5-2014

ASSESSING COLLEGE STUDENTS' UNDERSTANDING OF ACID BASE CHEMISTRY CONCEPTS

Yanjun Wan
Clemson University, ywan@clemson.edu

ASSESSING COLLEGE STUDENTS' UNDERSTANDING OF ACID BASE CHEMISTRY CONCEPTS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemistry Education

---

by
Yanjun Jean Wan
May 2014

Accepted by:
Dr. Melanie Cooper, Committee Chair
Dr. Gautam Bhattacharyya
Dr. Rhett Smith
Dr. Julie Martin

ABSTRACT

Typically most college curricula include three acid base models: Arrhenius', Bronsted-Lowry's, and Lewis'. Although Lewis' acid base model is generally thought to be the most sophisticated among these three models, and can be further applied in reaction mechanisms, most general chemistry curricula either do not include Lewis' acid base model, or quickly mention it at the end of the acid base chapter, because of the concern that Lewis' model may confuse general chemistry students (Shaffer 2006). While such a disconnection in curriculum might put students to disadvantage as they try to construct solid and coherent acid base mental models, there has not been any research data to favor one curriculum over another. The large sizes of general chemistry courses at most universities (from one hundred to several hundred students per lecture section) pose further challenges to the comparison of different general chemistry curricula on their effectiveness in helping students construct acid base mental models. In light of these challenges, the research questions I focused on were: 1) What are the important characteristics of activities that effectively promote and retain argumentation skills among college students? 2) In what ways is argumentation an effective assessment method for student understanding of acid base models? 3) How do different curricula affect students' acid base models? This dissertation presents promising results from using BeSocratic activities in promoting argumentation skills among college students and at the same time using their responses in the activities to understand aspects of their acid base mental models, and compare how two different general chemistry curricula affected students' acid base mental models.

DEDICATION

I love you, O Lord, my strength. – Psalm 18:1

Tis Grace that brought me safe thus far,

And Grace will lead me home.

– Amazing Grace

Whom have I in heaven but you?

And earth has nothing I desire besides you.

My flesh and my heart may fail,

But God is the strength of my heart and my portion forever.

– Psalm 73:25-26

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Dr. Cooper, for teaching me many things patiently throughout the past few years; and my advisory committee, Dr. Bhattacharyya, Dr. Smith, Dr. Potvin, and Dr. Martin, for giving me many advices in the past few years, for research and for the drafting of this dissertation; as well as members of the Cooper Research Group, for giving me many feedback and advice as I prepare different presentations in the past few years.

I would also like to thank Dr. Dieter and Dr. Stuart for the help and advice, making this dissertation possible; as well as other faculty members of the Chemistry Department, for everything you have taught me in the past few years.

Lastly, I would like to thank my parents and my brothers and sisters in Christ, for your prayers and encouragements throughout many years.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

List of Tables (Continued)

Table Page

LIST OF FIGURES

List of Figures (Continued)

CHAPTER ONE

INTRODUCTION

Acid base chemistry is an important area in different disciplines of chemistry. For example, many organic chemistry reactions can be considered as Lewis acid base reactions; inorganic chemists also frequently use d-block metals, which can be considered as Lewis acids, in coupling reactions and/or organometallic catalysts; the direction and extent of many biochemistry reactions are also determined by the comparative acid/base strength of different compounds.

Chemists have come up with many different acid base models to describe the reactions between acids and bases, because each model emphasize a particular aspect of the acid base reactions, and each model has its unique applications and limitations. However, when chemistry students were presented with these different acid base models, could they use those models flexibly, or would they contradict one model to another? How would their understanding and uses of different acid base models affect their ability to correctly solve acid base related problems, such as determining acid/base strength?

Before attempting to answer some of the above questions, we need to first take a look at the three acid base models most commonly taught in high school and college chemistry courses: Arrhenius', Bronsted-Lowry's, and Lewis' models.

Arrhenius' model is mostly taught in high school chemistry courses and it defines acids as compounds that can dissolve and dissociate in water to produce hydronium ions, while bases as compounds that can dissolve and dissociate in water to produce hydroxide ions. One big limitation of Arrhenius' model is that it requires a compound to first be able to dissolve and dissociate in water, while most organic compounds cannot meet this requirement. Thus, a large amount of compounds that can act as acids and/or bases would not be categorized as acids or bases according to Arrhenius' model.

Bronsted-Lowry's model is mostly taught in college general chemistry courses, and it defines acids as proton donors in a reaction, while bases as proton acceptors. Because all Arrhenius acids are proton donors in aqueous solutions, they are all Bronsted-Lowry acids as well. Similarly, because hydroxide ions are good proton acceptors in aqueous solutions, all Arrhenius bases are also Bronsted-Lowry bases. Bronsted-Lowry's model broadened Arrhenius' model largely, because now organic compounds can also be categorized as acids and/or bases based on whether they would lose or gain a proton in an organic reaction; and aqueous solution is no longer a limitation for Bronsted-Lowry acids and bases. Bronsted-Lowry's model is also the most frequently used acid base model, to the extent that when most chemists say "acid base" without specifying a particular model, they are automatically referring to Bronsted-Lowry acids and bases. This is because Bronsted-Lowry's acid base model directly associates acid with the concentration of hydronium ion, which is easily measureable. However, Bronsted-Lowry's model is still limited in the sense that only reactions involving proton transfer can be categorized as acid base reactions.

Lewis' model further broadened the definition of acids to include species that do not contain protons (and thus cannot be proton donors or Bronsted-Lowry acids). It defines acids as electron pair acceptors; thus, not only all the Bronsted-Lowry acids are included as Lewis acids (because any proton that can be easily donated would have a large partial positive charge, so it can be considered electron poor and a good electron pair acceptor), but compounds containing boron, aluminum, or transitional metal cations are also included as Lewis' acids. On the other hand, Lewis' model defined bases from a different angle than Bronsted-Lowry's model: instead of looking at the transfer of a proton for an acid base reaction, and defining a base as a proton acceptor (Bronsted-Lowry's model); Lewis' model looks at the donation of lone pair electrons into forming new bonds, and defines a base as a donor of lone pair electrons. Although Lewis' and Bronsted-Lowry's models look at bases from different angles, they do not contradict each other. For a compound to be categorized as a Lewis base, it must have at least one lone pair of electrons that it is willing to donate into a bond with an electron poor species (a Lewis acid). At the same time, for a compound to be categorized as a Bronsted-Lowry base, it must also have at least one lone pair of electrons that it is willing to donate into a bond, because that it what a Bronsted-Lowry base uses to accept the proton donated from a Bronsted-Lowry acid. Lewis' acid base model not only allows a boarder definition of acids and bases, but also allows many organic reactions to be considered as Lewis acid base reactions; thus it is the most frequently used acid base model among chemists. However, depending on the colleges and curricula, Lewis' acid base model might be covered briefly, if at all, in a college level general chemistry course.

Although there are many acid base models, and each define acids and bases differently; these models do not contradict each other in the essence of acid base behaviors. For example, HCl is an acid according to Arrhenius' model; and it is also an acid according to Bronsted-Lowry's model and Lewis' model. It is unlikely that one compound would be defined as an acid according to one model, but only a base according to another model (it might be defined as both an acid and a base because the second model broadened the definition of base from the first model). These different models look at acid base behaviors from different angles without contradicting in the essence of such behaviors, thus allowing chemists the flexibility to choose an appropriate model for each unique task. However, multiple models can pose a challenge to students, making it easier for them to confuse or contradict one model with another.

.

CHAPTER TWO

LITERATURE REVIEW

As mentioned in the Introduction, most college chemistry curricula include three different acid base models: Arrhenius', Bronsted-Lowry's, and Lewis'. While most high school chemistry curricula have covered Arrhenius' acid base model to some extent, most general chemistry curricula in college focus on Bronsted-Lowry acid base model because it is the most frequently used acid base model. Some colleges will also mention Lewis' model in their general chemistry curricula; whiles some other colleges are concerned that Lewis' model may confuse general chemistry students, and choose to teach Lewis' model in a higher level chemistry course (Shaffer 2006). While such a disconnection in curriculum might put students to disadvantage as they try to construct solid and coherent acid base mental models, there has not been any research data to favor one curriculum over another.

The following sections will review acid base chemistry related researches in several different categories. First of all, researches to understand students' ideas and beliefs related to acid base chemistry were mainly divided into two approaches: misconception research aims at identifying common misconceptions students have in the area of acid base chemistry; while mental model research attempts to identify different mental models students' use in describing acids and bases. Another type of research focuses on the uses of heuristics in solving specific acid base problems. Finally, the attempts to improve students' understandings of acid base chemistry concepts were also divided into two major categories: some researchers came up with different conceptual change frameworks to address the prevailing misconceptions identified in prior researches; while some others proposed and tested different interventions in and out of class, hoping to identify interventions that will significantly improve students' understanding of acid base chemistry.

Misconception Research

The misconception research in the specific area of acid base chemistry has mainly focused on high school students so far. Thus, a majority of misconceptions and alternative ideas reported were on surface levels.

For example, Demerouti et. al. designed a questionnaire consists of ten multiple-choice questions and eight open-ended questions covering seven different areas of acid base chemistry: "(a) dissociation and ionization, (b) definition of Brønsted–Lowry acids and bases, (c) ionic equilibria, (d) neutralization, (e) pH, (f) buffer solutions, and (g) degree of ionization" (Demerouti, Kousathana, & Tsaparlis 2004). This questionnaire was administered to one hundred and nineteen high school chemistry students; and students were asked to explain their choices for the multiple-choice questions. Then a total of four "experienced" high school teachers graded students' responses on a scale of 0-10, with Spearman ρ correlations ranging from 0.90 to 1.00 among the four graders. From the results Demerouti et. al. summarized a list of misconceptions and difficulties high school students experience in the area of acid base chemistry, for example, "a strong acid requires more moles of a strong base than a weak one for its neutralization because it is strong acid (and similarly for a strong base)” and “reactions of weak acids and bases as irreversible”.

In another study, Demircioglu et. al. designed and administered a twenty-item multiple-choice questionnaire to eighty-eight high school chemistry students as pre-test and post-test before and after instruction (Demircioglu, Ayas, & Demircioglu 2005). A list of popular misconceptions identified in the post-test (after instruction) include “at the end of all neutralization reactions, there are neither H^+ nor OH^- ions in the resulting solutions”, “in all neutralization reactions, acid and base consume each other completely”, “all salts are neutral”, “acids burn and melt everything”, “pH is a measure of acidity”, “as the number of hydrogen atoms increases in the formula of an acid, its acidity becomes stronger”, etc. This study also involved an intervention as an attempt to alleviate these misconceptions, which will be discussed later in the “Interventions” section.

Based on Demircioglu’s research (Demircioglu, Ayas, & Demircioglu 2005), Ozmen et. al. designed and administered a twenty-five item multiple-choice questionnaires to fifty-nine high school chemistry students as pre-test and post-test before and after instruction (Ozmen, Demircioglu, & Coll 2009), and found out some similar misconceptions such as “all salts are neutral”, “in all neutralization reactions, acid and base consume each other completely”, “at the end of all neutralization reactions, there is neither H^+ nor OH^- ions in the resulting solutions”, and “acids burn and melt everything”. In addition, several other popular misconceptions were identified, such as “strong acids can react with all metals to form H_2 gas”, “salts don’t have a value of pH”, “a strong acid is always a concentrated acid", "after all the neutralization reactions, the pH of formed solution is always 7", etc. This study also involved an intervention as an attempt to alleviate these misconceptions, which will be discussed later in the "Interventions" section.

At college level, Jasien designed a nine-question multiple-choice quiz to examine undergraduate students' understanding of acid base chemistry concepts (Jasien 2005). In this quiz, the first four questions were numerical; question five to eight were pictorial and paired with the first four questions, examining the same concepts from molecular-levels rather than from quantitative aspects; and question nine was another molecular-level question correlated to question five. A total of four hundred students participated in this study, coming from different colleges (a public university, a private university, and community college) and different levels (ranging from first-semester general chemistry to upper level biochemistry). Although the group from an upper level biochemistry class seemed to have higher averages on most questions, Jasien specified that the primary purpose of this study was not to compare the performance of different groups, due to the large variety of backgrounds among these groups. Instead, Jasien concluded that there was a positive correlation between the paired numerical and pictorial questions, although their causal relationship was uncertain. Jasien also noticed a "general confusion between the ideas of pH (i.e., free hydrogen ion concentration) and the overall concentration of the acid, HA, in solution", across all the groups, including the upper level biochemistry group.

Recently, McClary and Bretz developed a concept inventory to identify common misconceptions among organic chemistry students when they compare the acid strength between different compounds (McClary & Bretz 2012). This nine-item multiple-tier, multiple-choice concept inventory was constructed from previous qualitative studies (McClary & Talanquer 2011a&b), and then administered to one hundred and four undergraduate students at the beginning of their second semester organic chemistry course. The two common misconceptions identified through this concept inventory were "functional group determines acid strength" and "stability determines acid strength" However, because students can always guess in multiple-choice questions, the complete elimination of free response from an assessment will also miss the uniqueness of what each student truly believes.

Mental Model Research

Taking a different approach, several other research groups focused on qualitative research to understand different acid base mental models individual students use in solving different problems in the area of acid base chemistry (Bhattacharyya 2006, Halstead & Anderson 2009, and McClary & Talanquer 2011a).

Bhattacharyya interviewed ten organic chemistry doctoral students using a model-eliciting activity, in which students were given a list of pK_a values of different alcohols, and asked to "create a set of rules that could explain acidities of organic molecules from these data" (Bhattacharyya 2006). He concluded from the results that many "expert" students combine different theories freely to create their own models and highlight a particular aspect of a molecule's chemical behavior, and named such kind of mental models as "hybrid models".

Halstead and Anderson proposed the term "operational" to describe a type of mental models that "describes acids and bases in terms of macroscopic properties displayed by classes of substances or their solutions" (Halstead & Anderson 2009). For example, students who define acids as compounds with a pH value of 7 or less will be considered as having this operational mental model.

McClary and Talanquer interviewed nineteen first-semester undergraduate organic chemistry students and identified four distinct mental models students used in predicting acid strengths (McClary & Talanquer 2011a). They named these four mental models "Mental Model A through D", because although some of these mental models resemble the scientific acid base models commonly taught in college chemistry (specifically, Mental Model B resembles Arrhenius' acid base model; Mental Model C resembles Bronsted-Lowry's acid base model; Mental Model D resembles Lewis' acid base model), McClary and Talanquer believed that these mental models are "better characterized as synthetic models that combined assumptions from one or more scientific models". Mental Model A, on the other hand, represents a "rather underdeveloped conceptualization of acids and acid strength", according to McClary and Talanquer, because students expressing this mental model relied solely on "the presence of certain atoms or functional groups" to determine the acidic or basic property of a substance, rather than considering the acid base behavior from the molecular level. McClary and Talanquer also found out that some students used a single mental model to solve all the problems, while some other students changed mental models based on the nature of the problem.

Moreover, many researchers in the area of mental model research have come to agreement that the mental model of an individual student is incredibly complicated and unique, and different aspects of the mental model are exhibited based on different tasks. In a study of college students' understanding of structure-property relationships, Cooper et. al. observed that among the 17 interviewed students, "no two students used the same sets of ideas to perform the task at hand"; and thus proposed that "student understanding is best understood as a set of loosely connected ideas, skills, and heuristics" (Cooper, Corley, & Underwood 2013).

Although the above-mentioned qualitative researches offered much inside into the uniqueness and complicity of individual students' mental models, the need for an appropriate assessment for students' acid base mental models remains. The large size of most college-level general chemistry classes (from one hundred to several hundred students per lecture section) adds further challenge to the design of an appropriate assessment. We need an assessment that can strike a good balance between the retention of the individuality of each student's response and the relative easiness of administering the assessment and analyzing the data for large populations.

Uses of Heuristics

Maeyer and Talanquer reported that college students frequently use heuristics to aid their understanding of acid base behaviors (Maeyer & Talanquer 2010). In their study, Maeyer and Talanquer interviewed a total of thirty four second-semester general chemistry students, and asked each student to rank chemical substances based on the relative value of a physical or chemical property. The results of this study revealed that many students relied frequently on one or more types of heuristics to make their decisions. Maeyer and Talanquer then summarized the different heuristics students used into four categories: "recognition, representativeness, one-reason decision making, and arbitrary trend".

In a following study, McClary and Talanquer focused on students' uses of heuristics in making decisions about acid strength (McClary & Talanquer 2011b). By interviewing nineteen first-semester undergraduate organic chemistry students individually, McClary and Talanquer discovered a common trend that a number of students "thought of certain atoms, such as H, O, or Cl, or certain functional groups, such as hydroxyl (−OH) or carbonyl (−C=O), as intrinsically acidic or basic".

The frequent use of heuristics by many students can be further explained by the dual process theory (Gilovich, Griffin, & Kahneman 2002 and Evans 2003). Dual process theory categories the process of thinking into two types: system I thinking often uses heuristics (instructed or self-developed) to quickly solve a problem without engaging in detailed analysis; on the other hand, system II thinking is much slower and engages in detailed analysis. Although system II thinking is often more correct, it also takes significantly more time and requires a significantly larger cognitive load. Thus, even expects use heuristics in solving some problems. The challenge for many college students is, they do not always know when to use heuristics and how to use heuristics properly. So often when they use heuristics rather than system II thinking, they end up with wrong answers (Maeyer & Talanquer 2010). Furthermore, even after they were confronted with their inappropriate uses of heuristics, some students still would not go through system II thinking. For example, McClary and Talanquer found that many students use the presence of hydrogen atoms to determine acidity (the "more hydrogen means more acidic" heuristics). However, when a student came to the realization that the number of hydrogens in a compound does not determine the acid strength of that compound, he quickly resorted to a slightly different heuristics ("more chlorine means more acidic" rather than "more hydrogen means more acidic"): "It just seems like hydrogen usually doesn't play much in the like acidity thing…Okay, I'm gonna guess the more chlorine the more acidic", rather than approaching the initial problem with system II thinking and analyzing the acidity from a molecular level (McClary & Talanquer 2011b).

Conceptual Change Frameworks

After identifying common misconceptions, some researchers came up with different conceptual change frameworks to guide their further studies of how to alleviate such misconceptions. Although there are a few different conceptual change frameworks, they commonly agree that misconceptions are not merely "mistakes or false beliefs" (Posner, Strike, Hewson, & Gertzog 1982) but are "mental representations of concepts that are at variance with currently held scientific theories" (Demerouti, Kousathana, & Tsaparlis 2004).

In their conceptual change framework, Posner et. al. (Posner, Strike, Hewson, & Gertzog 1982) first pointed out that it is very difficult to modify some misconceptions once they are formed; because most conceptions do not exist alone, but are connected to other conceptions. Thus, when one misconception faces intellectual challenges, other related conceptions will serve as its "cognitive support group", and resist any modification of this misconception. In order to achieve a successful conceptual change, several crucial conditions must be met; including the dissatisfaction of their current conception, the ability to understand the new conception in a meaningful way, and the initial plausibility and fruitfulness of the new conception in solving previously unsolvable problems.

Chi proposed a different framework which separates conceptual changes into three categories: belief revision, mental model transformation, and categorical shift (Chi 2008). According to Chi, false beliefs are single ideas that are incorrect, and can be easily corrected by direct instruction of the corresponding correct ideas. On the other hand, a "flawed mental model" is an organized collection of individual beliefs, and it can be a coherent but incorrect representation of a concept (Chi, Slotta, & DeLeeuw 1994 and Chi 2008). When students were presented with different models, many of them often end up mixing different parts of different models into hybrids of models that are unique to each student. These hybrid mental models can be coherent but flawed. The reason this kind of models can be very appealing to many students is because they can generate explanations, make predictions, and answer questions in a consistent and systematic fashion (although such explanations and predictions are sometimes incorrect).

Eventually, many students fail to realize the different limitations and problems with each individual model. Consequently, misconceptions generated from these flawed mental models become more and more robust. Similar to conceptual change theories, the transformation of a flawed mental model calls for an accumulation of belief revisions, where critical false beliefs within a flawed mental model are refuted with correct information and explanations. According to Chi, the cumulative effect of many belief revisions will transform a flawed mental model into the correct model. However, Chi also admitted that "knowing and learning many correct beliefs does not guarantee successful transformation of a flawed mental model to the correct model". One can make numerous revisions in response to refutations of a flawed mental model, yet do not change the underlying core hypotheses. Thus, a flawed mental model can be "patched" multiple times, yet still does not transform into the correct model. In a recent study, Chi and her colleagues explored the effect of asking students to "contrasts his or her flawed mental model to an expert model", and concluded that it is a better method than simply giving students the expert model and asking them to explain it, in helping students build correct mental models (Gadgil, Nokes-Malach, & Chi 2012). Despite of the encouraging initial results, this study designed the post-test immediately after instruction, leaving a crucial question unanswered: how long can these students retain their corrected mental models?

Although Posner and Chi differ slightly in their categorizations of misconceptions, both of their conceptual change frameworks treat misconceptions as fairly coherent and reconstructable. Thus, both frameworks focus on how to help students reconstruct misconceptions into correct conceptions. On the other hand, some other researchers believe that students construct loosely woven explanations from smaller fragments (DiSessa 1993, 2006, & 2008, Hammer 1996, and Cooper, Corley, & Underwood 2013). DiSessa first proposed the term “phenomenological primitives” (or “p-prims”) to account for the existence of more fundamental, more abstract cognitive structures (DiSessa 1993). According to this framework, p-prims are not incorrect, but can be incorrectly activated to give incorrect final results. For example, “hydrogen atoms indicate acid” would be considered a misconception according to the previous frameworks, since not all compounds that contain hydrogen atoms are acids, and not all acids must contain hydrogen atoms. But according to DiSessa’s framework, “most acids contain hydrogen atoms” is considered a p-prim that itself is not incorrect. However, this p-prim can be activated incorrectly in some situations and give incorrect results (i.e. students that conclude alkanes are strong acids because they contain many hydrogens in their structures). If students do not hold coherent misconceptions, but rather construct loosely woven explanations according to different tasks, then different instruction approaches would be required to facilitate conceptual change.

Regardless of which specific framework a research adopts, assessing students’ prior knowledge and their understanding of acid base chemistry concepts after instruction is always a crucial step before any attempt to design different instruction approaches to improve students’ understanding in the area of acid base chemistry.

Interventions

Because there are very few of such kind of assessments, very few interventions have been reported on how to improve students' understanding of acid base chemistry concepts.

As mentioned earlier, Demircioglu et. al. developed a twenty-item multiple-choice questionnaire to identify some common misconceptions among high school chemistry students (Demircioglu, Ayas, & Demircioglu 2005). After administering this questionnaire as a pre-test, Demircioglu et. al. designed "new teaching material" for the treatment group based on "conceptual conflict strategy". The "new teaching material" targeted misconceptions students demonstrated in the pre-test, and designed "worksheets, demonstrations, and analogies" to help students active engage in confronting these misconceptions during class time. After instruction, a post-test was given, in which the treatment group with the "new teaching material" expressed significantly less misconceptions than the control group. Although the results were encouraging, how well the same strategy can be applied to college students and help them alleviate misconceptions related to structure property relationships remains a question. For example, the only part of the "new teaching material" demonstrated in this paper was a laboratory activity in which students used the pH paper and several other indicators to test the acidity/basicity of different samples. This activity specifically targeted the misconception "the only way to test a sample whether it is an acid or a base is to see if it eats something away, for example metal, plastic, animal, and us" identified during the pre-test. In the treatment group, 45% of the students expressed this misconception during the pre-test, but after the activity, none of them still held this misconception. However, this “misconception” alleviated by the intervention is more like a naïve idea – even without the laboratory activity, very few college students would still hold such a belief even after regular instruction in chemistry classes.

Based on Demircioglu’s research (Demircioglu, Ayas, & Demircioglu 2005), Ozmen et. al. developed a series of different laboratory activities and accessed the effectiveness of these activities using a twenty-five item multiple-choice questionnaires as pre-test and post-test (Ozmen, Demircioglu, & Coll 2009). They also reported that the intervention of these new laboratory activities helped students in the treatment group overcome significantly more misconceptions than the control group taught in a traditional lecture manner. However, its application in college level chemistry also remains a question.

Besides designing relatively short interventions in the hope of alleviating specific misconceptions – which becomes increasingly more difficult as we get into college level chemistry and structure property related misconceptions, another approach would be to redesign the entire curriculum to better foster meaningful learning. As mentioned earlier, in order to design different instruction approaches to improve students’ understanding, appropriate assessments must first be developed – assessments that can strike a good balance between the retention of the individuality of each student’s response and the relative easiness of administering the assessment and analyzing the data – for the size of the student populations we intend to study.

CHAPTER THREE

THEORETICAL FRAMEWORKS

Meaningful Learning

In order to assess students' understanding of a particular concept, we need to first understand how students learn in general. The overarching framework for this research is constructivism, which means learners construct their own knowledge rather than receiving the knowledge directly from the teacher in a passive way. However, pedagogy, curriculum, and other aspects of the learning environment still affect an individual's learning experience and consequently the knowledge construct (Vygotsky 1962, Ausubel 1968, Novak 1978, & Howe 1996). Thus, I have chosen the meaningful learning framework (Novak 1993) as a more specific theoretical framework for this proposed research. The meaningful learning framework proposed two extremes of learning: rote learning and meaningful learning. Then it suggested that meaningful learning can only occur when the learner has relevant prior knowledge, the new material is taught in a meaningful way, and the learner chooses to integrate the new knowledge into his existing knowledge construct (Novak 1993 & 2002). Because the learner will have to choose meaningful learning over rote learning, appropriate assessments are necessary to encourage meaningful learning (Ridley & Novak 1988, Pendley, Bretz, & Novak 1994, and Novak 2002). Appropriate assessments not only inform the educators whether the instruction has been successful in fostering meaningful learning, but also encourage students to understand the materials in a meaningful way and to synthesize the materials on their own, rather than to memorize and regurgitate information they were taught in class. For example, if an instructor promotes meaningful learning in his/her lecture, but only examines students in their ability to recall factual information; then students would be forced to put most of their efforts in memorizing and regurgitating factual information, in order to receive good grades. Furthermore, when students learn materials in a meaningful way, they are also more likely to discover problems with their current alternative conceptions, and more motivated to switch to more correct conceptions so they can better explain some questions they encounter. The assessment tool Novak and his colleagues employed a lot is concept mapping (Pendley, Bretz, & Novak 1994). Concept maps can be used to trace the concept development and change in an individual or a group, as well as to elicit misconceptions. However, it is a time consuming qualitative assessment tool and is not ideal for large-population undergraduate chemistry courses.

