International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

Development and Assessment of A Diagnostic Tool to Identify Organic Chemistry Students' Alternative Conceptions Related to Acid Strength

LaKeisha M. McClary [a] & Stacey Lowery Bretz [a]
[a] Department of Chemistry and Biochemistry, Miami University, Oxford, USA

Available online: 08 May 2012

International Journal of Science Education
2012, 1–25, iFirst Article

Development and Assessment of A Diagnostic Tool to Identify Organic Chemistry Students' Alternative Conceptions Related to Acid Strength

LaKeisha M. McClary and Stacey Lowery Bretz*
Department of Chemistry and Biochemistry, Miami University, Oxford, USA

The central goal of this study was to create a new diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. Twenty years of research on secondary and college students' conceptions about acids and bases has shown that these important concepts are difficult for students to apply to qualitative problem solving. Yet, few published studies document how students' prior knowledge of acids influences their understanding of acid strength in organic chemistry contexts. We developed a nine-item multiple-tier, multiple-choice concept inventory to identify alternative conceptions that organic chemistry students hold about acid strength, to determine the prevalence of these conceptions, and to determine how strongly these conceptions bias student reasoning. We identified two significant alternative conceptions that organic chemistry students hold about acid strength. Students who answered items incorrectly were more confident about their answers than peers who answered items correctly, suggesting that after one semester of organic chemistry, students do not know what they do not know. Implications for the teaching of acid strength are discussed.

Keywords: *Alternative Conception; Chemistry Education; Diagnostic Assessment; Organic Chemistry*

Introduction

Acid–base chemistry is a central concept in the discipline; it is useful to explain physical and chemical phenomena we observe in the natural world. Students must

*Corresponding author. Department of Chemistry and Biochemistry, Miami University, 651 E. High Street, Oxford, 45056, USA. Email: bretzsl@muohio.edu

ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/12/000001–25

http://dx.doi.org/10.1080/09500693.2012.684433

recognize macroscopic properties of acids, conceptualize how particles behave submicroscopically, and manipulate symbolic representations of acid–base reactions (Johnstone, 1993). Unfortunately, discussion of acid–base chemistry in introductory chemistry tends to center on algorithmic proficiency. Students in a typical introductory college chemistry course spend considerable time drawing initial-change-equilibrium tables and manipulating formulae to calculate pH, pK_a values, and concentrations of solutions. Relatively little time in general chemistry is devoted to helping students understand the limitations of scientific models of acids (i.e., Arrhenius, Brønsted–Lowry, and Lewis) or the interconnectedness of factors that affect acid strength (Furió-Más, Calatayud, Guisasola, & Furió-Gómez, 2005; de Vos & Pilot, 2001)

The limitations of the scientific models of acids and the nuanced relationship among the factors that affect acid strength, e.g., the structure of an acid and resonance, are necessary pre-requisite knowledge for students entering organic chemistry (Duis, 2011); such knowledge is required to make decisions about reactivity and reaction mechanisms (Anslyn & Dougherty, 2006; Bhattacharyya, 2006). Students who fail to develop coherent and relatively stable mental models of acids and acid strength often struggle to meaningfully understand reaction mechanisms (Bhattacharyya & Bodner, 2005; Ferguson & Bodner, 2008). To cope with the breadth of material covered in undergraduate organic chemistry curricula, many organic chemistry students rely on rote memorization without conceptually understanding chemical phenomena (Grove, Hershberger, & Bretz, 2008).

A number of studies have published findings that describe conceptions of acids held by secondary students (Hand, 1989; Hand & Treagust, 1988; Lin & Chiu, 2007; Rahayu, Chandrasegaran, Treagust, Kita, & Ibnu, 2011; Ross & Munby, 1991), general chemistry students (Carr, 1984; Cros et al., 1986; Maeyer & Talanquer, 2010; Nakhleh, 1994), and science teachers (Dreschsler & van Driel, 2008; Drechsler & Schmidt, 2005). Few studies have investigated conceptions of acids held by advanced college students, including organic chemistry undergraduates and organic chemistry graduate students. McClary and Talanquer (2011a) reported that undergraduates used implicit and explicit knowledge related to acids and acid strength to generate synthetic mental models of varying coherence (e.g., acids lose hydrogen atoms (H) or protons (H^+); the more easily H/H^+ is lost, the stronger the acid). Mental models and the saliency of features within qualitative ranking tasks guided students' reasoning strategies. Cartrette and Mayo (2011) recently reported that while students possessed declarative knowledge of acids and bases related primarily to general chemistry, few of them could use their knowledge to solve organic chemistry problems. Even with additional training in chemistry, difficulties with acid–base concepts persist into graduate school (Bhattacharyya, 2006).

We are particularly interested in alternative conceptions that organic chemistry undergraduates hold about acid strength. Most reactions taught in introductory organic chemistry can be classified as acid–base reactions or include at least one step involving an acid–base reaction (Bhattacharyya, 2006; Brown, Foote, Iverson, & Anslyn, 2009), yet few studies investigate these students' understandings of acids and bases (Cartrette & Mayo, 2011) or acid strength, in particular (McClary & Talanquer, 2011a, 2011b).

