the question. A score of "9" means that the question was not presented to the student. An example of an input file for 5 students and 9 questions is shown below:

```
          999999999
434235663 110991100
434255887 999110019
434255993 199999999
434257054 001101010
434257389 090991901
```

The first line with the spaces and the string of nines tell the BILOG MG 3.0 programs what characters represent questions that were not presented.

As an example, the student identified as 434235663 responded to questions 1, 2, 6, and 7 correctly, responded to questions 3, 8, and 9 incorrectly, and was not presented questions 4 and 5. Questions may not be presented for a variety of reasons such as: they were optional assignments, they were not randomly selected in mastery mode, or they were not assigned by the instructor.

Initially, BILOG MG 3.0 uses classical statistics to conduct a preliminary data analysis. Classical statistics calculated include the percentage correct, the Pearson correlation coefficient and the biserial coefficient. An example of the output from Phase 1 of BILOG MG 3.0 is shown below:

89285 OBSERVATIONS READ FROM FILE: TRY2_ALL_GRADES_V36.DAT

89285 OBSERVATIONS WRITTEN TO FILE: MF.DAT

ITEM STATISTICS FOR SUBTEST T2ATTUNL

						ITEM*TEST CORRELATION	
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT	PEARSON	BISERIAL
1	ITEM0001	31.0	20.0	64.5	-0.60	0.094	0.121
2	ITEM0002	30.0	17.0	56.7	-0.27	0.224	0.282
3	ITEM0003	29.0	7.0	24.1	1.15	0.420	0.577
4	ITEM0004	3841.0	2519.0	65.6	-0.64	0.285	0.367

5	ITEM0005	3896.0	2112.0	54.2	-0.17	0.283	0.355
921	ITEM0921	64.0	48.0	75.0	-1.10	0.034	0.046
922	ITEM0922	63.0	61.0	96.8	-3.42	-0.124	-0.303
923	ITEM0923	51.0	22.0	43.1	0.28	0.249	0.314
924	ITEM0924	79.0	47.0	59.5	-0.38	0.272	0.344
925	ITEM0925	62.0	42.0	67.7	-0.74	0.137	0.178
926	ITEM0926	35.0	30.0	85.7	-1.79	0.432	0.670
927	ITEM0927	35.0	28.0	80.0	-1.39	0.251	0.359
928	ITEM0928	31.0	17.0	54.8	-0.19	0.152	0.191
929	ITEM0929	33.0	20.0	60.6	-0.43	0.124	0.157

**** ITEM 232 WITH BISERIAL R LESS THAN -0.15 WILL NOT BE CALIBRATED. ****

**** ITEM 234 WITH BISERIAL R LESS THAN -0.15 WILLNOT BE CALIBRATED. ****

**** ITEM 920 WITH BISERIAL R LESS THAN -0.15 WILLNOT BE CALIBRATED. ****

**** ITEM 922 WITH BISERIAL R LESS THAN -0.15 WILLNOT BE CALIBRATED. ****

The biserial r for item j is calculated using Equation 3.5

Equation 3.5

Where: μ_x = mean total score for all students

μ_j = mean total score for students who responded correctly to item j

p_j = item difficulty index for item j

$q_j = 1 - p_j$

σ_x = standard deviation of the total score for all students

Questions with a biserial r of less than –0.15 are not calibrated by BILOG MG 3.0. These are questions where the slope is likely to be negative (i.e. lower ability students have a greater probability of answer the question correctly than higher ability students). BILOG MG 3.0 does not calibrate these items because the program thinks these items have been “miskeyed” (i.e. there is an error in the answer key).

Once the “a” and the “b” parameters are estimated, Marginal Maximum Likelihood Estimation (MMLE) is used to fit the data in Phase 2 of BILOG MG 3.0. In the BILOG MG 3.0 MMLE person parameters are initially treated as unknown quantities. An example of the MMLE output is shown below:

ITEM0926	1.582	1.217	-1.299	0.773	0.000
	0.524*	0.469*	0.000*	0.000*	0.000*
ITEM0927	1.105	1.039	-1.063	0.721	0.000
	0.454*	0.384*	0.000*	0.000*	0.000*
ITEM0928	-0.161	0.751	0.214	0.601	0.000
	0.399*	0.271*	0.000*	0.000*	0.000*
ITEM0929	0.081	0.801	-0.101	0.625	0.000
	0.393*	0.289*	0.000*	0.000*	0.000*

----

* STANDARD ERROR

LARGEST CHANGE = 0.147676

PARAMETER	MEAN	STN DEV
SLOPE	0.950	0.315
LOG(SLOPE)	-0.104	0.327
THRESHOLD	-0.863	1.540

-2 LOG LIKELIHOOD = 3000836.835

1

SUBTEST T2ATTUNL; ITEM PARAMETERS AFTER CYCLE 25

ITEM	INTERCEPT	SLOPE	THRESHOLD	LOADING	ASYMPTOTE
ITEM0001	0.967	0.633	-1.527	0.535	0.000
	0.416*	0.229*	0.000*	0.000*	0.000*
ITEM0002	0.705	0.784	-0.899	0.617	0.000
	0.425*	0.285*	0.000*	0.000*	0.000*
ITEM0003	-0.780	1.197	0.652	0.767	0.000

```
                 |  0.492*     |  0.447*     |  0.000*             |  0.000*     |  0.000*

                 |             |             |                     |             |
```

In addition to determining “a”, “b”, and “c” parameters, BILOG MG 3.0 also calculates the intercept and loading for each item. For our analysis, we are not concerned about these parameters.

Once initial parameters have been calculated for each of item, the Expectation a Posteriori (Bayes Estimation) method (11) is used to assign abilities for each of the students in the last steps of the program. An example of this output is shown below:

```
GROUP   SUBJECT IDENTIFICATION                        MARGINAL

 WEIGHT   TEST     TRIED  RIGHT  PERCENT    ABILITY   S.E.      PROB

 ------------------------------------------------------------------------

  1  104321CAF                                 |                  |

  1.00   T2ATTUNL   134      65        48.51   |  -0.4595   0.1147  |  0.000000

  1  104321CA6                                 |                  |

  1.00   T2ATTUNL   131      67        51.15   |  -0.4492   0.0651  |  0.000000

  1  104321C97                                 |                  |

  1.00   T2ATTUNL   134      88        65.67   |   0.4393   0.0676  |  0.000000

 ------------------------------------------------------------------------
```

CHAPTER 4

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

ANALYSIS OF OWL QUESTION MODE DATA

Section 4.1 – Introduction

OWL data involved approximately 10 million individual responses, consisting of 90826 students and 1203 questions. Question mode is the traditional method of presenting questions. Approximately 15% of all of the responses from students for the 2010 – 2011 academic year were responses from questions that were presented in question mode (12, 13).

Section 4.2 – Experimental

The data was sorted based on the number of attempts students were allowed for each question. Different categories were 1 or 2 attempts, 3 or 4 attempts, 5 or 6 attempts, 7 to 10 attempts, and 11 or more (or unlimited) attempts. Approximately 95% of the data involved questions where students were allowed 11 plus attempts, so the analysis concentrated on that data set. 1145 questions in this set of 1203 contained responses where 11 or more attempts were allowed.

Responses from the 11 plus attempts were converted into a format that could be read by the BILOG MG 3.0 program using the steps outlined in section 4.3. In converting the data, the following assumptions were made:

1. Each time a student attempted a question, it was counted as an attempt. This assumption was made to deal with students who saw the same question across two semesters. In the second semester, the OWL attempt number was "reset", so an individual student may have attempted a particular question three times during their first semester and five times during their second semester. OWL sets the attempt number as 1, 2, 3 for the first semester and 1, 2, 3, 4, 5 for the second semester. Since the student attempted the question three times in their first semester, the attempt count for the second semester was reset to 4, 5, 6, 7, 8.

2. Response patterns were converted into dichotomous data. A student responding correctly to a question or all parts of a multi-part question was assigned a score of "1". A student responding incorrectly to a question or any part of a multi-part question was assigned a score of "0".

3. Once a student answered a question correctly on a given attempt, the student was assigned a raw score of "1" for all subsequent "attempts".

These 1145 questions covered several different levels of chemistry including introductory chemistry, quick prep chemistry, mathematics, elementary organic chemistry, and general chemistry. The mathematics questions and organic chemistry questions were removed from the analysis leaving a total of 1007 questions. Further analysis showed that of the 1007 questions only 312 questions were clearly general chemistry course questions.

For the 312 general chemistry questions, responses from attempts 1, 2, and 3 were analyzed using IRT. Total information curves (TIC's) were generated for each attempt to determine which of the three attempts had the most normal distribution around the mean. This analysis revealed that the TIC for the second attempt had the most normal distribution around the mean.

Therefore, we based the remainder of our work on the students' second attempts. IRT parameters from the second attempt were further analyzed to identify difficult questions and their associated topics.

Employment of the three parameter model frequently caused BILOG MG 3.0 to fail to converge. This indicates that the three parameter model is inappropriate for these questions. This could be due to the lack of responses from low ability students (10). Therefore, the two parameter model was employed in the OWL question mode analysis.

Section 4.3 – Data Formatting

The data was formatted in a tab delimited table with the following fields: DatabaseID, CourseNumber, StudentID, ModuleNumber, IUNumber, QuestionNumber, RawScore, AnswerDate, AttemptNumber, Repeatquestion, Repeatmodule, Displayfeedback, Displayanswer, Partialcredit, Maxscore, Timepermitted, Questionresubmission, Displayhints. Arrangement of the raw data to a BILOG MG 3.0 format was performed in the following manner:

1) The data was separated into groups of approximately one million lines approaching the limit of the maximum number of Microsoft Excel rows. Each chunk of data was unique with respect to the questions numbers. 2) The data was sorted based on the number of attempts as previously mentioned into 1 or 2 attempts, 3 or 4 attempts, 5 or 6 attempts, 7 to 10 attempts, and 11 or more attempts. 3) The institution ID and student ID were combined to produce a unique 10 digit ID number for each student. 4) Unnecessary columns of data were deleted. 5) Because the OWL partial credit mode was employed for approximately 70% of the data, the raw data contained values greater than 0 and less than 1. We converted this to dichotomous data by setting all

attempt values equal to or greater than 0 and less than one to 0's. 6) The data for each attempt was put into a pivot table where each row was a different student and each column was a different question. Each cell contained a "0", a "1", or was blank. A "0" indicates the student did not successfully answer the entire question, a "1" means that the student successfully answered the entire question, and a blank means the student was not presented that question. 7) Just prior to submission of the data for BILOG MG 3.0 analysis, blank cells in the pivot table were replaced with a "9" to indicate to BILOG MG 3.0 that these questions were not presented to the student. 8) The table was saved as a .csv file then edited into a format acceptable for the BILOG MG 3.0 program (student ID, space, then a string of 0's, 1's, or 9's). 9) If several data sets were involved, a Perl script was used to combine the data sets for each of the attempts. The Perl script ensures that in the combined data set, students who did not answer any questions in a particular set were given all 9's for that set.

Inability of the IRT model to converge indicates that the data contains "bad" questions. Bad questions have one of these characteristics. 1) The question has less than 25 student responses. 2) The question has a high (>98%) correct response rate. 3) The question has a low (<5%) correct response rate. These bad questions were removed and the analysis performed again.

For step by step directions on how the data was formatted, see Appendix A.

Section 4.4 – Results – General Chemistry Questions – Total Information Curve (TIC) Analysis

To determine difficulty parameters for the questions, we need to determine which of the up to 11 attempts to analyze. We must choose that attempt which produces the most normal distribution around the mean. This required that we generate total information curves (TIC) for the 1st, 2nd, and 3rd attempts from the response patterns (figures 4.1, 4.2, and 4.3).

As previously mentioned, the peak location of the TIC curve will give us an idea of the overall assessment difficulty. We postulated that the overall assessment difficulty decreased with each attempt. As predicted the TIC curve peak shifts to lower difficulty (towards the left), as shown in figures 4.1, 4.2, and 4.3.

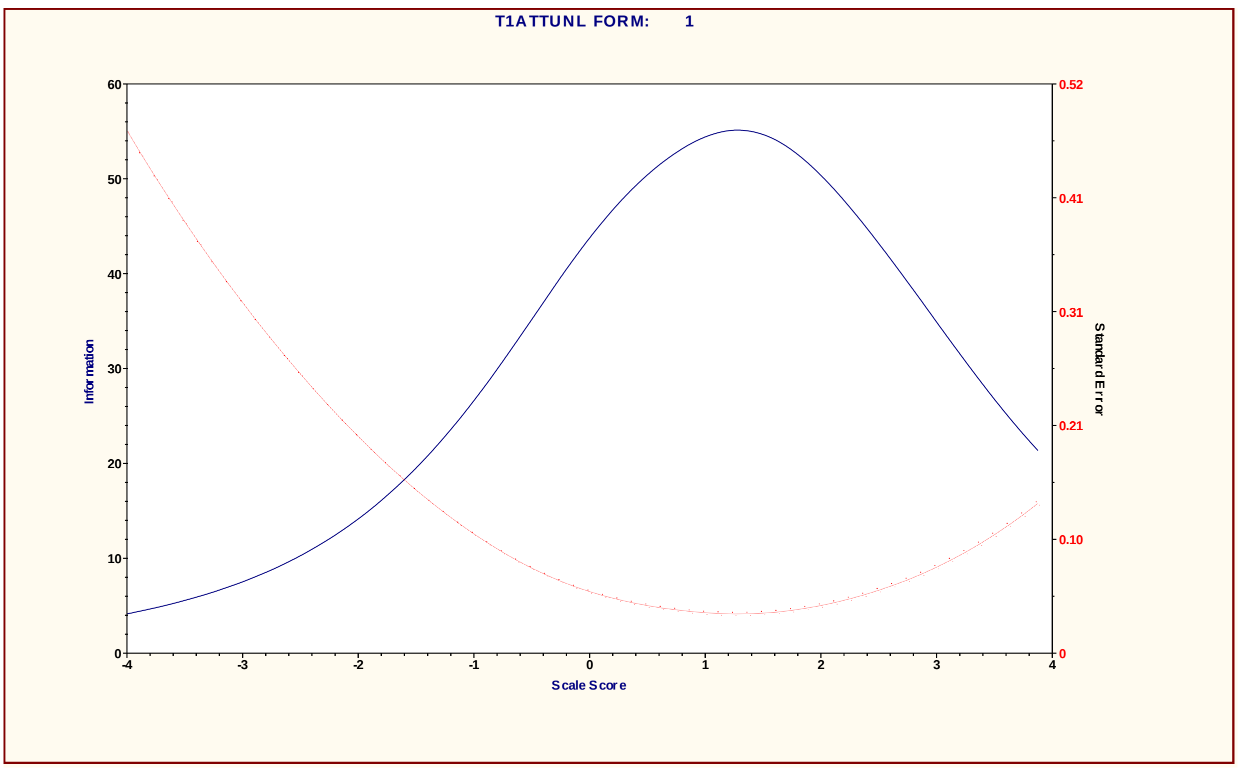


Figure 4.1 – Total Information Curve (TIC curve) – Attempt #1 – General Chemistry Questions Only. The TIC curve peak occurs at an ability of +1.250

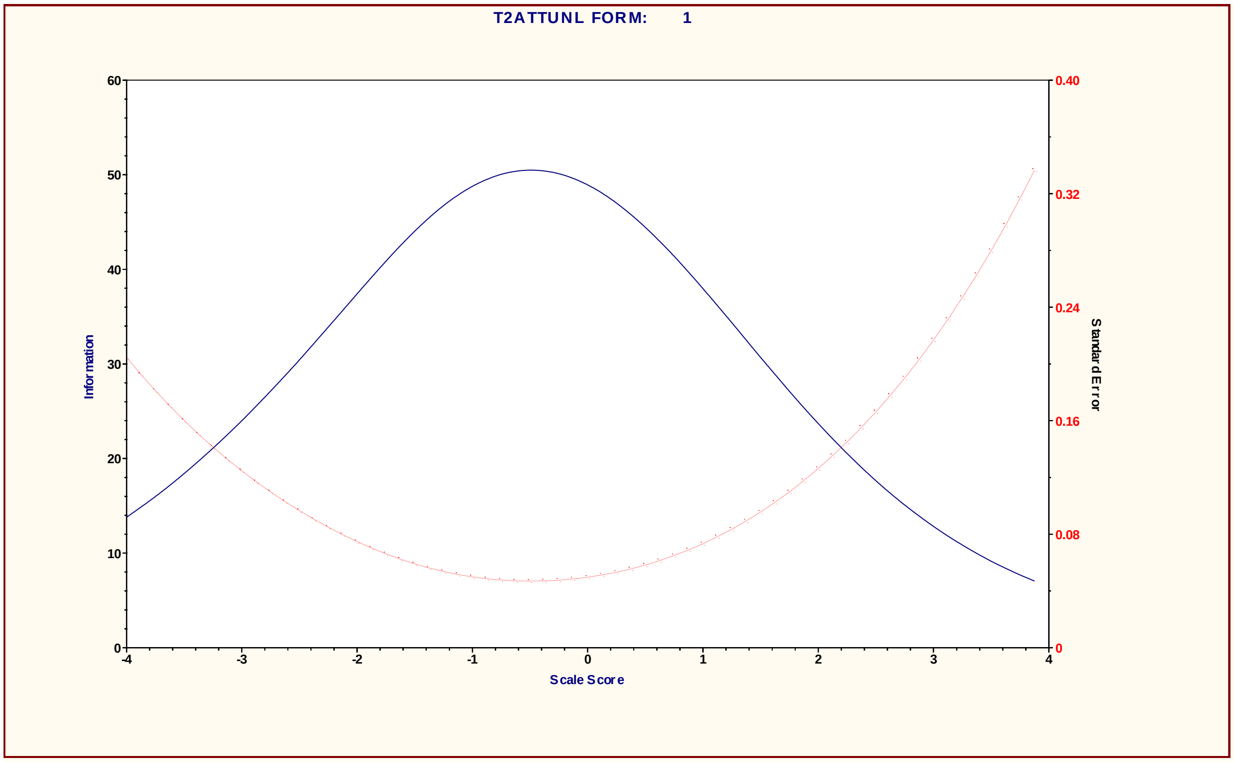


Figure 4.2 – Total Information Curve (TIC curve) – Attempt #2 – General Chemistry Questions Only. The TIC curve peak occurs at an ability of –0.500