Individual interview offers an in-depth understanding of a student's conception in a given area, and is an invaluable tool in exploring students' understandings and beliefs in not only the given topic, but also in related topics. However, interviews are not valid assessments for comparing the effectiveness of instruction in promoting meaningful learning in large populations. On the other hand, although quantitative assessments such as concept inventories composed of multiple-choice questions are relatively easy to administer and analyze, the reliability of the questions and choices for each question is doubtable, because it is very hard to capture the complexity of students' mental models with multiple-choice questions. Open-ended questions seem to be the most plausible approach because they can reduce the complexity of interviews, yet at the same time retain the rich information from different students without putting them in pre-labeled categories. However, even data from open-ended responses can be a far reach from students' real understanding if students are not trained to articulate their reasoning.

Toulmin's Argumentation Pattern

In fact, students' ability to "construct and defend their explanations" was a requirement according to the NRC Framework for Science Education. This research chose Toulmin's Argumentation Pattern (Toulmin 1958) as its methodological framework because it offers a good structure in teaching students how to articulate scientific reasoning, as well as in assessing the quality of a scientific argumentation. Toulmin identified several key components of a well-constructed argument: the claim, which is the purpose of the argument; the data, which includes evidence, example, and factual information about the claim; and the warrant, which bridges the claim and the data, and explains why the data lead to the claim. Other optional components of Toulmin's argumentation pattern also include backing, qualifier, and rebuttal, but are not necessary for all types of arguments (Toulmin 1958).

Based on this framework, some research has been conducted at K-12 level to study how to evaluate and improve scientific argumentation of individuals and groups (Erduran, Simon, & Osborne 2004, Osborne, Erduran, & Simon 2004, and Simon, Erduran, & Osborne 2006). Erduran and her colleagues (Erduran, Simon, & Osborne 2004 and Simon, Erduran, & Osborne 2006) first coded the different components of each students' scientific argumentation as "claim, data, warrant, backing, qualifier, or rebuttal", according to Toulmin's argumentation pattern. Then they "clustered" each argumentation by counting the number of components in each argumentation. For example, if an argument contains only claim and data, it would be a "cluster 2". An argument containing claim, data, and warrant would be a "cluster 3", as well as an argument containing claim, data, and rebuttal. After coding each student's argument into a cluster number, Erduran and her colleagues then traced a group of students over the course of two years, and found out that most individual students, as well as the group as a whole, improved significantly in their argumentation skills over the course of two years. Although this clustering method allows relatively easy coding and analyzing of data, two major downfalls of it include: 1) different components of an argument are not equal (some components such as qualifier and rebuttal are not necessary for all the arguments); and 2) this method merely counts the number of different components in an argument without assessing how well these components stand on their own and connect with each other. Realizing the problems, the same group of researchers explored a different method in coding the arguments (Osborne, Erduran, & Simon 2004), which they called the "rebuttal level method". According to this method, each individual argument was coded into one of the five levels, with level 1 being the weakest argument and level 5 being the strongest argument:

Level 1: Claim with no data, warrant, backing, or rebuttal;

Level 2: Claim with data, warrant, or backing, but no rebuttal;

Level 3: Claim with data, warrant, or backing, and a weak rebuttal;

Level 4: Claim with data, warrant, or backing, and a strong rebuttal;

Level 5: Claim with data, warrant, or backing, and multiple strong rebuttals.

Once each student's argument was coded according to the rebuttal level method, similar comparisons were performed over individual students as well as the group as a whole, and yielded similar results to the results from the cluster method. However, a downfall of this method is that not every argument needs one or multiple rebuttals.

Fewer studies have been conducted at college level to study how to evaluate and improve students' ability to construct scientific argumentation, because the idea of using Toulmin's argumentation pattern as a theoretical framework to study students' scientific reasoning skills has only recently come to the attention of the researchers in higher education. Cole and her colleagues borrowed the terminology of "as-if-shared idea" from mathematics education and used it to analyze the conceptual progress of a group of students in an undergraduate physical chemistry course (Cole, Becker, Towns, Sweeney, Wawro, & Rasmussen 2012). "As-if-shared ideas" are developed among a group of students when "warrants or backings are no longer required" for an argument, or when "previously justified claims function as data, warrant, or backing" to prove new claims (Rasmussen & Stephan 2008). Rather than first coding individual arguments, this approach looks at the conceptual shift in the entire class – when a claim no longer requires further explanation, or is quoted as data, warrant, or backing in supporting a new claim, it can be considered that the entire class has accepted the old claim as true and no longer needed explanation. Thus, the concept represented behind this old claim can be considered something this group of students have collectively learned and agreed upon.

Using this method, Cole and her colleagues analyzed a college level physical chemistry course, and found out that students collectively have developed several “as-if-shared ideas” such as “gas has the leas interaction”, “in solids, atoms are in a fixed position”, and “going from a solid to a liquid requires heat”, suggesting a collective conceptual growth in conceptions related to phases and phase changes (Cole, Becker, Towns, Sweeney, Wawro, & Rasmussen 2012).

This study employs Toulmin’s argumentation pattern as a training tool to improve students’ ability to articulate scientific explanations in writing form, in the hope that once students can articulate their reasoning, open-ended questions will be an appropriate and reliable reflection of their understanding of acid base chemistry and related concepts for a large population.

Research Questions

Based on the frameworks above, this study focused on three research questions:

RQ 1: What are the important characteristics of activities that effectively promote and retain argumentation skills among college students?

RQ 2: In what ways is argumentation an effective assessment method for student understanding of acid base models?

RQ 3: How do different curricula affect students' acid base models?

Most of the research to answer the research questions stated above was conducted at a public southeastern research university of approximately 20,000 undergraduate and graduate students. At this university, general chemistry courses are taught in lecture sections of between 100 and 150 students. Each semester, approximately 1500 students enroll in the on-semester general chemistry course. Two general chemistry curricula were offered by the Chemistry Department simultaneously: General Chemistry: Atom First by McMurry and Fay (students from this curriculum will be referred to as the "Traditional cohort" from now on); or Chemistry, Life, Universe, and Everything by Cooper and Klymkowsky (students from this curriculum will be referred to as the "CLUE cohort" from now on).

The next two chapters of this dissertation will describe two stages of this research in detail. Chapter four describes some preliminary research involving semi-structured interviews, open-ended questions, and multiple-choice questionnaires. Chapter five describes a research of students from two different general chemistry curricula in the course of two years. All the research was approved by Clemson University Institutional Review Board (IRB # 20124).

CHAPTER FOUR

PRELIMINARY RESEARCH

This chapter will be divided by the different methods used in this research: semi-structured interviews were first conducted to identify student beliefs about acids and bases for students of different levels. Some common ideas emerged from these interviews were further examined by open-ended questions administered on Ed's Tools. Popular student responses from these open-ended questions were then designed into a tiered multiple-choice questionnaire.

Semi-structured Interviews

To discover different conceptions concerning acid base chemistry from students of different levels (including general chemistry students, organic chemistry students, graduate students in chemistry-related majors, and graduate students in chemistry), semi-structured interviews were conducted. A total of eight volunteers from the public southeastern research university participated in the initial semi-structured interviews during the semester of Spring 2010. All the students were solicited by email. Among these eight participants, six were male and two were female; four were graduate students and four were undergraduate students; five majored in chemistry and three majored in chemical engineering, biology, and microbiology respectively. All participants have taken at least two semesters of general chemistry and one semester of organic chemistry. The majority of the participants (five out of eight) have also taken upper level chemistry courses after finishing two semesters of general chemistry and two semesters of organic chemistry. Semi-structured interviews with a few core questions ensured that the discussions stayed at higher levels, and at the same time allowed the freedom for different follow-up questions (See Appendix A – Interview Protocol). There were two major parts in each interview. In the first part, students were given different scenarios, where they had to explain acid and base concepts to audience of different levels (someone with no science background, their classmates, and their colleagues). Based on their explanations, different follow-up questions were asked to probe their understanding on the acid and base related concepts they used in their explanations. In the second part, students were provided with a list of chemical formulas and structures, and asked to identify each one of them as: 1) an acid; 2) a base; 3) both an acid and a base; or 4) neither an acid nor a base, and then explain each choice in a think-aloud manner, which means, students were encouraged to talk through their thought process. These chemical formulas and structures were discussed and carefully determined by one graduate student, one organic chemistry faculty, and one chemistry education faculty to represent different types of compounds and functional groups. Students at different levels were given different structures from the complete list. For example, the structures selected for undergraduate students did not include the most difficult organic compounds, whereas the structures selected for graduate students in chemistry did not include the most common acids and bases (i.e. HCl). Such selection allowed the study of a larger variety of compounds, yet prevented each individual student from being overwhelmed with extended interview time and questions. These interviews were transcribed and initially coded for any relevant acid and base ideas that emerged during the interview. Codes were generated during the coding process. A complete list of the codes generated during the coding process is shown below, followed by a few examples for each code. These codes were kept as-is rather than further clustered, because the initial interview only aims at getting a preliminary understanding of the common ideas about acids and bases college students have.

1. Incorrect ideas or fragments of ideas

    1.1 Wrong chemical formula, structure, or nomenclature

    1.2 Wrong reaction, expected product, mechanism, or explanation

    1.3 Incorrect/incomplete definition of acid/base

    1.4 Incorrect example of acid/base

    1.5 Incorrect/incomplete explanation of why something is an acid or a base

    1.6 Irrelevant Misconceptions

    1.7 Incorrect/incomplete explanation of acid/base related terms (acidity/basicity, pKa/pKb, neutralization, titration, electronegativity, etc)

2. Correct examples of acids/bases
3. Correct ideas of acids/bases
4. Correct acid/base strength comparison and/or reasoning
5. Incorrect acid/base strength comparison and/or reasoning
6. Strategies to identify acid/base
7. Correct identification of acid/base and correct reasoning in Part II
8. Incorrect identification of acid/base and/or incorrect reasoning in Part II

Table 4.1: Sample quotes from student interviews demonstrating each of the codes listed above.

Code	Student Quotes
1.1 Wrong chemical formula, structure, or nomenclature	(student drew "H_2PO_4" as an example of an acid)
	"Uh, and this…wow, um, H_2SO_4…oh no this is definitely wrong, haha." (student tried to draw the structure of H_2SO_4 by putting S in the center, which is then single bonded to 3 oxygens and 1 hydrogen, but then decided to scratch it off completely)
	(concerning BH_3) "Um, um, hydroboric acid? Or, yeah, hydroboric acid...um...probably boric acid…"(student scratched off "hydro" from "hydroboric acid")
1.2 Wrong reaction, expected product, mechanism, or explanation	(student drew "Fe^{3+} + MgO → Mg^{2+} + Fe_2O_3" on paper to demonstrate her belief that Lewis acids are also reductants and Lewis bases are also oxidants)
	(student wrote "HCl + CH_3COOH → neutral conjugated base + weak conjugated base" on paper)
1.3 Incorrect/incomplete definition of acid/base	"Bronsted means the other I think, that accepts OH^- I think…something like that."
	"…an acid is, a, substance that has a pH lower than 7."
1.4 Incorrect example of acid/base	"If they're like, 'go grab a weak base', like you would go get some, some NH_4."
	"Um…um…I guess a cation would make a good base."
1.5 Incorrect/incomplete explanation of why something is an acid or a base	(student gave "lemon juice" as an example of an acid, and was asked to explain why lemon juice is an acid) "Yeah, cuz everybody knows that lemon juice cleans things."
	(student drew a scale of pK_a values with NH_3 at 35) "Clearly over here (point to NH_3 at 35) I would call it a base."
1.6 Irrelevant Misconceptions	"…size (of an atom) increases going down and across (the periodic table)."
	"…so it's either tin hydroxide or tin oxide...haha…if it would have went from tin chloride to like tin hydroxide, then there would have been a reduction."
1.7 Incorrect/incomplete explanation of acid/base related terms (acidity/basicity, pKa/pKb, neutralization, titration, electronegativity, etc)	(concerning pK_a) "Um, well it's just a, it's just taking the equilibrium constant of an acid and putting it into a different scale, using the negative log."
	"Well I know the higher the pKa value, the stronger acid it is…"
	"Cuz the stronger the acid, the weaker the base, so you can look at, which has the, um…which is gonna be the stronger acid, so that would be your weaker base."

2 Correct examples of acids/bases	"Lemon juice, hydrofluoric acid…"
	"Bleach…sodium hydroxide…"
3 Correct ideas of acids/bases	"I would just say acid in water can split up into two ions basically, one of them being a proton, and the other whatever counter ion of that acid."
	(concerning acid) "…the fact that it donates proton...and accept an electron."
	"Acid is something that releases a proton, and base is something that will take up a proton."
	(concerning H_2SO_4) "...oxygens are very electronegative. And they, uh, pulling all the electron density towards this side...so, the bond between oxygen and the proton gets weaker, it breaks away, and then the proton gets released as H^+."
4 Correct acid/base strength comparison and/or reasoning	"HI would be more acidic (than HCl)...Um, I mean, one of the things that you hear is that, like the iodine, um, the iodide anion in this case, would be more polarizable, in solution, um…I mean, it'd be, it'd be like more stable anion, so I think therefore, this would be more likely to dissociate."
	"I mean if you have something like this (wrote "sp^2"), this would be more acidic than, let's just say, this (wrote "sp^3")...I mean one of the things that I've just learned along, is in organic chemistry, is that, um, carbons with more, um, s-character, um, hydrogens can act generally more acidic. Um, which is like an acetylene…those hydrogens are actually quite acidic."
5 Incorrect acid/base strength comparison and/or reasoning	"Actually, HI is strong, HBr, strong, HCl, strong, HF is weaker, because the higher up you go, the electronegativity...If I recall right, it's the electronegativity of the F, the fluorine. The fluorine is more electro...like one of the most electronegative ions there is. And, so, I know that has, that's the reason why it's more, it's less of a strong acid than the other ones is."
	(student explaining why HI is a weaker acid than HF, HCl, and HBr) "…it's something to do with the size of the bond and the bond strength. Um, this one (HI) seems to have like, um, a, a longer bond, or, the bond between it is really strong, so it causes, it makes it difficult for the hydrogen to leave, the, um, leave the halogen, so it makes it pretty weak acid."
6 Strategies to identify acid/base	"Oh it's easy (to identify a base), you know, like, uh, I look at the structures, and then see which one has OH group."
	"Well, it's kind of, OH in there, it's not an acid, it's a base. And then, if it's, if it's a hydrogen with a, bonded to a, to a halogen, like HI, HF, then those are all strong (acids)."

	"Well, I mean, I try to think of like how they would break down in solution. Um, and what, you know, when they dissociate, what they would look like. Um, and so, I guess, one of the very like, general things I was taught like a long time ago in general chemistry, and this is, this doesn't even always hold, but um, if there is, um, a hydrogen in front, and everything behind it, um, it's generally an acid. Um, if there's an OH, um, generally behind, it's usually a base."
7 Correct identification of acid/base and correct reasoning in Part II	"…hydrosulfuric acid, I see protons that'll dissociate. Um, lithium hydroxide, base, I see, um, hydroxide ion."
	"So…CH_3…this (acetic acid) is acid, weak acid. This is, this functional group, COOH, this is an acid group. It's an organic acid functionality, so, this breaks, because of the electrons can be delocalized between these two (oxygens)…so electron density will pull into this (oxygen)…"
	"Alright, um, I kind of used this one (pentane-2,4-dione) as one of my examples, um, these hydrogens right here are extremely acidic. You have two carbonyls, so two electron-withdrawing groups. Any kind of base, weak base, strong base like I just described, anything with a, um, you know, lone pair of electrons is gonna be able to take at least one of these hydrogens off. So, these hydrogens are definitely acidic, no questions about that."
8 Incorrect identification of acid/base and correct reasoning in Part II	"OK, I would say that (BF_3) is neutral. I don't see any group on there that would make it acidic or basic."
	"PH_3 have…hydrogens and lone pair, so…um…I'm not exactly sure how to solve this one…so, since it has hydrogens, I'm going to, actually, not only because it has hydrogens, but because if…if a hydrogen was taken away, and it gave its electrons to phosphorus, you would…um…you would get…PH_2, and that would have a -1 charge, as supposed to losing two electrons, that would give it a +2 charge, so, I'm going to say that PH_2^- is more stable, cuz it only has, uh, one, uh, negative, as oppose to two positives."

The initial interviews revealed problems with students' understandings of different acid base models and the related concepts. Most students had problem identifying acids and bases correctly, and/or predict their properties in particular reactions. Several students continuously used specific atoms and/or functional groups as the only means of identifying acids and bases. For example, one student recognized the chemical formula H_2SO_4 (structure c, see Appendix A for a list of structures) as sulfuric acid, but a few minutes later pointed at the Lewis structure of sulfuric acid (structure n) and concluded that it is a base because "it's got OH groups". This echoes with other research on how college students determine acid base strength (McClary & Talanquer 2011a&b): it is not uncommon to see students rely on specific atoms and/or functional groups as the only means to identify acids and bases and/or to determine acid base strengths. Two other common difficulties revealed during the initial interviews are:

- The identification of BF_3 as Lewis acids.

  Out of the eight students, only two (both are graduate students) correctly identified BF_3 as a Lewis acids because boron has an empty orbital to accept an electron pair. The rest of the students either identified it as a base because of the lone pairs of fluorine, or neutral because a lack of "functional groups".

- The correct explanation of how alcohols (structure j, methanol, was given to undergraduate students while structure k, 2-butanol, was given to graduate students) can act as either an acid (by donating the hydrogen connected to oxygen) or a base (by donating the lone pairs on oxygen).

  Again only two graduate students correctly explained how alcohol can act as either an acid or a base. A common misconception among the rest of the students is that the OH group off carbon can easily dissociate in water to produce OH^-, thus making methanol a base.

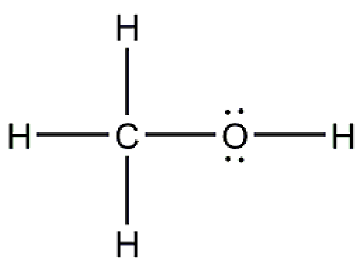
- The identification of PH_3 as Lewis bases because of its lone pairs.

Although this is a less common difficulty, two students out of six identified PH_3 as an acid because of the hydrogens. One specifically mentioned that "it also doesn't have an OH".

<u>Open-ended Questions</u>

Based on the initial interviews, five structures students commonly had difficulty identifying were selected and designed into open-ended questions, as shown below:

The Lewis structure of a compound is shown below. Is this compound a) an acid; b) a base; c) both an acid and a base; d) neither an acid nor a base? Please explain your choice in detail to receive full credit.



The other four structures chosen were PH_3, BF_3, H_2SO_4, and CH_4. Each question was worded in the same manner with the Lewis structure of the compound shown below (without naming the compound in the question).

These open-ended questions were then administered as chemistry education assessments during chemistry laboratory time to two groups of students taking Organic Chemistry I in Summer I of 2010. Chemistry education assessments were part of the general and organic chemistry laboratory assignments and counted towards students' laboratory grades, but were only graded by completion, taking the pressure of grade off the students as they complete the assessment, and thus allowing them to freely express what they really believed when answering the questions. The purposeful selection of chemistry laboratory time also separated the influence of instructor from the data collected. The selection of Organic Chemistry I students was based on the availability of classes during the summer. In Summer I session, only General Chemistry I and Organic Chemistry I are taught, and students from General Chemistry I have not learned the acid base chapter yet, so they are not suitable for the administration of these open-ended questions. In order not to overwhelm students with too many questions, these five structures were divided into two groups (the first group contains CH_3OH and PH_3, the second group contains the other 3 structures) and administered to different laboratory sections. Each group of questions was administered to two laboratory sections. All the questions were administered through Ed's Tools (http://edstools.colorado.edu), a free online tool for administering and coding open-ended questions.

Student responses were summarized by their reasoning of why a structure is an acid and/or a base, as shown in **Table 4.2** below.

Table 4.2: A numerical summary of how students categorized each compound in their open-ended responses (bolded categories are correct or reasonable)

Compound	**N**	**Acid only**	**Base only**	**Acid and Base**	**Neither acid nor base**
CH_3OH	19	5	4	**10**	0
PH_3	20	4	**5**	**5**	6
BF_3	19	**10**	9	0	1
H_2SO_4	17	**13**	4	**0**	0
CH_4	19	6	2	2	9

As shown in **Table 4.2**, out of a total of nineteen written responses, ten correctly categorized methanol as both an acid and a base, five categorized methanol as acid only, and four categorized methanol as base only. However, among the ten students who correctly categorized methanol as both an acid and a base, half of them did not offer an adequate explanation. For example, one student explained his choice as "Methanol can either be an acid or a base because it has hydrogen atoms that can be given off forming an acid, however it can also give up the OH group giving it basic characteristics". Another student who categorized methanol as only a base reasoned that "The Lewis definition of acids and bases is so handy! CH_3OH has a pK_a of 15, so it is probably best understood as being a Lewis base…" This explanation revealed that this particular student did not understand the definition of either pK_a or Lewis base. Out of these ten students, three expected methanol to act as a base by the dissociation of the OH group, another two cited the pK_a value of methanol to support that it is a base.

For phosphine, nine out of twenty students explained it correctly as either mainly a base or both an acid and a base; one categorized phosphine as a base but again incorrectly used a high pK_a value of phosphine as the reason. Six students categorized it as neither an acid nor a base, among which five explained their conclusions by the lack of polarity in the phosphine molecule. Four students categorized phosphine as an acid, mainly because of the presence of the hydrogens, with one student using a lack of OH as the reason.

For boron trifluoride, ten out of nineteen students categorized it as an acid, among which nine correctly explained it by Lewis' acid base model. Another eight students categorized it as a base, among which seven explained their conclusions by the multiple lone pairs on fluorine atoms, and the last one reasoned that boron trifluoride does not contain a hydrogen. The last student categorized boron trifluoride as neither an acid nor a base because it is nonpolar.

For sulfuric acid, most students (thirteen out of seventeen) recognized it as H_2SO_4 and concluded that it is an acid. The other four students concluded that it is a base because of the lone pairs.

For the structure of methane, nine out of nineteen students correctly categorized it as neither an acid nor a base. Six students categorized methane as an acid because of its hydrogens, and another two students who categorized methane as both an acid and a base reasoned that methane is an acid because of its hydrogens.

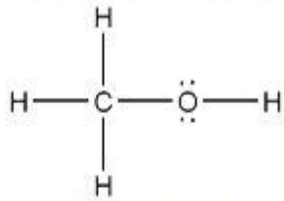
Common responses from these open-ended were developed into the Tiered Multiple-choice Questionnaire, as shown in **Figure 4.1** below. This questionnaire not only asked students to choose the best explanation for each question, but also asked them to explain why each option was correct or incorrect, allowing a better understanding of students' choices, since some students might choose the correct answer for a wrong reason.

Tiered Multiple-choice Questionnaires

The five structures examined in 4.2 were designed into tiered multiple-choice questions (**Figure 4.1**) and then separated into two questionnaires to reduce the amount of time required for completion. These two questionnaires were printed out and administered as chemistry laboratory assessments to laboratory sections of both General Chemistry II and Organic Chemistry II in Summer II, 2010, to see if students' categorization of acids and bases would change as they take organic chemistry courses. All groups of students were asked to first choose a correct statement and then to explain why it was correct and why the other choices were wrong. For the general chemistry laboratories, the structure PH_3 was replaced by NH_3 (without changing the wording of the questions and each option) because of the concern that the structure PH_3 might be too difficult to general chemistry students.

For each question, circle the best answer AND explain why you did or did not choose each answer.

1. The structure of methanol (CH_3OH) suggests that methanol is…? (6 points)



A. Both an acid and a base, because it can either donate or accept a proton.

____
____
____
____
____
____

B. Both an acid and a base, because it can dissociate in water and produce both H^+ and OH^-.

____
____
____
____
____
____

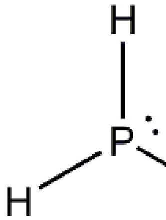
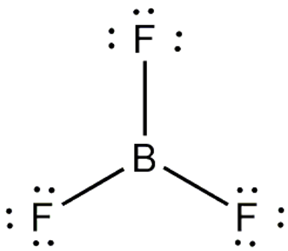
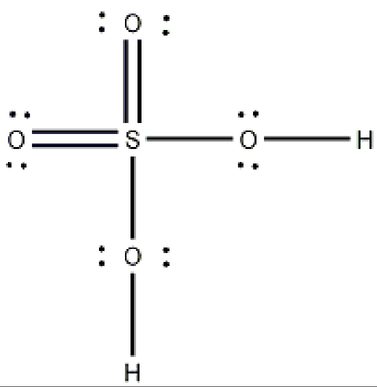
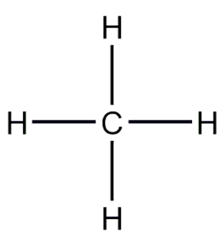
C. A base because only the OH group will dissociate in water.