To the best of our knowledge, no concept inventory to assess organic chemistry students' understanding of acid strength has been reported. Thus, the purpose of this study was to use our prior research to develop a diagnostic tool to measure organic chemistry students' alternative conceptions about acid strength. With evidence regarding the prevalence of students' alternative conceptions about acid strength, chemistry teachers could use the findings of this research to better design curricula to improve learning about the concept of acid strength and its fundamental role in the discipline.

Theoretical Framework

Students enter science classrooms with preconceived conceptions, beliefs, and expectations about the natural world (Driver & Erikson, 1983). Such prior knowledge guides how students learn science (Bodner, 1986). When conceptions differ from scientifically accepted ones, researchers refer to students' non-canonical ideas by many terms, including alternative conceptions (Hewson & Hewson, 1984), misconceptions (Driver & Easley, 1978), preconceptions (Clement, 1982), alternative frameworks (Driver & Erickson, 1983), and naïve beliefs (McCloskey, Caramazza, & Green, 1980). The variety of terms reflects different epistemologies about the nature and origin of students' ideas about science (Smith, diSessa, & Roschelle, 1993–1994). In this paper, our use of the term alternative conceptions suggests that the ideas differ from scientific principles agreed upon by a community of practice, in this case, chemists. Because alternative conceptions can interfere with learning, research has focused on identifying them (Duit, 2009).

Science education researchers have yet to reach consensus on how knowledge, including alternative conceptions, is organized in the mind. Some researchers argued that knowledge is organized in a stable, coherent framework (Carey, 1985; Vosniadou & Brewer, 1987). Consequently, students' alternative conceptions reportedly are widespread, held strongly, persist over time, and/or resistant to change (Smith et al., 1993–1994). Smith et al. (1993–1994, p. 123) considered such central assertions in alternative conceptions research to be 'inconsistent with constructivism'. The authors proposed a theoretical framework to better align with constructivist theories: knowledge consists of fragmented pieces called phenomenological primitives (p-prims), which diSessa (1993) described as mental abstractions from familiar events. P-prims, then, are not incorrect and do not need to be replaced. Rather, these intuitive commitments about the natural world facilitate learning because they provide continuity between novice and expert reasoning, allow students to functionally use their prior knowledge, and create complex systems of knowledge when needed.

Regardless of theoretical commitments, science education researchers agree that cognitive resources, such as prior knowledge, are units of information that one can access during learning and reasoning (McClary & Talanquer, 2011b; Taber, 2008).

Recently, research on college chemistry students' cognitive resources suggested that organization of conceptual knowledge varies by individual and depends greatly on context (Maeyer & Talanquer, 2010; McClary & Talanquer, 2011a; Taber, 2008, 2009; Taber & García-Franco, 2010; Talanquer, 2009). For example, McClary and Talanquer (2011a, 2011b) described organic chemistry students' expressed mental models and intuitive reasoning strategies to illustrate how intuitive ideas (similar to p-prims) guide reasoning about acids and acid strength. Specifically, 20 students participated in semi-structured interviews in which they were asked to think aloud as they performed three tasks requiring qualitative reasoning: (1) predict trends in acid strength for seven sets of three compounds, (2) explain one given trend in acid strength, and (3) justify three trends from the prediction task.

McClary and Talanquer (2011b) identified four mental models that undergraduate chemistry students used when asked to rank 2D skeletal representations of compounds from least acidic to most acidic. Mental models were defined as dynamic, internal representations of real or imagined situations, objects, events, or processes which were constructed spontaneously or retrieved from long-term memory (Gentner, 2002; Vosniadou, 2002). Mental models are unique to a situation and to an individual (Greca & Moreira, 2000). In learning, individuals organize their knowledge in ways that are meaningful to them. Ideas need not be factual or consistent with each other; mental models need to simply facilitate learning, build explanations, and predict phenomena (Greca & Moreira, 2000; Talanquer, 2010). Therefore, the mental models described by McClary and Talanquer (2011b) were based on underlying assumptions that were general in nature, rather than specific in nature. For example, some students made decisions and generated explanations based on the mental model that acids are substances that lose hydrogen (H) atoms or ions (H^+). Acid strength, therefore, depended on salient features like the number of hydrogen atoms or the polarity of bonds or molecules. As a cognitive resource, students' mental models then guided them in ranking compounds and provided, to varying success, explanatory power.

Students' mental models acted as a system of explicit and implicit knowledge that guided the use of intuitive reasoning (Cheng & Brown, 2010; Redish, 2004; diSessa, 1993; Stavy & Tirosh, 2000; Taber, 2008; Talanquer, 2006; Vosniadou, 1994). Therefore, prior knowledge and assumptions within mental models allowed students to make causal inferences about acids and factors that affected acid strength (McClary & Talanquer, 2011a). To make such inferences, students had to determine which implicit and/or explicit features within the 2D skeletal chemical structures were relevant to predicting trends in acid strength and then activate cognitive resources (e.g., mental models or p-prims) they perceived to be helpful in the decision-making process.