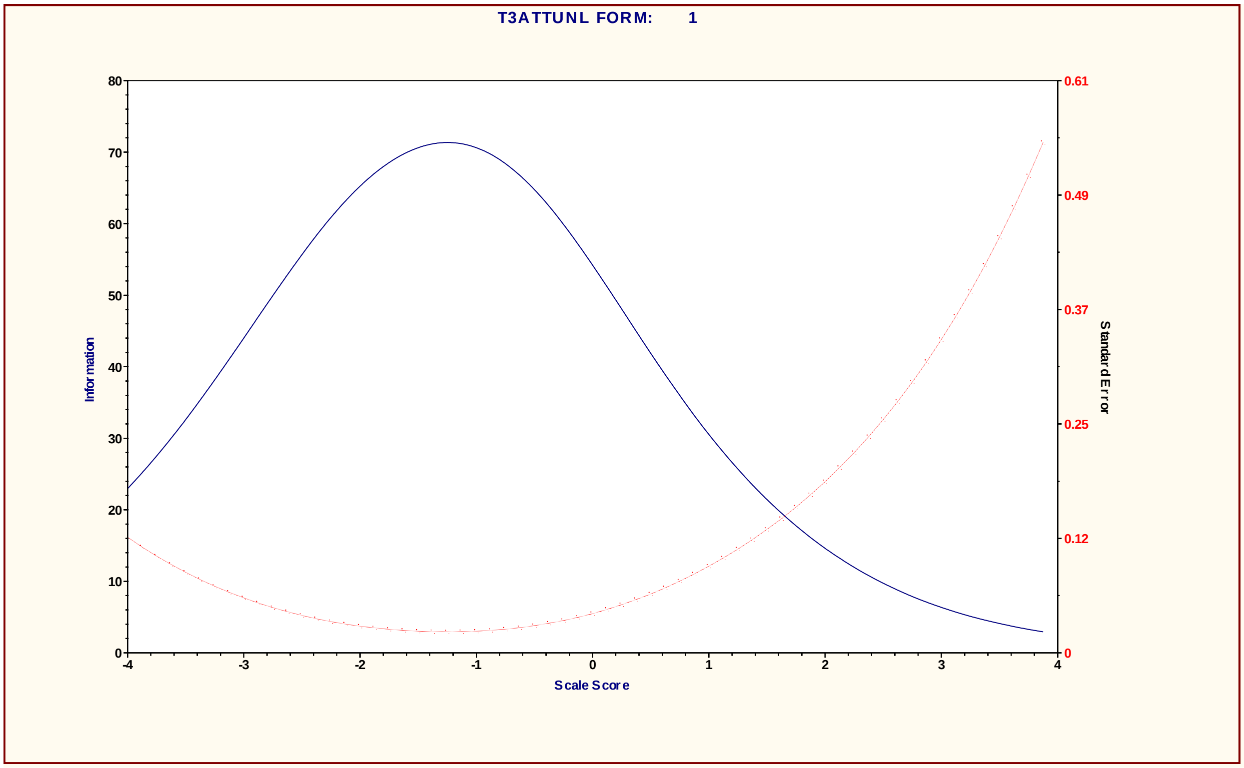


Figure 4.3 – Total Information Curve (TIC curve) – Attempt #3 – General Chemistry Questions Only. The TIC curve peak occurs at an ability of –1.250

The TIC curve that produces the most normal distribution around the mean is the TIC curve from the second attempt. A further indication that attempt #2 is the correct choice is shown in figure 4.4, a histogram of the student ability distribution. This histogram has a mean of 0 and a normal distribution. Based on this data, we will use the second attempt to compare student abilities as well as question and topic difficulties.

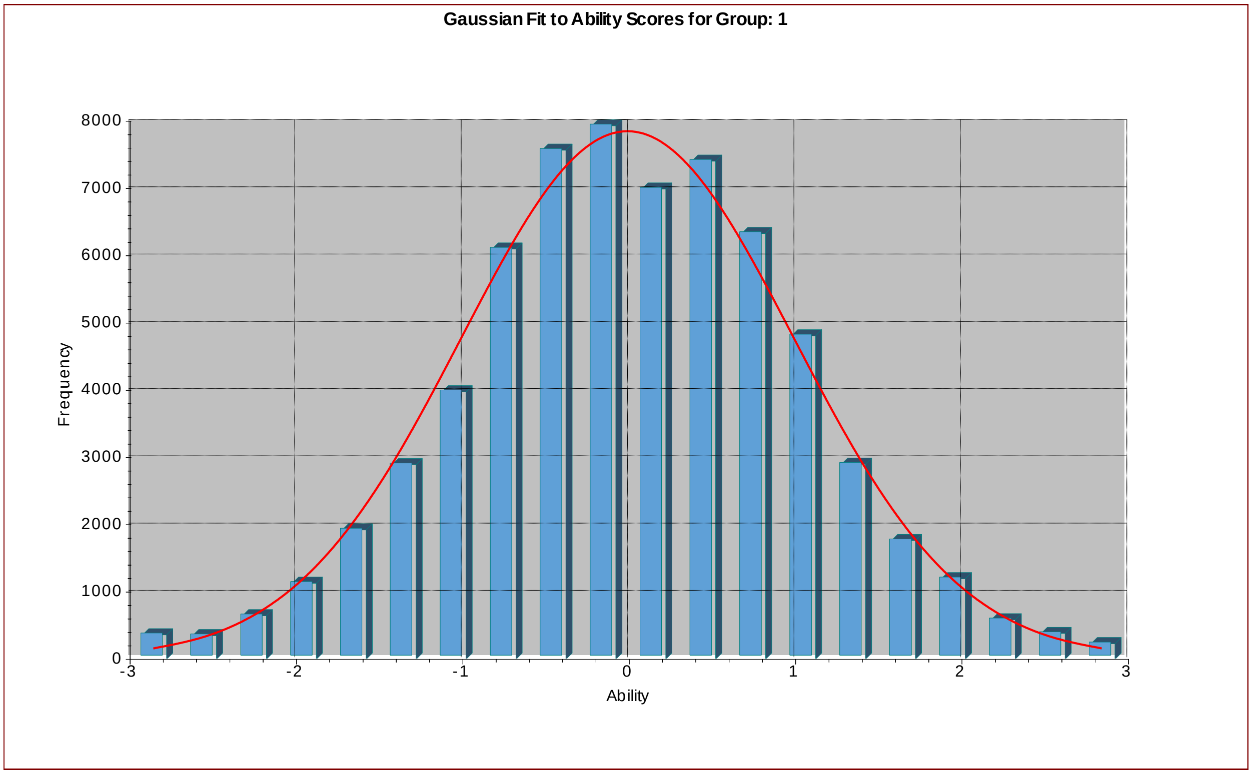


Figure 4.4– Histogram of student abilities for Attempt #2 – General Chemistry Questions only

Section 4.5 – Difficult Topic Analysis – General Chemistry Questions

For the purposes of this study we will define a difficult topic as one where the average ability of all the questions in that topic is greater than 1.067. This ability represents the top 10% of the question database.

Table 4.1 – Difficult OWL topics and question parameters in question mode

Difficult OWL topic (no of questions)	Tries	Avg. Diff. (b)	Avg. Discrim. (a)
Thermochemical equations (1)	26	3.506	0.861
Lewis acids and bases (1)**	5634	3.114	0.484
Redox Balancing – Basic solution (1)	30	2.740	0.890
Redox Balancing – Acidic solution (1)	34	2.569	0.761
Selective Precipitation (1)	37	2.531	0.988
Molarity of Ions in Solution (1)*	68	2.315	0.438
Crystal field splitting (3)	1135	2.314	0.614
Calculate K_{sp} Given Solubility (1)	4378	1.863	1.102
Strong Electrolytes in Aqueous Solution (1)*	98	1.825	0.614
Solubility of Insoluble Base in Acid (1)	36	1.791	0.744
Preparation of Buffer Solutions (2)	1434	1.688	1.038
Reaction Mechanisms (3)	10333	1.392	0.716
Allotropes and Diatomic Molecules (1)	148	1.383	0.528
Identify Species Oxidized and Reduced (2)	23283	1.369	0.744
Redox–Acidic Solution–Provide ½ reaction (1)	36	1.274	1.289

*Represents a question related to the key topic Particulate Nature of Matter

**Represents a question related to the key topic Lewis Acids and Bases

The most difficult question from each difficult topic in OWL is shown below. For topics where there is more than one question, the most difficult question will be shown.

For the OWL questions, **boldfaced** statements, words, number of chemical equations are parameterized indicating that the statement, word, number, or chemical equation may change from one attempt to another.

Topic – Thermochemical Equations:

When **NH_3(g)** reacts with **O_2(g)** to form **NO(g)** and **H_2O(g)**, **226** kJ of energy are **evolved** for each mole of **NH_3(g)** that reacts.

Write a balanced thermochemical equation for the reaction with an energy term in kJ as part of the equation.

Use the SMALLEST INTEGER coefficients possible and put the energy term in an appropriate box. If a box is not needed, leave it blank.

________ + ________ + ________ → ________ + ________ + ________

In each space, the student enters a substance or an energy term.

Topic – Lewis Acids and Bases:

Classify each of the following substances:

_______ **BCl_3**	**1)**	Lewis Acid
_______ **H^+**	**2)**	Lewis Base
_______ **CH_4**	**3)**	Can act as either a Lewis Acid or Lewis Base
_______ **O_2**	**4)**	Neither a Lewis Acid or Lewis Base
_______ **NH_3**		

Topic – Redox Balancing – Basic Solution:

When the following skeletal equation is balanced under basic conditions, what are the coefficients of the species shown?

_____ NO_2^- + _____ I_2 → _____ I^- + _____ NO_3^-

Water appears in the balanced equation as a _________(reactant, product, neither) with a coefficient of________. (Enter 0 for neither.)

Which **species** is the **reducing** agent? ______________

Topic: Redox Balancing – Acidic Solution

When the following equation is balanced properly under acidic conditions, what are the coefficients of the species shown?

_____ClO_3^- + _____Mn^{2+} → _____ClO_2 + _____MnO_2

Water appears in the balanced equation as a_______ (reactant, product, neither) with a coefficient of________. (Enter 0 for neither.)

Which **element** is **oxidized**?___________

Topic – Selective Precipitation:

In the laboratory you are given the task of separating **Pb^{2+}** and **Co^{2+}** ions in aqueous solution.

For each reagent listed below indicate if it can be used to separate the ions. Type "Y" for yes or "N" for no. If the reagent CAN be used to separate the ions, give the formula of the precipitate. If it cannot, type "No"

Y or N	Reagent	Formula of Precipitate if YES
1.	**K_2S**	
2.	**KCl**	
3.	**KOH**	

Topic – Molarity of Ions in Solution:

In the laboratory you dissolve **17.8** g of **sodium sulfate** in a volumetric flask and add water to a total volume of **500** mL.

What is the molarity of the solution? _______ M.

What is the concentration of the **sodium** cation? ______ M.

What is the concentration of the **sulfate** anion? ________ M.

Topic – Crystal Field Splitting:

Estimate the crystal field stabilization energy for the octahedral ion **tris(ethylenediamine)cobalt(III)**, if the wavelength of maximum absorption for the ion is **467** nm. [Note: This is a high-field (low-spin) complex.] Answer: ________ kJ mol^{-1}

Topic: Calculate K_{sp} Given Solubility:

For each of the salts on the left, match the salts on the right that can be compared directly, using K_{sp} values, to estimate solubilities.

1._____silver sulfate **A $Ca_3(PO_4)_2$**

2._____iron(III) sulfide **B. $Zn(OH)_2$**

C. $Ni(CN)_2$

D. $PbCl_2$

Write the expression for K in terms of the solubility, s, for each salt, when dissolved in water.

silver sulfate **iron(III) sulfide**

K_{sp} = _______ K_{sp} = ________

Note: Multiply out any number and put it first in the K_{sp} expression.

Combine all exponents for s (the molar solubility).

Topic – Strong Electrolytes in Aqueous Solution:

The compound **potassium chromate** is a strong electrolyte. Write the reaction when **potassium chromate** is put into water:

_________ ___ → ________ ___ + _________ ___

There are three boxes, each with two spaces. In the first space, students write the chemical formula of the species. The second space contains a drop-down menu where students choose the physical states (s, l, aq, g).

Topic – Solubility of an Insoluble Base in Acid:

Write a balanced net ionic equation to show why the solubility of **$Fe(OH)_3$**(s) increases in the presence of a strong acid and calculate the equilibrium constant for the reaction of this sparingly soluble salt with acid.

Use the pull-down boxes to specify states such as (aq) or (s).

__________ ___ + _________ ___ → ________ ___ + _________ ___

There are four boxes, each with two spaces. In the first space, students write the chemical formula of the species. The second space contains a drop-down menu where students choose the physical states (s, l, aq, g).

K = __________________

Topic – Preparation of Buffer Solutions:

Design a buffer that has a pH of **3.78** using one of the weak acid/conjugate base systems shown below.

Weak Acid	**Conjugate Base**	K_a	pK_a
$HC_2O_4^-$	$C_2O_4^{2-}$	6.4 x 10^{-5}	4.19
$H_2PO_4^-$	HPO_4^{2-}	6.2 x 10^{-8}	7.21
HCO_3^-	CO_3^{2-}	4.8 x 10^{-11}	10.32

How many grams of the **potassium** salt of the weak acid must be combined with how many grams of the **potassium** salt of its conjugate base, to produce **1.00** L of a buffer that is **1.00** M in the weak base?

grams **potassium** salt of weak acid = __________

grams **potassium** salt of conjugate base = ___________

Topic – Reaction Mechanisms:

The following mechanism for the gas phase reaction of H_2 and ICl that is consistent with the observed rate law is:

step 1 *slow*: **$H_2(g) + ICl(g) \longrightarrow HCl(g) + HI(g)$**

step 2 *fast*: **$ICl(g) + HI(g) \longrightarrow HCl(g) + I_2(g)$**

(1) What is the equation for the overall reaction? Use the smallest integer coefficients possible. If a box is not needed, leave it blank.

______________ + ___________ → __________ + ___________

(2) Which species acts as a catalyst? Enter formula. If none, leave box blank _________

(3) Which species acts as a reaction intermediate? Enter formula. If none, leave box blank __________

(4) Complete the rate law for the overall reaction that is consistent with this mechanism. Use the form $k[A]^m[B]^n$…, where '1' is understood (so don't write it) for m, n, etc.

Rate = __________________________________

Topic: Allotropes and Diatomic Molecules

Consider the elements below:

1. Aluminum

2. Neon

3. Sulfur

4. Hydrogen

Which elements have allotropes? _______________

Which elements form diatomic molecules? _________________

Enter in order the number (1-4) for each element above. If none, type "none".

Topic – Identify Species Oxidized and Reduced:

$\mathbf{H_2SO_3 + Mn^{2+} + H_2O \rightarrow SO_4^{2-} + Mn + 4\ H^+}$

In the above redox reaction, use oxidation numbers to identify the element oxidized, the element reduced, the oxidizing agent and the reducing agent.

name of the element oxidized _________ name of the element reduced __________

formula of the oxidizing agent ________ formula of the reducing agent _________

Topic – Redox – Acidic Solution – Provide ½ reaction:

The following skeletal oxidation-reduction reaction occurs under acidic conditions. Write the balanced **REDUCTION half reaction**.

$\mathbf{HClO} + \mathbf{H_2SO_3} \rightarrow \mathbf{Cl_2} + \mathbf{SO_4^{2-}}$

_________________ → __________________

reactants products

For comparison, the UGA questions most closely related to the OWL question are shown below.

Topic – The Particulate Nature of Matter (UGA) – Corresponds to the OWL topic “Molarity of Ions in Solution” (UGA parameters a = 1.325, b = 1.212)

How many mL of *x* M $Sr(OH)_2$ solution are required to make *y* mL of a *z* M $Sr(OH)_2$ solution? What is the molar concentration of the Sr^{2+} ions in the *z* M $Sr(OH)_2$ solution?

What is the molar concentration of the OH^{-} ions in the *z* M $Sr(OH)_2$ solution?

In this question, *x*, *y*, and *z* are parameters chosen at random by the JExam homework system.

Topic – The Particulate Nature of Matter (UGA) – Corresponds to the OWL topic “Strong Electrolytes in Aqueous Solution”

How many potassium, oxygen, carbon, and carbonate ions are present in one formula unit of this compound K_2CO_3? (UGA parameters a = 1.074, b = 1.104)

The questions are not exactly the same, but their topics are very similar. This is due to the fact that JExam cannot easily accept writing a reaction involving formulas and states.

Topic – Lewis Acids and Bases

There are no equivalent UGA questions that correspond to the OWL question. The UGA questions give the students a chemical reaction and ask the students to identify the Lewis acid and Lewis base. The OWL question gives you a list of five species and asks you to classify the compounds.

The OWL question related to the topic “Molarity of Ions in Solution” was more difficult than the corresponding UGA question. The increased difficulty may be due to an extra step in the OWL problem that asks the student to calculate the molarity of the solution. (The UGA question already gives the molarity). Unfortunately, we do not have access to the students' responses from OWL, so it cannot be determined if students incorrectly answer the last two parts of the question because they answer the first part of the question incorrectly, or they answered the first part of the question correctly then missed the question portion more directly associated with understanding dissociation.

The OWL question related to the topic "Strong Electrolytes in Aqueous Solution" is also more difficult than the corresponding UGA question. The difference with the OWL question is that the students input the products and reactants for the dissociation of the strong electrolyte whereas the UGA question asks for the number of each ion. However, to successfully answer either the JExam or OWL question, students must understand that polyatomic ions do not dissociate into individual ions (i.e. K_2CrO_4 does not dissociate into 2 K^+ ions, one Cr^{6+} ion and 4 O^{2-} ions).