____
____
____
____
____
____

D. An acid because it has four protons to donate.

____
____
____
____
____
____

(a)

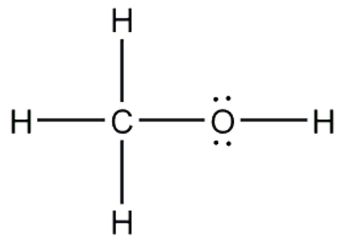
Structure	Options (with correct answer bolded)
	A. Not a base because it does not contain an OH group. B. An acid because it has three protons to donate. **C. A base because it can donate the lone pair electrons on P.** D. Neither an acid nor a base, because it is a non-polar molecule and does not dissolve in water.
	A. A base because it can lose one or more lone pairs on the fluorine atoms. **B. An acid because boron can accept electrons.** C. Either an acid or a base, because it can donate or accept electrons. D. Neither an acid nor a base, because it can neither donate nor accept a proton.
	**A. An acid because it can donate one or two protons.** B. A base because it will dissolve in water to produce OH^-. C. Either an acid or a base, because it can dissociate in water to produce either H^+ or OH^-. D. Both an acid and a base, because it can either accept or donate a pair of electrons.
	A. Both an acid and a base, because it can either donate or accept a proton. B. An acid because it has four protons to donate. C. Not a base because it does not contain an OH group. **D. Neither an acid nor a base, because it does not have lone pairs to donate, nor place to accept any lone pairs.**

(b)

Figure 4.1: The tiered multiple-choice questionnaire. a) Sample question on methanol. b) The rest of the questions.

Student choices were totaled and the percentages were calculated in the figures below. Percentages were used rather than raw numbers because the number of students in each group is slightly different (range from 25 to 39). Then each student's explanation of why a choice is correct or incorrect is also summarized. No statistical tests were performed on these data because of the small sample sizes.

For the first question, “The structure of methanol (CH_3OH) suggests that methanol is…?”



A. Both an acid and a base, because it can either donate or accept a proton.

B. Both an acid and a base, because it can dissociate in water and produce both H^+ and OH^-.

C. A base because only the OH group will dissociate in water.

D. An acid because it has four protons to donate.

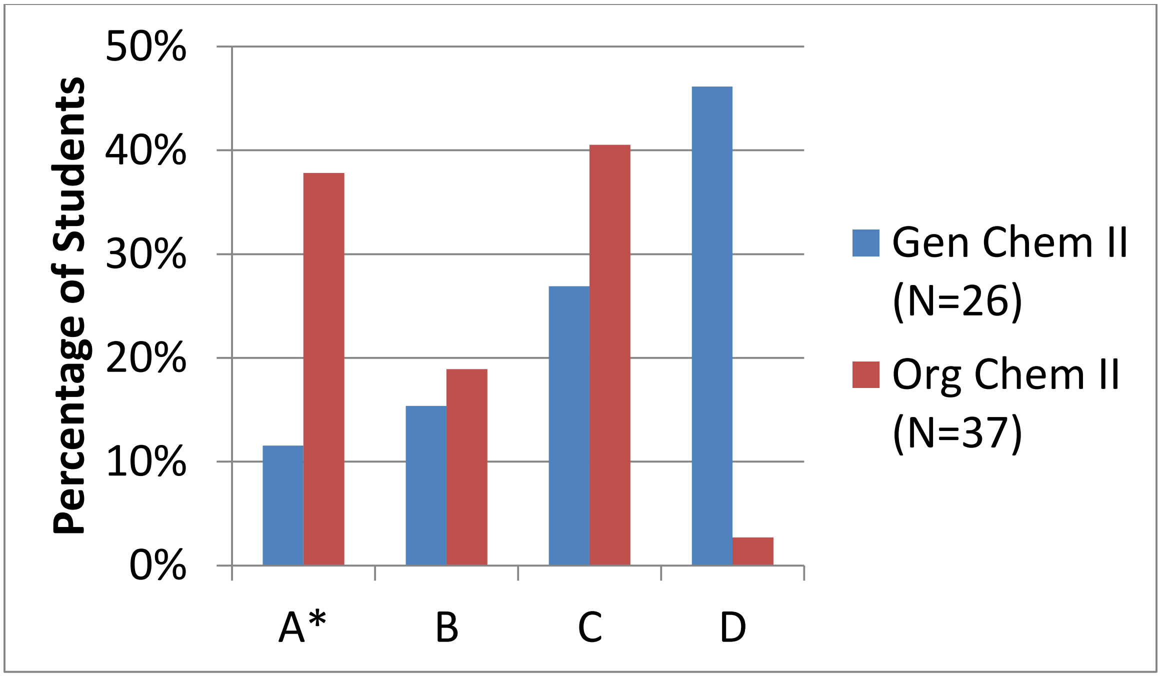
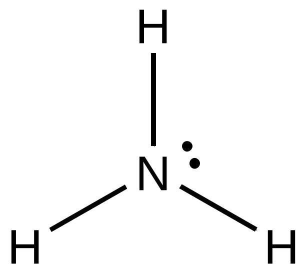


Figure 4.2: Student responses from the multiple-choice question on methanol.

As shown in **Figure 4.2**, a larger percentage of Organic Chemistry II students chose the correct answer A, but mainly because the majority of them have overcome the misconception that the C-H hydrogens in methanol can all be donated as protons. There was actually a slightly higher percentage of Organic Chem II students who thought methanol would act as a base by dissociating the OH group in water (B & C). Regardless, over half of the students in either group believed that methanol can act as a base by dissociating the OH group in water. Even among the seventeen students (fourteen from Organic Chemistry II and three from General Chemistry II) who chose A, only nine of them offered an adequate explanation for their choice; the other eight students either did not explain their choice, or had an obviously wrong explanation, such as "Lone pairs on O can act as base, all 4 H's can dissociate", "CH_3OH can either release H^+ or OH^-"

(essentially agreeing with B), “Losing OH^- group changes CH_3OH from a base to an acid”, “Water is in the compound, so it could be either an acid or a base”.

For the second question, “The structure of ammonia (NH_3) (or phosphine, PH_3 for the Organic Chemistry II students) suggests that ammonia is…?”



A. Not a base because it does not contain an OH group.

B. An acid because it has three H’s that will dissociate in water.

C. A base because it can donate the lone pair electrons on N.

D. Neither an acid nor a base, because it is a non-polar molecule and does not dissolve in water.

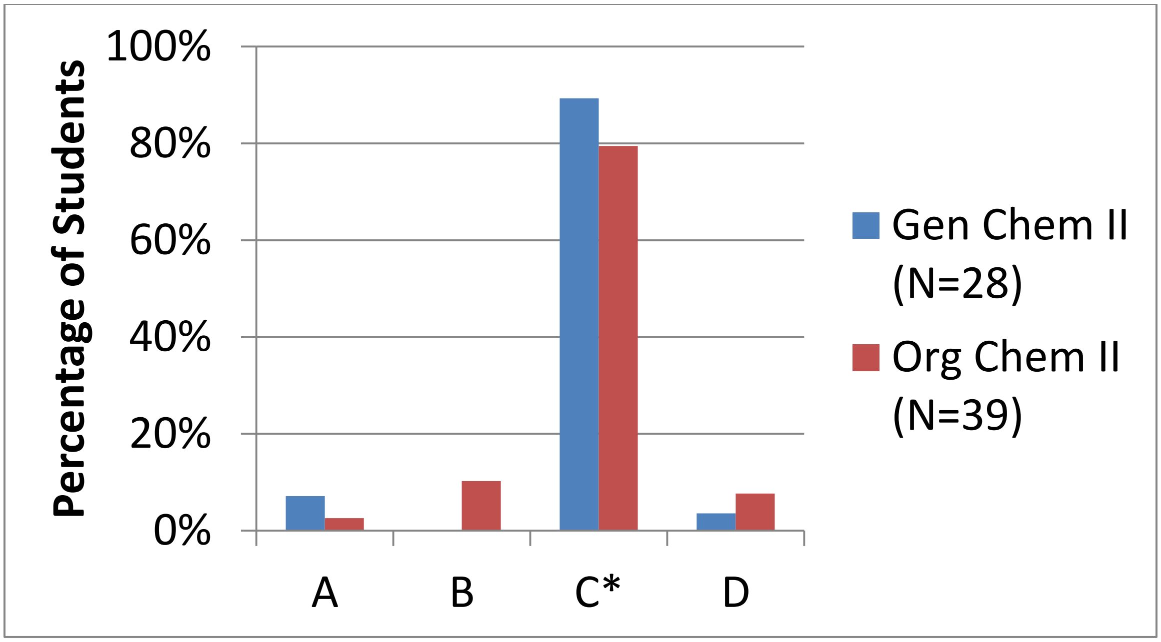
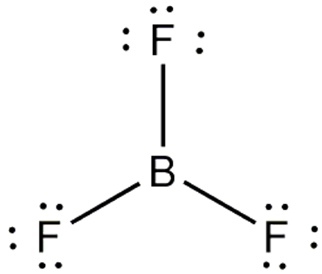


Figure 4.3: Student responses from the multiple-choice question on ammonia/phosphine.

As shown in **Figure 4.3**, a majority of both the General Chemistry II and Organic Chemistry II students correctly identified ammonia (or phosphine) as a base because of the lone pair on nitrogen (phosphorous).

For the third question, "The structure of boron trifluoride (BF^3) suggests that boron trifluoride is…?"



A. A base because it can lose one or more lone pairs on the fluorine atoms.

B. An acid because boron can accept electrons.

C. Either an acid or a base, because it can donate or accept electrons.

D. Neither an acid nor a base, because it can neither donate nor accept a proton.

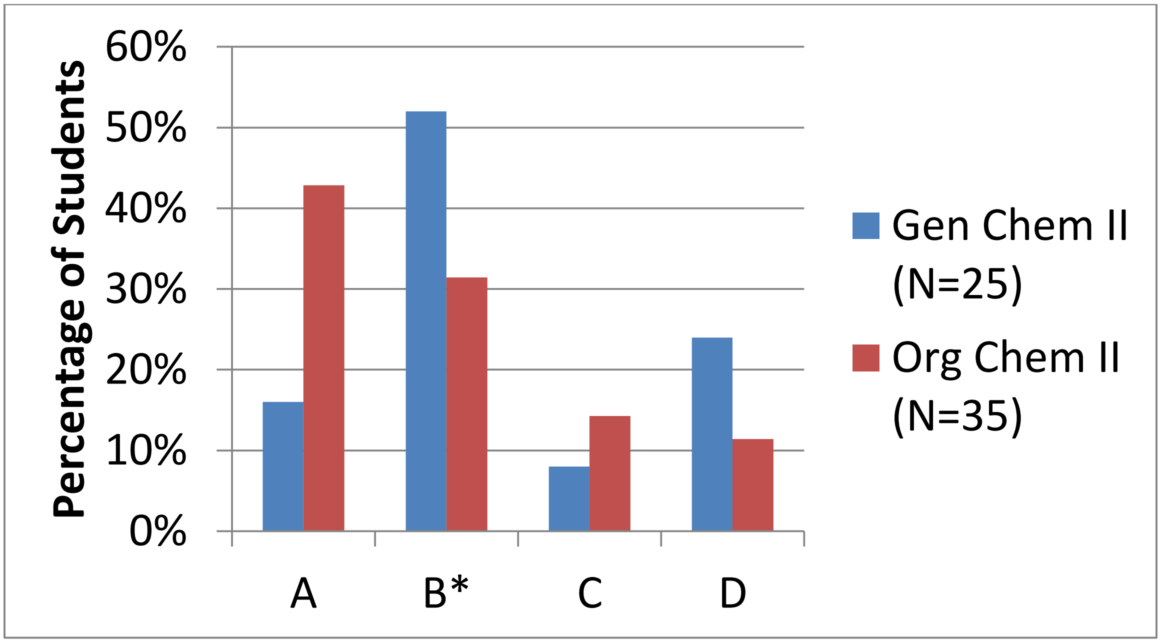
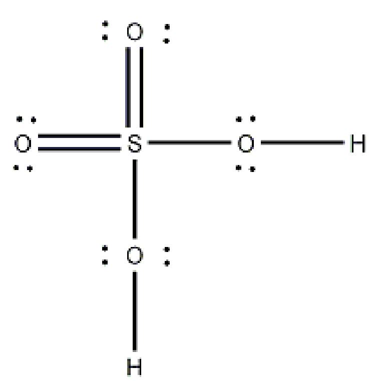


Figure 4.4: Student responses from the multiple-choice question on boron trifluoride.

As shown in **Figure 4.4**, the General Chemistry II students seemed to have a higher success rate on this question than the Organic Chemistry II students. Among the General Chemistry II students, thirteen out of twenty-five chose the correct answer B, among which eleven correctly explained their choice by the empty orbital boron has. However, although only four out of twenty-five students agreed with statement A ("BF_3 is a base because it can lose one or more lone pairs on the fluorine atoms"), most of the remaining students did not correctly explain why fluorines are not electron donors. Among the twenty-one students who did not choose A, the most popular reason was because "the fluorines are stable with a full octet, so they will not want to lose electrons" (eight out of twenty-one), followed by the correct explanation, "fluorine is highly electronegative and will not give up its lone pairs" (six out of twenty-one). The other seven students either did not explain why they thought A was wrong, or offered an inadequate explanation, such as "Fluorine can either lose or gain electrons, so it can be either an acid or a base", "This cannot be a base because there is no OH", "A base needs a H^+". On the other hand, a higher percentage of Organic Chemistry II students categorized BF_3 as Lewis base because of the lone pairs on fluorine atoms. Out of the fifteen students who chose A, thirteen of them clearly stated that fluorine atoms have multiple lone pairs they can donate. Even among the eleven students who chose the correct answer B, only eight of them explained it as boron has an empty orbital to accept electrons; the other three explanations were not adequate, such as "BF_3 is an acid because it can donate electrons", "BF_3 acts with water", "Boron is stable with an octet of 6 electrons".

For the fourth question, "Judging from the Lewis structure of the compound below, this compound is most likely to act as…?"



A. An acid because it can donate one or two protons.

B. A base because it will dissolve in water to produce OH-.

C. Either an acid or a base, because it can dissociate in water to produce either H+ or OH-.

D. Both an acid and a base, because it can either accept or donate a pair of electrons.

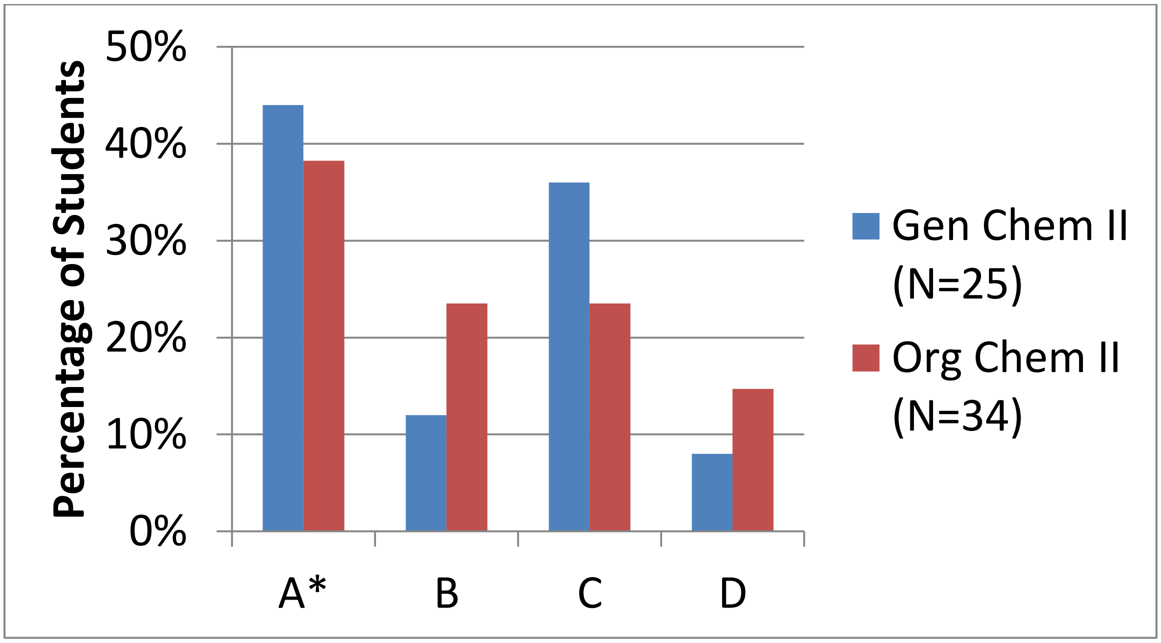
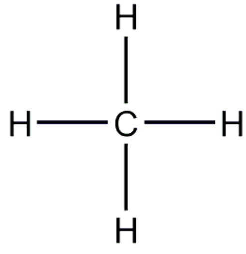


Figure 4.5: Student responses from the multiple-choice question on sulfuric acid.

As shown in **Figure 4.5**, approximately half of the General Chemistry II students chose the correct answer for this question. However, out of those eleven students who chose A, only six offered an adequate explanation of why sulfuric acid is a good proton donor; the other five recognized the structure as H_2SO_4 and thus chose A. Still about half of the students (twelve out of twenty-five) chose B or C, making a similar assumption as in the previous methanol that a compound with an OH group can always dissociate into OH^- in water and thus acting as a base. Out of those twelve students, eight specifically agreed in their explanations that the OH group can dissociate into OH^-, and another two thought both OH groups would dissociate into OH^-. Organic Chemistry II students' choices and answers were similar – thirteen out of thirty-four chose A, but only eight explained why the hydrogen off oxygen can dissociate, the other five simply recognized the structure as H_2SO_4 and chose A because they recognized sulfuric acid. Another three students out of the five that chose D also offered an adequate explanation – this compound can act as an acid by donating a proton or donating the lone pair on oxygen. Since this question did not specifically ask whether this compound is more likely going to act as an acid or a base, these three explanations were considered acceptable. However, there were still a total of sixteen students out of thirty-four that chose B or C, among which seven agreed in their explanations that this compound can dissociate into both H^+ and OH^- in water, and another seven reasoned that this compound would dissociate into either H^+ or OH^- in water, depending on the pH of the solution, but not both H^+ and OH^- at the same time. The results from this question and the previous one on methanol seem to suggest that a number of students believe any compound containing an OH group could act as a base by dissociating the OH group into OH^- in water, regardless of what the OH group was connected to.

For the fifth question, "The structure of methane (CH_4) suggests that methane is…?"



A. Both an acid and a base, because it can either donate or accept a proton.

B. An acid because it has four protons to donate.

C. Not a base because it does not contain an OH group.

D. Neither an acid nor a base, because it does not have lone pairs to donate, nor place to accept any lone pairs.

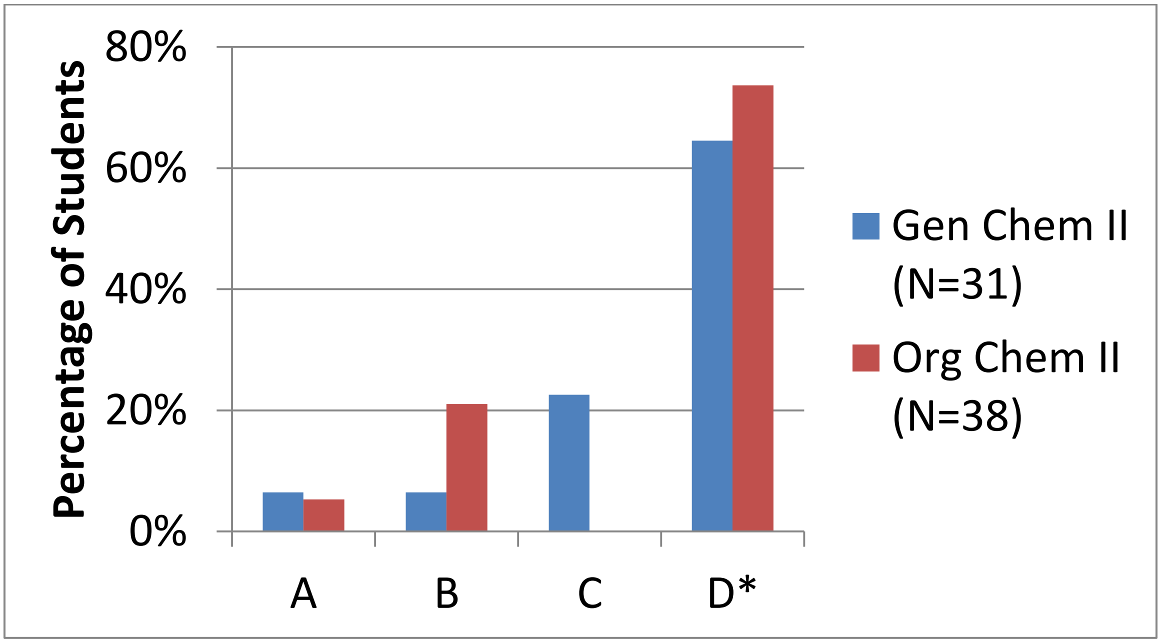


Figure 4.6: Student responses from the multiple-choice question on methane.

As shown in **Figure 4.6**, the majority of both the General Chemistry II and Organic Chemistry II students chose the correct answer. However, out of the twenty General Chemistry II students who chose D, only five offered an adequate explanation why methane is neither an acid nor a base; eight arrived at their conclusions because methane is nonpolar, thus it cannot be an acid or a base; another four reasoned that methane is stable with a full octet. Out of the seven students who chose C, five clearly agreed in their explanations that a compound cannot be a base without an OH group. Among the twenty-eight Organic Chemistry II students who chose D, only seven clearly explained why methane is neither an acid nor a base; eight reasoned specifically that methane is stable with a full octet and thus is neither an acid nor a base, another eight also used stability as their reasoning but did not specifically attribute the stability of methane to a full octet. Distractor C did not present a problem for Organic Chemistry II students, as it did to the seven General Chemistry II students. None of the Organic Chemistry II students chose C, and twenty-nine students pointed out that bases do not necessarily contain OH groups in their explanations of why C is wrong.

Although these tiered multiple-choice questions revealed more of what concepts students understand and what concepts students still struggle with, its limitation also became more obvious – most students, although given plenty of space, would only explain their choices briefly. Although the multiple-choice part offers quantitative data, students' further explanations were not very helpful in confirming their understandings because most students do not articulate their explanations. This lead to another research described in the next chapter, with an initial focus of teaching students how to articulate their explanations.

CHAPTER FIVE

BESOCRATIC ACTIVITIES

BeSocratic is a web-based software developed by a collaboration of a number of people in different disciplines to provide intelligent feedback and tutorials to students, with the purpose of fostering meaningful learning (NSF funding # 1122472).

One key feature of the BeSocratic system lies in its ability to record all student responses, so that researchers can review and/or analyze these responses later. Even the part of a response that a student initially typed/drew and then deleted will be recorded completely, so that researchers can later see that this student initially typed an answer, then deleted it and typed a new answer. Furthermore, more and more features are being developed in the BeSocratic system as a current project in the Cooper research group.

Argumentation Training

The initial BeSocratic activity was designed in order to promote students' argumentation skills. As explained in the theoretical frameworks, open-ended questions seemed to be the approach for studying students' different acid base mental models. But many open-ended responses were not as informative as we would like because students were not trained to articulate their reasoning. Thus, the initial thought was to design an activity that could promote and retain their argumentations skills. Appendix B contains screenshots of different steps in the BeSocratic activity. During the activity, students were first asked to determine the stronger acid between ammonia and water, and explain why. Their initial responses were recorded by the system (this response will be referred to as “Q1-pre” in data analysis). Then they were introduced to the different components of a complete scientific explanation one by one: the claim, the data, and the explanation. After the introduction, they were reminded of all these components again on a single screen. Then a biology example was introduced and students were asked to identify the different components of a complete scientific explanation in that given example. A biology example was chosen to ensure that students understood the different components before moving forward, without influencing students’ thoughts about acid base chemistry with the example. Then students were shown the initial question about ammonia and water, and asked to identify each component step by step. After students had identified the components step by step, they were presented with their initial response to the ammonia water question, and given the opportunity to make changes to that response (this response will be referred to as “Q1-post” in data analysis). Finally, a different question was asked (which one is the stronger base between methanol and methylamine) and students’ responses were collected (this response will be referred to as “Q2” in data analysis). This question was given immediately after students had revised their answers to the first question, but students were not reminded of the components of a complete scientific explanation, nor guided step by step to compose their answers.

This BeSocratic activity was first administered to General Chemistry II students from both the traditional and the CLUE curricula in Spring 2012. For students in the traditional curriculum, this activity was administered during chemistry laboratory time. As explained above, chemistry education assessments were part of the general chemistry laboratory assignments and counted towards students' laboratory grades, but were only graded by completion. The purposeful selection of chemistry laboratory time separated the influence of instructor from the data collected, and allowed a random sampling of students from multiple general chemistry instructors in the traditional curriculum. The same BeSocratic activity was assigned to students from the CLUE curriculum as homework assignments graded by completion, because those students were separated into many general chemistry laboratories, making it impossible to administer the BeSocratic activity to the CLUE cohort in the same way it was administered to the Traditional cohort. However, all student-completed assessments were only graded for completion, taking the pressure of grade off the students as they complete the assessment, and thus allowing them to freely express what they really believed when answering the questions. The same activity was administered again in Spring 2013 to ensure the reproducibility of data. **Table 5.1** below summarized the sizes of the different cohorts from different years who participated in this BeSocratic activity.

Table 5.1: Summary of the cohorts participated in the BeSocratic activity.

Year	Curriculum	Number of Students
Spring 2012 (General Chemistry II)	Traditional	70
	CLUE	107
Spring 2013 (General Chemistry II)	Traditional	91
	CLUE	115

Because of the richness of the data collected in these BeSocratic activities across two years, the analysis of data below will be divided into several parts.

First of all, the BeSocratic activity was initially designed to promote argumentation skills among the students, and train them how to articulate their reasoning in open-ended questions. Initially, students' responses in Q1-pre (before instruction) and Q1-post (after instruction) were compared in Microsoft Word to see if they edited their responses after being instructed on the different components of a complete scientific explanation. **Table 5.2** below shows the first ten responses from the CLUE cohort and the first ten responses from the Traditional cohort, in Spring 2012.

Table 5.2: Selective student responses to the first question before and after instruction, "Which is the stronger acid between water and ammonia?"