According to dual-process theories, to make decision, individuals process information by using two reasoning systems: System 1 and System 2 (Evans, 2008; Osman, 2004; Sloman, 1996). While there is disagreement about whether System 1 and System 2 run concurrently or sequentially, dual-process theories are consistent in describing a system for intuitive, quick, and impulsive information processing (i.e., System 1, heuristic) and a system for systematic, slow, and reflective information processing (i.e., System 2, analytical; Evans, 2008). Humans employ heuristic reasoning to simplify tasks and reduce cognitive effort (Shah & Oppenheimer, 2008; Todd & Gigerenzer, 2000). Quick and automatic reasoning, however, sometimes leads to biases in decision-making (Talanquer, 2006)

One heuristic that college chemistry students rely on to make decisions about chemical properties like acid strength is the *representativeness heuristic* (McClary & Talanquer, 2011a; Maeyer & Talanquer, 2010). To use a representativeness heuristic, one must judge the extent to which a target object resembles a prototypical instance of the class. If the target and the prototype share commonalities, then one assumes the two objects are from similar classes. McClary and Talanquer (2011a) reported that organic chemistry students often used functional groups to determine or to explain relative acid strength. In this way, if a compound contained a functional group that the student assumed conveyed acid-like properties (e.g., –COOH), then a representativeness heuristic was typically used to rank the compound as more acidic than at least one other compound. Similarly, a compound that contained a functional group that was not assumed to be acid-like was typically predicted to be the least acidic in its set. For example, one student used the representativeness heuristic to determine the trend 2,4-pentadione $<$ phenol $<$ acetic acid, offering as an explanation

> Well, two of them have OHs which are definitely more acidic than [2,4-pentadione], which has CH_2. My guess would be, I'm not entirely certain, but I'm gonna guess that the carboxylic acid would probably be the stronger [one, because it] has acid in the name.

Heuristic reasoning biased students' abilities to predict correct trends in acid strength. Students also applied heuristic reasoning on the explanation and justification tasks.

Diagnostic Tools in Science Classrooms

A typical college student in the USA whose degree program requires organic chemistry enrolls in introductory organic chemistry in the second year. Few students, if any, have had formal instruction about functional groups or synthesis prior to beginning an organic chemistry course. Thus, most organic chemistry students' prior knowledge and assumptions about chemistry were learned prior to college or in introductory college chemistry (Duis, 2011); it is these ideas that students access when constructing their knowledge about organic chemistry (Driver & Erickson, 1983; Lin & Chiu, 2007).

Few research studies exist in the literature that pertain to conceptions that undergraduate organic chemistry students have about acid strength in organic chemistry contexts (Cartrette & Mayo, 2011; & McClary Talanquer, 2011a, 2011b). Participants in these studies generally relied on surface features to cue appropriate prior knowledge, but they failed to apply their knowledge correctly when asked to qualitatively determine acid strength for a set of molecules represented by skeletal structures.

Previous studies' findings were elicited from think-aloud interviews with at most 20 students who were enrolled in second-semester organic chemistry at the time of their interviews. The smaller sample sizes in these studies make it difficult to generalize the findings to larger populations of students, hence the need for this study. Diagnostic tools have the advantage of helping educators and researchers gain insight into conceptions held by larger samples of students (Treagust, 1988).

Diagnostic assessments covering a range of topics in chemistry have been reported, including particulate nature of matter (Nyachwaya et al., 2011), kinetic particle theory (Treagust et al., 2010), acids and/or bases (Lin & Chiu, 2007; Rahayu et al., 2011), and chemical equilibrium (Voska & Heikkinen, 2000). Some diagnostic assessments broadly investigate students' understanding of general chemistry concepts (Mulford & Robinson, 2002; Pavelich, Jenkins, Birk, Bauer, & Krause, 2004; Schmidt, 1997; Villafañe, Bailey, Loertscher, Minderhout, & Lewis, 2011; Villafañe, Loertscher, Minderhout, & Lewis, 2011), reasoning abilities (Cloonan & Hutchinson, 2011), and even attitudes about chemistry (Bauer, 2005; Brandreit, Xu, Bretz, & Lewis, 2011).

The formats of diagnostic assessments vary depending on the developers' goals, but most are multiple-choice so that they can be quickly administered and reliably scored. Since the publication of Treagust's (1988) seminal paper on two-tier assessments, many multiple-choice instruments have included an answer tier followed by a reason tier. Treagust recommended that designers use prior research, concept maps, and propositional knowledge to build two-tier items. The items on two-tier instruments create an opportunity to gain insight into students' mental models more efficiently than think-aloud interviews. Rather than simply test students' declarative or procedural knowledge, the addition of a reason tier allows educators to determine if students possess a deeper understanding of the concept(s) being addressed within a particular item. While two-tier diagnostic instruments can tell educators more about students' conceptions related to a given topic, they do not provide any information on how such strongly alternative conceptions bias students' reasoning.