The OWL question related to the topic "Lewis Acids and Bases" cannot be directly compared to an equivalent UGA question. However, without a chemical reaction, students are not identifying Lewis acids and bases but species that are likely to be Lewis acids or Lewis bases.

Section 4.6 – Questions That Discriminate Poorly

A question discriminates poorly if it has a discrimination parameter (slope) of 0.500 or less on the first attempt. There are several reasons that questions discriminate poorly. These reasons are:

1. The question is not simple, but there is a probability that it may be inadvertently answered correctly. One such example occurs when a student leaves all the boxes blank (per OWL

instructions) and clicks "Check answer" when there is no net ionic equation and therefore gets the question correct.

2. The question asks for information that is easily memorized, such as "The symbol for carbon is ____."

3. There is an error in the question.

4. The wording of the question is confusing, or it is unclear what the question is asking.

5. It is not clear how the answer should be entered, or the correct format for entering the answer is unclear (i.e. the order of elements for an empirical formula, or H^+ vs. H_3O^+ in representing the hydrogen ion).

6. The parameters for a question are such that two sets of parameters for the same question produce questions that vary greatly in difficulty.

7. The question may have been updated in the middle of the semester (or between semesters) resulting in the question's difficulty changing.

Questions that poorly discriminate for reasons 1 or 2 (especially #2) often have a low ability (easy question on the first attempt). Questions that poorly discriminate for reasons 3, 4, or 5 often have a high ability (difficult question on the first attempt).

Questions that poorly discriminate for reasons 6 or 7 cannot be discerned just by looking at the difficulty parameter. One needs to look at the question, examine all possible parameters, and see what date and time the question was updated.

The one drawback to question mode is not all of the topic were represented. There were no questions on the topics of freezing point depression, boiling point elevation, integrated rate law calculations, and nuclear chemistry.

CHAPTER 5

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

ANALYSIS OF OWL MASTERY MODE DATA

Section 5.1 – Introduction

The previous chapter discussed questions that were answered in question mode. However, approximately 85% of the questions in the academic year 2010 – 2011 were presented in mastery mode (12). Therefore, the question mode analysis did not involve all of the questions. Some of the topics previously mentioned had no question mode responses.

Section 5.2 – Experimental:

Since approximately 2/3 of the questions in the question mode data were questions from introductory chemistry, modules containing only general chemistry questions were included. The name of the textbook was also included so each response can be classified as a general chemistry response or a non-general chemistry response. In the question mode analysis, the institution and type of course were considered to determine which responses were from general chemistry students.

The data from OWL involved approximately 45 million individual responses, comprising 94290 students and 4027 questions. All of the responses involved questions presented in mastery mode.

Mastery mode is where a set number of questions (usually three) are chosen from a mastery pool, and students need to get a given number (usually two) of those questions correct to master the module.

The data was initially sorted based on the number of attempts in the same manner as the question mode data was sorted. Approximately 95% of the data involved questions where students were allowed 11 or more attempts, so the analysis concentrated on that data set. Once the responses were sorted based on the number attempts, all non-general chemistry responses were removed. These not only included responses from textbooks that were not general chemistry textbooks, but included math questions, organic chemistry questions, and introduction to OWL questions. After removing those questions, 1731 questions and 66402 students remained.

One problem with mastery mode data is that the attempts were numbered based on modules. However, since a student may not see a question for the first time until their third attempt, the questions were renumbered based on the number of question attempts, rather than the number of module attempts. Table 5.1 and 5.2 illustrate how the renumbering process works for a module containing eight questions where the module was attempted nine times. Three questions are presented for each module attempt.

Table 5.1 – Presentation of Questions

Attempt Number	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8
1			X			X	X	
2	X	X					X	
3				X	X	X		
4	X			X				X
5			X				X	X
6	X	X			X			
7					X	X		X
8		X		X			X	
9				X			X	X

"X" – The question was presented

Table 5.2 – Data for Presentation Table 5.1

Question Number	Module Attempt	Question Attempt
1	2, 4, 6	1, 2, 3
2	2, 6, 8	1, 2, 3
3	1, 5	1, 2
4	3, 4, 8, 9	1, 2, 3, 4
5	3, 6, 7	1, 2, 3
6	1, 3, 7	1, 2, 3
7	1, 2, 5, 8, 9	1, 2, 3, 4, 5
8	4, 5, 7, 9	1, 2, 3, 4

For the 1731 general chemistry questions, the responses from attempts 1, 2, and 3 were analyzed using IRT. Total information curves (TIC's) were generated for each attempt in order to determine which attempt gave us the most information about the most students. In the analysis, even though there were questions where there was some probability of a low ability student guessing the answer to a question, the two parameter model was used. The reasons for using the two parameter model are the same reasons the two parameter model was used in question mode.

Section 5.3 – Data Formatting

The data was formatted in a tab delimited table with the following fields: Textbook, DatabaseID, CourseNumber, StudentID, ModuleNumber, IUNumber, QuestionNumber, RawScore, AnswerDate, AttemptNumber, Repeatquestion, Repeatmodule, Displayfeedback, Displayanswer, partialcredit, Maxscore, Timepermitted, Questionresubmission, Displayhints. The same steps were employed in formatting the data into a BILOG MG 3.0 compatible format (see section 4.3). One key difference is that the attempt numbers were renumbered so that each time a student was presented with a question, it counted as an attempt. In question mode, the only time the attempts were renumbered was for a student that took a course in a later semester where the same question was presented. In mastery mode, this renumbered all the questions because the attempt number was based on the number of times the module was attempted, not the individual question.

For step by step directions on how the data was formatted, see Appendix A.

Section 5.4 – Results – General Chemistry Questions – Overall Data Distribution

In determining the difficulty parameter of a question, one needs to determine which attempt to use. The attempt which gives us a Total Information Curve (TIC) with the most normal distribution centered on the mean is the attempt that will be used to determine the question parameters. Figure 5.1 shows the distribution of the students' abilities.

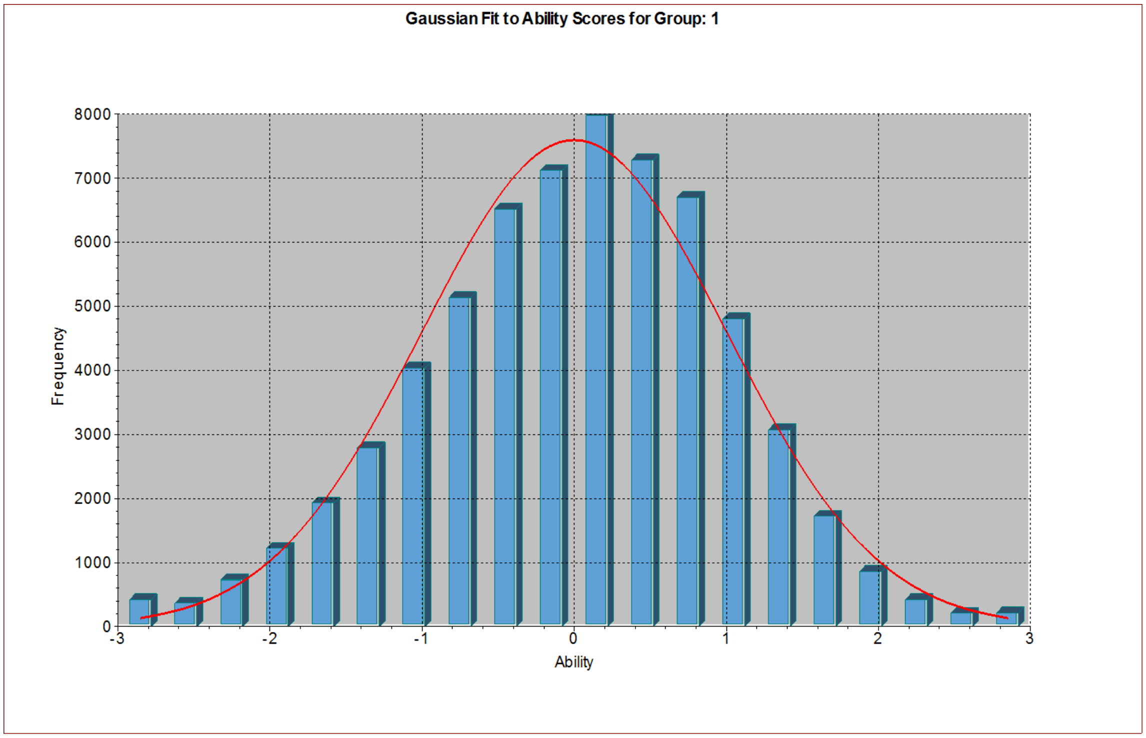


Figure 5.1 – Histogram of student abilities – General Chemistry Questions only

The total information curves (TIC) for the 1st, 2nd, and 3rd attempts were derived from the response patterns for all 1731 questions.

As previously mentioned, the location of the peak of the TIC curve can give us an idea of the difficulty of the overall assessment. One would postulate that the overall assessment decreases in difficulty with each attempt. This can be shown by the peak of the TIC curve shifting to the left, toward more negative abilities, as shown in figures 5.2 through 5.4.

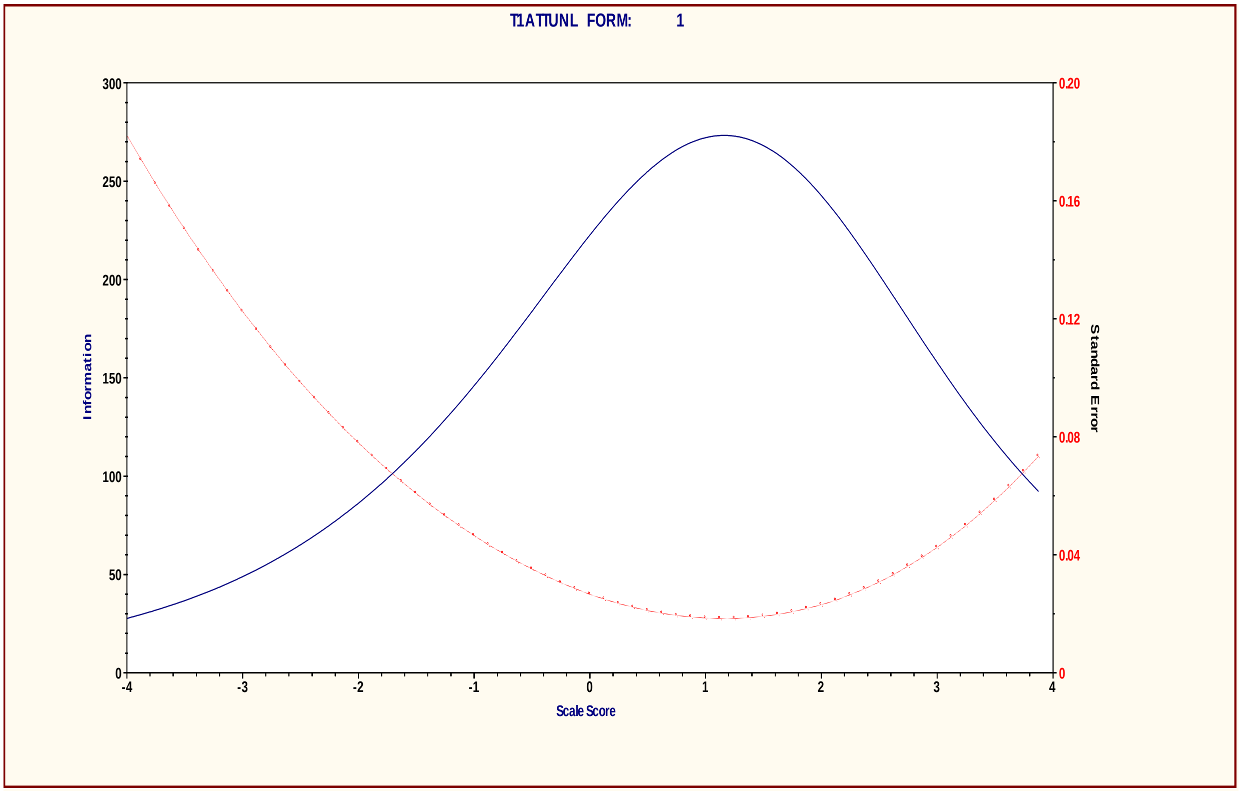


Figure 5.2 – TIC curve – Attempt #1 – General Chemistry Questions Only. The maximum of the TIC curve is at an ability of +1.125

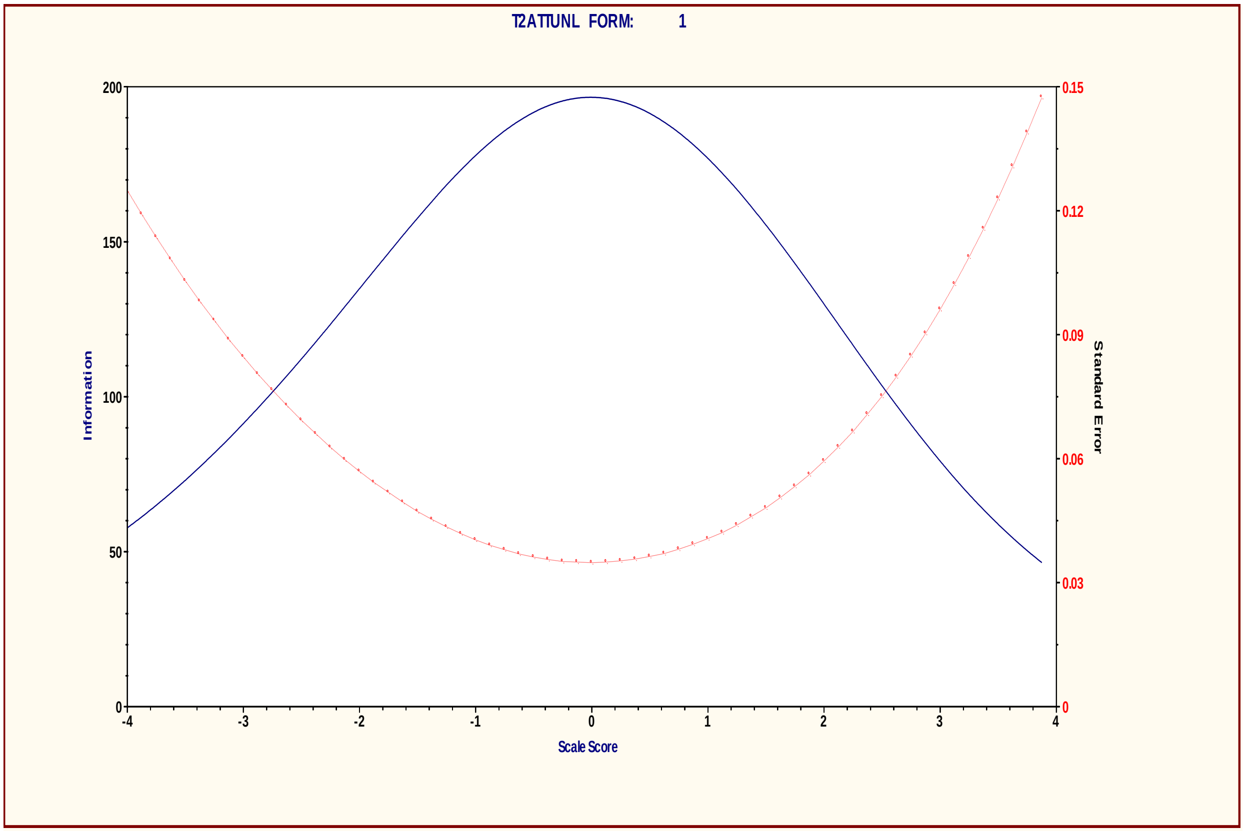


Figure 5.3 – TIC Curve – Attempt #2 – General Chemistry Questions Only. The maximum of the TIC curve is at an ability of 0.000

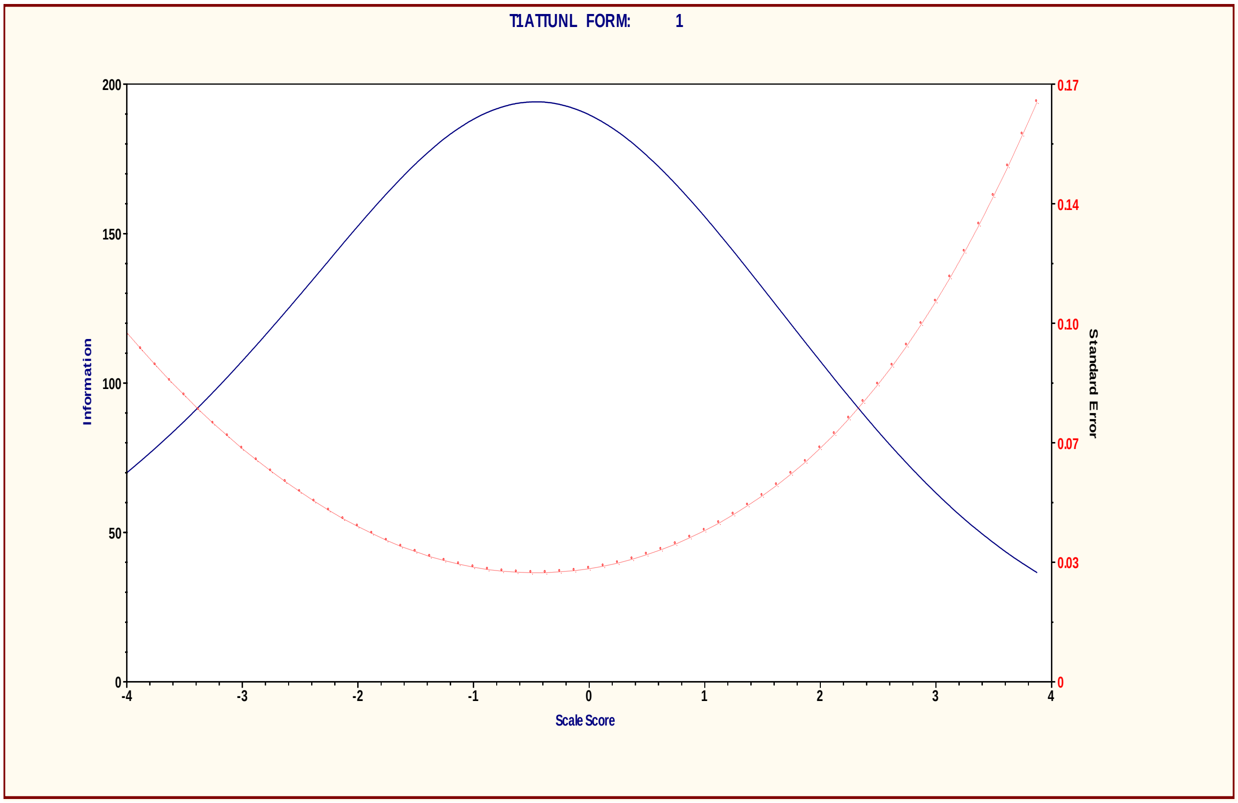


Figure 5.4 – TIC Curve – Attempt #3 – General Chemistry Questions Only. The maximum of the TIC curve is at an ability of –0.500

The TIC with the peak closest to zero is the TIC from the second attempt. Therefore, the second attempt will be used to compare question abilities to determine question and topic difficulties.