CLUE 1	Ammonia. It has a single lone pair~~.~~, which allows the molecule to dissociate more easily.
CLUE 2	H20 is a stronger acid because in this reaction Ammonia is a weak base. ammonia is a base because it will except a hydrogen from the water and it will donate its electrons to the water.pdons
CLUE 3	I assume that HN3 is a stronger base by looking at the Kb value
CLUE 4	nh3 is the acid because it is more likely to give up a hydrogen... wrong h20 is the strong acid in the readction because it goes through and gives up a hydrogen therefore it is the strogn acid
CLUE 5	H2O because it has more lone pair
CLUE 6	The H2O would be the stronger ~~base~~acid because NH3 would actually be considered a base. ~~You can also say that the~~The more stable compound would be more basic because the less stable (ie. more lone pairs) the compound, the more it will seek free Hydrogens. With H2O having more lone pairs, it would be able to bind with more available Hydrogens that would make the resulting compound acidic.
CLUE 7	NH3 is the stronger acid, as it is the electron pair reciever and water is the electron pair donator. The reason for this is that oxygen is more electronegative than nitrogen, causing the hydrogen to be attracted to the oxygen to create a more stable system.
CLUE 8	H20 is the stronger acid. H2O accepts electrons and has a stronger conjugate base, making it the stronger acid. According to Lewis acid theory, acids are electron pair acceptors and ~~H20 accepts the electron and~~since H2O has an empty orbital, it is able to accepts those electrons. It then creates -OH, a more stable conjugate base than ammonium would~~.~~ ~~Since~~ and acid strength and conjugate base strength are inversely ~~proportionate, H20 is the stronger acid.~~proportional.
CLUE 9	Water because it will be more willing to accept and electron pair to form a bond.

CLUE 10	Water is a stronger acid because it is more likely to accept an electron and lose a hydrogen than the ammonia. Because of this, it dissociates more fully in water and is a stronger acid.
Traditional 1	NH3 because it has more hydrogen ions to donate. Lewis acids are based on the hydrogen content of an acid and the more hydrogen ions that are in the compound the better it can donate them in solution. The more hydrogen ions in solution the more acidic a solution will be.
Traditional 2	The stronger acid would be H2O because it has more lone pairs, which allows the molecule to accept more H+ ions.
Traditional 3	~~H2O is the stronger acid because O has 2 free loan pairs of electrons.~~NH3 is the stronger acid because H2O is more polar than NH3. The higher polarity in H2O will cause the bonds between the Hydrogens and the oxygen to be stronger than the bond between hydrogen and nitrogen. Because of this, it is easier for NH3 to donate an H+
Traditional 4	NH3 because it has an extra hydrogen and as far as I know H2O is not acidic.
Traditional 5	Ammonia is a stronger acid because it has a single lone pair. Also, water has a neutral pH and can act as either an acid or a base, so it is not strong either way. A single lone pair of electrons signifies the presence of a bronsted-lowry acid.
Traditional 6	~~ammonia is the stronger acid because it only has one pair of unbonded atoms.~~I would change it to the answer i wrote in previous problems stateing the claim the evidence and the data.
Traditional 7	~~NH3~~ The ammonia molecule would be the stronger acid ~~based on~~. The Bronsted-Lowry theory claims that an acid is a molecule that will accept an H+ ion. From the ~~number~~recent studies in my chemistry class, I know that water will donate one of ~~hydrogen bonds~~ its H+ ions, making NH3 the ~~structure has~~acid.
Traditional 8	NH3 is a stronger structure because it is less electronegative which means it will be a stronger acid.
Traditional 9	Water is the stronger acid because ~~as~~oxygen is more electronegative than nitrogen. As electronegativity increases, acid strength increases. ~~Since oxygen is more electronegative than nitrogen, the acid containing oxygen would be more stronger.~~
Traditional 10	water is a stronger acid because water may react with water to make hydroium or hydroxide ions whereas ammonia can only react with water to make ammonium ion and hydroxide which generates a more basic solution than water.

As shown in **Table 5.2**, most students edited their responses after the instruction (red underline text shows what they added to their initial responses and red strikethrough text shows what they deleted from their initial responses). But are those students making their explanations more complete or simply adding more words to their explanations? To answer this question, a coding scheme must be developed to categorize the levels of students' explanations. Two graduate students who designed different BeSocratic activities based on Toulmin's Argumentation Pattern to help students articulate their reasoning in different areas of chemistry together came up with the initial coding scheme, based on Toulmin's Argumentation Pattern. The initial coding scheme included four levels:

Level 0: No claim. Student did not even make a claim as to which one is a stronger acid.

Level 1: Claim only. Student did make a claim but did not offer any explanation to support the claim.

Level 2: Claim and Data. Student not only made a claim but also supported the claim with at least one piece of data, or multiple pieces of data that were not supporting or explaining each other.

Level 3: Claim, Data, and Explanation. Student made a claim, supported the claim with at least one piece of data, and offered further explanation for at least one piece of data to support the data.

It is very important to distinguish between an incomplete response with a claim and multiple data and a complete response with claim, data, and explanation. A student might offer several reasons (data) that were not connected to each other in order to support the claim; and if one reason did not explain another, the "explanation" component was still considered missing. **Table 5.3** below shows some sample responses from students and how they were coded.

Table 5.3: Sample responses from Spring 2012 students and explanations of how they were coded.

St. #	Response (Q1-pre)[1]	Code[2]	Explanation
1723	Oxygen has a greater electronegativity than nitrogen, so the non-bonded electron pair on the nitrogen atom is more available for sharing than the non-bonded electron pair on the oxygen atom.	Level 0	This student did not make a claim as to whether ammonia or water is the stronger acid.
2680	H_2O	Level 1	This student made a claim that water is the stronger acid, but did not explain why.
1554	Ammonia. It has a single lone pair.	Level 2	This student made a claim and supported it with one piece of data (although both are wrong[2]).
2693	Water is a stronger acid because ammonia is a weak base. Oxygen is more electronegative than nitrogen.	Level 2	This student made a claim and supported it with two pieces of data: "ammonia is a weak base" and "oxygen is more electronegative than nitrogen". The fact that ammonia is a weak base does not explain, and is not explained by, the fact that oxygen is more electronegative than nitrogen. Thus, both pieces were coded as data, and the entire argument was coded as Level 2, although multiple pieces of data were listed, there was no explanation to support either piece of data.
1700	Water is the stronger acid because it can more easily donate a proton. It can do this because it is more electronegative than nitrogen and therefore can more easily hold the negative charge.	Level 3	This is an example of a complete argument. The fact that water can donate a proton easier is further explained by electronegativity of water, which resulted in a more stable conjugate base.

Notes:

1. All responses were students' initial responses before instruction on complete scientific explanations.

2. At this point, only the completeness of each response was coded; the correctness of each response was not taken into consideration (but will be analyzed and presented later).

Once this coding scheme was agreed upon and finalized, it was used to code all the data from both Spring 2012 and Spring 2013. After coding, it appeared that very few students fell into the first two levels (Level 0 and Level 1). So the first three levels were combined to simplify the statistical comparison of two cohorts and data presentation. The two final categories remained are:

Level 0-2: Incomplete explanation;

Level 3: Complete explanation.

Figure 5.1 below presents the percentage of complete student responses in Q1-pre (before instruction), Q1-post (after instruction), and Q2, from both years and both cohorts.

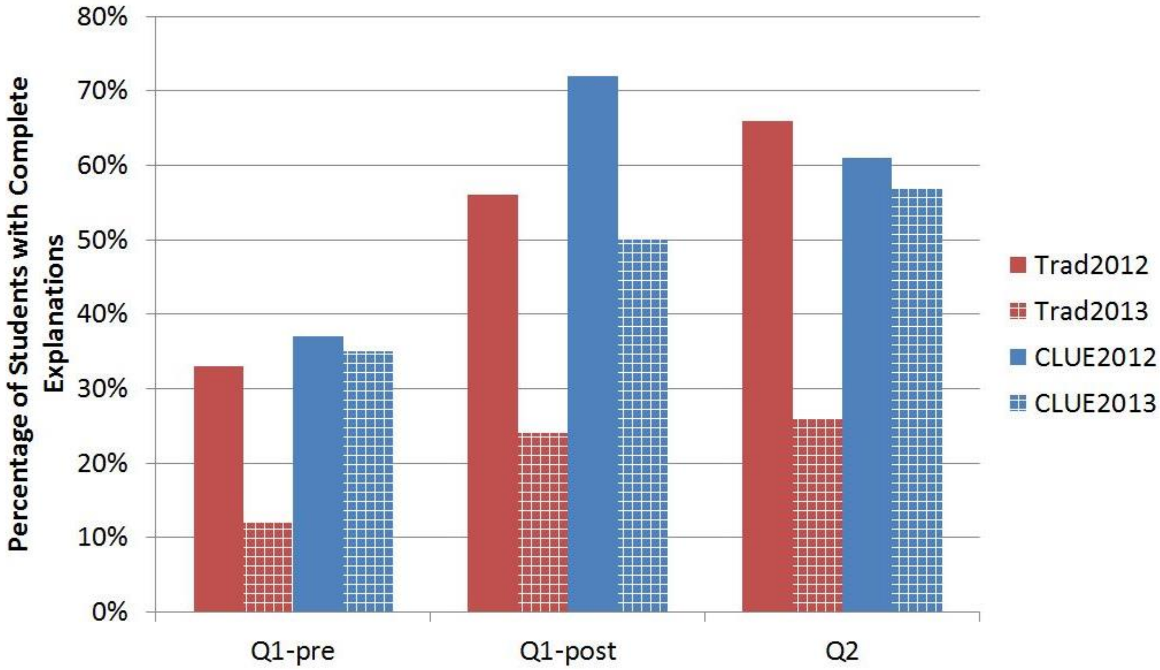


Figure 5.1: The percentage of student responses that were complete, during the initial question (before the instruction), after editing the initial response, and during the second question; for a total of four cohorts across two years and two different general chemistry curricula.

For Spring 2012, both the Traditional and the CLUE cohorts demonstrated very similar trends. First of all, during their initial response, most students only supported their claims with data, but did not explain why their data can lead to the claim (such as student 1554 who answered “Ammonia. It has a single lone pair.”). After instruction on how to make a complete scientific explanation, most students edited their responses and a considerable amount of the edited responses moved from “Incomplete” (Level 0-2) to “Complete” (Level 3). Moreover, students retained the argumentation skills for the immediate question that follows, even though no hint or guidance was given in the second question. Both cohorts demonstrated very similar trends at each stage (Q1-pre, Q1-post, and Q2), suggesting that the two different curricula did not affect students' argumentation skills differently. Both cohorts needed instruction on how to articulate argumentations; and the quick instruction embedded in the BeSocratic activity was successful at least for a short amount of time.

McNemar's Chi-square test was performed on both cohorts, comparing if there is a difference between students' responses to the first question before and after instruction on scientific argumentation (comparing Q1-pre to Q1-post), and if the instruction has an immediate lasting effect (comparing Q1-post to Q2). McNemar's Chi-square test was chosen because data was categorical (treating incomplete explanations as "0" and complete explanations as "1") and the samples are dependent (pre-post testing rather than comparing two groups). Phi effect sizes were calculated from Chi-square values according to the following equation:

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

In which N is the total sample size (70 x 2 for the Traditional 2012 cohort, and 107 x 2 for the CLUE 2012 cohort). Phi effect size was chosen because data was categorical; it is a simplified situation for Cramer's V:

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Cramer's V is the effect size used for all categorical data, calculated from Chi-square values, in which k is the less of the number of rows and the number of columns for the categorical data. In this case, the table only has two rows and two columns (pre/post, 0/1), so k = 2, and k-1 = 1. So Cramer's V is simplified into Phi. For both Cramer's V and Phi effect size, the conventional standard is ~0.1 means small effect size, ~0.3 means medium effect size, and ~0.5 means large effect size.

Table 5.4: McNemar's Chi-square tests for Traditional 2012 cohort and CLUE 2012 cohort.

Cohort	Comparison	McNemar's Chi-square	p-value	Phi
Traditional 2012	Q1-pre Vs Q1-post	19.05	**<0.001**	**0.369**
	Q1-post Vs Q2	0.346	0.556	0.050
CLUE 2012	Q1-pre Vs Q1-post	31.03	**<0.001**	**0.404**
	Q1-post Vs Q2	3.704	0.054	0.140

As shown in **Table 5.4**, both Traditional 2012 cohort and CLUE 2012 cohort improved significantly from Q1-pre (their initial responses) to Q1-post (their edited responses after the instruction on scientific argumentation), both with a p-value of less than 0.001, and medium to large effect sizes (0.369 for the Traditional 2012 cohort and 0.404 for the CLUE 2012 cohort). Both cohorts also retained the argumentation skills on the second question immediately after they finished editing their first responses: comparing Q2 to Q1-post, neither cohort had a significant difference (p-values are larger than 0.05).

Chi-square tests without Yates' correction were also performed between the Traditional 2012 cohort and the CLUE 2012 cohort at all three stages (Q1-pre, Q1-post, and Q2) to see if two cohorts are significantly different at either stage. Again Chi-square test was chosen because data was categorical, but this time McNemar's test was not chosen because the comparison is between two independent samples. Two other similar tests for categorical data from independent samples are Fisher's exact test, and Chi-square test with Yate's correction, both of which are more suitable for small sample sizes. With the large sample sizes this research concerns, Chi-square tests without Yates' correction would be sufficient. Phi effect sizes were calculated according to the same equation shown above (N = 70 + 107 = 177).

Table 5.5: Chi-square tests for Traditional 2012 cohort and CLUE 2012 cohort, on Q1-pre, Q1-post, and Q2.

Question	Chi-square	p-value	Phi
Q1-pre	0.352	0.553	0.045
Q1-post	5.556	**0.018**	**0.177**
Q2	0.539	0.463	0.055

As shown in **Table 5.5**, the Traditional 2012 cohort and the CLUE 2012 cohort were not significantly different in their initial responses. After the instruction, the CLUE 2012 cohort had a significantly higher percentage of complete explanations, with a p-value of 0.018 and a small effect size (0.177). However, this difference soon disappeared in the second question. Overall, both cohorts were fairly similar in their argumentation skills before and after the activity, and both cohorts benefited from the activity.

To confirm the above findings, the same activity was administered again in Spring 2013 to a Traditional cohort (N=91) and a CLUE cohort (N=115) at the same university. The percentages of complete explanations were plotted in **Figure 5.1** together with the percentages from Spring 2012, for better comparison. The results from the CLUE cohort in Spring 2013 showed very similar trends as both Traditional and CLUE cohorts in Spring 2012, but the Traditional cohort in Spring 2013 showed lower percentage complete responses across all three questions. One possible explanation is because the activity was administered right before the spring break (because of the availability in general chemistry laboratory schedule), and some students might be rushed to leave. McNemar's Chi-square tests were performed to see if both the Traditional 2012 cohort and the CLUE 2013 cohort improved significantly after the instruction on scientific argumentation; Chi-square tests without Yate's correction were performed to see if the two cohorts are significantly different at any stage (Q1-pre, Q1-post, and Q2).

Table 5.6: McNemar's Chi-square tests for Traditional 2013 cohort and CLUE 2013 cohort.

Cohort	Comparison	McNemar's Chi-square	p-value	Phi
Traditional 2013	Q1-pre Vs Q1-post	9.091	**0.003**	**0.223**
	Q1-post Vs Q2	0.071	0.789	0.020
CLUE 2013	Q1-pre Vs Q1-post	12.19	**<0.001**	**0.230**
	Q1-post Vs Q2	1.829	0.176	0.089

Table 5.7: Chi-square tests for Traditional 2013 cohort and CLUE 2013 cohort, on Q1-pre, Q1-post, and Q2.

Question	Chi-square	p-value	Phi
Q1-pre	14.05	**<0.001**	**0.261**
Q1-post	13.85	**<0.001**	**0.259**
Q2	19.87	**<0.001**	**0.311**

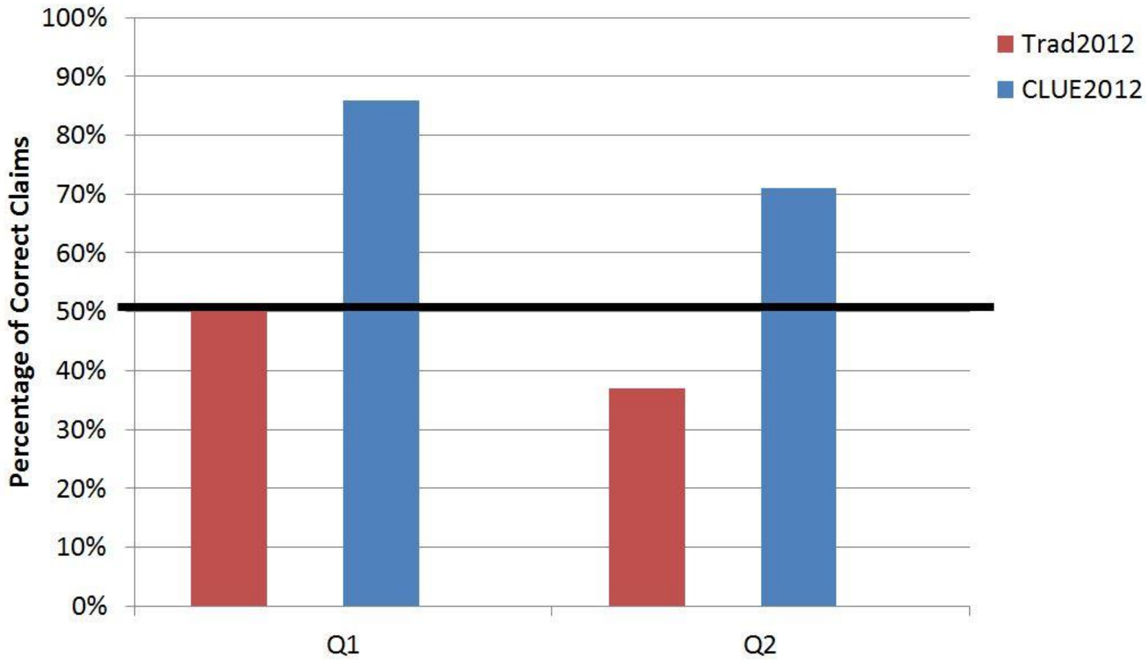
As shown in **Table 5.6**, both cohorts benefited from the activity, just as the previous year, but with smaller effect sizes (0.223 and 0.230). Both cohorts also retained the argumentation skills on the second question immediately after they finished editing their first responses: comparing Q2 to Q1-post, neither cohort had a significant difference (p-values are larger than 0.05). Different from the previous year, the CLUE 2013 cohort showed significantly more complete explanations at all stages (Q1-pre, Q1-post, and Q2), with p-values of less than 0.001 and medium effect sizes (range from 0.259 to 0.311), as shown in **Table 5.7**. However, this is mainly due to the fact that Traditional 2013 cohort had a significantly smaller percentage of complete responses than all other cohorts, as shown in **Figure 5.1**. CLUE 2013 cohort did not outperform Traditional 2012 cohort or CLUE 2012 cohort according to **Figure 5.1**.

Overall, the BeSocratic activity have helped students improve and temporarily retain their argumentation skills: all four cohorts demonstrated significantly higher percentage of complete explanations after the instruction embedded in the BeSocratic activity, with medium effect sizes ranging from 0.223 to 0.404; all four cohorts also maintained their argumentation skills in the second question, with p-values all larger than 0.05.

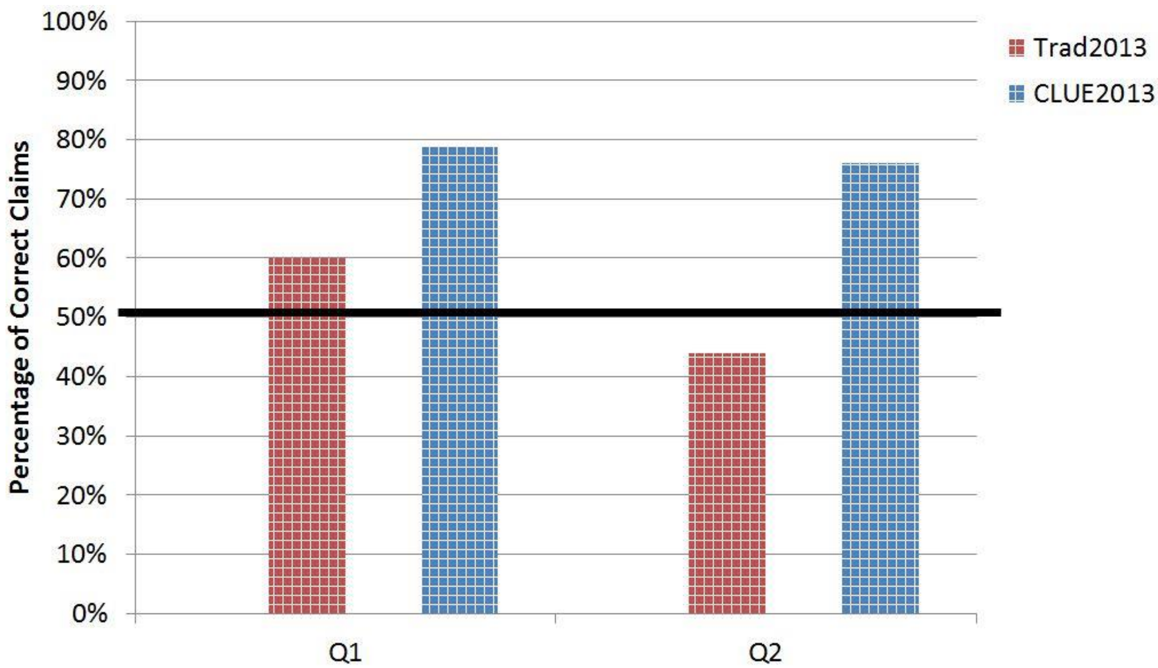
Analysis of Students' Acid Base Concepts

Because the BeSocratic activity succeeded in promoting students' argumentation skills at least for the length of the activity, data collected from Q1-post (after students edited their responses) and Q2 were well-articulated to a level that can be analyzed to compare the differences between two different general chemistry curricula at the large southeastern university, while data from Q1-pre was disregarded for this part of analysis. Thus, to simplify the representation, data from Q1-post will be labeled as "Q1" from now on.

The first and easiest question to ask is, are students able to make the correct claim in each question? Although each claim could be explained in a few different ways, water is a stronger acid than ammonia, and methylamine is a stronger base than methanol. Can students from either curriculum make the correct claim? Is there a difference between the two different cohorts? **Figure 5.2** below shows the comparison of two curricula in the correctness of claim. Chi-square tests without Yate's correction were performed to see if there is a significant difference between the two curricula (comparing Traditional 2012 cohort to CLUE 2012 cohort, and comparing Traditional 2013 cohort to CLUE 2013 cohort); and to see if each curriculum changes across two years (comparing Traditional 2013 cohort to CLUE 2013 cohort, and comparing CLUE 2012 cohort to CLUE 2013 cohort). Chi-square test was chosen because data was categorical (treating correct claim as "1" and incorrect claim as "0"). Fisher's test or Chi-square test with Yate's correction was not chosen because the sample sizes involved were large enough.



(a)



(b)

Figure 5.2: Comparison of Traditional cohort and CLUE cohort by percentage of correct claim for both Question 1 and Question 2. a) Spring 2012 cohorts; b) Spring 2013 cohorts.

Table 5.8: Chi-square tests for the correctness of claims on both questions, for all four cohorts.

Comparison	Q1			Q2		
	Chi-square	p-value	Phi	Chi-square	p-value	Phi
Traditional 2012 Vs CLUE 2012	23.60	**<0.001**	**0.365**	16.59	**<0.001**	**0.306**
Traditional 2013 Vs CLUE 2013	8.597	**0.003**	**0.204**	21.59	**<0.001**	**0.324**
Traditional 2012 Vs Traditional 2013	1.307	0.253	0.090	0.759	0.384	0.069
CLUE 2012 Vs CLUE 2013	1.313	0.252	0.077	1.519	0.218	0.083

As shown in **Figure 5.2a** and **Table 5.8**, the CLUE 2012 cohort had significantly higher percentages of correct claims in both questions than the Traditional 2012 cohort (p-values for both questions were less than 0.001), with medium effect size (0.365 and 0.306, respectively). It should also be noted that for each question, students had 50% chance of guessing correctly. As shown in **Figure 5.2a**, the Traditional 2012 cohort did no better than guessing (denoted by the black line in **Figure 5.2a**) in question 1, and worse than guessing in question 2, suggesting some kind of common alternative conception in solving question 2 (this will be further analyzed later). The CLUE 2012 cohort, on the other hand, did better than guessing in both questions. This result was reproduced in Spring 2013 with the Traditional cohort and the CLUE cohort. As shown in **Figure 5.2b**, the same trend was demonstrated with slightly smaller effect sizes (0.204 and 0.324). Again, a later section will compare the explanations each cohort of students provided. So far, looking only at the students' ability to correctly predict the stronger acid/base in a pair of compounds, students in the CLUE curriculum outperformed their counterparts in the Traditional curriculum.

Traditional 2012 and Traditional 2013 cohorts were also compared by the Chi-square test without Yate's correction, as well as CLUE 2012 and CLUE 2013 cohorts. The purpose of this comparison is to see whether the two groups of students who went through the same curriculum in two years would perform similarly. As shown in **Table 5.8**, there was no significant difference between the Traditional cohorts or between the CLUE cohorts from two different years (all four p-values were larger than 0.05), further proving the reproducibility of this data.

Based on the comparison of merely the correctness of claims, students in the CLUE curriculum seemed to outperform their counterparts in the Traditional curriculum. However, are they merely better at the "guessing game"? To answer this question, it is necessary to take a further look into the explanations students offered to support their claims.

In the initial process of coding students' responses, different codes were generated as different themes appear in students' responses. As shown in **Table 5.9**, every time a new type of reasoning appeared, a new code was generated. Later, some of the codes were condensed into a few different categories based on the similarity of the reasoning (while some other codes remained in their own categories), in order to compare which broad categories students in different curricula used to compare acid/base strength. In **Table 5.9**, the first few codes, "Proton Transfer", "Electron Donation", "Electronegativity", "Conjugate Acid/Base", "Definitions of Acid/Base", and "Heuristics", remained in their own categories; while several initial codes were condensed into a category called "Recognition", and several other initial codes that appeared less often were condensed into a category called "Other". Four initial codes were condensed into one general category called "Recognition", because students using this type of argumentation were all basing their decisions on the recognition of something – whether it is an atom, a functional group, or a molecule. Also, codes with less than 10% students in all four cohorts were gradually condensed into a general code "Other" (if a code appeared more than 10% in any cohort, it would not be condensed into "Other"), because these ideas were not prevalent. Although such an approach sacrificed the richness of each individual student's reasoning, it makes the purpose of this research plausible – to summarize some commonalities among students' understandings and to compare the overall impact of different curricula on students' understandings of acid base concepts.