Recently, Caleon and Subramaniam (2010a, 2010b) reported on their use of a concept inventory which included a confidence tier. After *both* the answer tier and the reason tier, students were asked to indicate their confidence on a Likert scale from just guessing to absolutely confident. The addition of a confidence tier to both the answer tier and the reason tier allowed the authors to quantify the strength of the students' alternative conceptions. Stronger alternative conceptions were thought to have a greater influence on students' reasoning than weaker alternative conceptions, which were considered relatively transient. To mitigate students selecting socially desirable confidence values, the authors informed the students that their knowledge and confidence would be used to identify learning difficulties and to design future instruction accordingly.

Prior Research

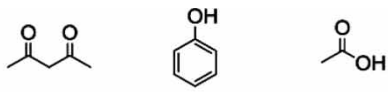
To appreciate the methods employed in developing and assessing ACID I, it is important to summarize the prior research on which the diagnostic tool was based. As described earlier, McClary and Talanquer (2011a, 2011b) interviewed 20 students who were taking first-semester organic chemistry for the first time at a research-intensive university in the southwestern United States. Students were recruited and interviewed 2–4 weeks following formal instruction and an exam covering acid strength in organic chemistry, either during Fall, 2008 or Fall, 2009. Participants were asked to think aloud as they ranked seven sets of three chemical substances from least acidic to most acidic. One set of acids (i.e., a prediction task) was shown at a time on a laptop screen. Students had to predict a trend in acid strength before moving on to the next set of acids; they were not allowed to return to previous screens.

The acids represented in the prediction tasks were Brønsted–Lowry acids, which by definition donated protons (H^+) to Brønsted–Lowry bases that accepted them. In a dynamic, solvent-based process, the products of a proton transfer between an acid (HA) and a base (:B) are referred to as a conjugate base (A^-) and a conjugate acid (BH^+). In cases where the conjugate base and conjugate acid (which alternatively can be defined as the deprotonated acid and protonated base, respectively) are stronger than the acid and base reactants, equilibrium will shift toward reactants. Conversely, when the conjugate base and conjugate acid are weaker than the acid and base reactants, equilibrium will shift toward products. In the Brønsted–Lowry acid model, acid strength can be determined by considering the relative strength of the conjugate base for one acid and the conjugate base of another acid. A stronger acid, when deprotonated, will form a weaker conjugate base.

Of the seven sets of substances, prediction tasks 3, 4, and 7 were very difficult for students to rank correctly (Figure 1). Task 3 was designed to elicit first-semester organic chemistry students' understandings of resonance—an important factor that affects acid strength. Resonance occurs when two or more plausible molecular structures exist in such a way that electrons can move (or delocalize) among fixed atoms (Holum, 1969, p. 77). The classic example of resonance is benzene (C_6H_6). The molecule is then said to resonate among several structures or to have a structure that is a 'resonance hybrid' of the Lewis structures. The energy calculated for a resonance hybrid is lower than the energies of any of the alternative structures; the molecule is then said to be stabilized by resonance.

Because organic chemistry students are taught to consistently associate benzene with resonance, phenol (structure B) was included to cue students' focus on resonance rather than on other factors such as structure or inductive effect, though certainly structure and inductive effect influence the correct trend in acid strength for 2,4-pentadione (structure A), phenol, and acetic acid (structure C).

Interviewees correctly chose acetic acid as the most acidic compound simply because it was a carboxylic acid. Few students were able to move beyond their *representativeness heuristic* and provide a scientifically valid explanation (McClary & Talanquer, 2011b). Using a representativeness heuristic as an anchor (Todd & Gigerenzer, 2000), 85% of the 20 students incorrectly predicted that phenol was more acidic than 2,4-pentadione, in part because phenol and acetic acid both have oxygen as the acidic atom (i.e., the structure is similar). 2,4-Pentadione is actually more acidic ($pK_a = 8.95$;



A B C

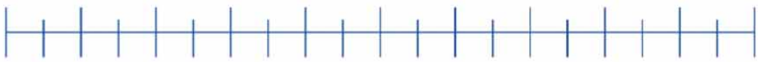
Question 2

Compound C is the most acidic of the above three structures. Which of the following represents the correct trend in acid strength for compounds A and B?

(I) A < B

(II) B < A

How confident are you about the ***answer*** you chose? (place an X anywhere on the scale):



0 % 20 % 40 % 60 % 80 % 100 %

Not confident (just guessing) Confident (absolutely certain)

Question 3

Select or provide the ***best*** reason to justify your answer for question 2:

(I) B is more acidic than A because B has a more electronegative acidic atom than A.

(II) B is more acidic than A because the benzene better stabilizes the conjugate base than the carbonyl groups of A.