Section 5.5 – Difficult Topic Analysis – General Chemistry Questions (UGA comparison):

The definition of a difficult topic was presented in Chapter 4. That same definition will be used.

However, in the mastery mode analysis, the top 10% of the questions had a difficulty parameter of 1.32 or greater

Table 5.3 – Difficult Questions Based on the OWL Homework Database in Mastery Mode

Topic (Number of questions)	Tries	Average Difficulty	Average Discrimination
Lewis acids and bases (1)*	5508	3.866	0.538
Polyprotic acids – [A^{2-}] calculation (3)	10472	3.011	0.671
pH calculation – Titration Curves – Equivalence Point (2)	942	2.718	0.713
Enthalpy of Dissolution (1)***	6770	2.270	0.523
Selective Precipitation (1)	1660	2.239	1.048
Write ½ reaction (redox) – acidic soln. (8)	13167	2.230	1.061
pH calculation – Titration curves – Beyond the equivalence point (2)	891	2.147	1.277
pH calculation – Titration curves – Before the equivalence point (2)	928	2.112	0.328
Write ½ reaction (redox) – basic soln. (8)	11302	2.089	1.146
Solubility of an insoluble base in acid (3)	11076	2.007	0.924
Strong electrolytes in aqueous solution (1)*	25093	1.923	0.856
Nuclear Binding Energy Calculation (2)	7090	1.865	0.698
Bomb calorimetry (7)	27567	1.813	0.338
Redox Balancing – Basic Soln. (18)	29133	1.746	1.044
Redox Balancing – Acidic Solution (36)	42649	1.626	1.084
Nomenclature of Coordination Comp. (6)	2428	1.416	0.814

Reaction mechanisms (2)	12497	1.328	0.834

The number in parentheses after each topic denotes the number of questions that were analyzed for that particular topic

*Represents a question related to the key topic Particulate Nature of Matter

**Represents a question related to the key topic Lewis Acids and Bases

***Represents a question related to the key topic Solution Calorimetry

In the OWL questions, any statement, word, number, or chemical equation that is **boldfaced** is parameterized. That means the statement, word, number, or chemical equation may change from one attempt to another attempt. For those topics where there is more than one question, the most difficult question is included.

Topic – Lewis Acids and Bases (this is same question that was presented in question mode for this topic):

Classify each of the following substances:

_______ **BCl_3** **1)** Lewis Acid

_______ **H^+** **2)** Lewis Base

_______ **CH_4** **3)** Can act as either a Lewis Acid or Lewis Base

_______ **O_2** **4)** Neither a Lewis Acid or Lewis Base

_______ **NH_3**

Topic – Polyprotic Acids – Calculation of [A^{2-}]

Calculate the concentration of $\mathbf{C_6H_6O_6^{2-}}$ in an aqueous solution of **0.0382** M **ascorbic acid**, $\mathbf{H_2C_6H_6O_6}$ (aq).

[$\mathbf{C_6H_6O_6^{2-}}$] = _______ M.

Topic – pH Calculation – Titration Curves – At the Equivalence Point

What is the pH at the equivalence point in the titration of a **16.9** mL sample of a **0.372** M aqueous **nitrous acid** solution with a **0.384** M aqueous **barium hydroxide** solution?

pH = _______

Topic – Enthalpy of Dissolution

When a solid dissolves in water, heat may be evolved or absorbed. The ***heat of dissolution*** (dissolving) can be determined using a coffee cup calorimeter.

In the laboratory a general chemistry student finds that when **3.95** g of $\mathbf{CsClO_4}$(s) are dissolved in **112.00** g of water, the temperature of the solution drops from **22.15** to **19.97** °C.

The heat capacity of the calorimeter (sometimes referred to as the *calorimeter constant*) was determined in a separate experiment to be **1.75** J/^{o}C.

Based on the student's observation, calculate the enthalpy of dissolution of $\mathbf{CsClO_4}$(s) in kJ/mol.

Assume the specific heat of the solution is equal to the specific heat of water.

$\Delta H_{dissolution}$ = _______ kJ/mol

Topic – Selective Precipitation (this is the same question for this topic that was presented in question mode).

In the laboratory you are given the task of separating **Pb^{2+}** and **Co^{2+}** ions in aqueous solution.

For each reagent listed below indicate if it can be used to separate the ions. Type "Y" for yes or "N" for no. If the reagent CAN be used to separate the ions, give the formula of the precipitate. If it cannot, type "No"

Y or N	Reagent	Formula of Precipitate if YES
1.	**K_2S**	
2.	**KCl**	
3.	**KOH**	

Topic – Write Half Reaction (redox) – Acidic Solution

The following skeletal oxidation-reduction reaction occurs under acidic conditions. Write the balanced **OXIDATION half reaction**.

$$\mathbf{Mn^{2+} + CrO_4^{2-} \rightarrow MnO_4^{-} + Cr^{3+}}$$

__________________________ → ___________________________

Reactants Products

Topic – pH Calculation – Titration Curves – Beyond the Equivalence Point

When a **28.3** mL sample of a **0.454** M aqueous **hypochlorous acid** solution is titrated with a **0.494** M aqueous **potassium hydroxide** solution, what is the pH after **39.0** mL of **potassium hydroxide** have been added?

pH =________

pH Calculation – Titration Curves – Before the Equivalence Point

A **39.3** mL sample of a **0.595** M aqueous **acetic acid** solution is titrated with a **0.259** M aqueous **potassium hydroxide** solution. What is the pH after **51.5** mL of base have been added?

pH = ______

Topic: Write Half Reaction (redox) – Basic Solution

The following skeletal oxidation-reduction reaction occurs under basic conditions. Write the balanced **REDUCTION half reaction**.

$$\mathbf{Zn + Fe(OH)_3 \rightarrow Fe(OH)_2 + Zn(OH)_2}$$

____________________________ → _____________________________

Reactants Products

Topic – Solubility of an Insoluble Base in Acid

Write a balanced net ionic equation to show why the solubility of **$CuCO_3$** (s) increases in the presence of a strong acid and calculate the equilibrium constant for the reaction of this sparingly soluble salt with acid.

Use the pull-down boxes to specify states such as (aq) or (s).

__________ ___ + _________ ___ → ________ ___ + _________ ___

There are four boxes, each with two spaces. In the first space, students write the chemical formula of the species. The second space contains a drop-down menu where students choose the physical states (s, l, aq, g).

K = __________________

Topic – Strong Electrolytes in Aqueous Solution:

The compound **potassium chromate** is a strong electrolyte. Write the reaction when **potassium chromate** is put into water:

_________ ___ → ________ ___ + _________ ___

There are three boxes, each with two spaces. In the first space, students write the chemical formula of the species. The second space contains a drop-down menu where students choose the physical states (s, l, aq, g).

Topic – Nuclear Binding Energy Calculation

What is the binding energy in kJ/mol **nucleons** for the formation of **indium-113**?

_______ kJ/mol nucleons

$\mathbf{49}\ {}^{1}_{1}\mathrm{H} + \mathbf{64}\ {}^{1}_{0}\mathrm{n} \rightarrow {}^{113}_{49}\mathrm{In}$

The required masses (g/mol) are: $^{1}_{1}H = \mathbf{1.00783}$; $^{1}_{0}n = \mathbf{1.00867}$; $^{113}_{49}In = \mathbf{112.90430}$

Topic – Bomb Calorimetry

When 1.000 g CH_4 burns in a bomb calorimeter containing 8.060 kg of water, the temperature rises 1.520°C. Under these conditions, 885.3 kJ of heat is evolved per mole of methane burned. Calculate the calorimeter's heat capacity.

The specific heat of H_2O is 4.184 J/g°C.

a) 2.512 x 10^3 J/°C

b) 2.566 x 10^3 J/°C

c) 2.603 x 10^3 J/°C

d) 2.658 x 10^3 J/°C

e) 2.719 x 10^3 J/°C

Topic – Redox Balancing – Basic Solution

When the following skeletal equation is balanced under basic conditions, what are the coefficients of the species shown?

_____ $\mathbf{SiO_3^{2-}}$ +_____ $\mathbf{H_2PO_2^-}$ → _____ $\mathbf{HPO_3^{2-}}$ +_____ **Si**

Water appears in the balanced equation as a_________ (reactant, product, neither) with a coefficient of________. (Enter 0 for neither.)

Which **species** is the **reducing** agent? ____________________

Topic – Redox Balancing – Acidic Solution

When the following skeletal equation is balanced under acidic conditions, what are the coefficients of the species shown?

______ O_2 + ______ **Br^-** → ______ **H_2O_2** + ______ **BrO_3^-**

Water appears in the balanced equation as a ________ (reactant, product, neither) with a coefficient of ________. (Enter 0 for neither.)

Which **species** is the **oxidizing** agent? ______________________________

Topic – Nomenclature of Coordination Compounds

Assign a systematic name to the following coordination compound.

Compound	Name
$K_4[Mo(CN)_7H_2O]$	

Topic – Reaction Mechanisms

A gas phase reaction between nitrogen dioxide and fluorine is proposed to occur by the following mechanism:

step 1 *slow*: **$NO_2 + F_2 \longrightarrow NO_2F + F$**

step 2 *fast*: **$F + NO_2 \longrightarrow NO_2F$**

(1) What is the equation for the overall reaction? Use the smallest integer coefficients possible. If a box is not needed, leave it blank.

_____________ + __________ → _________ + __________

(2) Which species acts as a catalyst? Enter formula. If none, leave box blank ________

(3) Which species acts as a reaction intermediate? Enter formula. If none, leave box blank ______

(4) Complete the rate law for the overall reaction that is consistent with this mechanism. Use the form $k[A]^m[B]^n$…, where ‘1’ is understood (so don’t write it) for m, n, etc.

Rate = ________________________________

The three key topics where there was overlap between the OWL homework questions and the UGA JExam question are Lewis Acids and Bases, the Particulate Nature of Matter, and Solution Calorimetry. The first two topics will not be discussed since they were discussed in Chapter 4. Just like in Chapter 4, in order to compare the OWL question with the corresponding UGA question, the UGA question is shown below:

Topic – Solution Calorimetry (UGA) – Corresponds to the OWL topic “Heat of Dissolution” (UGA parameters: a = 1.191*, b = 1.656*)

*The question was presented in Spring 2008, Spring 2009, Spring 2010, and Spring 2011. The UGA parameters are the mean parameters for those four semesters

A coffee-cup calorimeter having a heat capacity of 350 J/^{o}C is used to measure the heat evolved in this aqueous reaction.

$PbBr_{2(aq)} + 2\ NaCl_{(aq)} \rightarrow PbCl_{2(s)} + 2\ NaBr_{(aq)}$

300.0 mL of 0.300 *M* $PbBr_2$ are mixed with 700.0 mL of 0.300 *M* NaCl, both solutions are initially at 20.00^{o}C. After thorough mixing the temperature of the mixture is 19.60^{o} C.

Assume that the solutions have a density of 1.00 g/mL and a specific heat of 4.18 J/g^{o}C. Find the amount of heat, in kJ, liberated or absorbed in this sample.

Be sure to enter a sign with BOTH your q and ΔH answers!

q = __________ kJ

How many moles of reaction were consumed in the sample? __________ moles

What is the ΔH for this reaction? Remember to use + or - signs! ________ kJ / mol

It was initially thought that the UGA question was difficult due to the moles of reaction. However, classical analysis of the question shows that 61.07% correctly answer the number of moles of reaction and only 53.47% answer the amount of heat correctly. We hypothesize the reason for the difficulty may be due to the math involved in calculating the amount of heat given a calorimeter constant, especially if the reaction is endothermic. The OWL question does not involve moles of reaction where the student needs to find the limiting reactant. The other difference between the two questions is that the OWL question involves the heat of dissolution, whereas the UGA question involves the heat of reaction. It would be interesting to break up the OWL question for the topic "Heat of Dissolution" into two questions, one where there is a temperature increase and one where there is a temperature decrease, and see if there is a difference in the difficulty of the two questions.

Section 5.6 – Easy OWL topics

It is not only important to identify difficult topics, but also to identify easy topics, especially topics that are easy in OWL. These are topics that are not only easy, but topics where OWL is doing an excellent job at helping the students master the topic. Table 5.4 lists topics that are easy for students in the OWL homework system

Table 5.4 – Easy Topics in Mastery Mode OWL Homework

Topic (Number of questions)	Tries	Average Difficulty	Average Discrimination
Temperature Conversions (10)*	75011	–4.373	0.449
Number of each Atom in a Compound (3)	69252	–4.281	0.693
Interpret Vapor Pressure Curves (6)	68738	–3.615	0.559
Chemical vs. Physical Properties (2)	56906	–3.309	0.425
Electronegativity Trend (4)	69952	–2.935	0.511
Ionization Energy Trend (2)	25577	–2.708	0.531
Percent by Mass of Element in Compound (6)	57360	–2.608	0.594
Number of Protons, Neutrons, Electrons (10)*	122520	–2.584	0.684
Pressure Unit Conversions (5)*	51267	–2.296	0.749
Boyle’s Law (4)	42565	–2.215	0.806
Average Atomic Mass Calculation (4)	66284	–2.103	0.885
Energy Unit Conversions (2)*	19067	–2.062	0.732
Vapor Pressure and Heat of Vaporization (5)	52200	–2.026	0.492
Atomic Radii Trend (1)*	12463	–1.950	0.548
Molecular and Empirical Formulas (9)*	93696	–1.747	0.752
Electron Configuration “Boxes” (4)*	53023	–1.722	0.667
Nomenclature – ID Cation & Anion (3)	65857	–1.639	0.716
Molar Mass Calculation (5)	58998	–1.637	0.783
Formation of Ionic Compounds from Ions (6)	84030	–1.561	0.852

Write Electron Configurations (2)*	22782	–1.518	0.621
Drawing Lewis Structures (22)	229736	–1.492	0.672

*This topic was also easy for UGA students

There a lot of overlap between easy OWL topics and easy topics at UGA. Nomenclature was not among the easy topics at UGA. One possible reason is that students can Google the answer for the OWL homework, but cannot do that for the exam. We were surprised that some of the other trends were not identified as easy at UGA. The ionization energy trends questions in OWL were easier than UGA questions about the same topic because OWL tells the students "not to worry about exceptions to the trend" whereas the UGA students were tested on the Group IIA and Group VA exceptions.

In the next chapter, we will compare the question mode data and the mastery mode data. We will also compare the data from OWL homework with the data from the UGA JExam homework. In addition, we will also utilize the fact that JExam can not only tell us if the question was answered correctly or not, we can look at each part of a question to determine which of the parts of a multi-part question was the most difficult. This may give us insight into why the OWL questions are difficult.

CHAPTER 6

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

COMPARISON OF OWL MASTERY MODE DATA, OWL QUESTION MODE DATA, AND JEXAM HOMEWORK DATA

Section 6.1 – Introduction

Results from the IRT analysis of question mode data and mastery mode data were presented separately in the previous two chapters. In this chapter, those results will be compared with one another as well as to IRT results from the UGA JExam homework system. The average, median, maximum, and minimum discrimination (and difficulty) parameters of the two modes and the UGA JExam homework will be compared. Difficult topics in the two modes will also be compared to see which topics were identified as difficult for each of the modes, and which topics were identified as difficult in both OWL modes. Discrimination and difficulty parameters will be compared for topics with questions that appeared in both question mode and mastery mode databases.

Section 6.2 – Parameter Comparison

We have looked at the OWL homework TIC curves for each the first three attempts in question mode (chapter 4) and mastery mode (chapter 5). If JExam homework is going to be part of the comparison, we must also look at the TIC curve and the student ability histogram produced from the IRT analysis of the JExam homework. Since there are three attempts in JExam homework, we will initially look at the responses from the first attempt. Figure 6.1 shows the distribution of the students' abilities and figure 6.2 shows the TIC for the first JExam homework attempt.