Table 5.9: Initial codes and condensed categories from students' responses to Question 2, Spring 2012.

Student Quotes[1]	Initial Code	Final Category
…because it is more likely to accept a hydrogen.	Proton Transfer	Proton Transfer
…which allows the compount to donate more H^+ to an acid.		
…able to react in water to steal Hydrogen ions from water…		
…it will hold on to its lone pairs much more than nitrogen will. Nitrogen is more likely to donate its lone pairs.	Electron Donation	Electron Donation
…because during the reaction it will lose electrons.		
… and is not as good of an electron donor.		
… because oxygen is more electronegative than nitrogen…	Electronegativity	Electronegativity
… because nitrogen is less electronegative than oxygen.		

When a hydrogen is given off to form the acid, the acid that methanol forms is stronger than the acid methanamine forms. A strong acid always has a weak base and a weak acid has a strong base. Therefore, since methanamine forms the weakest acid it is the strongest base.	Conjugate Acid/Base	Conjugate Acid/Base
…and can hold on to the negative charge and distribute it through resonance…		
…becasue is conjugate acid is very weak.and if the N in CH_3NH_2 accepts protons the molecule will become a very weak acid…		
(bases)… and based on the bronsted lowery definition the are proton accepotors.	Definition of Acid/Base	Definition of Acid/Base
According to Lewis acid/ base theory, a base is an electron pair donor.		
Bases accepts and proton in the form of hydrogen…		
…because it has two lone pairs as compared to methanamine's one.	Heuristics	Heuristics
Because it has more H…		
The methanol also contains more oxygens…		
…because methanol is an alcohol.Alcohols are generally strong bases.	Recognition of Molecule	Recognition
I would say methanl is more basic because I know amines are weak bases.		
…oxygen is commonly found in acids…	Recognition of Atoms	
Methanamine is a stronger base than methanol because of the amount of nitrogen that is added to the element from methanol and because it doesnt consist of any oxygen elements also.		
…because it has an OH group on it. Bases typically have an OH on them which makes them a dead give away for being a base.	OH Means Base	
methanol because it has an OH group. Lewis bases require an OH group to be consired a base and they donate the OH group when in solution.		
Nitrogen has more Hydrogen atoms or H^+ ions which are acidic…[2]	H Means Acid	
…because the NH bond is weaker than the OH bond…	Bond	Other
… making the bonds tighter and closer, and allowing for more bonds to be formed.		

The nitrogen hydrogen bond would be stronger than a oxygen hydrogen bond because it would contain less energy.		
…because Nitrogen normally only forms three bonds.		
… because nitrogen is less polar than the oxygen…	Polarity	
the stronger base has a greater polarity…		
…This means that methanol is more stable and, thus, less likely to participate in an acid-base reaction.	Stability	
Methanamine would be the stronger base because nitrogen is more stable with four bonds, while oxygen is stable with two bonds.		
It (methylamine) is closer to being a stable compound than methanol.		
…The more electronegative a compound is, the stronger base it is.	Periodic Table Trend	
As electronegativity increases, acid strength increases.		
… a higher pH level, causing it to be more basic.	pH	
… as the concentration of H+ ions increases, the closer the pH gets to neutral. Strong bases have a pH greater than neutral, which is seven.		
The stronger base would be the methanamine. This compound has a more nucleophilic attraction to molecules than does the methanol.	Effective Nuclear Charge	
This is because the Nitrogen has a stronger effective nuclear charge then the Oxygen…		

Notes:

1. All examples were taken from students responses from Spring 2012, to question 2 (which is a stronger base between methylamine and methanol?). Because each response usually had more than one code, the examples were only the fragments of responses corresponding to the code. Typos in students' original responses were kept as is.
2. This quote was also coded as "Heuristics". The quote was coded as "Recognition" because this student was using the presence of hydrogen atoms to identify acids; it was

also coded as "Heuristics" because this student further argued that the more hydrogen atoms a compound has, the more acidic it must be.

Several other ways of analyzing this rich data were explored but found unsuccessful. An initial attempt to code student responses by the correctness and completeness found that very few student responses were completely correct. The majority of the students made different types of mistakes in their responses, from terminology issues (such as calling the O-H or N-H bond as "hydrogen bonding"), to mistakes in memorization (such as "nitrogen is more electronegative than oxygen"), to mistakes in acid base conceptions (such as describing electron donation as an atom permanently loses those electrons). Ten student responses on Question 1 were shown in Appendix C to demonstrate why the attempt to code student responses by the correctness and completeness turned out to be unsuccessful. First of all, student responses with incomplete explanations and/or incorrect claims were taken out. Among the responses that were coded as complete explanations with correct claims, the first five responses from the Traditional 2013 cohort and the first five responses from the CLUE 2013 cohort were included in Appendix C to demonstrate the point without making this section too tedious.

As shown in Appendix C, nine out of the first ten responses were not completely correct and complete, even though those are responses selected from students with correct claims and complete explanations (including claim, data, and explanation). If student responses were coded by whether it is complete and completely correct, overwhelming majority of students in any cohort would not have complete and completely correct explanations. This would again void the purpose of this research – to study commonalities among students' concepts and compare the differences between two different curricula.

Another possible approach was to code the data with a finer "grain size", retaining more individuality of each student's responses. However, since individual students tend to have very different and unique mental models, a lot more codes would be needed to retain the individuality of these mental models; making it harder to compare the statistical difference of two cohorts later.

Thus it was determined that this research would use category codes from **Table 5.9** to try to understand the approaches students take when solving a specific problem, and later compare the collective approaches from two different cohorts of students taking different general chemistry curricula. Once the coding method was determined and the category codes were generated, the same codes were used to code the rest of the student responses. Each student response was coded with one or more category codes. Category codes were later totaled as either 0 (not present) or 1 (present), rather than by frequency. This is because a student might repeat a piece of supporting information twice; and repetition of the same information multiple times does not suggest a better understanding. Below are a few examples of complete student responses to question 2, and how they were coded. These responses were chosen as examples because they were the earliest responses coded and because they covered different category codes explained above.

"Methanamine is a stronger base because oxygen is more electronegative than nitrogen. Because oxygen is more electronegative, it is less likely to donate an electron pair and is therefore a weaker base." – Codes: Electronegativity and Electron Donation

"Methanol is a stronger base. This is because it contains an alcohol group. Alcohol groups are able to react in water to steal Hydrogen ions from water and leave hydroxide ions in solution. This generates a pH of greater than 7 and is the stronger base." – Codes: Recognition, Proton Transfer, and Other

"Methanol would be the stronger base, because it has two electron lone pairs instead of just one." – Code: Heuristics

To ensure the validity of these category codes, a second coder coded a total of 20 responses to Question 1 from both Traditional 2013 cohort and CLUE 2013 cohort. This coder was provided with the category codes, and then coded blindly from her understanding of these category codes. The inter-rater reliability was calculated based on the number of codes two coders agree or disagreed on, rather than the number of students responses, because some student responses contained multiple codes while some other student responses contained only one code. Two coders agreed on 31 codes while disagreed on 16 codes, leading to an initial inter-rater reliability of 66%, before two coders had any discussion.

The biggest disagreement was about the use of the code "Definitions of Acid/Base". The first coder initially created this code to describe students who refer to the definitions of acids and/or bases according to Arrhenius', Bronsted-Lowry', or Lewis' acid base model, in supporting their answers. **Table 5.9** showed three of such occasions under the code “Definitions of Acid/Base”. The second coder instead used this code “to capture the students who do not use the properties, but only familiarity by classifying something just based on their belief that it will be more likely to donate protons”. For example, “water is a better acid because it donates protons” would be coded as both “Proton Transfer” and “Definitions of Acid/Base” by the second coder, because this student “did not explain how the properties of the substance related to its ability to behave as an acid or base”. On the other hand, “since oxygen is more electronegative, it is more likely to accept electrons than to give them away, which is an acidic property” would not be coded as “Definitions of Acid/Base” because the acidic behavior was explained by the higher electronegativity of oxygen. This response was coded as “Electronegativity” and “Electron Donation”. After consulting a Chemistry Education faculty, the second coder’s proposal was accepted and the first coder recoded all the responses accordingly.

The other initial differences were resolved relatively quickly after two coders discussed with each other. The first coder was initially “on the safer side” as not to interpret students’ responses at all. For example, response “H_2O, because it has two lone pairs making it the stronger acid” was initially coded as “Recognition” by the first coder, because the student did not clearly compare the two lone pairs on water to the one lone pair on ammonia. However, because the second coder also interpreted this sentence as a hidden “Heuristics” that the student meant to say water is a stronger acid because it has more lone pairs than ammonia, both coders agreed on coding this response as

“Heuristics”, and the first coder also went back and recoded all the similar occurrences as “Heuristics”.

The first coder recoded all the responses after the discussion and agreement with the second coder on the twenty responses coded by both coders. Below is a list of the final categories after two coders agreed with each other, and what each category is capturing. Appendix D includes the numbers and percentages of each code from each cohort and each question, after the first coder recoded all the responses based on the final categories agreed upon.

Electronegativity*: This category captures students who used the electronegativity difference between two atoms as a support for their conclusions.*

Electron Donation*: This category captures students who looked at the acid/base behaviors by the donating/accepting of electron pairs.*

Proton Transfer*: This category captures students who looked at the acid/base behavior by the donating/accepting of protons.*

Conjugate Acid/Base*: This category captures students who used the stability of the conjugate bases/acids to determine the relative acid/base strength of a given pair of compounds.*

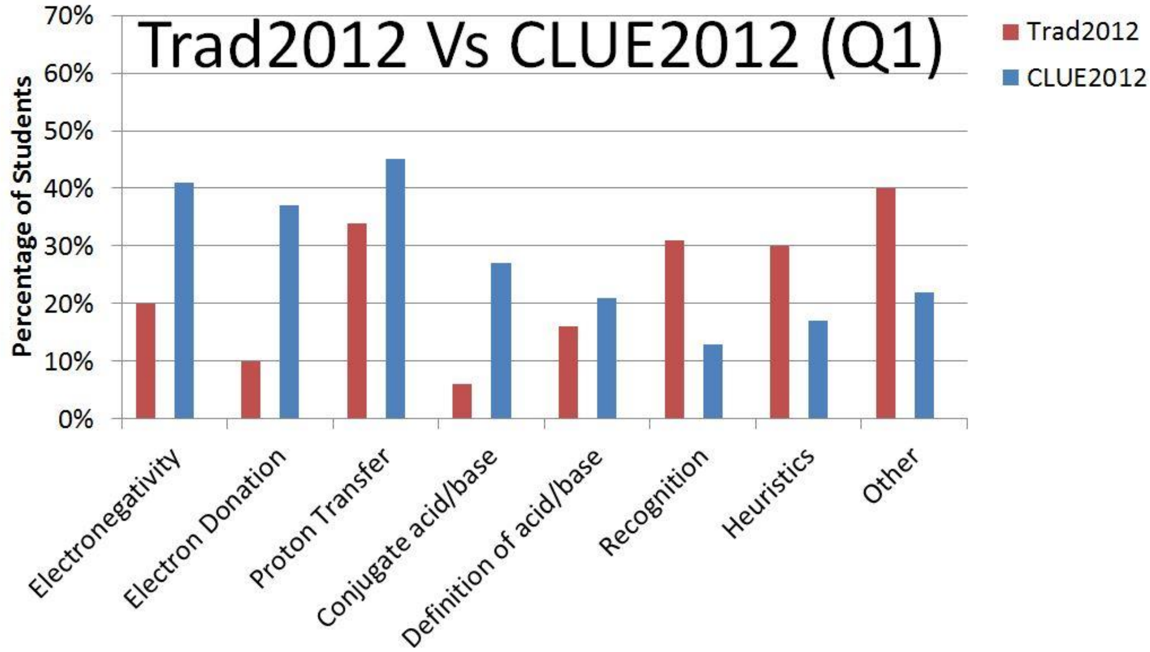
Definitions of Acid/Base*: This category captures students who did not explain how the properties of the substance related to its ability to behave as an acid or base.*

Recognition*: This category captures students who rely on the recognition of particular atoms, functional groups, or molecules to determine relative acid/base strength.*

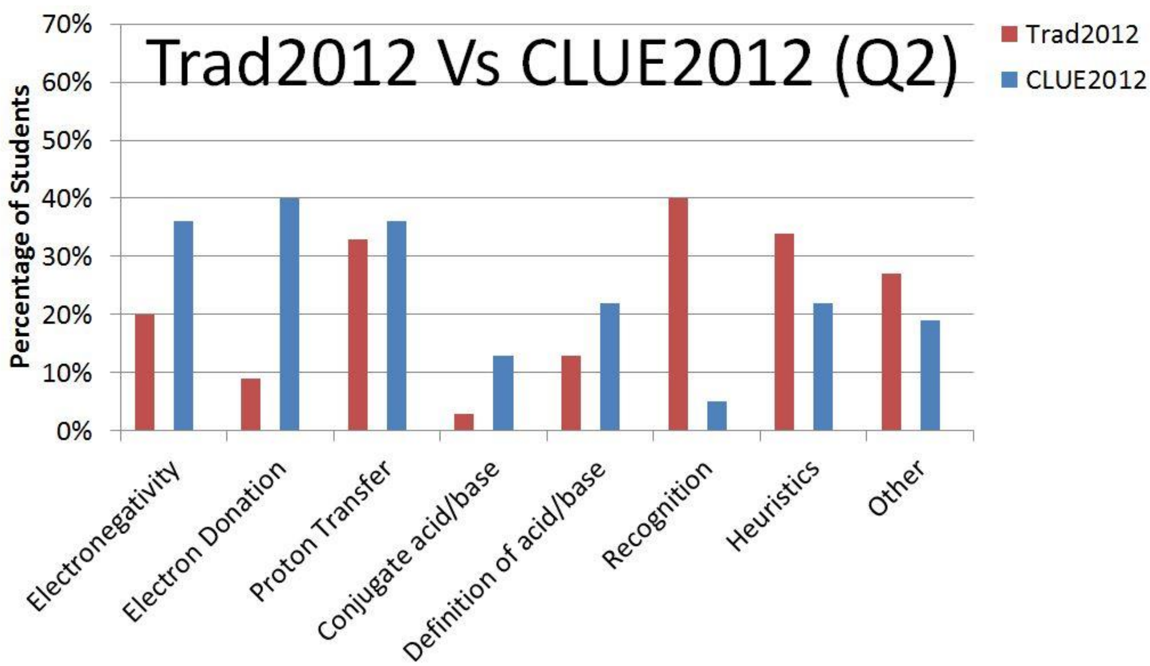
***Heuristics**: This category captures students who use the "more A, more B" heuristics to support their conclusions, as shown in **Table** **5.9**.*

***Other**: This category contains less prevalent explanations such as "Bond", "Stability", Polarity", "pH", "Periodic Table Trend", "Effective Nuclear Charge", etc (as shown in **Table 5.9**). Code with more than 5% appearance in any question of any cohort was kept until all responses were coded, and then combined into the "Other" category if it did not have more than 10% appearance in any of the questions in any cohorts (for example, the code "Bond" was not condensed into the "Other" category until all data were coded). This avoids condensing a code initially that would later have more than 10% appearance in the data from Spring 2013.*

After all the coding was finalized, the Traditional and CLUE cohorts from Spring 2012 were compared to see if there was any difference in their approaches to the two questions that might render the difference in the percentage of correct claims between the two cohorts. **Figure 5.2a** above indicated that the Traditional 2012 cohort did no better than guessing in question 1, and worse than guessing in question 2; while the CLUE 2012 cohort did better than guessing in both questions, and significantly better than the Traditional 2012 cohort with both p-values of less than 0.001 and medium effect sizes (0.365 on Q1 and 0.306 on Q2). **Figure 5.3** below illustrated the differences between the two cohorts in the approaches they took in solving each question. Chi-square tests without Yate's correction were performed and Phi effect sizes were calculated for each category code, as shown in **Table 5.10**.



(a)



(b)

Figure 5.3: Category comparison between Traditional 2012 cohort and CLUE 2012 cohort. a) Question 1 responses; b) Question 2 responses.

Table 5.10: Chi-square tests for each category code in both questions, for Traditional 2012 cohort and CLUE 2012 cohort.

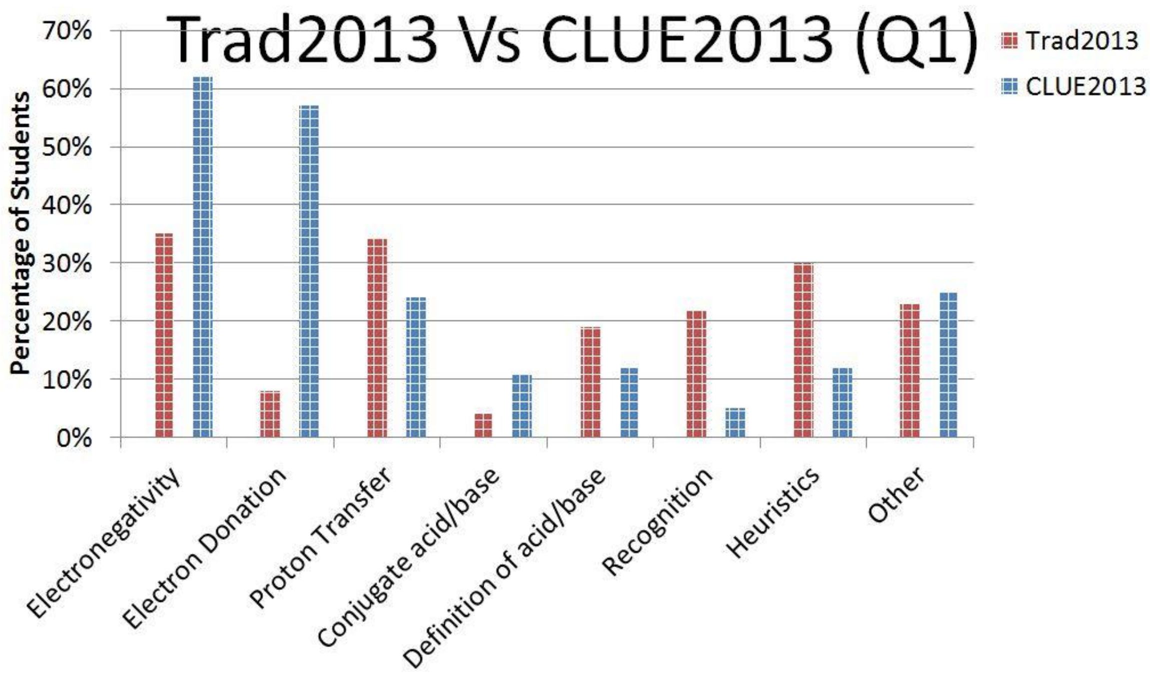
Category	**Q1**			**Q2**		
	Chi-square	**p-value**	**Phi**	**Chi-square**	**p-value**	**Phi**
Electronegativity	8.569	**0.003**	**0.220**	5.458	**0.019**	**0.176**
Electron Donation	16.27	**<0.001**	**0.303**	21.13	**<0.001**	**0.345**
Proton Transfer	1.961	0.161	0.105	0.240	0.624	0.037
Conjugate Acid/Base	12.76	**<0.001**	**0.269**	5.383	**0.020**	**0.174**
Definitions of Acid/Base	0.655	0.418	0.061	2.132	0.144	0.110
Recognition	8.789	**0.003**	**0.223**	34.82	**<0.001**	**0.444**
Heuristics	4.277	**0.039**	**0.155**	3.009	0.083	0.130
Other	7.064	**0.008**	**0.200**	1.759	0.185	0.100

As shown in **Figure 5.3a** and **Table 5.10**, when solving Question 1 – "Which is the stronger acid between water and ammonia?" – the CLUE 2012 cohort was significantly more likely to think about electronegativity, electron donation, and conjugate acid/base, with medium effect sizes ranging from 0.220 to 0.303. All these approaches can be "good starts" to successfully solve the given question. On the other hand, the Traditional 2012 cohort was significantly more likely to rely on recognition, heuristics, and other approaches, with small effect sizes ranging from 0.155 to 0.223. These results seem to suggest that the CLUE 2012 cohort outperformed the Traditional 2012 cohort in the correctness of claim for question 1 not because the CLUE 2012 cohort was better at guessing, but because the CLUE 2012 cohort was more likely to reach their conclusion from considering electronegativity, electron donation, and conjugate acid/base, rather than from recognition and heuristics.

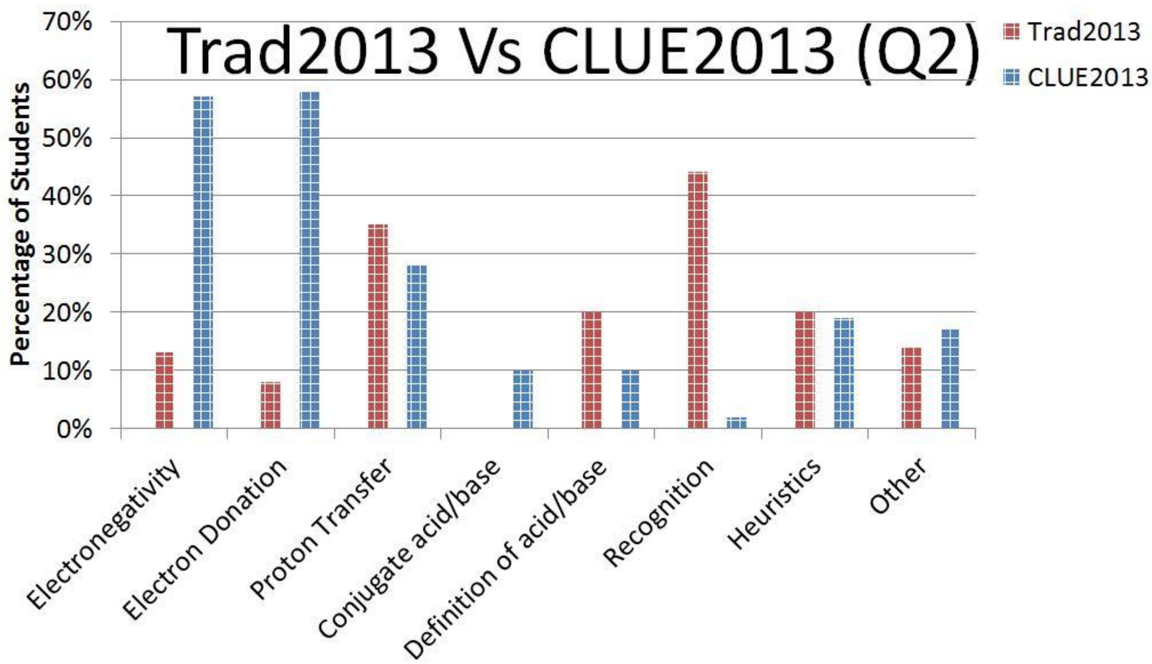
Similar trends were illustrated between these two cohorts in their explanations to question 2 – "Which is a stronger base between methanol and methylamine?" As shown in **Figure 5.3b** and **Table 5.10**, the CLUE 2012 cohort again was significantly more likely to think about electronegativity, electron donation, and conjugate acid/base, with small to medium effect sizes ranging from 0.174 to 0.345; while the Traditional 2012 cohort was significantly more likely to rely on recognition with a medium-large effect size, 0.444. Another interesting finding is that recognition was the most frequent approach the Traditional 2012 cohort took; and among those students most of them were basing their decisions solely or mainly on the recognition of the OH group in methanol, leading them to the wrong claim for this question. Prior researches have also suggested that students who rely on recognition of certain atoms or functional groups, as well as students who rely on heuristics, could often reach the wrong conclusion (McClary & Talanquer 2011a&b and McClary & Bretz 2012). This may explain why the Traditional 2012 cohort performed worse than guessing on this question with less than 40% correct claim (**Figure 5.2a**).

Summarizing both questions, the CLUE 2012 cohort outperformed the Traditional 2012 cohort significantly in the correctness of claim, not because students in the CLUE 2012 cohort were better at "guessing", but because students in the CLUE 2012 cohort were significantly more likely to approach both questions from electronegativity, electron donation, and conjugate acid/base; while students in the Traditional 2012 cohort were significantly more likely to rely on recognition. Analysis of which categories led to significantly more correct or incorrect answers will be presented later in the chapter.

The same study was repeated in Spring 2013 with the Traditional cohort and the CLUE cohort, as shown in **Figure 5.4** and **Table 5.11**.



(a)



(b)

Figure 5.4: Category comparison between Traditional 2013 cohort and CLUE 2013 cohort. a) Question 1 responses; b) Question 2 responses.