(III) A is more acidic than B because the carbonyl groups better stabilize the conjugate base than the benzene of B.

(IV) A is more acidic than B because A has two oxygen atoms instead of one oxygen atom.

How confident are you about the ***reason*** you chose? (place an X anywhere on the scale):

0 % 20 % 40 % 60 % 80 % 100 %

Not confident (just guessing) Confident (absolutely certain)

Figure 1. An answer–reason pair from ACID I showing a trio of acids with related prediction and reason response options

Eidinoff, 1945) than phenol (pK_a=9.98; Liptak, Gross, Seybold, Feldgus, & Shields, 2002) because the oxygen atoms of the two carbonyl groups lower the energy of the conjugate base despite the acidic atom being carbon. The strong association students have with benzene and resonance biased them toward predicting that phenol was more acidic than 2,4-pentadione.

Another bias students generally had is that carbon acids were considered to be weaker than oxygen acids. Task 4 was included to explore students' understanding of carbon acids. The task required them to predict the trend in acid strength for 2,4-pentadione (structure A), acetone (structure B), and acetaldehyde (structure C) (Figure 1). To focus students' attention toward inductive effect and resonance, compounds containing carbonyl groups were included. Inductive effect is similar to resonance in that an atom or groups of atoms can donate or withdraw electrons. Inductive effect is limited by distance and occurs through sigma (σ) bonds, whereas resonance is limited by the length of bond conjugation and requires pi (π) bonds (Anslyn & Dougherty, 2006).

First-semester organic chemistry students generally are unfamiliar with the nuances of mechanisms involving carbonyl chemistry, so they were expected to draw upon less sophisticated prior knowledge to decide which of the hydrogen atoms was most acidic—particularly for the dione and the aldehyde, both of which had more than one kind of hydrogen atom. Findings showed that many students relied heavily on either structure/composition features (e.g., the number of oxygen atoms or the explicit representation of the hydrogen atom in acetaldehyde) or on less sophisticated electronic properties (e.g., bond/molecular polarity), rather than more sophisticated electronic properties (e.g., inductive effect and resonance) to explain their trends for the carbon acid series.

The third task that students had difficulty predicting a correct trend for was Task 7. Task 7 was designed for students to consider structure, inductive effect, and resonance when making a decision about *p*-nitrophenol (structure A), *p*-methylphenol (structure B), and phenol (structure C) (Figure 1). Intentionally increasing the number of factors to which students must attend increased the complexity of the task. A majority of the 20 students (55%) correctly predicted that *p*-nitrophenol was most acidic primarily because the nitro ($-NO_2$) group was a substituent that was most salient to them. Explanations for students choices varied from memorization of a pK_a table (i.e., nitro groups make acids more acidic without knowing why), to the number of electronegative atoms the substituent contained, to correct conceptions involving inductive effect and/or resonance. Students were expected to consider the additional resonance stabilization offered by the nitro ($-NO_2$) group to the conjugate base and the destabilization provided by the methyl ($-CH_3$) group.

Based on the findings of this prior research, Tasks 3, 4, and 7 were classified as *deep structure* tasks because they required students to understand the underlying principles related to acid strength without over-relying on surface features as they did with the other four prediction tasks (Chi, Feltovich, & Glaser, 1981). Along with students' transcripts, deep structure tasks were the bases for designing and interpreting data collected in the present study.

Methods

Goals and Research Questions

The purpose of this study was to design a valid paper-and-pencil instrument that can be quickly administered and reliably scored. Specifically, we designed a multiple-tier, multiple-choice (MTMC) concept inventory using students' prior knowledge and assumptions about acids and acid strength, expressed in the students' own words from our previous work (McClary & Talanquer, 2011a, 2011b), to design distracters. We sought to answer the following research questions:

- What alternative conceptions do organic chemistry students have about acid strength?
- With what frequency do the alternative conceptions appear in our population?
- How strongly do the alternative conceptions in students' mental models bias their reasoning?

Setting and Participants

The instrument, ACID I, was administered to 104 students in two sections of second-semester organic chemistry during a regularly scheduled lecture period at a medium-sized liberal arts university in the United States. The students were enrolled in a course which used a spiral curriculum: topics were covered broadly during the first semester and more in depth during the second semester (Grove et al., 2008). Most students in the course were second year students intending to pursue careers in health-related professions such as human or veterinary medicine, optometry, and physician assistants. The professor for both sections signed a letter of support as part of Institutional Review Board approval.

ACID I was administered for 15 min during the second week of the Spring 2011 semester, during the week prior to instruction on electrophilic aromatic substitution. This timing was important because students learn more refined explanations about how substituents affect acid strength (e.g., electron-donating vs. electron-withdrawing) when discussing electrophilic aromatic substitution. Therefore, the sample was essentially equivalent to students who have completed just one semester of organic chemistry.