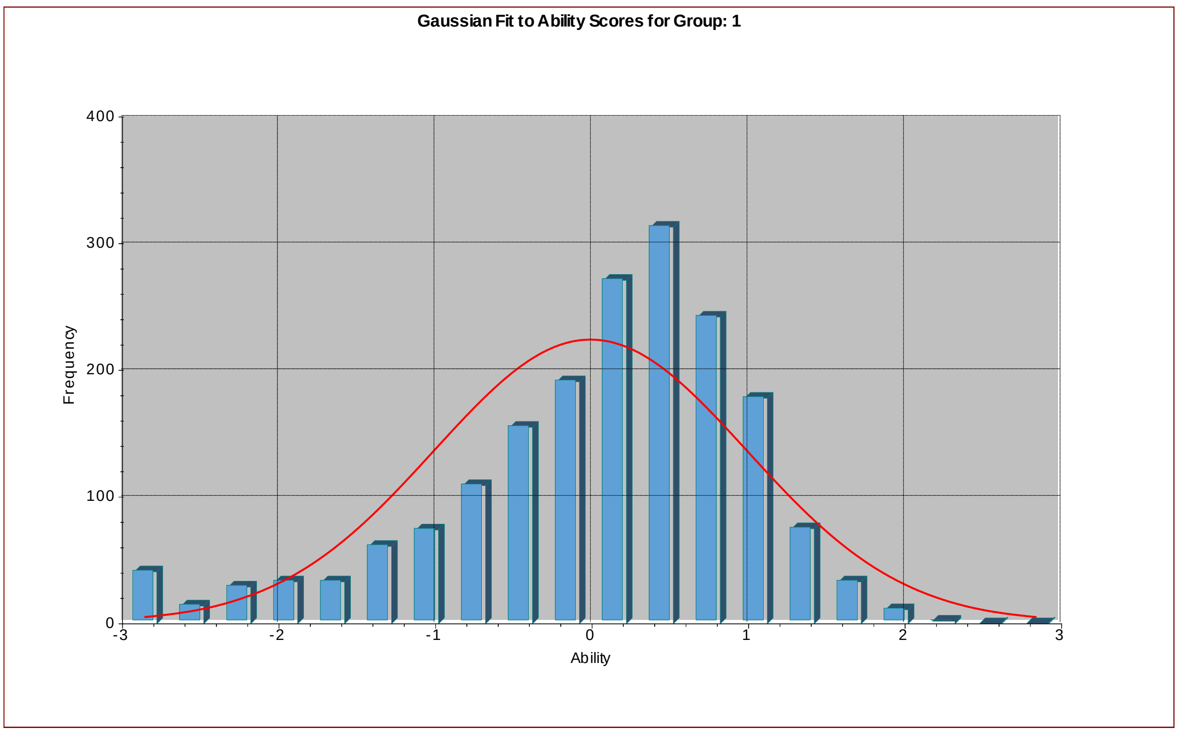


Figure 6.1 – Histogram of students' abilities for JExam homework – Attempt #1

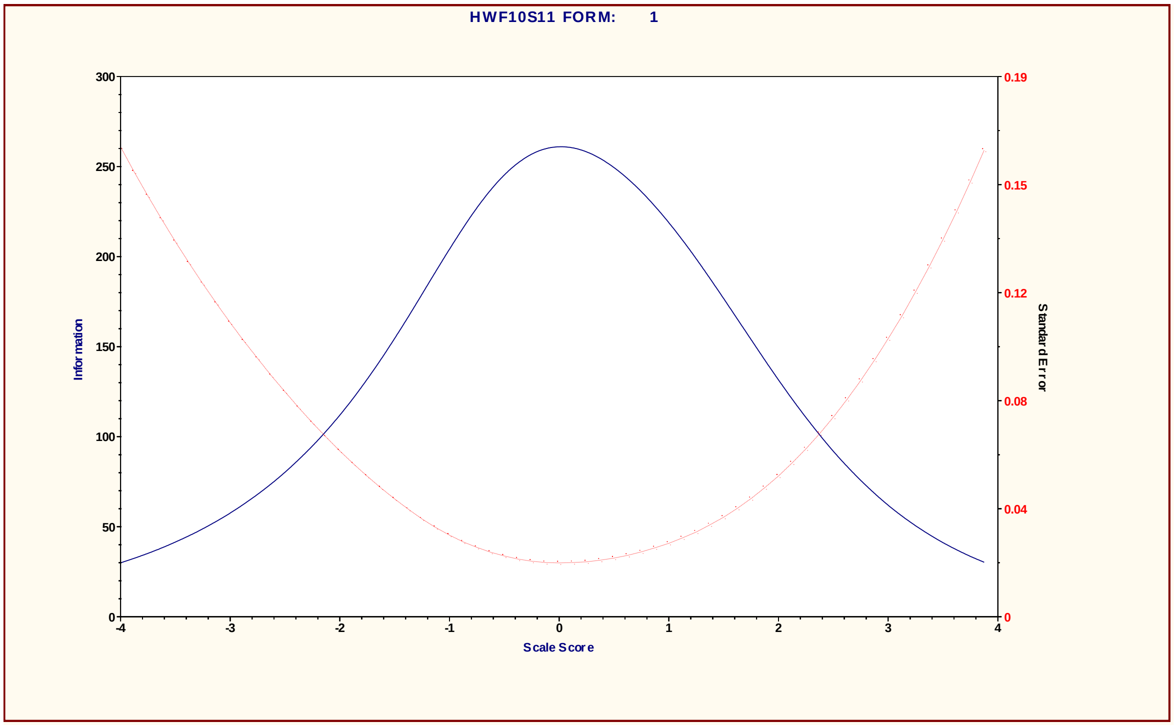


Figure 6.2 – TIC curve for JExam homework – Attempt #1 – The maximum of the TIC curve is at an ability of 0.000

The TIC curve of the 1st attempt on JExam not only is the most normally distributed. Therefore, the 1st attempt parameters will be used for the JExam homework analysis. The overall parameters for the two OWL homework modes and the JExam homework are in table 6.1 below.

Table 6.1 – Overall parameters for question mode, mastery mode, and JExam homework

Parameter	Mastery Mode Try 2 (N = 1711)	Question Mode Try 2 (N = 302)	JExam Homework Try 1 (N = 1143)
Average slope	0.720	0.911	1.017
Median slope	0.713	0.885	0.979
Maximum slope	2.018	1.866	2.415
Minimum slope	0.045	0.240	0.282
Average difficulty	–0.657	–0.599	–0.127
Median difficulty	–0.364	–0.431	–0.024
Maximum difficulty	4.533	3.506	5.306
Minimum difficulty	–28.464	–5.565	–6.643

The more extreme range of slopes and difficulties occurs in the OWL mastery mode data. It is uncertain if this is due to the greater number of questions, or the greater variety (i.e. a greater mixture of easy and difficult questions) of questions compared to question mode. The average and median discrimination parameters are greatest in the JExam homework, followed by question mode, then mastery mode. It is reasonable that the question mode data has higher abilities and discrimination than mastery mode since questions are randomly chosen in mastery mode from a question pool. It is possible that some students will see a given question in both their first and second attempt in the module; however, other students may not see that same question until their fourth or seventh attempt in the module. Those students who did not see a particular question until a later (third or greater) module attempt will find that question easier compared to the students who saw the same question in an earlier (first or second) module attempt since the student who saw the question in a later module attempt answered more questions about the topic, and will likely have greater knowledge of the topic. This may result in a student with an overall lower ability having a greater probability of answering a particular question due to the knowledge gained by previously answering related questions. However, the reason for the JExam questions having the highest discrimination parameter is unclear. One possible reason is the fixed number of attempts that are allowed. Students only get three attempts to do the JExam homework. Therefore, all students will legitimately attempt the question on the first attempt. With the OWL homework, some students (even good ones) may not really put effort into a question until the third attempt.

If one looks at the average difficulty, then the question mode questions are slightly more difficult. However, when one looks at the median difficulty, the mastery mode questions overall are more difficult. This is due to the average difficulty being skewed by several questions with a difficulty less than –10.0 in mastery mode. The questions in JExam are overall more difficult having an average difficulty of –0.127. This could be due to the "difficulty" of the JExam multiple answer questions that have an average difficulty of 1.380. In fact, 22 out of the 25 most difficult JExam homework questions are multiple answer questions.

Multiple answer questions are difficult due to the fact that some students put no effort into answering the question. Perhaps they do this because they get full credit for a question in JExam homework if they answer the question correctly on the first, second, or third attempt. The students check all but one of the boxes and submit the question. If the question is incorrect, the program tells the student which answers are correct. The student leaves the correct answers checked, unchecks the incorrect answers then submits the question a second time. If the question is still incorrect, the one box that was not checked on the first attempt is checked. Unless a student makes a mistake, they are guaranteed to get the question correct. This skews the JExam multiple answer question parameters and renders them unreliable.

Section 6.3 – Difficult Topic Comparison (BOTH modes)

We will first look at topics that were difficult in BOTH question mode and mastery mode. The parameters for the JExam homework are also presented if there are JExam homework questions in that topic.

Table 6.2 – Summary of Topics Difficult in BOTH modes

Topic	Total Mastery Mode Tries	Avg. Mastery Mode Parameters	Total Question Mode Tries	Avg. Question Mode Parameters	Total JExam Homework Tries	Avg. JExam Homework Parameters
Lewis acids and bases	5508	a = 0.538 b = 3.866	5634	a = 0.484 b = 3.114	5625	a = 0.777 b = –0.353
Separation of Ions (precipitation)	1660	a = 1.048 b = 2.239	37	a = 0.988 b = 2.531	1306	a = 1.368 b = 1.386*
Write ½ reaction (redox) – acidic soln.	13167	a = 1.061 b = 2.230	36	a = 1.289 b = 1.274	N/A	No equivalent

Solubility of an insol. base in acid	11076	a = 0.924 b = 2.007	36	a = 0.744 b = 1.791	N/A	No equivalent
Strong electrolytes in aqueous solution	25093	a = 0.856 b = 1.923	98	a = 0.614 b = 1.825	1666	a = 0.829 b = 1.373 #
Redox Balancing – Basic Soln.	29133	a = 1.044 b = 1.746	30	a = 0.890 b = 2.740	3342	a = 1.390 b = 0.226
Redox Balancing – Acidic Solution	42649	a = 1.084 b = 1.626	34	a = 0.761 b = 2.569	3342	a = 1.113 b = 0.172
Reaction mechanisms	12497	a = 0.834 b = 1.328	10333	a = 0.716 b = 1.392	2904	a = 1.338 b = 0.160

*The equivalent question is a multiple answer question

#The question is very particular about how the answers are entered

The topics “Lewis Acids and Bases” and “Reaction Mechanisms” are the only two topics having sufficient statistics to claim that those topics are difficult in both modes. Even though there are a sufficient number of tries in mastery mode for redox balancing and the other topics, there are not enough students attempting those questions in question mode to conclude that the topics are also difficult in question mode. However, Lewis Acids and Bases & Reaction Mechanisms are not difficult for UGA students in JExam homework. Possible reasons for Lewis Acids and Bases not being difficult in UGA JExam homework include 1) the JExam questions ask students to identify the Lewis acid and base in a chemical reaction, unlike the OWL question which asks students to identify potential Lewis acids and bases and 2) JExam Lewis acid-base questions are not parameterized, so students who have previously answered the question may tell other students the correct answer. The reaction mechanism questions are also not parameterized, so again there is the potential for "contamination" from students who previously answered the question, making the question easier. Also, there are not as many separate parts of each question to answer in the JExam homework questions about reaction mechanisms when compared to the OWL homework questions. Bill Vining (OWL author) presented data at the 2013 New Orleans ACS meeting with respect to two kinetics questions involving a reaction mechanism. The data is summarized in Table 6.3 (14).

Table 6.3 – Summary of two reaction mechanism questions present

Question	Information Asked	How Asked	Avg. # attempts	Time amount (min.)
1	Molecularity Role of each substance	Drop down box Drop down box	2.9	3.4
2	Molecularity Role of each substance Equation for net reaction	Drop down box Formula entry Formula entry	3.5	6.0

Notice that in the second question where more information was asked, the students took on average 0.6 more attempts to answer the question when the equation for the net reaction was asked. One could also claim that the different method of entry may have contributed to the difficulty which will be addressed later in this chapter with respect to reaction mechanisms.

Section 6.4 – Topics Only Difficult in Question or Mastery Mode

We will now look at topics that are either difficult in question mode or mastery mode. We will only look at topics that had questions presented in BOTH modes. The parameters for the JExam homework are also presented if there were JExam homework questions in that topic.

Table 6.4 – Topics That Are Difficult in One of the Two Modes

Topic	Total Mastery Mode Tries	Avg. Mastery Mode Parameters	Total Question Mode Tries	Avg. Question Mode Parameters	Total JExam Homework Tries	Avg. JExam Homework Parameters
pH calculation – Titration curves (before Eq. pt.)*	928	a = 0.328 b = 2.112	12827	a = 1.064 b = 0.794	N/A	No equivalent
pH calculation – Titration Curves (at Eq. Pt.)*	942	a = 0.713 b = 2.718	11862	a = 0.987 b = 0.856	N/A	No equivalent
pH calculation – Titration curves (past Eq. pt.)*	891	a = 1.277 b = 2.147	10996	a = 1.079 b = 0.180	N/A	No equivalent
Enthalpy of Dissolution**	6770	a = 0.523 b = 2.270	66	a = 0.811 b = –3.247		a = 1.192 b = 1.121[#]

Bomb calorimetry**	27567	a = 0.338 b = 1.813	11727	a = 0.989 b = 0.084	N/A	No equivalent

*Topic where there was complete overlap with question and mastery mode questions

**Topic where there was no overlap with question and mastery mode questions

#The JExam questions involved Solution Calorimetry rather than the enthalpy of dissolution

The three titration questions are difficult in mastery mode, but not in question mode. The three difficult titration curve questions are part of a set of five questions that guide students along the titration curve. Table 6.5 summarizes the titration curve questions.

Table 6.5 – Summary of Titration Curve Questions

Question number	Description of Question
1	Calculate pH before the strong substance was added
2	Calculate pH after a given volume of the strong substance was added
3	Calculate pH at the ½ equivalence point
4	Calculate pH at the equivalence point
5	Calculate pH beyond the equivalence point

When doing the titration curve problems in question mode, students can use the feedback to help determine the type of problem that is being presented on their first attempt. If a student realizes that question #5 is asking a student to calculate the pH beyond the equivalence point, they can use that knowledge on their second attempt knowing that question #5 is always going to ask for the pH calculation beyond the equivalence point in question mode. However, in mastery mode, questions are randomly chosen from a mastery pool. Therefore, the “beyond the equivalence point” question may be presented as question number #1 on their first module attempt, presented as question number #3 on their second module attempt, and not even presented on the third module attempt This randomness may explain the increase in the question difficulty.

When looking at the enthalpy of dissolution topic, there is a significant difference between the two questions. The question mode calculation does not involve a calorimeter constant whereas the mastery mode question involves the use of a calorimeter constant in the calculation. When a calorimeter constant is involved, the question becomes more complex involving more steps. The two equations below illustrate the increased difficulty with the calorimeter constant (C_{cal}).

$$q_{process} = -(c_{solution} \times m_{solution} \times \Delta T_{solution})$$

$$q_{process} = -[(c_{solution} \times m_{solution} \times \Delta T_{solution}) + (C_{cal} \times \Delta T_{solution})]$$

Where:

$q_{process}$ is the amount of heat

$c_{solution}$ is the specific heat of the solution

$m_{solution}$ is the mass of the solution

$\Delta T_{solution}$ is the temperature change of the solution (final temperature – initial temperature)

C_{cal} is the calorimeter constant

One must be very careful when doing the math, especially for endothermic reactions, where ΔT is negative (ΔT = final temperature – initial temperature). Unfortunately, the OWL question is highly parameterized so there is a possibility of a student getting an endothermic or an exothermic process. JExam tests only used endothermic solution calorimetry problems and the JExam homework only used exothermic solution calorimetry problems. Because of that, we cannot conclude whether exothermic or endothermic solution calorimetry problems are more difficult. The other issue with the JExam test questions was whether students had more trouble with calculating the amount of heat lost/gained, or if students had more trouble calculating the number of moles of reaction since the mole concept was also presented as a topic that was difficult for UGA students on JExam tests. For the 2010 – 2011 academic year, the results of the solution calorimetry JExam test questions were examined. 53.47% of the students correctly calculated the amount of heat gained/lost, and 61.07% of students calculated the moles of reaction correctly. Therefore, the more difficult part of the question appears to be the amount of heat calculation.

Section 6.5 – Difficult topics in UGA JExam Homework

In comparing difficult topics based on JExam homework multiple answer questions were ignored and not included in the analysis for the reasons outlined above. For the JExam homework, the threshold corresponding to the top 10% of the questions is an ability of 1.072. Table 6.6 lists the topics that are difficult based on the IRT analysis of JExam homework. Even though JExam homework was not the main emphasis of our analysis, comparing the two homework systems may yield more information, especially since we also have the capability (in JExam) to determine which parts of the question students would answer incorrectly. We did not have that information available for the OWL homework. The more difficult question for each topic is included, followed by possible reasons why the topic may be difficult.

Table 6.6 – UGA Homework Difficult topics

Topic (Number of questions)	Tries	Average Difficulty	Average Discrimination
Significant figure calculations involving mixed operations (2)	1670	2.094	0.499
Nomenclature – Binary Acidic Gases (4)	2672	1.472	0.992
Mass percent of solutions (6)	3220	1.449	1.062
Electron configuration – trans. metal cations (4)	2056	1.303	0.693
Work, Internal Energy and Gibb's Free Energy (6)	2921	1.228	1.159
Hybridization – clickable models (17)	7794	1.171	1.210
Formation of Basic Salts (7)	2250	1.150	1.227
Equilibrium concepts (4)	968	1.128	0.792
Solution calorimetry with moles of reaction (4)	1950	1.121	1.192

The number in parentheses after each topic denotes the number of questions that were analyzed

Topic – Significant figure calculations involving mixed operations

> Perform this arithmetic operation and enter the correct answer in the box below. Pay special attention to the rules of significant figures when entering your answer.
>
> A. (306.2 - 5.200) / 2.32298 x 10.3688 = ?

B. The answer contains ____ significant figures

The above question is the most difficult; however, it is poorly discriminating and should be classified as a bad question. Of the two questions on significant figures one of them discriminates poorly (a = 0.412). It is unclear why this question is difficult.

Topic – Nomenclature – Binary Acidic Gases

Each of the four questions on nomenclature of binary acidic gases has four parts. Those four parts are presented below along with the percentage of students that answered each part of the question correctly on their first attempt. This will help us understand why the question(s) are difficult.