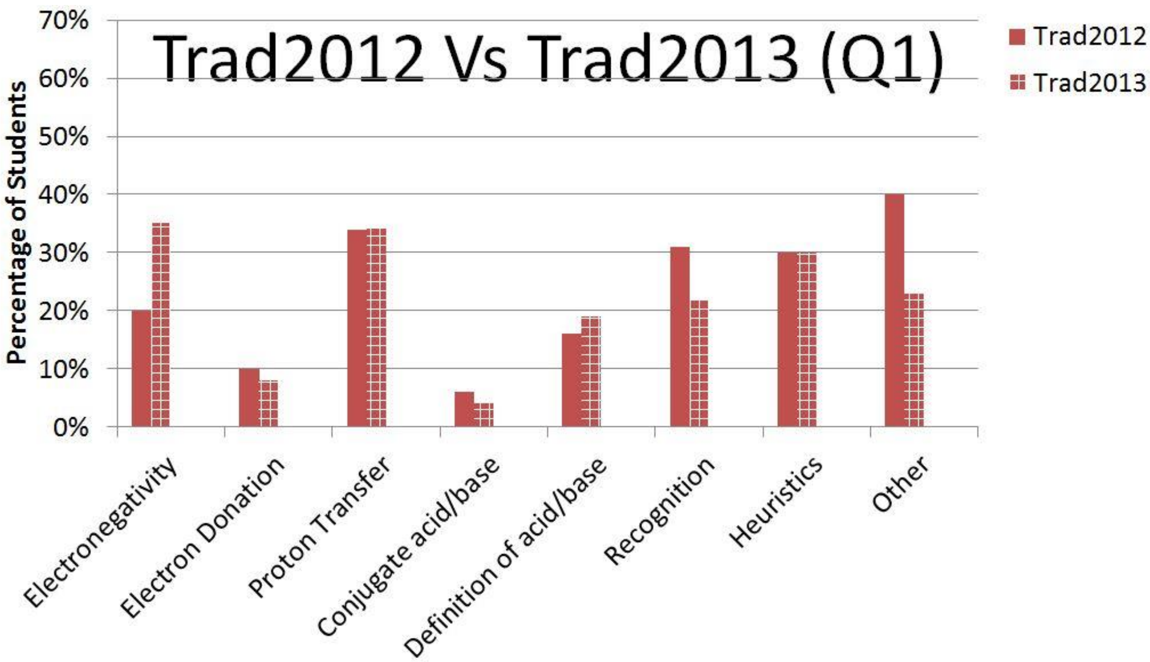
Table 5.11: Chi-square tests for each category code in both questions, for Traditional 2013 cohort and CLUE 2013 cohort.

Category	**Q1**			**Q2**		
	Chi-square	**p-value**	**Phi**	**Chi-square**	**p-value**	**Phi**
Electronegativity	14.35	**<0.001**	**0.264**	42.19	**<0.001**	**0.453**
Electron Donation	54.84	**<0.001**	**0.516**	56.44	**<0.001**	**0.523**
Proton Transfer	2.815	0.093	0.117	1.278	0.258	0.079
Conjugate Acid/Base	3.203	0.074	0.125	10.08	**0.001**	**0.221**
Definitions of Acid/Base	1.683	0.195	0.090	3.566	0.059	0.132
Recognition	12.94	**<0.001**	**0.251**	55.78	**<0.001**	**0.520**
Heuristics	9.755	**0.002**	**0.218**	0.014	0.907	0.008
Other	0.127	0.722	0.025	0.364	0.546	0.042

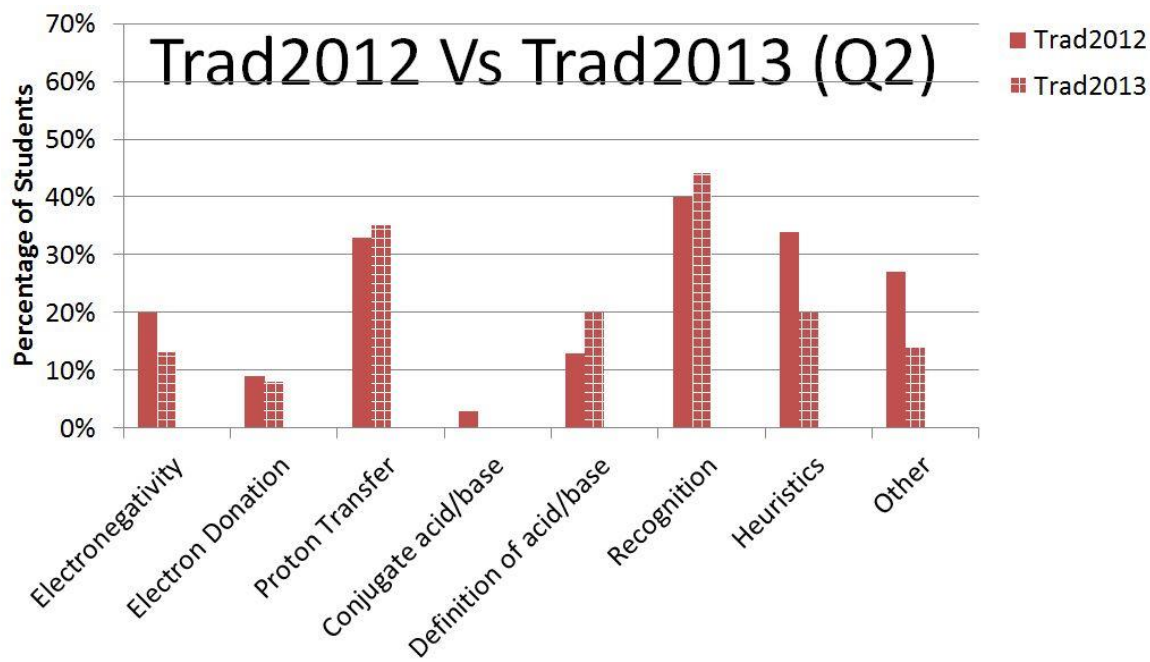
Again similar trends were observed, showing that the results were reproducible. The CLUE cohorts from both years consistently outperformed their Traditional counterparts because they were consistently more likely to reach their conclusion from considering electronegativity, electron donation, and conjugate acid/base, rather than from recognition and heuristics. However, the difference of using electron donation as an approach to both questions increased from 2012 to 2013. Although the CLUE 2012 cohort was significantly more likely to consider electron donation than the Traditional 2012 cohort (close to 40% to less than 10%), it only corresponds to medium effect sizes (0.303 and 0.345). This difference was enlarged between the CLUE 2013 cohort and the Traditional 2013 cohort (close to 60% to less than 10%), with large effect sizes (0.516 and 0.523). While the Traditional cohorts in both years had less than 10% of students who considered electron donation in solving both questions, the CLUE cohorts increased from approximately 40% to approximately 60% in considering electron donation in solving both questions. Another significant difference with large effect size came from the use of recognition in Question 2. As mentioned earlier, the Traditional 2012 cohort was significantly more likely to rely on recognition in solving Question 2 (40%), than the CLUE 2012 cohort (less than 5%); which explains why the Traditional 2012 cohort not only performed significantly worse than the CLUE 2012 cohort in Question 2, but also performed worse than guessing in this question. In 2013, this difference was further enlarged: 44% of the Traditional 2013 cohort used recognition in solving this question, while less than 2% of the CLUE 2013 cohort did so, leading to a large effect size of 0.520.

Since the CLUE cohorts increased in their usage of electron donation from 2012 to 2013, as mentioned in the paragraph above, another comparison was conducted between two cohorts in the same curriculum to see whether students in the same curriculum across two years would perform similarly. Chi-square tests without Yate's correction were chosen for the same reasons explained above, and the corresponding p-values and Phi effect sizes were calculated.

The Traditional 2012 cohort and the Traditional 2013 cohort were first compared side by side as shown in **Figure 5.5** and **Table 5.12**. Despite of a few differences with small effect sizes ranging from 0.160 to 0.182, the two cohorts showed very similar results in most categories, suggesting that the difference in curricula outweighs the difference among students from different years in affecting students' approaches to the questions studied here.



(a)



(b)

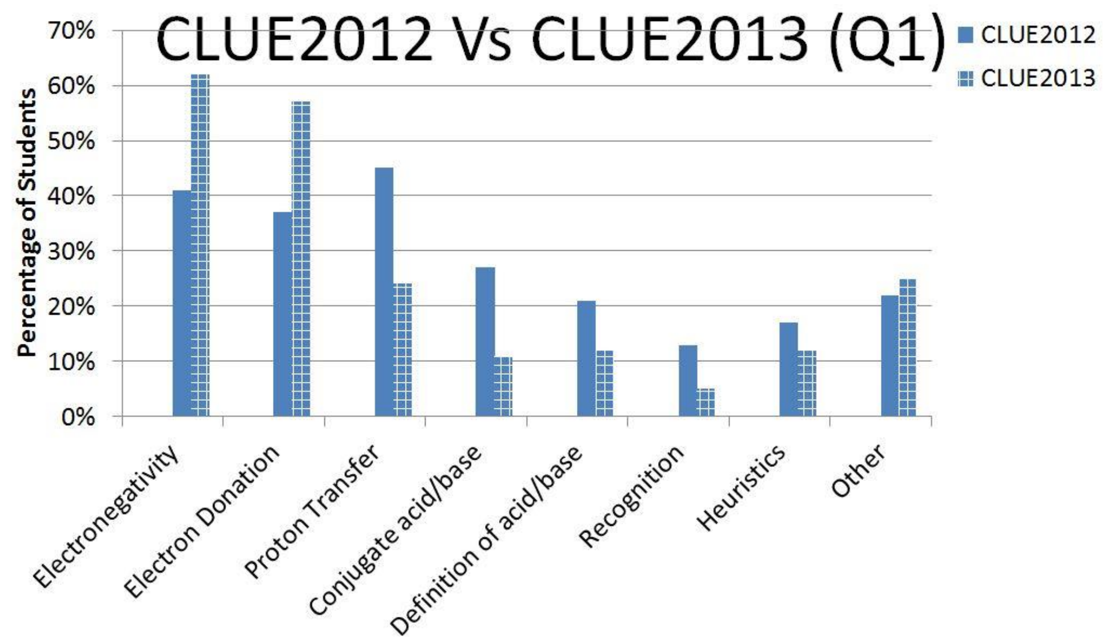
Figure 5.5: Category comparison between Traditional 2012 cohorts and Traditional 2013 cohort. a) Question 1 responses; b) Question 2 responses.

Table 5.12: Chi-square tests for each category code in both questions, for Traditional 2012 cohort and Traditional 2013 cohort.

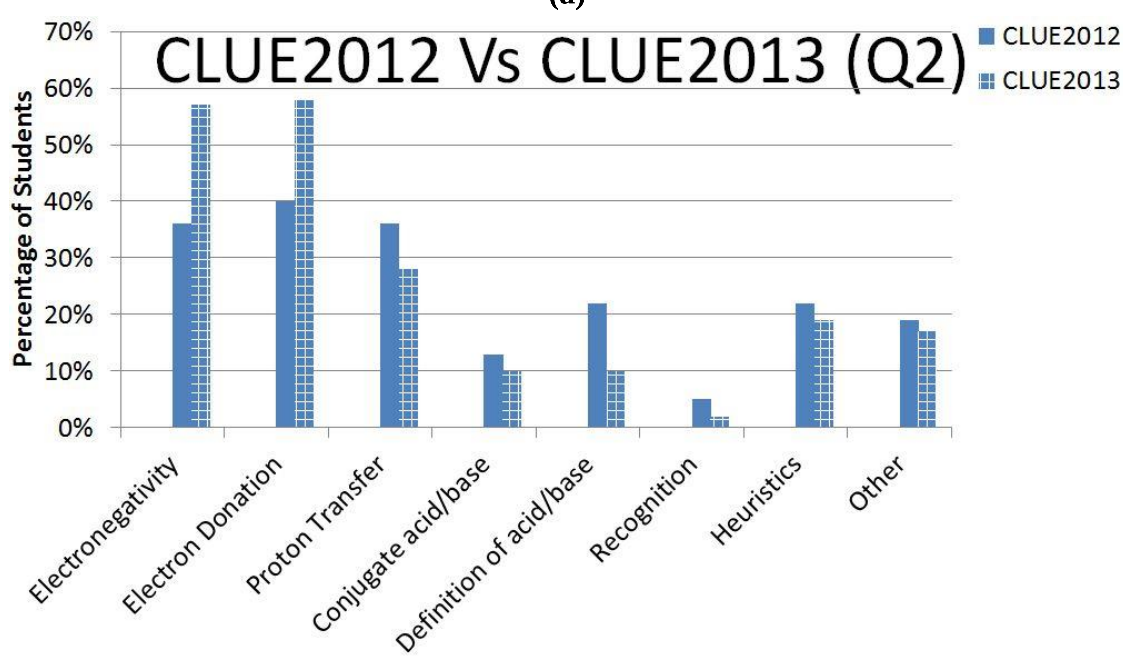
Category	Q1			Q2		
	Chi-square	p-value	Phi	Chi-square	p-value	Phi
Electronegativity	4.458	**0.035**	**0.166**	1.356	0.244	0.092
Electron Donation	0.265	0.606	0.041	0.041	0.839	0.016
Proton Transfer	0.001	0.977	0.002	0.094	0.760	0.024
Conjugate Acid/Base	0.146	0.703	0.030	2.633	0.105	0.128
Definitions of Acid/Base	0.242	0.622	0.039	1.359	0.244	0.092
Recognition	1.833	0.176	0.107	0.254	0.614	0.040
Heuristics	0.002	0.964	0.004	4.318	**0.038**	**0.164**
Other	5.352	**0.021**	**0.182**	4.107	**0.043**	**0.160**

Similarly, the CLUE 2012 cohort and the CLUE 2013 cohort were also compared side by side, as shown in **Figure 5.6** and **Table 5.13**. The CLUE students seemed to increase in electronegativity and electron donation, while decreasing in proton transfer and conjugate acid/base. But all these changes only correspond to small effect sizes ranging from 0.137 to 0.226. These small changes may suggest that in Spring 2013 more students were moving from the Bronsted-Lowry acid base model to the Lewis acid-base model, even for question 1 (which can be easily solved by either theory). Since there was no significant difference in the correctness of claims between the two CLUE cohorts (as shown in **Table 5.8**), it is equally acceptable whether students switch between Bronsted-Lowry's acid base model and Lewis' acid base model, or stick to Lewis' acid base model.

Overall, the differences across two years in either curriculum appear to be negligible, suggesting that the analysis method based on category coding can be used to assess different groups of students – when the sample sizes are large (such as the groups involved in this research), the differences between individual students would not significantly affect the group as a whole.



(a)



(b)

Figure 5.6: Category comparison between CLUE 2012 cohorts and CLUE 2013 cohort. a) Question 1 responses; b) Question 2 responses.

Table 5.13: Chi-square tests for each category code in both questions, for CLUE 2012 cohort and CLUE 2013 cohort.

Category	**Q1**			**Q2**		
	Chi-square	**p-value**	**Phi**	**Chi-square**	**p-value**	**Phi**
Electronegativity	8.991	**0.003**	**0.201**	9.753	**0.002**	**0.210**
Electron Donation	8.894	**0.003**	**0.200**	7.243	**0.007**	**0.181**
Proton Transfer	11.33	**<0.001**	**0.226**	1.894	0.169	0.092
Conjugate Acid/Base	9.019	**0.003**	**0.202**	0.376	0.540	0.041
Definitions of Acid/Base	2.870	0.090	0.114	5.106	**0.024**	**0.152**
Recognition	4.184	**0.041**	**0.137**	1.562	0.211	0.084
Heuristics	0.971	0.324	0.066	0.367	0.544	0.041
Other	0.428	0.513	0.044	0.063	0.801	0.017

Another approach to analyze the data was to see whether students' mental models change based on different questions. The initial design of two questions moved from a pair of more commonly seen structures to a pair of less commonly seen structures, and from comparing acid strength to comparing base strength, but did students change their approaches based on the change of questions? McNemar's Chi-square tests were performed on each cohort, comparing the difference between their responses in Q1-post and Q2. McNemar's Chi-square tests were chosen because data was categorical and dependent (considering Q1-post as "pre" and Q2 as "post since each pair of responses came from the same student).

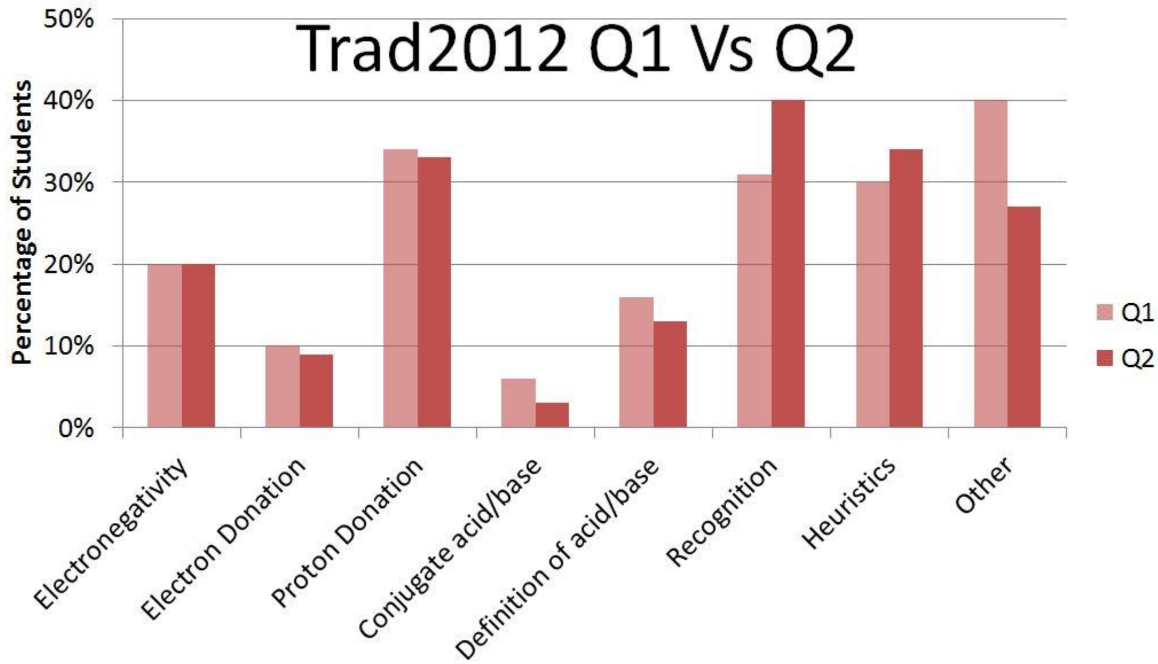


Figure 5.7: Category comparison between the two questions for the Traditional 2012 cohort.

Table 5.14: McNemar's Chi-square tests for Traditional 2012 cohort, from Q1 to Q2 on each category code.

Category	McNemar's Chi-square	p-value	Phi
Electronegativity	0.083	0.773	0.024
Electron Donation	0.000	1.000	0.000
Proton Transfer	0.000	1.000	0.000
Conjugate Acid/Base	0.617	0.250	0.066
Definitions of Acid/Base	0.083	0.773	0.024
Recognition	1.042	0.307	0.086
Heuristics	0.174	0.677	0.035
Other	2.370	0.124	0.130

As shown in **Figure 5.7** and **Table 5.14**, students in the Traditional 2012 cohort demonstrated very similar approaches in solving both problems (with no significant difference).

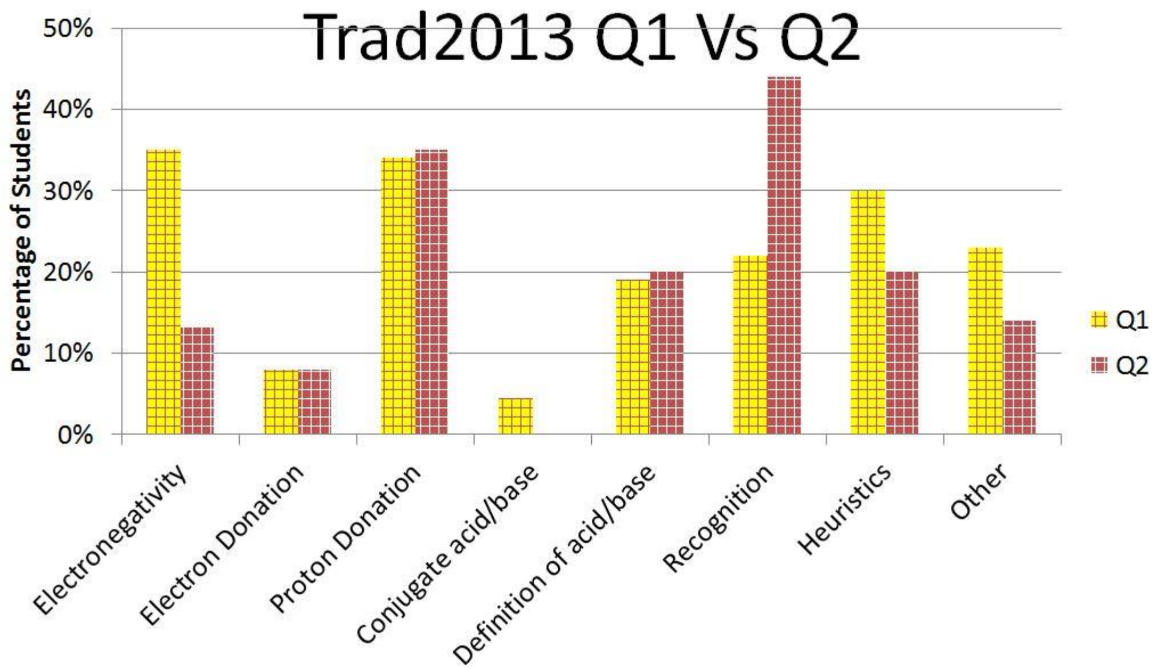


Figure 5.8: Category comparison between the two questions for the Traditional 2013 cohort.

Table 5.15: McNemar's Chi-square tests for Traditional 2013 cohort, from Q1 to Q2 on each category code.

Category	McNemar's Chi-square	p-value	Phi
Electronegativity	12.89	**<0.001**	**0.266**
Electron Donation	0.125	0.724	0.026
Proton Transfer	0.000	1.000	0.000
Conjugate Acid/Base	2.250	0.134	0.111
Definitions of Acid/Base	0.000	1.000	0.000
Recognition	8.595	**0.003**	**0.217**
Heuristics	2.783	0.095	0.124
Other	2.042	0.153	0.106

As shown in **Figure 5.8** and **Table 5.15**, students in the Traditional 2013 cohort approached the second and less familiar question significantly more from recognition and less from electronegativity, but only with small effect sizes (0.266 and 0.217).

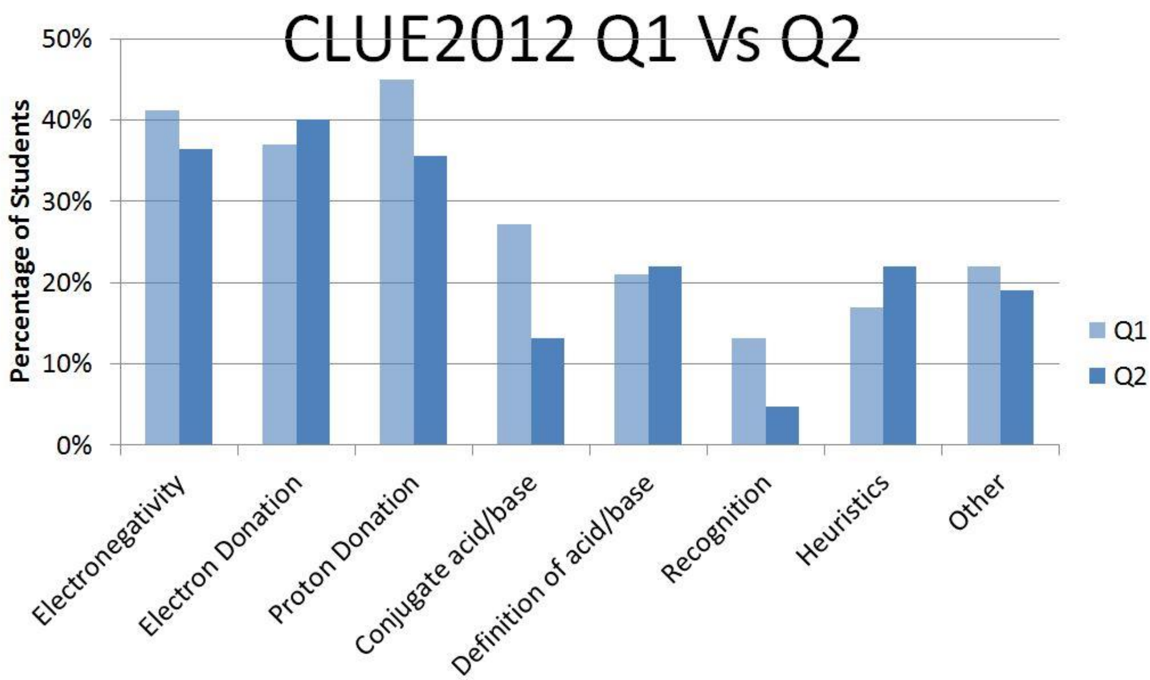


Figure 5.9: Category comparison between the two questions for the CLUE 2012 cohort.

Table 5.16: McNemar's Chi-square tests for CLUE 2012 cohort, from Q1 to Q2 in each category code.

Category	McNemar's Chi-square	p-value	Phi
Electronegativity	0.516	0.473	0.049
Electron Donation	0.129	0.719	0.025
Proton Transfer	1.641	0.200	0.088
Conjugate Acid/Base	7.840	**0.005**	**0.191**
Definitions of Acid/Base	0.000	1.000	0.000
Recognition	4.923	**0.027**	**0.152**
Heuristics	2.083	0.149	0.099
Other	0.129	0.719	0.025

As shown in **Figure 5.9** and **Table 5.16**, students in the CLUE 2012 cohort approached the second question significantly less from conjugate acid/base but also less from recognition, both with small effect sizes (0.191 and 0.152).