Instrument Design

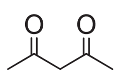
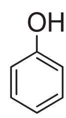
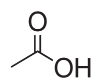
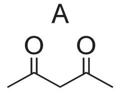
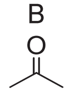
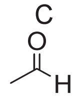
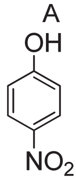
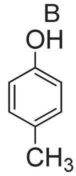
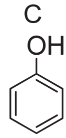
Using the three deep structure prediction tasks and student interview data from our previous work (McClary & Talanquer, 2011a, 2011b), we developed a nine-item MTMC instrument to investigate the prevalence and strength of organic chemistry students' alternative conceptions about acids and acid strength. Students in the prior research generated verbal explanations for these prediction tasks, and in doing so, provided rich descriptions of their understanding about acids and acid strength. Thus, these students' conceptions were used to create response options for each item on ACID I. Response options were written to best capture students' conceptions for each prediction task.

Three items were created for each prediction task. These three items and the corresponding prediction task constituted a *set* (Table 1, Figure 1). Few students in the prior research correctly predicted a trend in acid strength for at least one deep structure task. Most students, however, were able to identify the most acidic compound in each set. These results from the prior research shaped the format of items within a set. For the first item in each set, students were told which of three compounds was most acidic and asked to select the best reason to explain this fact. The second item in each set asked students to predict the trend in acid strength for the two remaining compounds. Selecting the reason for the trend constituted the third item of each set. Each of these three items was accompanied by a confidence scale. Caleon and Subramaniam's (2010a, 2010b) confidence scale was modified from a Likert scale to an interval scale, ranging from 0% (just guessing) to 100% (absolutely confident). Quantifying students' confidence in their responses facilitated an analysis of the strength of alternative conceptions, as described below.

Validity

The three deep structure tasks used to frame ACID I were originally part of an instrument that was previously validated by organic chemistry faculty who teach undergraduate and graduate organic chemistry courses (McClary & Talanquer, 2011a). ACID I was given to the organic chemistry faculty member whose classes were sampled in this study. He agreed that the instrument was a fair assessment of his

Table 1. Skeletal structures of the sets of acids used on ACID I to elicit students' conceptions of acid strength

Set	Skeletal structures		
1	 A	 B	 C
2	 A	 B	 C
3	 A	 B	 C

students' understanding of acid strength. To further validate the inventory, a fifth response option was provided to allow students to write in their own answers (Voska & Heikkinen, 2000). To mitigate students inflating their confidence due to social desirability, students were told that their responses would be used to adjust instruction on acid strength later in the semester (Caleon & Subramaniam, 2010a, 2010b; Turner, VanderHeide, & Fynewever, 2011).

Data Analysis

Students who consented to participate in the research were assigned an identification number to maintain confidentiality. Students' answer choices and confidences were entered into SPSS. Answers were scored 0 for incorrect and 1 for correct. Any student who failed to answer an item or to provide a confidence level was excluded from data analysis. Thus, 89 students were included in the final analysis reported in this paper.

ACID I was assessed using psychometrics commonly reported in the literature (Adams & Wieman, 2011; Ding & Beichner, 2009), relying primarily on elements from classical test theory: reliability (Cronbach's α), item difficulty (p), item discrimination (D), and point biserial coefficients (r_{pbi}). The reliability coefficient was based on internal consistency, i.e., the average of how well each of the nine items correlated to others. Item difficulty (p) measured how many students answered the item correctly. Ranging from 0 to 1, a higher p ($=N_{correct}/N_{total}$) ratio indicated an easier item because more students answered correctly. Item discrimination (D), which also ranges from 0 to 1, measured how well an item distinguished between a group of higher scoring students and a group of lower scoring students. A larger D suggested that higher scoring students were better at selecting the right answer than lower scoring students. Finally, a point biserial coefficient was calculated for each item to determine the correlation of a given test item to a participant's total score. Ding and Beichner (2009) referred to this psychometric as *item reliability.* Students who did well—and presumably had a greater understanding of the concept(s) assessed—should have answered more items correctly, resulting in a higher score. Similarly, students who answered fewer questions correctly earned lower scores.

Psychometric analyses identified not only students' alternative conceptions about acids and acid strength, but also determined the prevalence with which they appeared in the population. While items with lower p and lower r_{pbi} values suggested alternative conceptions, such analyses offered little information about how functional the alternative conceptions were as cognitive resources. Following analyses by Caleon and Subramaniam (2010b), alternative conceptions were classified as *significant* when a distracter was chosen by at least 10% of participants above the probability of selecting the distracter by chance. For a multiple-choice item with five response options, 30% or more of the participants must have selected a distracter to be considered a significant alternative conception.

Students' confidence was used further to classify significant alternative conceptions as either *genuine* or *spurious*. Alternative conceptions were considered genuine if the average confidence mean was above 50% or spurious case if the average confidence mean was below 50%. Conceptions held with greater confidence suggested that such ideas were more strongly held in students' mental models (i.e., they were genuine) rather than being spontaneously created as a result of completing the assessment (i.e., they were spurious). More strongly held conceptions were expected to bias reasoning to greater extents than less strongly held conceptions.