Answer the following questions about this inorganic chemical compound: HI(g)

(other question variations had HF(g), HBr(g), and H_2S(g))

A. This is a(n) (ionic/covalent) compound. (77.23%)

B. This is a(n) example of a ______________ (choices were: ionic compound containing a metal that exhibits only one oxidation number, ionic compound containing a metal that exhibits more than one oxidation number, pseudobinary ionic compound, binary molecular compound, binary acid, ternary acid). (45.80%)

C. How many ions are present in one formula unit of the compound? (47.94%)

D. Using the IUPAC nomenclature system (do not use any older methods) the correct name of the compound is ____________ (68.83%)

The most difficult parts of the question involved classifying the compound and determining the number of ions per formula unit. Identifying the number of ions per formula unit has shown to be a difficult topic for UGA students on tests, so it is not surprising that students also have trouble with this topic on JExam homework.

The easiest part of the question is identifying whether the question is ionic or covalent. This is quite easy due to the 50% chance of a student randomly guessing the answer. However, the next most difficult part is the nomenclature itself. The nomenclature of compounds is easy in JExam homework since students can Google (or look up) the formula of the compound (and in fact they do) (15).

Topic – Mass Percent of Solutions

Different variations of the questions used different compounds and varied the mass percent of the solution. In the question, \`a\` and \`b\` are variable numbers that are different for each student, even if the compound is the same.

A. The density of 26.0% aqueous barium chloride solution is 1.279 g/mL. What mass of barium chloride is required to make \`a\` mL of this solution? (63.03%)

B. A reaction requires \`b\` g of barium chloride. What volume of the above solution do you need if you want to have a 30.0% excess of barium chloride? (24.10%)

Clearly the second part of the question is more difficult than the first. We speculate that this is because the students do not know how to use the "30.0% excess" in the problem. Unfortunately, there is no related JExam homework problem that asks the same question but does not involve the statement "30.0% excess".

Topic – Electron Configuration of Transition Metal Cations

Choose the correct ground-state electron configuration for Fe^{2+} (the three other questions asked for the electronic configuration of V^{2+}, Mn^{2+}, and Ni^{2+}). The order of the answers in JExam is shuffled.

A. [Ar] $3d^6$

B. [Ar] $4s^2 3d^3$

C. [Ar] $4s^2 3d^4$

D. [Ar] $3d^5$

E. [Ar] $4s^2 3d^8$

Most likely, the reason students have trouble with this topic is that students do not realize when transition metal cations form ions the “s” electrons are the first to leave rather than the “d” electrons.

Topic – Work, Internal Energy and Gibb’s Free Energy

There are two types of questions. The first type only involves work and internal energy, the second type involves Gibb’s free energy in addition to work and internal energy.

(Type 1) Given this information about the generic chemical reaction (\`a\` and \`b\` are whole number variables, \`c\` is a non-integer variable):

\`a\` $A_{(g)} + B_{(g)} \rightarrow$ \`b\` $C_{(g)} + D_{(\ell)}$ ΔH^0_{rxn} =\`c\` kJ/mol

What are the values of w, q, and ΔE^0_{rxn} for this chemical reaction at constant temperature and pressure (assume the constant temperature is 298 K).

A. w = ____________ kJ (67.57%)

B. q = _____________ kJ (89.64%)

C. ΔE^0_{rxn} = __________ kJ (64.44%)

Also choose the appropriate responses for each Choose Box (D and E).

Be sure to enter the appropriate sign with each answer.

D. In this reaction work is being done (by the surroundings on the system / by the system on the surroundings) (83.58%)

E. In this reaction heat is flowing from the (surroundings to the system / system to the surroundings) (87.76%)

Looking at parts A. and B. of this question type, it appears that most of the UGA students have grasped that ΔH^0_{rxn} and q are synonymous at the stated conditions. However they must perform a calculation to determine the value of w. The calculation requires recognizing physical states of the species involved, correctly performing a "products – reactants" calculation, and determining the appropriate sign. Students can slip up on any of these steps but we suspect that one of the more common mistakes is to ignore the physical states of the species involved. It is interesting to note that the percentage of students correctly determining the value of ΔE^0_{rxn} closely tracks the percentage of students correctly determining w. This is expected because the value of ΔE^0_{rxn} is determined by adding q and w.

(Type 2) At 25^{o}C and under constant pressure, ΔE^o for the following reaction is `a` kJ/mol rxn. ΔS^o_{rxn} = `b` J/mol.K. Answer the following questions. Variables `a` and `b` are randomly chosen whereas variable `c` is constrained to be a whole number.

2A(g) + 3 B(s) → C(g) + `c`D(g)

A. In this reaction, the work is (done by the surroundings on the system / done by the system on the surroundings) (86.27%)

B. The amount of work involved in the reaction is ___ J (Be sure to use + or - sign.) (72.53%)

C. Calculate the ΔH^o_{rxn} (kJ) for this reaction (65.47%)

D. Calculate the ΔG^o_{rxn} (kJ) for this reaction (63.44%)

E. This reaction is (spontaneous at this temperature / non spontaneous at the temperature / at equilibrium at this temperature) (93.75%)

F. This reaction is (spontaneous at all temperatures / not spontaneous at all temperatures spontaneous at high temperatures / spontaneous at low temperatures) (87.92%)

This question seems difficult due to the amount of cognitive load involved, especially considering the interrelationship of all the thermodynamic terms.

One difference between the two types of questions is the calculation for work is more difficult in the first type of question. There are two possible unproven reasons for this. The first type of question was on homework #5 and the second type of question was on homework #6. Students could look at the solution for the problems on homework #5 to get help on the similar questions on homework #6. The other possible reason is that the questions on homework #5 (the first type) ask for the amount of work in kilojoules (which requires an extra step) and homework questions on homework #6 (the second type) ask for the amount of work in Joules.

Topic – Hybridization – Clickable Models

Shown to the left is a molecular model. The common notation for molecular models is as follows:

White balls = H atoms

Grey balls = C atoms

Red balls = O atoms

Blue balls = N atoms

Yellow balls = S atoms

Light green balls = Cl atoms

Dark red balls = Br atoms

Beige balls = F atoms

Pink balls = Xe atoms

Orange balls = P atoms

Dark green balls = Be atoms

The molecular model in your question will have only a few of these elements present.

Click on all of the sp^3 hybridized atoms in this molecule.

(Move the molecular model in order to see it from different angles.)

The bond between the O and C atoms, the red/gray bond, is formed by the overlap of which <u>hybrid</u> orbitals on the two atoms?

C atom orbital _______ O atom orbital ________

It is unclear why this question is difficult. One possible reason could be the fact that hybridization is usually discussed in terms of the central atom. If one looks at the molecule acetone, the central carbon atom is sp^2 hybridized and the outside two carbon atoms are sp^3 hybridized. However, the oxygen atom is also sp^2 hybridized (two lone pairs and a double bond to carbon). Another possible reason could be that students count the double bond as two regions of electron density and therefore think that the central carbon atom is sp^3 hybridized. However, if that were the case, then the errors would also show up in other problems involving hybridization, which is not true. The questions involving geometry (both molecular and electronic), hybridization and polarity have an average difficulty of 0.091. These questions included several molecules containing double and triple bonds. Another possible reason for these questions being difficult is students may not always realize that the models need to be turned so that all of the atoms are visible (16).

Topic – Formation of Basic Salts

Predict the salt produced when 0.30 moles of iron(III) hydroxide react with 0.60 moles of hydrobromic acid.

Other variations of the problem use a divalent hydroxide + a monoprotic acid, but both are in equimolar amounts so the base is not completely neutralized.

One probable reason this question is difficult is the input is very specific. Parentheses are used in the writing of the formulas, even when there is only one hydroxy, such as Ba(OH)Cl. There are also no examples for students to consult, unlike acidic salts which are more common (i.e. Na_2HPO_3). Another possible reason is the difficulty of the chemistry involved. Students learned previously how to predict the products for neutralization reactions, and do not understand that you do need to pay attention to the mole ratio.

Topic – Equilibrium Concepts

The equilibrium concepts question is a three part question. In part C, \`a\` is a randomly generated number.

A. For the reaction $A_{(g)} \leftrightharpoons 2B_{(g)}$ the correct form of K_c is

1. $[B]^2 / [A]$

2. $[A] / [B]^2$

3. $[B] / [A]$

4. $[A] / [2B]^2$

5. $[2B]^2 / [A]$

B. The Kc for this reaction has a value of 0.00485 at 25°C. We can conclude that the reaction is ________ favored.

1. reactant

2. product

3. neither product nor reactant

C. If the initial nonequilibrium concentrations are [A] = [B] = `a` *M* , the reaction will proceed to the ____ side.

1. right (product)

2. left (reactant)

It is unclear why these questions are difficult. Both equilibria are A(g) ⇌ 2 B(g). It is possible that students forget to square the numerator when calculating Q_c. Another problem deals with the magnitude of the equilibrium constants. One of the equilibrium constants is 7.20. Even though an equilibrium constant greater than 1 is product favored, there are many cases where the concentrations of the reactants could be greater than the concentrations of products, especially if the moles of reactant gas are not equal to the moles of product gas.

Topic – Solution calorimetry with moles of reaction

The percentages after each question are the percent of students that got that part of the question correct.

A coffee cup calorimeter is used to measure the heat of neutralization for this acid-base reaction.

$HCl_{(aq)} + LiOH_{(aq)} \rightarrow LiCl_{(aq)} + H_2O_{(\ell)}$

`a` mL of `b` *M* HCl are reacted with `c` mL of `d` *M* LiOH. Both solutions are initially at 20.500°C. After the reaction has finished, the temperature of the mixture is `l`° C. The heat capacity of the calorimeter is `h` J/°C and the specific heat of the mixture is 4.184 J/g°C. The density of the solution is 1.02 g/mL. Use the information given above to answer the following questions (`a`, `b`, `c`, `d`, `f`, `h`, and `l` are all numerical variables).

(Number of moles must have *3* significant figures when it is used to calculate the heat of neutralization. Include the appropriate "+" or "-" sign for <u>heat of neutralization</u>.)

Total heat generated in experiment _______ J (60.32%)

Number of moles of reaction _______ mol (82.74%)

Heat of neutralization for this reaction = _______ kJ / mol (41.45%)

(Enter + or - sign and numerical answer.)

This topic appeared consistently, not only in UGA exam responses, but also in OWL homework, even though the moles of reaction may not have been involved. Previous analysis of JExam test questions (see Chapter 5) demonstrated that the amount of heat may be a more difficult calculation than the moles of reaction calculation.

For homework, the difficulty difference between the amount of heat calculation and the moles of reaction calculation is greater than that on the JExam tests. One possible reason for the larger difference is due to the fact that students are not asked to enter the sign for the amount of heat released. For an exothermic reaction (all homework problems are exothermic), the amount of heat technically should be negative. This may confuse students, especially since the question does not explicitly mention to not include the sign for the amount of heat.

We have compared the two modes in the OWL homework, in addition to JExam homework. In the next chapter, we will look at how small changes in a question make a big difference in the question's difficulty, how the method of entering the answer affects the difficulty of the question, and how the presentation of the answers affects the question difficulty.

CHAPTER 7

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

DIFFICULTY DIFFERENCES WITHIN TOPICS

Section 7.1 – Introduction

In the previous three chapters, difficult topics were examined. However, many topics are not difficult overall, but a specific question involving an easy topic may itself be quite difficult. One major reason for difficulty differences is cognitive load. One example of this involves calculating the molar mass of sodium carbonate vs. calculating the molar mass of Na_2CO_3. Both questions involve using the formula to calculate the molar mass but in the first question, we have the added step of converting the name to the formula. In other cases, it may not be a cognitive load issue that increases the difficulty of a question. One possible reason involves a small change or subtlety in a problem that a student may overlook when reading the question. An example of this is the difference between a Calorie (capital "C") and a calorie (lower-case "c"). Another possible reason is that there is an extra step involved that is usually not encountered in other similar problems, such as having to reduce the coefficients after balancing an oxidation-reduction reaction. How the data is entered could also account for an increase in difficulty, such as entering a number vs. entering a rate law. Differences in entering the data do not always change the question difficulty (14). Another reason is the presentation of the answers. A question having one correct answer out of five choices is easier than a question that may have more than one correct answer out of five possible choices.

Section 7.2 – Problematic subtopics

This section will examine small changes that make a large difference in difficulty. The results are summarized in Table 7.1. The individual questions are presented after the table. Any text that is **boldfaced** denotes parameters that will change the next time the question is presented.

Table 7.1 – Summary of problematic subtopics

Topic (No. of questions)	Description	Total Tries	Discrimination parameter	Difficulty parameter
Energy Unit Conversions (2)	No food Calories are involved	19067	0.731	–2.062
Energy Unit Conversions (1)	Food Calories are the given units	9636	0.368	2.313
LeChatelier's Principle (1)*	No solids are in the equilibrium	19321	1.277	0.340
LeChatelier's Principle	There is a solid in the equilibrium, and the solid is added (equil. effects)	18813	0.565	2.760
Q_c and K_c relationship (2)	No solid is present in the equilibrium	32482	1.197	–0.838
Q_c and K_c relationship (1)	A solid is present in the equilibrium (and the number of moles is given)	15182	0.450	1.256

Entropy Change Calculation (2)	Entropy of fusion and vaporization	8053	0.511	–0.805
Entropy Change Calculation (2)	Entropy of condensation and freezing	7842	0.528	1.021
Lewis structure drawing (2)	No double bonds involved	24430	0.652	– 2.427
Lewis structure drawing (1)	At least one double bond involved	11877	0.461	0.408
Nuclear chemistry (2)	Radioactive decay (only one answer for each question)	5873	0.730	–1.397
Nuclear chemistry (2)	Radioactive decay (two correct answers for each question)	5665	0.275	0.587
Calculate K_{sp} Given Ion Conc. (1)	Given cation concentration (cation coefficient = 1)	5079	0.718	0.142
Calculate K_{sp} Given Ion Conc. (1)	Given anion concentration (anion coefficient > 1)	4464	0.975	1.977
Relationship between K_c and K_p (1)	No solids in the equilibrium	10205	0.945	–0.117
Relationship	A solid is present in the equilibrium	10482	0.520	1.310

between K_c and K_p (1)				
Conceptual Common Ion Effect (1)	Solubility in Different Solutions, No Acids	9216	1.216	0.120
Conceptual Common Ion Effect (1)	Solubility in Acid vs. Water	8488	0.674	2.144

*There were other questions about LeChatelier's principle, but this question was the only question that was virtually the same as the LeChatelier's principle question involving solids in the equilibrium.

Topic: Energy Units – No Food Calories

Regular calorie question: Expressing amounts of energy in different energy units is necessary to solve many chemistry problems. For practice, complete the following table. The Joule (J) is the SI unit of energy. 1 calorie (cal) = 4.184 J

J	kJ	kcal
475		
	0.843	
		0.545

Topic: Energy Units – Food Calorie question

> **A list of the calorie content of foods** indicates that **a croissant** contains **259** Calories.
>
> Express this value in kJ and in J 1 cal = 4.184 J

It is unclear why the food Calorie question is more difficult. Two possible reasons for the difficulty are students are not reading the question clearly and are missing the capital "C" in Calorie, or maybe students have not been taught the difference between a calorie and a Calorie. The Whitten 9th edition general chemistry textbook only mentions the food calorie one time referring to the food calorie as a "large calorie". Note that the word "calorie" is not capitalized. With the change from calorie to Calorie, the question not only becomes more difficult, but also goes from an average discriminating question to a poorly discriminating question. It would be interesting to see if the lower discrimination is due to nutrition (food science) majors more likely to answer the question correctly due to their knowledge of a calorie vs. a Calorie or that both high ability and low ability students answer the question incorrectly.

Topic: LeChatelier's Principle – No Solids in the Equilibrium

> Consider the following system at equilibrium where K_c = **1.20E-2** and ΔH^o = **87.9** kJ/mol at **500** K.
>
> $\mathbf{PCl_5}(g) \rightleftharpoons \mathbf{PCl_3}(g) + \mathbf{Cl_2}(g)$
>
> The production of $\mathbf{PCl_3}(g)$ is favored by:
>
> <u>Indicate True **(T)** or False **(F)** for each of the following:</u>
>
> **1. increasing** the temperature.
>
> **2. decreasing** the pressure (by changing the volume).

3. decreasing the volume.

4. adding PCl_5.

5. removing Cl_2.

Topic: LeChatelier's Principle – Solids in the Equilibrium

Consider the following system at equilibrium where K_c = **1.80E-4** and Δ H° = **92.7** kJ/mol at **298** K.

NH_4HS(s) ⇌ **NH_3**(g) + **H_2S**(g)

The production of **NH_3**(g) is favored by:

Indicate True **(T)** or False **(F)** for each of the following:

1. decreasing the temperature.

2. decreasing the pressure (by changing the volume).

3. decreasing the volume.

4. adding NH_4HS.

5. removing H_2S.

The only difference between the questions is the second question involves a solid as part of the equilibrium. If a solid is present, the equilibrium does not shift due to the fact that the pure solid is not included in the equilibrium constant expression.

We cannot conclude with certainty that part #4 was the part that students were answering incorrectly; however, the other parts of the difficult question are similar to the parts of the easier question. Partial credit mode was turned off for this question, so we could not determine how many of the parts the students answered correctly. Even if partial credit mode were turned on, we were not able to get the individual responses for the various parts of the question. Further research would be needed to definitively determine if part #4 (the presence of a solid) is the reason this question is difficult.

Topic: Q_c and K_c Relationship – No Solids in the Equilibrium

Consider the following reaction where K_c = **10.5** at **350** K.

2 CH_2Cl_2(g) ⇌ **CH_4**(g) + **CCl_4**(g)

A reaction mixture was found to contain **1.31E-2** moles of **CH_2Cl_2**(g), **3.14E-2** moles of **CH_4**(g), and **4.52E-2** moles of **CCl_4**(g), in a 1.00 Liter container.

Is the reaction at equilibrium?

If not, what direction must it run in order to reach equilibrium?

The reaction quotient, **Q_c**, equals _________.