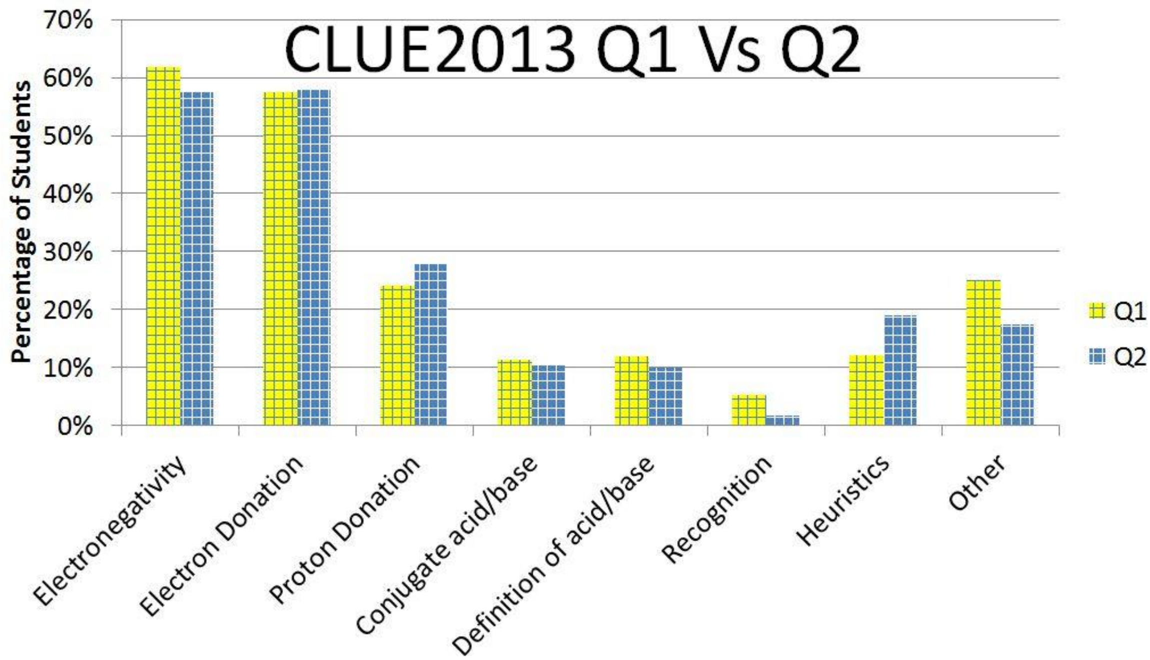


Figure 5.10: Category comparison between the two questions for the CLUE 2013 cohort.

Table 5.17: McNemar's Chi-square tests for CLUE 2013 cohort, from Q1 to Q2 in each category code.

Category	McNemar's Chi-square	p-value	Phi
Electronegativity	0.640	0.424	0.053
Electron Donation	0.000	1.000	0.000
Proton Transfer	0.696	0.404	0.055
Conjugate Acid/Base	0.000	1.000	0.000
Definitions of Acid/Base	0.083	0.773	0.019
Recognition	1.125	0.289	0.070
Heuristics	3.063	0.080	0.115
Other	2.560	0.110	0.106

As shown in **Figure 5.10** and **Table 5.17**, students in the CLUE 2013 cohort demonstrated very similar approaches in solving both problems (with no significant difference).

Overall, most students seemed to approach both problems with similar approaches, suggesting that while a student may have a very complicated and unique acid base mental model, this student might use a few similar initial approaches in solving different acid base related questions.

Because many students seemed to be using similar approaches in solving both questions, and students from different curricula seemed to favor different approaches, data from all four cohorts were combined to analyze whether a particular approach would lead to significantly higher/lower percentage of correct claims. Instead of focusing on comparing two different curricula, this analysis focuses on comparing the "success rate" of each category code. Earlier in this chapter, it was noticed that CLUE cohorts were significantly more likely to use categories such as Electronegativity, Electron Donation, Conjugate Acid/Base in solving both questions, while the Traditional cohorts were significantly more likely to use categories such as Recognition and Heuristics. It was hypothesized that the categories CLUE cohorts preferred were "good starts" that were more likely to lead to correct conclusions, while categories Traditional cohorts preferred were more likely to lead to incorrect conclusions. This analysis below tries to test this earlier hypothesis by calculating the percentage of correct claims for each question based on the use of different categories. For example, the first line in **Table 5.18** indicates that for Question 1, the total number of responses collected from all four cohorts was three hundred and eighty-three, among which one hundred and sixty-one students used "Electronegativity" in their explanations, while the rest of two hundred and twenty-two students did not. Among the one hundred and sixty-one students who used electronegativity in their explanations, eighty-two percent of them reached the correct conclusion (water is a stronger acid than ammonia); but among the students who did not use electronegativity in their explanations, only sixty-four percent of them reached the correct conclusion. Chi-square test without Yate's correction was performed to see if these two groups had significantly different percentage of correctness, and the result confirms this difference with a small effect size of 0.202.

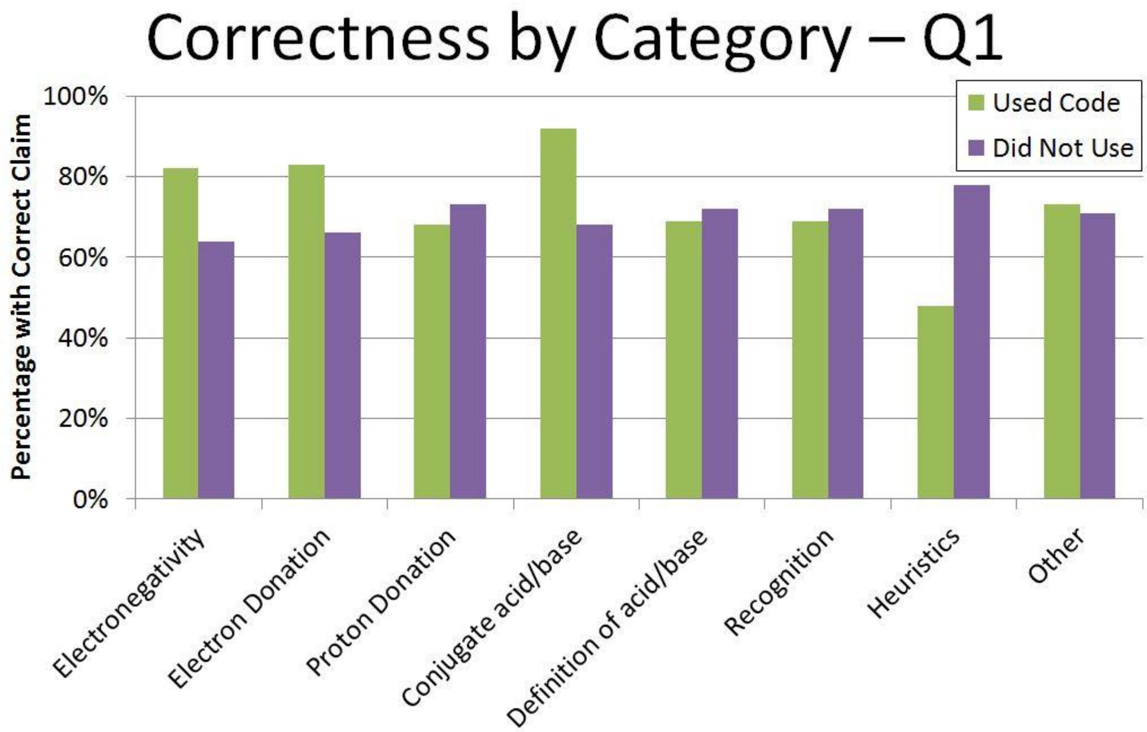


Figure 5.11: Percentage of correct claims by the use or lack of each category in Question 1.

Table 5.18: Chi-square tests for each category in Question 1, with p-values and Phi effect sizes.

Code (N = 383)	Used Code		Did Not Use Code		Chi-square	p-value	Phi
	Number of Students	Correct %	Number of Students	Correct %			
Electronegativity	161	82%	222	64%	15.16	**<0.001**	**0.202**
Electron Donation	120	83%	263	66%	10.75	**0.001**	**0.168**
Proton Transfer	130	68%	253	73%	0.763	0.382	0.045
Conjugate Acid/Base	50	92%	333	68%	12.06	**<0.001**	**0.177**
Definitions of Acid/Base	64	69%	319	72%	0.240	0.624	0.025
Recognition	62	69%	321	72%	0.134	0.715	0.019
Heuristics	80	48%	303	78%	27.93	**<0.001**	**0.270**
Other	101	73%	282	71%	0.265	0.607	0.026

Similarly, all other categories were compared for Question 1, and the results were shown in **Figure 5.11** and **Table 5.18**. Besides Electronegativity, Electron Donation and Conjugate Acid/Base were also categories that lead to higher percentages of correct claims when students used such categories; but all three categories only correspond to small effect sizes ranging from 0.168 to 0.202. On the other hand, the use of Heuristics led to a significantly lower percentage of correct claims, with a medium effect size of 0.270.

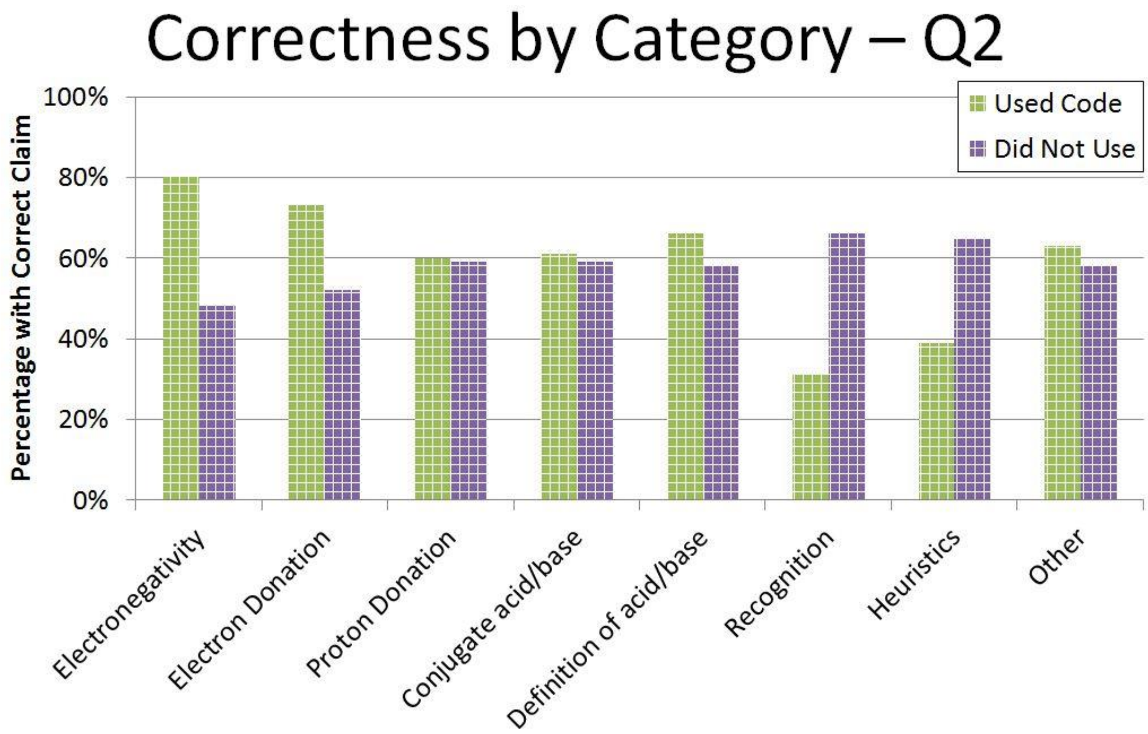


Figure 5.12: Percentage of correct claims by the use or lack of each category in Question 2.

Table 5.19: Chi-square tests for each category in Question 2, with p-values and Phi effect sizes.

Code (N = 383)	**Used Code**		**Did Not Use Code**		**Chi-square**	**p-value**	**Phi**
	Number of Students	**Correct %**	**Number of Students**	**Correct %**			
Electronegativity	131	80%	252	48%	36.80	**<0.001**	**0.310**
Electron Donation	123	73%	260	52%	20.35	**<0.001**	**0.231**
Proton Transfer	126	60%	257	59%	0.021	0.886	0.007
Conjugate Acid/Base	28	61%	355	59%	0.036	0.849	0.010
Definitions of Acid/Base	62	66%	321	58%	1.551	0.213	0.064
Recognition	75	31%	308	66%	30.97	**<0.001**	**0.284**
Heuristics	88	39%	295	65%	19.60	**<0.001**	**0.226**
Other	72	63%	311	58%	0.447	0.504	0.034

Similarly comparisons were performed for all the responses to Question 2, and the results were shown in **Figure 5.12** and **Table 5.19**. Again, the use of Electronegativity and Electron Donation led to significantly higher percentages of correct claims, with small to medium effect sizes (0.310 and 0.231); while the use of Recognition and Heuristics led to significantly lower percentage of correct claims, with small to medium effect sizes (0.284 and 0.226). While Recognition did not significantly affect the percentage of correct claims in Question 1, it led to significantly lower percentage of correct claims in Question 2. This can be explained by the fact that the two compounds involved in Question 1 are more common than the two involved in Question 2. Some students who used recognition in solving Question 1 recognized that ammonia was a base, and were able to reach the correct claim that water must be the stronger acid; while some other students recognized water as neutral, and wrongly deducted that ammonia must be the stronger acid. But in Question 2, almost all the students who used recognition were recognizing the OH group as a sign of base, so the use of recognition in Question 2 led to a significantly lower percentage of correct claims.

Lastly, students were also divided into three levels based on the categories they used in solving each question, to see if different levels have different percentages of correct claims in either question:

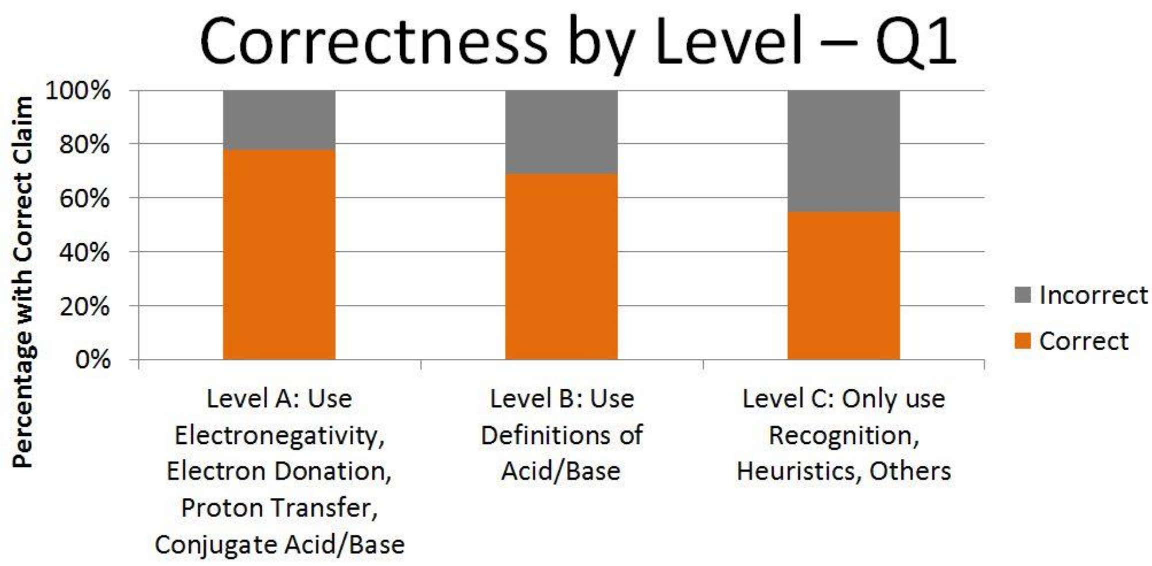
***Level A**: Students in this level used one or more categories from Electronegativity, Electron Donation, Proton Transfer, and Conjugate Acid/Base in their explanations, suggesting that they take one or more of these aspects into consideration when determining the acid base behaviors.*

***Level B**: Students in this level did not use any of the above categories, but used the category "Definition of Acid/Base". Initially, the category code "Definition of Acid/Base" was created to capture students who did not explain how the properties of the substance related to its ability to behave as an acid or base; so this is an "intermediate level" where students seem to have a vague idea of acid base behaviors.*

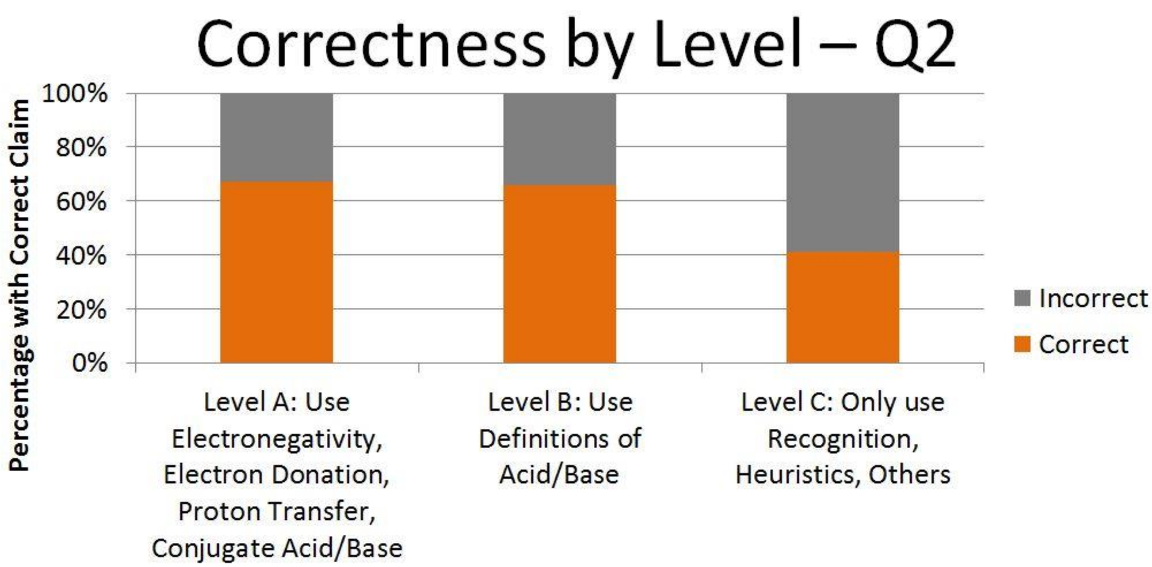
***Level C**: Students in this level only used one or more categories from Recognition, Heuristics, and Others. Rather than deducting acid base strength from related structure features, students in this level rely solely on surface features in reaching their conclusions.*

The differences in percentages of correct claims in each category were plotted in **Figure 5.13**. Chi-square tests without Yate's correction were performed, but with three groups instead of two (resulting in a three by two table instead of a two by two table, with three levels and two outcomes – correct claim or incorrect claim). Cramer's V was again simplified into Phi effect size, because k in the formula is the less of the number of rows and the number of columns for the categorical data (in this case, it is again 2 because there were only two possible outcomes).

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$



(a)



(b)

Figure 5.13: Percentage of correct claims by level. a) Question 1; b) Question 2.

Table 5.20: Chi-square tests for three levels with p-values and Phi-effect sizes, for both questions.

Level		Use Electronegativity, Electron Donation, Proton Transfer, Conjugate Acid/Base	Use Definitions of Acid/Base	Only use Recognition, Heuristics, Others
Q1	**Number of Students**	227	64	92
	Correct%	78%	69%	55%
	Chi-square	17.13		
	p-value	**<0.001**		
	Phi	**0.211**		
Q2	**Number of Students**	206	62	115
	Correct%	67%	66%	41%
	Chi-square	22.37		
	p-value	**<0.001**		
	Phi	**0.242**		

As shown in **Table 5.20**, Chi-square tests suggest that these three levels led to significantly different outcomes. Initially, Chi-square tests were performed on all three levels together, because the initial hypothesis was that all three groups are equivalent. However, now that overall Chi-square tests revealed significant differences, each two levels were compared to see where the differences came from.

Table 5.21: Chi-square tests comparing each two levels for both questions.

Question	**Levels Compared**	**Chi-square**	**p-value**	**Phi**
Q1	Level A and Level B	2.578	0.108	0.094
	Level B and Level C	2.810	0.094	0.134
	Level A and Level C	17.07	**<0.001**	**0.231**
Q2	Level A and Level B	0.016	0.900	0.008
	Level B and Level C	10.28	**0.001**	**0.241**
	Level A and Level C	20.62	**<0.001**	**0.253**

As shown in **Table 5.21**, for Question 1, students in Level A (78% correct claim) significantly outperformed students in Level C (55% correct claim) with a small-medium effect size of 0.231; for Question 2, students in both Level A (67% correct claim) and Level B (66% correct claim) significantly outperformed students in Level C (41% correct claim), with small-medium effect sizes (0.241 and 0.253).

Summarizing the last two analyses, students using some category codes listed in Level A are more likely to reach the correct claims than students using some other category codes listed in Level C. This finding is not surprising because Level A represents students who associate acid base behaviors more with a molecular explanation, whereas Level B and C represents students who infer acid base behaviors more from a surface level. This result further confirmed the earlier hypothesis that CLUE students outperformed their Traditional counterparts in both questions across two years because they approached these questions more from categories that are more likely to lead to correct claims, such as Electronegativity and Electron Donation, and less from categories that are more likely to lead to incorrect claims, such as Recognition and Heuristics. So far, analyses of both the claims and the explanations of students' responses suggested that students in the CLUE curriculum significantly outperformed students in the Traditional curriculum in both years and both questions.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

The three research questions asked at the beginning of this study have been answered to some extent, as summarized below.

RQ 1: What are the important characteristics of activities that effectively promote and retain argumentation skills among college students?

The BeSocratic activity seemed to significantly promote argumentation skills among different groups of college students for a short among of time. Two important characteristics of this activity are: 1) the instruction of complete scientific explanation based on Toulmin's argumentation pattern; and 2) allowing students to edit their initial responses. First of all, many students did not know what kinds of answers are complete or incomplete; the instruction based on Toulmin's argumentation pattern gave them a clear framework as to how to construct complete scientific explanations, and an example (in biology) to ensure they understood all the terminologies in the Toulmin's argumentation pattern correctly. Then students were immediately offered the opportunity to review their initial answers, and use Toulmin's argumentation pattern to judge if their initial answers were complete or not, and make changes to their initial answers if deemed incomplete. This step not only helps students to immediately apply the theory they have just learned, but also offers a contrast using each student's own answers. The results seem to echo with Gadgil, Nokes-Malach, & Chi's recent report (Gadgil, Nokes-Malach, & Chi 2012) that asking students to "contrasts his or her flawed mental model to an expert model" seem to better help students acquire a correct mental model at least in short term, in the sense that both methods asked students to confront their initial answers with a better answer as a part of the instruction plan; and in both studies, students performed significantly better in the post-test immediately following instruction.

On the other hand, whether it is possible to move this significant change into a long-term effect through several short interventions (10-15 minutes each), and if so how to do that, still remain as questions for future research.

RQ 2: In what ways is argumentation an effective assessment method for student understanding of acid base models?

Students' responses from the BeSocratic activity can also be used to assess how students in different curricula approach the same question differently. The correctness of their claims in both questions can be used as a quick quantitative measurement of students' content knowledge in the area of acid base chemistry. A coding scheme was developed with eight category codes to further analyze the initial approaches students in different curricula were more likely to take in solving a particular problem. Using these parameters, data collected during the BeSocratic activity in both Spring 2012 and Spring 2013 from students in CLUE curriculum as well as Traditional curriculum was analyzed. The same activity was administered two years in a row to determine the reproducibility of its results. The fact that the CLUE cohorts from both years performed very similarly to each other, and the Traditional cohorts from both years also performed very similarly to each other, suggest that this BeSocratic activity and the analysis methods developed from it is a reliable method to assess students' understanding of acid base models and to compare the differences between student groups from different general chemistry curricula. The fact that responses from several hundred students were analyzed by this method also answered the earlier call for an effective assessment that can be used on large sample sizes in research studies of how differences among instructors, teaching styles, teaching interventions, etc affect large groups of students on average. It was indicated in the earlier chapters that in order to design different interventions or even curricula to facilitate more meaningful learning among students, we need to first be able to assess such large populations with a reliable yet effective method that could be administered to large populations and still maintain some degree of individual students' beliefs. The development of the BeSocratic activity described in this research serves as a good example of the types of assessments we could design in future for further studies of large populations of students.

RQ 3: How do different curricula affect students' acid base models?

As a whole, students who associated acid base behaviors more with molecular explanations such as electronegativity, electron donation, conjugate acid base strength, etc were more likely to reach the correct conclusions about acid base strengths than students who inferred acid base behaviors from surface level features such as the existence/absence and/or the number of particular atoms and/or functional groups. This is not surprising because associating acid base behaviors with molecular explanations requires system II thinking. As mentioned in Chapter two, the process of thinking was categorized into two types according to the Dual process theory: system I thinking often uses heuristics (instructed or self-developed) to quickly solve a problem without engaging in detailed analysis; whereas system II thinking is much slower and engages in detailed analysis. Although more time consuming, it is also more often correct. On the other hand, inferring acid base behaviors from surface level features such as the presence or absence of a particular atom or functional group only requires system I thinking – although faster, it is more often incorrect. The challenge, however, is how to encourage students to employ system II thinking when solving such problems, rather than simply resort to system I thinking because it is much faster.

The results from this study suggest that students from the CLUE curriculum were more likely to consider factors such as electron negativity, electron donation, and conjugate acid base strengths (system II thinking), whereas their Traditional counterparts were more likely to rely on recognition and/or heuristics (system I thinking). As a result, students from the CLUE curriculum consistently outperformed students from the Traditional curriculum in the two questions investigated. These results were fairly consistent and reproducible across two years. Such a significant and consistent difference may arise from the different design of the CLUE curriculum: instead of arranging the chapters and topics of the curriculum based on a conventional order (Johnstone 2010), the CLUE curriculum connects structures with properties, and encourages students to construct the new knowledge in a more relevant and meaningful way. Such an approach seemed to help and encourage students to develop more advanced acid base models and rely less on simple heuristics.

Several potential future directions based off this research include:

1) Continue to explore methods that can help students retain argumentation skills and get into a habit of articulating their answers in open-ended questions. Students'

ability to construct and defend their explanations was highlighted in the National Research Council (NRC) Framework for Science Education; however, students' initial responses to the first question in the BeSocratic activity (prior to the instruction of making scientific argumentations) revealed that most students from either curriculum were not able to offer well-constructed explanations to a scientific question. Although short interventions such as the BeSocratic activity designed in this study can temporarily help students in constructing scientific argumentations, it is hard to imagine the effect of such a short intervention lasting for a long time. As a continuation of this study, we could design a new study to see how long the effect of a single BeSocratic activity would last on average, and even to test whether repeating similar activities could prolong such an effect. However, it is possible that a lasting effect of improved scientific argumentation abilities among students can only be achieved when multiple courses adopt curricula that would teach, encourage, and continue to remind students to construct new knowledge to their existing knowledge, and offer students opportunities to construct, revise, and defend their scientific explanations.