Overall mean confidence (CF) was calculated per item and per answer–reason (A–R) pair. The mean confidence for students who answered an item or A–R pair correctly (CFC) and incorrectly (CFW) was calculated to determine the confidence mean quotient (CDQ), which is mathematically defined as (CFC – CFW)/SD of confidence. According to Caleon and Subramaniam (2010b), the CDQ 'indicates whether subjects can discriminate between what they know and what they do not know'. In other words, the CDQ may be considered a measure of students' metacognitive ability (Potgieter & Davidowitz, 2011).

Results and Discussion

The central goal of developing ACID I was to identify alternative conceptions that organic chemistry students have about acid strength, to determine the frequency with which the conceptions appeared in the sample, and to determine how strongly the conceptions in students' mental models biased their reasoning for specific cases. The nine items on our concept inventory identified two significant alternative conceptions which were held by at least 30% of our participants: *functional group determines acid strength* and *stability determines acid strength* (Table 2). Each significant alternative conception included specific cases that were genuine (i.e., CF $>$ 50%) and spurious (i.e., CF $<$ 50%).

Functional Group Determines Acid Strength

Students frequently selected a distracter that was based on the presence of a functional group, which is a collection of atoms that have a defined connectivity. Functional groups, such as amines (RNH_2), alcohols (ROH), and carboxylic acids (RCOOH), help organic chemists make decisions when designing syntheses or when proposing reaction mechanisms because the structure/composition of a substance provides valuable information regarding implicit, electronic properties that generally are better predictors of chemical behavior than structure/composition alone. Chemists are adept at processing structural/compositional information to recall relevant and applicable information related to a given task (e.g., proposing a mechanism). Students learning chemistry for the first time, however, tend to overgeneralize and rely mostly on structural/compositional features of substances (rather than implicit, electronic properties) to predict chemical properties (Maeyer & Talanquer, 2010). Reliance on structural/compositional features can bias learning and decision-making in organic chemistry, where such features are often functional groups.

Table 2. Seven items on ACID I identified two significant alternative conceptions

Set	Item	Type of Tier	Specific case in which an alternative conception (AC) was elicited	Strength of AC	Frequency of AC (%)
1	1	Reason	Acetic acid is more acidic than both phenol and 2,4-pentadione because it is a carboxylic acid. *Functional group determines acid strength*	Genuine	61.8
	2	Answer	Phenol is more acidic than 2,4-pentadione.	Genuine	77.5
	3	Reason	Phenol is more acidic than 2,4-pentadione because the benzene better stabilizes the conjugate base than the carbonyl groups of 2,4-pentadione. *Stability determines acid strength*	Genuine	56.2
2	4	Reason	2,4-pentadione is more acidic than the acetone and acetaldehyde because 2,4-pentadione has two carbonyl groups. *Functional group determines acid strength*	Spurious	32.6
	6	Reason	Acetaldehyde is more acidic than acetone because it has a hydrogen atom instead of another methyl group. *Functional group determines acid strength*	Spurious	34.8
3	7	Reason	*p*-Nitrophenol is more acidic than *p*-methylphenol and phenol because *p*-nitrophenol has a nitro ($-NO_2$) group. *Functional group determines acid strength*	Genuine	49.4
	9	Reason	*p*-Methylphenol is more acidic than phenol because the methyl ($-CH_3$) group destabilizes the conjugate base of *p*-methylphenol. *Stability determines acid strength*	Spurious	31.5

Note: The significant alternative conceptions are written in italics. Students primarily relied on structure/composition properties of acids to predict and explain trends in acid strength.

On four items, 32.6–61.8% of our participants used *functional group determines acid strength* to choose a distracter. Items 1, 4, 6, and 7 belong to each of the three sets on ACID I, suggesting that students use this alternative conception to justify trends in acid strength for molecules across a variety of functional groups. The mean confidence (CF) for Item 1 and Item 7 was greater than 50%, while CF was less than 50% for Item 4 and Item 6 (Table 4).

Item 1 in Set 1 (Table 1) was based on Task 3 from our previous study (McClary & Talanquer, 2011a, 2011b). In the present study with ACID I, students were rather confident (Table 2, CF = 58.7%) about their answers for Item 1. Sixty-two percent of students chose acetic acid as most acidic because it is a carboxylic acid, consistent with the previous study. Only 15.7% chose the best available explanation: acetic acid is most acidic because it has the most positive acidic hydrogen. In solution, the bond polarization of the O–H bond—increased by the proximity of the carbonyl group—reduces the electron density on the hydrogen atom relative to the acidic hydrogen on 2,4-pentadione and phenol. A reaction of acetic acid with a base is more energetically favorable than the same reaction with either of the other two acids because of the greater electron deficiency on the acidic hydrogen of acetic acid. Interestingly, only 32.6% of participants who completed ACID I chose a response that considered the implicit, electronic property of acid strength, even though two of the four choices provided for them related to an implicit, electronic property (i.e., bond polarization or inductive effect).