The reaction ______

A. must proceed in the forward direction to reach equilibrium.

B. must proceed in the reverse direction to reach equilibrium.

C. is at equilibrium.

Topic: Q_c and K_c Relationship – A Solid is in the Equilibrium

Consider the following reaction where K_c = **5.10E-6** at **548** K.

NH_4Cl(s) ⇌ **NH_3**(g) + **HCl**(g)

A reaction mixture was found to contain **5.50E-2** moles of **NH_4Cl**(s), **3.16E-3** moles of **NH_3**(g), and **2.26E-3** moles of **HCl**(g), in a 1.00 Liter container.

Is the reaction at equilibrium?

If not, what direction must it run in order to reach equilibrium?

The reaction quotient, $\mathbf{Q_c}$, equals ____________.

The reaction

A. must proceed in the forward direction to reach equilibrium.

B. must proceed in the reverse direction to reach equilibrium.

C. is at equilibrium.

Similar to the LeChatelier's principle question, having a solid present in the equilibrium makes the question more difficult. Again, we cannot say for certain if the students are using the moles of the solid in calculating Q_c. This is another topic where further research is needed to conclusively identify why the question is difficult.

Topic: Entropy Change Calculation – Entropy of Fusion / Vaporization

For **bismuth**, **Bi**, the heat of fusion at its normal melting point of **271**^{o}C is **11.0** kJ/mol.

The entropy change when **2.06** moles of solid **Bi** melts at **271**^{o}C, 1 atm is ________ JK^{-1}.

Topic: Entropy Change Calculation – Entropy of Freezing / Condensation

For **mercury**, **Hg**, the heat of vaporization at its normal boiling point of **357**^{o}C is **59.3** kJ/mol.

The entropy change when **1.70** moles of **Hg** vapor condenses at **357**^{o}C, 1 atm is ______ JK^{-1}.

The suspected reason that the entropy of freezing or condensation calculation is more difficult is due to students forgetting that the entropy of freezing or condensation is negative, since the system is becoming more ordered. For all questions, students are given either the enthalpy of fusion or the enthalpy of vaporization. Again, we cannot conclusively prove that students are entering the correct number but forgetting the negative sign since we do not have access to the students' individual responses.

Topic: Lewis Structure Drawing – No Double Bonds

> Draw a Lewis structure for **NHF_2** in which the central **N** atom obeys the **octet rule**, and answer the following questions based on your drawing.
>
> The number of **unshared pairs** (lone pairs) on the central **N** atom is: __________
>
> The central **N** atom forms ______ **single** bonds.
>
> The central **N** atom forms ______**double** bonds.

Topic: Lewis Structure Drawing – Double Bonds Present

> Draw a Lewis structure for **NO_3^-** in which the central **N** atom obeys the **octet rule**, and answer the following questions based on your drawing.
>
> The number of **unshared pairs** (lone pairs) on the central **N** atom is: _________
>
> The central **N** atom forms_______ **single** bonds.
>
> The central **N** atom forms _______ **double** bonds.

When drawing a Lewis structure, one of the last steps involves appropriately placing electrons if the central atom has less than an octet around it. There are two possible reasons why students have trouble with Lewis structures involving double bonds. One reason is that students forget to check the central atom for an octet (the outside atoms are given enough electrons to form an octet early in the Lewis structure drawing process). Another possible reason is that students will add extra electrons to the Lewis structure. Sometimes the extra electrons are added to the central atoms; in other cases the extra electrons are added to a single bond to make it a double bond (this causes one of the outside atoms to have more than an octet). We have observed the second reason (adding extra electrons) when tutoring students; however, we cannot say with certainty why the question becomes more difficult when double bonds are involved.

Topic: Radioactive Decay – One correct answer

When the nuclide **chromium-56** decays to **manganese-56**, what kind of decay does **chromium-56** undergo?

Choose all that are possible.

____ positron emission

____ beta decay

____ alpha decay

____ electron capture

Topic: Radioactive Decay – Two correct answers

Whether or not the process is observed in nature, which of the following could account for the transformation of **magnesium-20** to **sodium-20**?

Choose all that apply.

____ alpha decay

____ electron capture

____ positron emission

____ beta decay

The first question has only one possible answer – beta decay. The other related question also has one possible answer: alpha decay. The question about alpha decay is much easier than the beta decay question (b = –2.119 for alpha decay vs. b = –0.676 for beta decay). The question involving the topic "Radioactive Decay – Two correct answers" with two correct answers (electron capture and positron emission) is more difficult. Again, the reason for the increased difficulty cannot be determined with absolute certainty. Are the students only entering one of the two correct answers? If so, which of the two answers is being entered more frequently? This is where having the ability to examine the students answers would give further insight as to why the question is difficult. Another possible reason (again this cannot be proven) is the generic term "beta decay" may include negative (beta) and positive (positron) decay (16).

Topic: Calculate K_{sp} Given Concentration of One Ion – Cation Concentration Given, Cation Coefficient = 1

> A student measures the **Ca^{2+}** concentration in a saturated aqueous solution of **calcium hydroxide** to be **1.22E-2** M.
>
> Based on her data, the solubility product constant for **calcium hydroxide** is__________.

Topic: Calculate K_{sp} Given Concentration of One Ion – Anion Concentration Given, Anion Coefficient > 1

A student measures the **OH$^-$** concentration in a saturated aqueous solution of **nickel hydroxide** to be **8.03E-6** M.

Based on her data, the solubility product constant for **nickel hydroxide** is ____________.

There are two possible ways for students to answer this question incorrectly. Students may take the given anion concentration and multiply the given anion concentration by the coefficient (in this case 2) instead of dividing by 2 in calculating the cation concentration. Students may also not be raising the anion concentration to the appropriate power (in this case 2) when calculating the K_{sp}. Either of these reasons is plausible; however, more research is needed to determine which of these two errors are more common, or if there is a different reason students are having trouble with this problem.

Topic: Relationship between K_p and K_c – No solids present in the Equilibrium

The equilibrium constant, K_p, for the following reaction is **0.636** at **600** K.

Calculate K_c for this reaction at this temperature.

$\mathbf{COCl_2}(g) \rightleftharpoons \mathbf{CO}(g) + \mathbf{Cl_2}(g)$

K_c = ____________________________

Topic: Relationship between K_p and K_c – Solids present in the Equilibrium

The equilibrium constant, K_p, for the following reaction is **1.04E-2** at **548** K.

Calculate K_c for this reaction at this temperature.

$\mathbf{NH_4Cl}(s) \rightleftharpoons \mathbf{NH_3}(g) + \mathbf{HCl}(g)$

K_c = ___________________________

It is postulated that this question is more difficult due to the presence of the solids. In calculating Δn, solids and liquids are not included in the calculation. Again, we cannot positively conclude that students are getting the question incorrect due to including the solids and liquids in the Δn calculation.

Topic: Common Ion Effect – No solubility in acids

Compare the solubility of **calcium phosphate** in each of the following aqueous solutions:

_______	0.10 M $\mathbf{NaCH_3COO}$	**1)**	More soluble than in pure water.
_______	0.10 M $\mathbf{NH_4NO_3}$	**2)**	Similar solubility as in pure water.
_______	0.10 M $\mathbf{(NH_4)_3PO_4}$	**3)**	Less soluble than in pure water.
_______	0.10 M $\mathbf{Ca(NO_3)_2}$		

Topic: Common Ion Effect – Solubility in Acids

Each of the insoluble salts below are put into **0.10 M hydrobromic acid** solution. Do you expect their **solubility to be more, less, or about the same** as in a pure water solution?

_________	**magnesium hydroxide**	**1)**	More soluble than in pure water.
_________	**silver bromide**	**2)**	Similar solubility as in pure water.
_________	**lead chloride**	**3)**	Less soluble than in pure water.
__________	**calcium sulfite**		

The two questions are not exactly the same, but are similar enough to compare them. In the first question, none of the compounds is an acid. Therefore, the only thing that affects the solubility is the presence of one of the common ions. However, when you look at the solubility of a compound in acid, you need to see if the anion reacts with a weak acid and therefore increases the solubility of the salt. This question is related to the topic "Solubility of an Insoluble Base in Acid", which is a difficult topic. This helps to verify that students have difficulty with the concept of how certain compounds whose anions have conjugate acids that are weak are more soluble in acidic solution.

Section 7.3 – Effects of Answer Entering on Question Difficulty

We have examined how small changes can make a very large difference in the difficulty of a question. In this section, we will look at how the method of entering the answer affects the question difficulty.

One of the OWL homework difficult topics involved entering the rate law. Bill J. Vining presented some research that involved the entry of rate laws (14). The two questions presented involved the determination of the rate law and rate constant by the initial rates method. One question asked the student to only enter the order with respect to each reactant and calculate the rate constant. The other question asked the student to enter the actual rate law expressions and calculate the rate constant.

The question where less information was entered only took 5.5 minutes vs. 11.2 minutes for the question where more information was entered; however, both questions were completed in an average of 4.5 attempts. At least on a per attempt basis, there is no difference in entering the entire kinetic rate law or just the orders for each of the reactants. This implies that the reaction mechanism questions are difficult because the topic is difficult and not due to the difficulty in entering the information correctly. The table below compares methods of entry for paired questions that have different methods of entering the information (such as entering a number vs. entering a formula)

Table 7.2 – Similar questions with different methods of entry

Topic	Method of Entry	Avg. Discrim.	Avg. Ability
Balancing acidic soln. half reaction	Enter reactants and products with coefficients (formulas)	1.061	2.230
Balancing acidic soln. half reaction	Enter coefficients only	0.802	0.102
Balancing basic soln. half reaction	Enter reactants and products with coefficients (formulas)	1.146	2.089
Balancing basic soln. half reaction	Enter coefficients only	0.703	–0.195
Hess's Law Involving Equilibria	Number entry	0.822	0.461
Hess's Law Involving Equilibria	Formula entry	0.891	1.536
Hydrogen Bonding – Pure Substances	Multiple Choice	0.651	–1.765

Hydrogen Bonding – Pure Substances	Multiple Answer	0.592	1.131
Hydrogen Bonding – With Water	Multiple Choice	0.045	–21.653
Hydrogen Bonding – With Water	Multiple Answer	0.327	1.676

Topic: Balancing Redox Half Reactions (both Acidic and Basic Solutions) – Number Entry

When the following half reaction is balanced under acidic conditions, what are the coefficients of the species shown?

_____ SO_4^{2-} + _____ H^+ ⟶ _____ H_2SO_3 + _____ H_2O

In the above half reaction, the oxidation state of **sulfur** changes from _____ to _____.

Topic: Balancing Redox Half Reactions (both Acidic and Basic Solutions) – Enter reactants and products

The following skeletal oxidation-reduction reaction occurs under basic conditions. Write the balanced **REDUCTION half reaction**.

$Zn + Fe(OH)_3 \rightarrow Fe(OH)_2 + Zn(OH)_2$

____________________________ → _____________________________

Reactants Products

These questions are very similar, but the entry method is different. The OWL system is sometimes very picky about entering reactions. Dr. Atwood had mentioned that some students at the University of Utah had trouble entering reactants and products for reactions (16).

When I attempted this problem, the system was forgiving of things like the order in which substances were entered and the spacing. One thing students need to be aware of is the use of superscripts and subscripts. When you want to go to subscript or superscript, you need to click the superscript button when you want to write superscript text. However, you also need to click the same superscript button to go from superscript text to regular text. This could be a possible reason for the difference in difficulty.

Topic: Hess's Law with Equilibria – Number Entry

At 298K, the equilibrium constant for the following reaction is **4.20E-7**:

$\mathbf{H_2CO_3}(aq) + H_2O \rightleftharpoons H_3O^+(aq) + \mathbf{HCO_3^-}(aq)$

The equilibrium constant for a second reaction is **4.80E-11**:

$\mathbf{HCO_3^-}(aq) + H_2O \rightleftharpoons H_3O^+(aq) + \mathbf{CO_3^{2-}}(aq)$

Use this information to determine the equilibrium constant for the reaction:

$\mathbf{H_2CO_3}(aq) + 2H_2O \rightleftharpoons 2H_3O^+(aq) + \mathbf{CO_3^{2-}}(aq)$

K = ______________________________

Topic: Hess's Law with Equilibria – Formula Entry

Consider the reaction:

$\mathbf{PCl_3(g) + Cl_2(g) \rightleftharpoons PCl_5(g)}$

Write the equilibrium constant for this reaction in terms of the equilibrium constants, $\mathbf{K_a}$ and $\mathbf{K_b}$, for reactions **a** and **b** below:

a.) **$P(s) + 3/2\ Cl_2(g) \rightleftharpoons PCl_3(g)$** .... $\mathbf{K_a}$

b.) **$P(s) + 5/2\ Cl_2(g) \rightleftharpoons PCl_5(g)$** .... $\mathbf{K_b}$

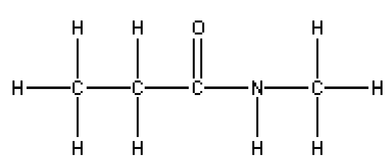
K = ____________________

The problem with the formula entry question is how precisely the formula needs to be entered. The correct answer for the above question (according to OWL) is $K = K_b / K_a$. When the answer $K = K_a^{-1} K_b$ was submitted (which is technically correct), it was marked as incorrect. The question was not set up to handle possible negative exponents or alternate answers. This problem may also cause misconceptions later because K_a and K_b are specific acid equilibria and base equilibria, respectively.

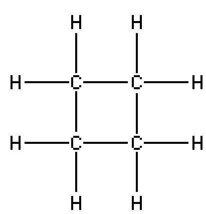
Topic: Hydrogen Bonding in Pure Substances – Multiple Choice

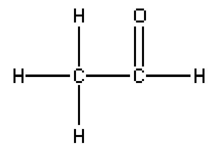
In which of the following pure substances would hydrogen bonding be expected?

N-methylpropanamide



cyclobutane



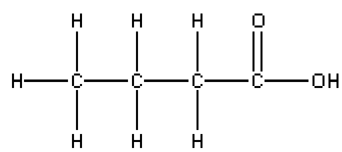
acetaldehyde 

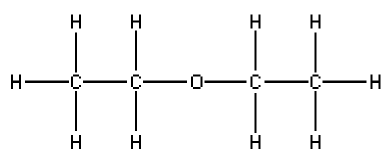
All of the Above

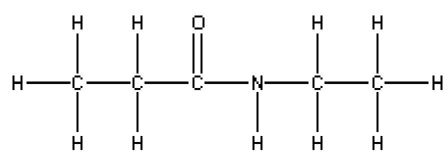
Topic: Hydrogen Bonding in Pure Substances – Multiple Answer

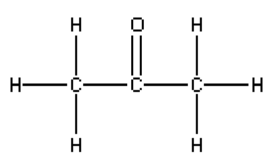
In which of the following pure substances would hydrogen bonding be expected?

Choose all that apply.

butanoic acid 

diethyl ether 

N-ethylpropanamide 

acetone 

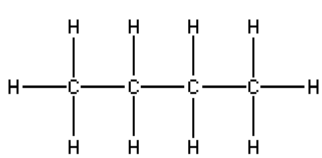
☐ None of the Above

In the multiple choice question, the three choices are a (cyclo)alkane, a compound containing an acceptor atom (O or N) that is not attached to a hydrogen (such as an ester, ether, ketone or aldehyde), and a compound containing both a donor atom and an acceptor atom (the correct choice will either be an alcohol, a non-substituted amide, or a carboxylic acid). This question seems to be a question where a student could eliminate two of the answers immediately, the (cyclo)alkane and the response "All of the Above", making this question easier. In the multiple answer question, there are always two correct answers out of the five possible answers. One of the correct answers is an alcohol, carboxylic acid, or an amine and the second correct answer is always an amide. The two distracters (in addition to "None of the Above") are an ether or an ester plus an aldehyde or a ketone. Both of the distracters have an acceptor atom, but no hydrogen that is directly attached to that atom. This may be a case where the difficulty is not due to the method of entering the data, but due to the nature of the distracters.

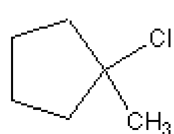
Topic: Substances that Hydrogen Bond with Water – Multiple Choice

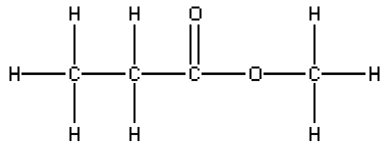
Which of the following would be expected to form hydrogen bonds with water?

○ **butane**



○ **1-chloro-1-methylcyclopentane**



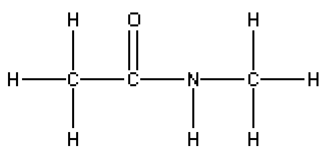
methyl propanoate 

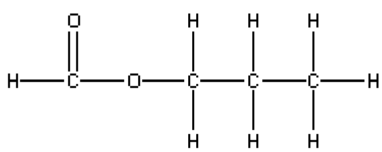
None of the Above

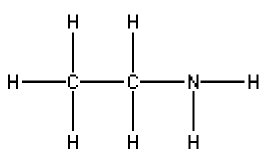
Topic: Substances that Hydrogen Bond with Water – Multiple Answer

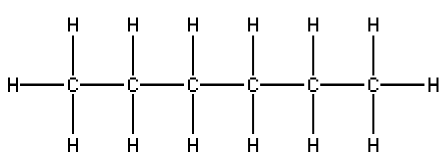
Which of the following would be expected to form hydrogen bonds with water?

Choose all that apply.

N-methylacetamide 

propyl formate 

ethylamine 

hexane 

None of the Above

The multiple choice question is extremely easy (b = –21.653). It may be due to the nature of the distracters, or the fact that only one of the answers contains an atom that is an acceptor atom. Another possible reason for the low difficulty is due to the fact that in all four of the questions, one answer must be selected (i.e. you cannot just enter "Check Answer" and look at the feedback on the first attempt). It is proposed that even a poor student will likely get the answer correct on their first or second attempt due to the fact that two of the distracters do not have an acceptor atom. This multiple answer question has three correct choices. The only choice that is not correct is an alkane, since almost all of the organic molecules that contain oxygen or nitrogen have an acceptor atom and therefore hydrogen bond with water.