2) Further exploring of the rich qualitative data collected in this study. Because this study aimed at finding an effective assessment to compare the differences between two different curricula with large student populations, some degrees of individuality were sacrificed in the analysis of data. For example, less popular responses (codes with less than 10% appearance in any cohort) were combined into one category. Although less frequently used, detailed analysis of some of these strategies could reveal important beliefs students hold, and we may even be able to trace back and see where such beliefs

initially came from. For example, quite a few students used polarity of a molecule to support their conclusion – was that because they confused the polarity of a molecule with the polarizability of a bond? Several students also generated their own “trends” to help them determine acid/base strengths faster – where did those “trends” come from? Another possible angle to further explore the qualitative data would be to further separate some of the categories developed in this research. For example, among the students who used the likelihood of proton transfer to determine acid strengths, some described the transfer as the base “robbing” or “stealing” the proton from the acid, some others described the transfer as the proton “hopping” onto the base, yet some others described the transfer as the proton “wandering around” in the solution, then suddenly “found” the base. These beliefs about the proton transfer process are quite different from each other and worth further exploring. An inevitable limitation of open-ended questions is the limited amount of information they can elicit from each student – if a student did not mention something in his answer, we could not know whether it was because the student did not know the particular information, or the wording of the question did not elicit the particular information from the student. Although the argumentation training can temporarily prompt students to elaborate their ideas more, many beliefs students hold may still not be elicited. Individual interviews could best elicit each student’s beliefs, but as discussed earlier, it will not be an effective assessment for a large student population. Alternatively, questions can be worded more “specifically” by telling students exactly what concepts we are looking from them; but that could defeat the purpose of eliciting what they truly believe – when provided with a “guided route”, many students will try to fit their answers accordingly, rather than freely construct their answers. Different methods as discussed above can provide different insights into students' understandings in the area of acid base chemistry from different angles. Not one of them is "the correct one" or "the complete one". This research chose the angle to understand the averages of large groups of students; other future researches can choose different angles based on the research interest.

3) Further investigate how students' acid base mental models can affect their understanding of reaction mechanisms and ability to predict correct mechanisms and products for novel organic acid base reactions. Prior research in the Cooper Research Group revealed that when asked to draw reaction mechanisms, many undergraduate organic students only add the curved mechanism arrows after they had finished the entire reaction (Grove, Cooper, & Rush 2012). It is not a far reach to hypothesize that a good understanding of Lewis acid base models can help students build many organic chemistry mechanisms upon it when they move from general chemistry to organic chemistry, since nucleophiles and electrophiles can be considered Lewis bases and Lewis acids, respectively. However, little research has been reported to prove such a connection. This research revealed the plausibility of using BeSocratic system to assess students' acid base mental models for large populations, making it possible to follow up on large populations of general chemistry students as some of them move into organic chemistry to see how their acid base mental models change as they move through these courses, and how their acid base mental models affect their learning and ability to use reaction mechanisms.

CHAPTER SEVEN

IMPLICATIONS FOR TEACHING

Both students' open-ended responses during the preliminary studies and students' initial responses at the beginning of the BeSocratic activity suggested that a majority of students were not trained to articulate their ideas. Although the NRC Framework for Science Education specifies that "students need to construct and defend their explanations, the interpretations that they offer based on data or the solutions they propose", students in large college science classes lack the opportunities and trainings to construct and defend their explanations, as most examinations only contain multiple-choice questions. Although some short interventions such as the BeSocratic activity described in chapter five of this research can temporarily improve students' argumentation skills, it is very difficult to design short interventions with lasting effects when students were not taught and/or encouraged to articulate their ideas on a consistent base. To meet the standards in the NRC Framework, students would first need to be instructed on how to properly articulate their ideas, and then be consistently reminded of such practice and encouraged to articulate their ideas. While the first part can be done relatively easily (using short interventions such as the BeSocratic activity designed in this study), the second part requires a lot more effort from different departments collectively. Even if students occasionally come across a few more classes in which they are encouraged to articulate their ideas in class, as long as the majority of their classes do not require such a practice from them, it is very hard for them to form a habit of articulating their ideas. Yet it is far from an easy task to create an environment in which students are consistent encouraged and expected to articulate their ideas – this may require a change of curriculum and/or teaching method, etc. This becomes even more challenging for the large size general education classes college students take in the first couple of years of their programs. Furthermore, even if all those classes could together create an environment to promote scientific argumentation on a consistent base, students still may not give scientific argumentation its due priority as long as their performances in those classes are not evaluated with adequate assessments (namely, if the examinations that determine students' grades in those classes are still multiple-choice only). This poses an even bigger challenge in the administrative aspect of course designing since many courses have very large student populations per section (for example, several hundred students per section and several thousand students per semester for a general education requirement course at a large university). So far, a common approach is to "keep it simple" by giving multiple-choice examinations and even use technologies such as scantron readers to make grading faster and easier. However, we have to ask the fundamental question: is this "simple" method getting the students to where we want them to be? If the answer is no, then we will have to move to a "not so simple" approach that would actually get us to the place we want to be. Some teachers have started using a combination of open-ended questions and multiple-choice questions in their examinations, which is a good start to encouraging students to articulate their ideas. However, as mentioned above, as the size of the class increase, the challenge also increases with incorporating open-ended questions in examinations. More resources (teachers, TAs, etc) are needed to grade such open-ended questions, and a detailed grading rubric for each open-ended question must be developed if multiple graders are involved in grading the same examination for large classes.

Secondly, this research revealed that many students have fragmented ideas about acid base concepts – very few students can offer a complete and completely correct explanation to either of the two questions investigated in this study (as shown in Appendix C). Many students, however, were fairly consistent in the approaches they take to infer acid base behaviors (**Figures 5.7-5.10**, **Tables 5.14-5.17**). Moreover, this research revealed that students who associated acid base behaviors more with a molecular explanation (considering factors such as electronegativity, electron donation, conjugate acid base strength, etc) were more likely to reach the correct conclusions about acid base strengths than students who inferred acid base behaviors more from a surface level (relying on recognition and/or heuristics). Thus, an immediate question for the teachers to consider would be: how to help students approach acid base problems from molecular level behaviors instead of relying on heuristics and/or recognition?

Although individual students will finally have to make the choice of whether to use system I thinking or system II thinking – namely, whether to make a quick guess based on surface features or to reason acid base behaviors from a molecular level – two factors can affect their choices. The first factor ties with the construct of scientific arguments. If a student is asked to not only pick an answer, but also construct a scientific explanation to support this answer, then the student might be more encouraged to think through and consider more factors that would support the answer, rather than making a quick guess based on one surface feature and then immediately move on to the next question. To help students form such a habit in class, teachers could show examples of how to predict acid base behaviors from molecular levels, and even contrast the results with what a "quick guess" based on the surface level would most likely be. Through such comparison, students may take notice that many times quick guesses based on a particular surface feature are likely going to lead to the wrong answer, and consequently be more encouraged to approach similar problems from molecular levels when solving similar problems on their own.

The second factor is the integration of knowledge – namely, if a student cannot construct everything he learned in class about acids and bases in a meaningful way in his own understanding, then even if he attempts to construct a scientific argumentation to support his conclusion about an acid base question, the scientific argumentation he constructs would be limited by the fragmented knowledge he has in this area. This is where the effect of different curricula comes in. A curriculum that constantly connects structures with properties, such as the CLUE curriculum, would encourage students to construct the new knowledge in a more relevant and meaningful way. Some other research in the Cooper Research Group has reported students in the CLUE curriculum demonstrated "an improved understanding of structure-property relationships" (Cooper, Underwood & Hilley 2012). Similarly, students in the CLUE cohorts in this study also demonstrated more frequent uses of molecular level acid base properties in solving the questions examined, and thus resulting in much higher success rates of arriving at the correct claims for these questions. It should be noticed that both the CLUE cohorts and the Traditional cohorts from Spring 2012 and Spring 2013 involve multiple instructors.

The fact that the two CLUE cohorts performed very similarly, and the two Traditional cohorts also performed similarly, suggest that the effects different curricula have on students cannot be replaced by different teaching methods or pedagogies. While teachers can explore different teaching methods and pedagogies to see if any particular method(s) would help students form the habit of approaching acid base problems from molecular levels rather than macroscopic or surface levels, the choice of an appropriate curriculum is also important, and can affect students' understandings from a different aspect than the teaching methods and pedagogies. Curricula based on educational research findings are thus preferable. Both pedagogy and curriculum must go hand in hand to help students develop better understandings in the area of acid base chemistry.

APPENDICES

Appendix A

Interview Protocol

1. Please briefly talk about your major, year, and the chemistry courses you've taken.

2. Are you taking any chemistry course this semester? If so, what course are you taking and why are you taking this particular course?

The first two questions served the purposes of gathering background information about the student (such as level of chemistry courses taken) and putting the students at ease for the interview.

Part I:

3. If you are going to describe what an acid is to a friend who has never taken any chemistry courses, what would you say?

- Could you give an example? Could you draw the structure?
- Why do you think it is an acid?
- What are the features in your drawing that makes it acidic?

4. If your friend who is in the same chemistry class with you had missed the acid/base chapter and is asking you to explain what is an acid, what would you say?

- Could you draw a couple of examples?
- What are the features in your drawing that makes it acidic?

5. Concluding from your examples, what are some essential features of acids?

6. If you were given a list of structures to identify the acids among them, how would you do it?

- Could you draw a couple of examples? (follow up on each "strategy" a student mentioned)

Question 3-6 were then repeated for base.

Part II:

7. Below are a list of compounds, please go through one by one and identify whether it is an acid. Please think out loud as you go through.

For Part II, if a student did not use Lewis structures in the reasoning, the interviewer would later ask the student to draw the Lewis structures, and help the student with getting the correct Lewis structure if necessary, and then ask the student whether the Lewis structure has changed their previous decision on the compound.

Below is the complete list of all the compounds used in different interviews. As stated in section 4.1, students at different levels were given different structures from the complete list to allow the study of a larger variety of compounds without overwhelming each individual student with extended interview time and questions.

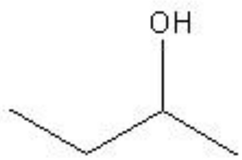
a. HCl
b. HI
c. H_2SO_4
d. BH_3
e. BF_3
f. NH_3

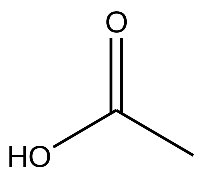
g. PH_3

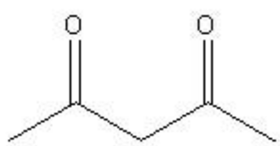
h. CH_4

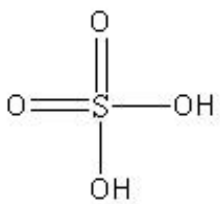
i. LiOH

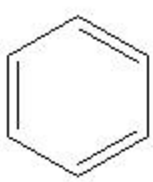
j. CH_3OH

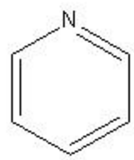
k. 

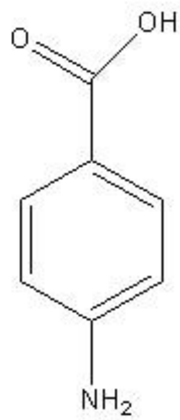
l. 

m. 

n. 

o. 

p. 

q. 

<u>Appendix B</u>

<u>Screenshots of the Steps in the BeSocratic Activity</u>

Based on the Lewis structures of ammonia (NH3) and water (H2O) below, which one of them is a stronger **<u>acid</u>**? Please explain your reasoning.

Water is a stronger acid because of electronegativity.

When finished, click Next to continue.

(a) Initial question and student responses collected.

Making a Scientific Explanation

It is important when answering scientific questions or making a scientific explanation that you provide a **complete** answer.

There are three main parts when you construct a complete scientific explanation:

1. You must clearly state your answer (claim) to the question;

2. You must provide the reason(s) for making your claim (evidence or data);

3. Then you have to link the evidence to the claim (the answer) – by explaining **why** that evidence supports your claim.

When finished, click Next to continue.

(b) Students were then instructed on how to make a complete scientific explanation.

Now let's look at an example:

What color are plant's leaves in the summer?

A plant's leaves are green in the summer because they contain chlorophyll. Chlorophyll is a natural pigment that absorbs all wavelengths in the visible spectrum except green. The reflected green light is the color our eyes see.

Can you identify the claim, the data or evidence, and the explanation in the above example respectively?

When finished, click Next to continue.

(c) A biology example was used to ensure students understood the different components of a complete scientific explanation.

Here is your initial response to the question "Based on the Lewis structures of ammonia (NH3) and water (H2O) below, which one of them is a stronger **acid**? Please explain your reasoning."

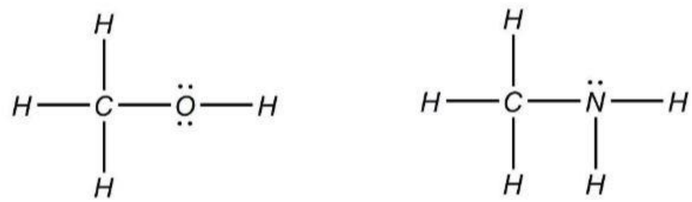
Now that you have the three steps of making an explanation, how would you change your previous answer? You may edit the response directly.

Water is a stronger acid because of electronegativity. Oxygen is more electronegative than nitrogen, so the hydrogens attached to oxygen are more partially positive and can be donated easier.

When finished, click Next to continue.

(d) Students were given the opportunity to revise their initial responses (the highlighted parts were added).

Now let's look at two slightly more complicated compounds. Based on the Lewis structures of methanol (CH3OH) and methanamine (CH3NH2) below, which one of them is a stronger **base**? Please explain your reasoning.



methanol *methanamine*

When finished, click Submit.

(e) Students were given a new question without any reminder of making a complete scientific explanation, and their responses were collected.

Appendix C

First Five Complete Responses with Correct Claims From Traditional 2013 Cohort and From CLUE 2013 Cohort

Student Response to Q1	Problem(s) with Response
I would say that based on the Lewis Structure that water is the stronger acid. This is because the hydrogen atoms are more likely to leave the water molecule than the ammonia molecule. This is due to the fact that oxygen is more electronegative than nitrogen and therefore attracts electron more. Therefore, the hydrogen atoms are more likely to leave the water molecule than the ammonia molecule. Since an acid is a proton donor, water is the better proton donor and is the better acid.	N/A.
H2O is the stronger acid. It is stronger because oxygen is more electronegative than nitrogen and therefore can disperse the negative charge better thus making it a stronger acid, where the hydrogen atom is harder to remove.	Student did not clearly explain how the negative charge got on oxygen (conjugate base after donating a proton), and seems to be confused at the end (hydrogen atom should be "easier" to remove, not "harder").
H2O is the stronger acid because it the oxygen is more electronegative. Because of this it steals electrons from the hydrgens and makes them more willing to break off.	The term "stealing" does not describe the uneven share of electrons in the O-H bond appropriately, and the hydrogen atoms would not "break off" on their own, but need to be attracted by another atom willing to donate its lone pair.
The water is the stronger acid. This is because the oxygen is more electronegative than the nitrogen and therefore has stronger bonds with its hydrogens. You can determine the electronegativty of the oxygen by looking at the trend on the peridoic table and determining that it is stronger as it is farther to the right.	Electronegativity does not determine bond strength, nor does bond strength determine acid strength.

Water is the stronger acid because it has two pairs of electrons it can donate. Also, ammonia has hydrogen bonding, so those hydrogens would be more difficult to donate. This would cause water to allow more dissociation which will result in a lower pH.	Student mistook acids as electron donors, and failed to realize that water also has hydrogen bonding. Regardless, hydrogen bonding would not determine acid strength.
water is a stronger acid than the ammonia molecule because the O-H bond is more polarized so it is more likely to attract a electron pair from a base (electron donor)	Student did not explain why the O-H bond is more polarized (because oxygen is more electronegative than nitrogen), and did not clearly explain why a more polarized bond is more likely to attract an electron pair.
Water is the stronger acid because oxygen is more electronegative than nitrogen, so it would be less likely to give up its electrons the form a bond with a proton from ammonia. Therefore, the electron pair on the nitrogen would form a bond with a hydrogen atom from water and the electrons in the bond between that hydrogen and the oxygen would become a lone pair on the oxygen.	Student seemed to consider electron donation as losing or giving up the lone pairs – a common alternative concept among many students. But what percentage of students hold this concept and the reason(s) why students hold this concept would call for a different study. Nevertheless, this student response was rendered "not completely correct".
The stronger acid would be H2O because water is more likely to accept an electron pair than ammonia. This is due to the fact that oxygen is more electronegative than nitrogen and more likely to hold on to more electrons. This is turns causes the conjugate base of water to be more stable than the conjugate base of ammonia.	Student seemed to be using Lewis' acid base model in the first half and Bronsted-Lowry's acid base model in the second half, without a clear explanation how "more likely to accept an electron pair" would lead to a more stable conjugate base.
H20 is the stronger acid because Oxygen is the more electronegative atom. Because of this, it wants to hold on to its electrons more than nitrogen. Because the oxygen wants to hold on to its electrons, it is the electrophile in this situation and is a stronger acid than ammonia.	Student did not clearly explain why oxygen being the electrophile and holding onto its electrons would render water a stronger acid.
H2O is the stronger acid because it more electronegative so it is willing to give it a lone pair to N	Since oxygen is more electronegative, it should be less willing to share its lone pairs.

Appendix D

Totaling of Codes in Each Cohort

Cohort	N	Question	Electronegativity	Electron Donation	Proton Donation	Conjugate Acid/Base	Definitions of Acid/Base	Recognition	Heuristics	Other
Traditional 2012	70	Q1	14 (20.0%)	7 (10.0%)	24 (34.3%)	4 (5.7%)	11 (15.7%)	22 (31.4%)	21 (30.0%)	28 (40.0%)
		Q2	14 (20.0%)	6 (8.6%)	23 (32.9%)	2 (2.9%)	9 (12.9%)	28 (40.0%)	24 (34.3%)	19 (27.1%)
CLUE 2012	107	Q1	44 (41.1%)	40 (37.4%)	48 (44.9%)	29 (27.1%)	22 (20.6%)	14 (13.1%)	18 (16.8%)	23 (21.5%)
		Q2	39 (36.4%)	43 (40.2%)	39 (36.4%)	14 (13.1%)	23 (21.5%)	5 (4.8%)	24 (22.4%)	20 (18.7%)
Traditional 2013	91	Q1	32 (35.2%)	7 (7.7%)	31 (34.1%)	4 (4.4%)	17 (18.7%)	20 (22.0%)	27 (29.7%)	21 (23.1%)
		Q2	12 (13.2%)	7 (7.7%)	32 (35.2%)	0 (0.0%)	18 (19.8%)	40 (44.0%)	18 (19.8%)	13 (14.3%)
CLUE 2013	115	Q1	71 (61.7%)	66 (57.4%)	27 (23.5%)	13 (11.3%)	14 (12.2%)	6 (5.2%)	14 (12.2%)	29 (25.2%)
		Q2	66 (57.4%)	67 (58.3%)	32 (27.8%)	12 (10.4%)	12 (10.4%)	2 (1.7%)	22 (19.1%)	20 (17.4%)

REFERENCES

Ausubel, D. P. (1968). *Educational psychology: a cognitive view*. New York, NY: Holt, Rinehart, & Winston, 1968.

Bhattacharyya, G. (2006). Practitioner development in organic chemistry: how graduate students conceptualize organic acids. *Chemistry Education Research and Practice*, 2006, 7 (4), 240-247.

Chi, M. T. H., Slotta, J. D., & DeLeeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Learning and Instruction*, 1994, 4 (1), 27-43.

Chi, M.T.H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. *Handbook of research on conceptual change* (edited by Vosniadou). Hillsdale, NJ: Erlbaum, 2008.

Cole, R., Becker, N., Towns, M., Sweeney, G., Wawro, M., & Rasmussen, C. (2012). Adapting a methodology from mathematics education research to chemistry education research: documenting collective activity. *International Journal of Science and Math Education, 2012, 10(1), 193-211*.

Cooper, M. M., Corley, L. M., & Underwood, S. M. (2013). An investigation of college chemistry students' understanding of structure–property relationships. *Journal of Research in Science Teaching*, 2013, 50 (6), 699-721.

Cooper, M. M., Grove, N. P., Pargas, R., Bryfczynski, S. P., & Gatlin, T. (2009). OrganicPad: an interactive freehand drawing application for drawing Lewis structures and the development of skills in organic chemistry. *Chemistry Education Research and Practice*, 2009, 10 (4), 296–301.

Cooper, M. M., Sandi-Urena, S., Stevens, R. (2008). Reliable multi method assessment of metacognition use in chemistry problem solving. *Chemistry Education Research and Practice*, 2008, 9 (1), 18–24.

Cooper, M. M., Sandi-Urena, S. (2009). Design and Validation of an Instrument to Assess Metacognitive Skillfulness in Chemistry Problem Solving". *Journal of Chemical Education*, 2009, 86 (2), 240-245.

Cooper, M. M., Underwood, S. M., & Hilley, C.Z. (2012). Development and validation of the implicit information from Lewis structures instrument (IILSI): Do students connect structures with properties? *Chemical Education Research and Practice*, 2012, 13 (3), 195-200.

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, 1993, 10 (2-3), 105-225.

DiSessa, A. A. (2006). A history of conceptual change research: threads and fault lines. *The Cambridge Handbook of the Learning Sciences* (edited by Sawyer). New York, NY: Cambridge University Press, 2006.

DiSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (from the "pieces" side of the fence). *International Handbook of Research on Conceptual Change* (edited by Vosniadou). New York, NY: Routledge, Taylor, & Francis, 2008.

Demerouti, M., Kousathana, M., & Tsaparlis, G. (2004). Acid-base equilibria, Part I. Upper secondary students' misconceptions and difficulties. *The Chemical Educator*, 2004, 9 (2), 122-131.

Demircioglu, G., Ayas, A., & Demircioglu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice*, 2005, 6 (1), 36-51.

Erduran, S., Simon, S., & Osborne, J. F. (2004). TAPping into argumentation: developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, 2004, 88 (6), 915–933.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 2003, 7(10), 454-459.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, 2012, 22 (1), 47-61.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: the psychology of intuitive judgment*. Cambridge: Cambridge University Press, 2002.

Grove, N. P., Cooper, M. M., & Rush, K. M. (2012). Decorating with arrows: toward the development of representational competence in organic chemistry. *Journal of Chemical Education*, 2012, 89 (7), 844-849.

Halstead, S. E. & Anderson, T. R. (2009) What can we learn from studies on students' conceptual difficulties? The case of propositional knowledge of acid-base models.

Proceedings of the Seventeen Annual Meeting of the Southern African Association for Research in Mathematics, Science, and Technology Education, 2009, 1, 82-93.

Hammer, D. (1996) Misconceptions or P-Prims: How May Alternative Perspectives of Cognitive Structure Influence Instructional Perceptions and Intentions? *The Journal of the Learning Sciences,* 1996, 5(2), 97-127.

Howe, A. C. (1996). Development of science concepts within a Vygotskian framework. *Science Education*, 1996, 80 (1), 35-51.

Jasien, G. P. (2005). Numerical and pictorial understanding of acids among science students. *The Chemical Educator*, 2005, 10 (5), 400-405.

Johnstone, A. H. (2010). You can't get there from here. *Journal of Chemical Education*, 2010, 87 (1), 22-29.

Maeyer, J. & Talanquer, V. (2010). The role of intuitive heuristics in students' thinking: ranking chemical substances. *Science Education*, 2010, 94 (6), 963–984.

McClary, L. M. & Bretz, S. L. (2012). Development and Assessment of A Diagnostic Tool to Identify Organic Chemistry Students' Alternative Conceptions Related to Acid Strength. *International Journal of Science Education*, 2012, 34(5), 2317-2341.

McClary, L. M. & Talanquer, V. (2011a). Heuristic reasoning in chemistry: making decisions about acid strength. *International Journal of Science Education*, 2011, 33 (10), 1433-1454.

McClary, L. M. & Talanquer, V. (2011b). College chemistry students' mental models of acids and acid strength. *Journal of Research in Science Teaching*, 2011, 48 (4), 396-413.

Novak, J. D. (1978). An alternative to Piagetian psychology for science and mathematics education. *Studies in Science Education*, 1978, 5 (1), 1-30.

Novak, J. D. (1993). Human constructivism: a unification of psychological and epistemological phenomena in meaning making. *International Journal of Personal Construct Psychology*, 1993, 6 (2), 167-193.

Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 2002, 86 (4), 548-571.

Osborne, J. F., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 2004, 41 (10), 994–1020.

Ozmen, H., Demircioglu, G., & Coll, R. K. (2009). A comparative study of the effects of a concept mapping enhanced laboratory experience on Turkish high school students' understanding of acid-base chemistry. *International Journal of Science and Mathematics Education*, 2009, 7 (1), 1-24.

Pendley, B. D. & Bretz, R. L. (1994). Concept maps as a tool to assess learning in chemistry. *Journal of Chemical Education*, 1994, 71 (1), 9-15.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Science Education*, 1982, 66(2), 211-217.

Rasmussen, C. & Stephan, M. (2008). A methodology for documenting collective activity. *Design Research in Education* (edited by Kelly & Lesh). Mahwah: Erlbaum, 2008.

Ridley, D. R. & Novak, J. D. (1988). Assessing student learning in light of how students learn. Paper prepared for The AAHE Assessment Forum. American Association for Higher Education, 1988.

Shaffer, A. A. (2006). Let Us Give Lewis Acid-Base Theory the Priority It Deserves. *Journal of Chemical Education*, 2006, 83 (12), 1746-1748.

Simon, S., Erduran, S., & Osborne J. F. (2006). Learning to teach Argumentation: Research and Development in the Science Classroom. *International Journal of Science Education*, 2006, 28(2-3), 235-260.

Toulmin, S. E. *The uses of argument*. Cambridge: University Press, 1958.

Vygotsky, L. (1962). *Thoughts and language*. Cambridge, MA: MIT Press, 1962.