We examined cases where small changes made a big difference in the difficulty of the question. We also looked at how the method of input affected the difficulty of the question. Some of the cases involved entering reactants and products vs. coefficients, but a few of the questions involved multiple choice questions vs. multiple answer questions. The multiple answer questions would be expected to be more difficult, but not by the magnitude that was observed. We expected the difference in the difficulty parameter would be only 1 or 2. However, the difference in the difficulty parameter between the multiple choice and multiple answer questions was greater than 2, and in one case, was approximately 22.

We also examined how small changes can make a big difference in the difficulty of a question. A majority of the cases involved equilibrium problems when a pure solid or pure liquid was involved. It is not clear if students have misconceptions about heterogeneous equilibrium, or they are not paying attention to the phases in the equilibrium reaction (i.e. not reading the question carefully).

CHAPTER 8

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

CONCLUSIONS AND FUTURE DIRECTION

Section 8.1 – Identification of Difficult Topics

We have identified several different topics that students find difficult based on IRT analysis of OWL question mode homework responses, OWL mastery mode homework responses, and JExam homework responses. Some topics were difficult in one of the homework sets but not in the others. The results are summarized in table 8.1.

Table 8.1 – Summary of Difficult Topics

Topic	Difficult in OWL Question Mode?	Difficult in OWL Mastery Mode?	Difficult in UGA JExam Homework?
Lewis Acids and Bases*	Yes	Yes	No
Balancing Redox Reactions (Acid & Base)	Yes	Yes	No
Selective Precipitation	Yes	Yes	Yes**
Reaction Mechanisms	Yes	Yes	No

Titration Curves – pH Calculations	No	Yes	Not presented
Bomb Calorimetry	No	Yes	Not presented
Coordination compound nomenclature	Not presented	Yes	Not presented
Nuclear Binding Energy Calculation	Not presented	Yes	Not presented
Mass Percent of Solution	No	No	Yes
Equilibrium Concepts	No	No	Yes
Enthalpy of Dissolution*	No	Yes	Not presented
Strong Electrolytes in Aqueous Solution*	Yes	Yes	Yes
Sig. Fig. Calculation – Mixed Operations	Not presented	No	Yes
Nomenclature – Binary Acid Gases	Not presented	No	Yes
Work, ΔE^{o}, and ΔG^{o}	No	No	Yes
Hybridization – Clickable Models	Not presented	Not presented***	Yes
Soln. Calorimetry with Moles of Reaction**	No	Yes	Yes
Polyprotic Acids – $[A^{2-}]$ Calculation	Not presented	Yes	Not presented
Molarity of Ions in Solution*	Yes	No	No
Crystal Field Theory	Yes	Not presented	Not presented
Transition Metal Electron Configuration	Not presented	No	Yes

Calculate K_{sp} Given Solubility	Yes	Yes	Not presented
Identify Oxidized and Reduced Species	Yes	No	No

*Topic Relates to a Key Topic at UGA

**The question was a multiple answer question

***The topic was covered in OWL, but the questions were not clickable model questions

The three topics that are difficult for UGA students on exams and difficult for students nationwide in OWL are: Lewis acids and bases, the particulate nature of matter (the OWL question was classified as "Strong Electrolytes in Aqueous Solution"), and solution calorimetry (the OWL question was classified as "Enthalpy of Dissolution"). The topic Lewis acids and bases was quite difficult with the question in both mastery and question mode having a difficulty parameter greater than 3. More emphasis on teaching Lewis acids and bases should help students better understand this topic. Students that have a better understanding of Lewis acids and bases should also do better in organic chemistry, since the concept of Lewis acids and bases occurs frequently in organic chemistry.

The only topic that was difficult in all three homework sets was the topic "Selective Precipitation". However, the JExam question for that topic was a multiple answer question. It is uncertain whether the difficulty was due to the topic difficulty, or students applying the algorithm that guarantees answering any multiple answer question correctly within three tries.

Other topics that were difficult nationally in OWL mastery mode include: calculation of $[A^{2-}]$ in diprotic acids, pH calculations for titration curves (before, at and beyond the equivalence point), balancing redox reactions (acidic and basic solutions) where H^+, OH^-, and H_2O have to be added, solubility of an insoluble base in acid, relationship of K_{sp} to solubility, calculating nuclear binding energy, bomb calorimetry, nomenclature of coordination compounds, and reaction mechanisms.

Other topics that were difficult nationally in OWL question mode (with at least 500 students) were: reaction mechanisms, crystal field theory, relationship of K_{sp} to solubility, identifying species oxidized and reduced in redox reactions, and preparation of buffer solutions. There were no questions about crystal field theory in mastery mode and no questions about nomenclature of coordination compound in question mode. In determining difficult topics, we only looked at questions answered by at least 500 students.

In addition to difficult topics, the method an answer is entered can affect the question difficulty. This is especially true when entering chemical formulas, or when a question does not accept an alternate, but technically correct answer ($K = K_2 / K_1$ is scored as correct whereas $K = K_1^{-1} K_2$ is scored as incorrect). In some cases the difficulty parameter changes by 2 ability units.

There are also some cases where a small change in a question can make a big difference. One needs to be aware of those difficult topics and emphasize those subtopics in class. Additionally, when writing exams with several versions, an instructor needs to be aware of possibly difficult subtopics to ensure that all versions of the exam have approximately the same difficulty.

This research identified what the difficult topics are but does not identify why these topics and subtopics are difficult. If class instructional methods are to be modified to help students understand these difficult topics, the reason for the difficulty needs to be fully understood.

Section 8.2 – Mastery Mode vs. Question Mode

Mastery mode questions overall are less discriminating than question mode questions since there is more randomization of the mastery mode questions. Therefore, students may have received different amounts of feedback by the time they get to the second attempt of a particular question.

The randomization of questions also affects the difficulty for subsequent attempts. In question mode, there was a dramatic shift of the TIC from +1.25 on the first attempt to –1.25 on the third attempt. However, in mastery mode, the shift was less dramatic, from +1.125 on the first attempt to –0.500 on the third attempt. The shift from the third attempt to the fourth attempt was small going from –0.500 on the third attempt to –0.750 on the fourth attempt. In question mode, the questions are not scrambled, so if a student gets question #2 (index number 383) incorrect in an instructional units, they can use the feedback as a guide since on the second attempt a very similar type of question (index 383) will be presented for question #2. Even though the questions are parameterized, the parameters are often constrained. One question in a precipitation reaction question could involve insoluble sulfates whereas another question in that same instructional unit could involve insoluble phosphates. This was the reason that titration curves questions were not difficult in question mode. If a student knew that question #3 was going to be a titration curve pH calculation at the equivalence point, they could use the feedback to help on the second attempt. When they attempt question #3 for the second time, the feedback "cues" the student to the correct answer (in effect, the student "memorizes" how to do the problem). However, in mastery mode, OWL question number 383 could be question #1 in the first module attempt and question #3 on the second module attempt. Therefore, students cannot “memorize” the feedback to solve question #1 since it has some differences.

Section 8.3 – Recommended number of attempts

In question mode, students can use the feedback to help them answer the question correctly on subsequent attempts since the feedback is more closely aligned with the question. Subsequent attempts on a question mode question are similar enough to the question on the first attempt so that feedback given to the students can be effectively used to answer the subsequent attempts correctly. By the third attempt, only the low ability students are unable to answer the question successfully. Because of this, it is recommended that question mode questions be limited to three attempts.

Mastery mode questions are randomly chosen from a question pool, so feedback from a particular question such as question #1 on the first module attempt may not be relevant to a student on question #1 for the second module attempt since the question may be different from the first attempt question.

Due to this randomization, allowing for an unlimited number of attempts will help students learn the material, considering that the overall homework difficulty decreases slowly with each attempt after the first attempt.

If an instructor wants to assess a group of students using the OWL homework questions, those questions should be run in question mode. Instructors wanting to give student practice with chemistry problems should run those questions in mastery mode.

REFERENCES

(1) Schurmeier, Kimberly, Analyzing and Improving General Chemistry Teaching and Assessment Using Item Response Theory. Ph.D. Thesis, University of Georgia, Athens, GA, 30602

(2) Schurmeier, Kimberly D.; Shepler, Carrie G.; Lautenschlager, Gary J.; Atwood, Charles H. In *Using Item Response Theory To Identify and Address Difficult Topics In General Chemistry*. ACS Symposium Series; American Chemical Society; Washington DC, 2011; pp 137–176.

(3) Oliver-Hoyo; Allen. *J. Chem. Ed*. **2004**, 81, 441 – 448.

(4) http://www.ncsu.edu/project/stereotypes/OliverHoyo.htm

(5) http://www.udel.edu/pbl/curric/chem103-prob.html

(6) https://engineering.purdue.edu/CATME

(7) Layton, Richard A., et. al. *Advances in Engineering Education*, **2010**, 1 – 28.

(8) Ohland, M. W., et. al. *Academy of Learning Management Learning & Education*, **2012**, 11, 609 – 630.

(9) Ohland, M. W., et. al. *Educational and Psychological Measurement*, **2007**, 67, 505 – 524.

(10) de Ayala, R. J., *The Theory and Practice of Item Response Theory*, Guilford Press: New York NY, 2009.

(11) du Toit, Mathilda, IRT from SSI, Scientific Software International: Lincolnwood, IL, 2003.

(12) Personal conversation with Lisa Lockwood of Cengage Learning

(13) Personal conversation with Gordon Anderson

(14) Personal conversation with William Vining

(15) Personal conversation Joel Caughran

(16) Personal conversation with Charles Atwood

APPENDICIES

IDENTIFYING DIFFICULT TOPICS AND PROBLEMATIC SUBTOPICS USING ITEM RESPONSE THEORY (IRT) ANALYSIS OF MASTERY MODE HOMEWORK QUESTIONS IN A NATIONAL ONLINE WEB-BASED LEARNING (OWL) DATABASE

Appendix A – Step by step instructions for data analysis:

The data was separated into groups of approximately one million lines (called groups). The data was separated in such a way that all of the responses for a given question were in the same group. Once this is done, do the following steps in order to convert the data into a form where BILOG MG 3.0 can analyze it.

1. Sort each data group into the following categories based on number of attempts allowed: 1 or 2 attempts, 3 or 4 attempts, 5 or 6 attempts, 7 to 10 attempts, and 11 or more (includes unlimited) attempts. If any of the categories in a group have less than one million lines, combine the groups within the same category, as long as the total number of lines is one million or less.

2. Save the file as "Atx_Groupz" where x is the minimum number of attempts permitted in the data set (1, 3, 5, 7, or UNL) and z is the group number. Be sure to save all files in steps 2 – 12 as Microsoft Excel files.

3. Delete all columns except for the following: Institution ID, Course ID, Module No., Student ID, Question Number, Date & Time, Raw Score, and Attempt No. Save the file as "Atx_Groupz_v1".

4. To account for Institution ID's with less than four digits, insert 2 columns to the left of the Institution ID column. Use the command "=IF(A2<1000,A2+9000,A2)" in the first column. Fill in all of the cells. Copy and paste special (values) the column of formulas into the second column. Pasting formulas as values prevents Excel from "freezing" due to its tendency to recalculate formulas. Delete the two left most columns. This step ensures that all institution ID's have four digits.

5. Insert a column to the left of the index column. Label the column "Inst+StuID". Assuming the maximum number of digits in a student ID is six, type the formula "=InstID*1e6+StuID" and fill the cells. Insert a column to the right then copy and paste special. Delete the column of formulas. Delete the Institution ID and Student ID columns. Save the file as "Atx_Groupz_v2". These steps ensure that two students from two different institutions with the same ID number are counted as two separate records.

6. Insert a column to the left of the Attempt number column. Label the column "Combine" and type the formula "=concatenate(Inst+StuID, Index). Fill the cells. Insert another column and copy and paste special (values). Delete the column of formulas.

7. Select all of the columns and "Remove Duplicates". Sort the "Answer Date" or the "Attempt Number" column in ascending order (click on "Expand the sort" when that box pops up), then sort the "Combine" column. Add "New.Att." column to the right of the Attempt column. Type the formula "=countif(Combine$2:Combine2,Combine2)". Fill 400 of the cells in the column (the actual minimum number is 2 x the maximum attempt number). Go to the last cell in the column with a formula. Remove the "$" symbol from the formula. Fill in the remainder of the cells (this calculation is complex and may take a minute or two). Copy and paste special (values)

the “New.Att.” column into another column. Delete the column of formulas. Save the file as “Atx_Groupz_v3”. This renumbers the attempts so the attempt number is question based rather than module based.

8. Insert two columns to the left of the “Raw Score” column. In the first column, labeled “Score”, type in the formula “=floor(RawScore,1)”. Fill the cells. Copy and paste special (values) the “Score” column into the second column.

9. Delete all columns except for the following: Inst+StuID, Index, Score, and New.Att. Save the file as “Atx_Groupz_v4”.

10. Sort the “New.Att.” column in ascending order. Copy all responses from attempt #1 into a different sheet and rename the sheet “Try1”. Copy all responses from attempts #1 and #2 into another sheet and rename the sheet “Try2”. Do the same thing up to attempt #6. Save the file as “Atx_Groupz_v5”.

11. Delete the sheet containing all of the response data, and the “New.Att.” column for all six tries. Save the file as “Atx_Groupz_v6”.

12. Click on the “Try1” sheet. Click on “Insert” and choose “Pivot Table”. Be sure that all of the data is selected. A new worksheet will be inserted. Drag the “Inst+StuID” to the “Row Labels” box. Drag the “Index” to the “Column Labels” box. Drag the “Score” to the “Values” box. The label “Sum of Score” will appear in the “Values” box. Click on the “Sum of Score” label. Select “Values of Field”. A box will appear. Select “Max”, and then click “Ok”. This ensures that a student getting a score of “1” on a particular attempt will receive a score of “1” on all subsequent attempts.

13. Create a new worksheet. Copy and paste special (values) the contents of the entire pivot table into the new worksheet. Save the new worksheet (in the filename, list the try number first). The data will likely consist of 0's and 1's (except for the row and column labels). Insert a column to the right of the row labels. Fill all of the cells in the column with the letter "Z". In the new worksheet, do a find and replace. Find "" (blank cells) and replace with "9". In order to prevent Excel from crashing, do only 1 million cells at a time. Once that is done, save the file as "Tryx_Groupz_v1". In this case, save the file as a Commas Separated Values (.csv) file. Do this for each of the first 6 attempts. Create folders for each of the attempts (i.e. Try1, Try2, etc.)

14. Open each .csv file in Notepad++. Save each file in the format "inputx.dat" where x is the number of the data set. Be sure to save the file in the correct folder based on the attempt. Replace all of the commas with nothing. Replace all of the "Z"s with a single space. Do that for all of the files. Be sure to save each file.

15. Once all of the files have been converted to a format that can be read by the IRT program, the files need to be merged. A Perl script has been developed that merge all of the individual files into one large text file containing all of the grades. The Perl script can only be used if each file has a unique set of question numbers.

16. At this point, the data is in the correct format for IRT analysis. After the IRT is run, there is the possibility that the data may not converge, or a few questions may not fit the IRT model. Those questions need to be removed. In addition, any questions where all or none of the students responded correctly need to be removed.

17. Insert the “counter” text file at the top of the data file. This helps with matching up each column with its item number. Then open the file into Microsoft Excel. Select “Fixed Width” for the file type that best describes your data. This option allows you to break up the data into individual cells. Click on “Next>”. Any item numbers that are bad questions are to be put into an individual column of cells, so be sure to break up the columns in order to isolate those bad questions. In addition, be sure the space between the identification numbers and the response patterns is also put into an individual column of cells. Once you have separated all of the bad questions, click on “Next>”. For the third step, select each column individually. Select “Text” under the heading “Column data format”. BE SURE TO DO THIS FOR ALL COLUMNS. Once you have done that, click on “Finish”.

18. Save the file as an Excel file (you can use the same filename, but the extension will be “.xlsx”. You will notice that this decreases the size of the file. Remove all columns that represent response data for bad questions (note that the item number is contained in the cells above each column). Rename Sheet1 to “a”. Insert another worksheet. Rename that worksheet “B”

19. In some cases, removing some of the questions may cause some of the students to have 0 responses. Those students with 0 responses need to be removed. Steps 19 – 22 will identify and remove those students with 0 responses.

20. Copy the first column from SheetA to SheetB. In column B (SheetB), put “Z” in all of the cells in that column. In column “C” in SheetB, type “=concatenate(‘A’!c1,’A’!d1…). Be sure to include all of the cells that contain data. Fill all of the cells in SheetB.

21. In SheetB, type the following:

In column D, "=search(0,c1,1)

In column E, "=search(1,c1,1)

In column F, "=iferror(d1,2000)

In column G, "=iferror(e1,2000)

In column H, "=f1+g1"

In column I, "=Len(c1)"

22. Fill all of the cells. Copy and paste special (values) the contents of column H into column J. Check be sure that all of the values in column I are the same. You do that by typing in the command =stdev(I1:I"y") where "y" is the number of students. If the result is 0, proceed. If not, check to ensure that you have included all of the columns.

23. Delete columns D – I. Sort column D (formerly column J) in ascending order. Any row that has a value of 4000 in column D refers to a student (ID in column A) that has 0 responses. Remove those rows. Delete column D. You should have a column ID in column A, a "Z" in column B, and a string of 0's, 1's, and 9's in column C.

24. Save SheetB as a .csv file. Be sure to save it as a unique file name, such as "Try3_all_grades_vxxx", where xxx is increased by 1 every time question(s) are removed. You then edit the file in notepad (change the extension to .dat), removing the commas and the "Z" as you did before.

25. If, after running the IRT, you still get a bivariate plot with huge error bars, a correlation less than 0.400, and a histogram that does not have a proper distribution, then you need to go to the .sco (score) file, and look for students that did not attempt a single question. In this case, what has happened the student answered one or more questions that were not calibrated due to having a biserial R of less than –0.150. Remove those students that did not attempt any of the questions from the .dat files.

When you are finished, you should have the results from the first 6 attempts.