Another genuine case of *functional group determines acid strength* was identified by Item 7, which was based on Task 7 from prior research (*vide supra*). One genuine alternative conception (Item 7) and one spurious alternative conception (Item 9) were detected in Set 3. Forty-nine percent of students chose that *p*-nitrophenol was more acidic than both *p*-methylphenol and phenol because *p*-nitrophenol had a NO_2 group (Table 2). Similar to the genuine case identified in Item 1, students were drawn to choices that focused on structure/composition features of the acid rather than on implicit, electronic properties.

Students were generally less confident about employing *functional group determines acid strength* to justify reasons for Items 4 and 6 than they were for Items 1 and 7. Students had conceptions of acid strength similar to students in the prior study. When asked to select the response that best explained why 2,4-pentadione is more acidic than acetone and acetaldehyde on ACID I Item 4, 32.6% of participants chose a reason based on structure/composition features: 2,4-pentadione has two carbonyl functional groups (Table 2). On Item 5, a majority of students (69.7%) correctly predicted that the aldehyde was more acidic than the ketone. However, half of them incorrectly justified their decision on Item 6 by selecting a reason based on structure/composition rather than on an electronic property, i.e., that acetaldehyde is more acidic than acetone because acetaldehyde has a hydrogen atom instead of another methyl group. Students who hold this conception think the aldehyde hydrogen (which is the only hydrogen atom explicitly represented in the skeletal structures for Set 2) is the acidic hydrogen rather than a methyl hydrogen (McClary & Talanquer, 2011a). Even though participants were permitted to draw on ACID I, no one drew the conjugate base of acetaldehyde to verify which of the two types of hydrogen atoms would be most acidic.

Stability Determines Acid Strength

Stability is a difficult concept for chemistry students to understand (Taber, 2009). In the context of acid–base chemistry (i.e., chemical equilibrium), the concern is thermodynamic stability, the result of which is a decrease in free energy (G) of a chemical system. For Brønsted–Lowry acids, a stronger acid has a more stable conjugate base (A^-). In general chemistry, many students conceptualize stability in terms of chemical bonding. Taber (2008) reported an *octet rule heuristic* that students overgeneralized to make decisions about the existence of configurations of atoms like sodium or chlorine. Organic chemistry students, on the other hand, consider a variety of concepts related to stability when making decisions about relative acid strength. For example, McClary and Talanquer (2011a) described how students reasoned that stable molecules were more or less likely to react, and therefore, were stronger or weaker acids depending on their prior knowledge. Students with more coherent mental models of acids and acid strength articulated how conjugate base stability influenced acid strength, and used their knowledge to predict, explain, and justify many or all of their trends. In this study, it was not possible to provide a thorough description of students' understanding of stability based on responses to ACID I as the inventory was not designed to do so. Consequently, *functional group determines acid strength* may be operating rather than, or in addition to, *stability determines acid strength* as students compared benzene and carbonyl groups (Item 3) and methyl and a hydrogen atom (Item 7).

On ACID I, 31.5–77.5% of students used *stability determines acid strength* to predict a trend (Item 2) or to justify the reason for a trend (Item 3, Item 9). Mean confidence (CF) on both Item 2 and Item 3 was greater than 50%, while CF was less than 50% for Item 9. Students in the previous qualitative study used implicit assumptions about stability to make inferences about acid strength.

As with Items 1, 4, 6, and 7, the responses for Item 2 and Item 3 were also consistent with the previous research study (vide supra). Seventy-eight percent of participants who completed ACID I incorrectly predicted that phenol (structure B) was more acidic than 2,4-pentadione (structure A). Most of these participants (56.2%) justified their prediction by selecting a response for Item 3 that phenol was more acidic than 2,4-pentadione because the benzene better stabilized the conjugate base than the carbonyl groups of 2,4-pentadione.

In addition to the genuine alternative conception identified on Item 7, the trio of acids in Set 3 elicited the use of *stability determines acid strength* for Item 9, though with less confidence (CF = 45.7%) than *functional group determines acid strength* reported on Item 7 (Table 3). Most of the participants (57.3%) correctly predicted

Table 3. Students' self-reported confidence for each item on ACID I on a scale of 0% (just guessing) to 100% (absolutely confident)

Set	Item	Mean confidence (CF, %)	Mean confidence when correct (CFC, %)	Mean confidence when incorrect (CFW, %)	Mean confidence quotient (CDQ)
1	1	58.72	54.78	59.46	**−0.21**
	2	65.09	53.91	68.33	**−0.63**
	3	52.13	47.88	52.99	**−0.22**
2	4	49.78	55.40	44.75	0.48
	5	52.79	52.65	53.10	**−0.02**
	6	45.59	44.19	46.34	**−0.10**
3	7	64.62	63.23	65.15	**−0.07**
	8	60.37	64.67	54.61	0.39
	9	45.74	40.16	48.30	**−0.35**

Note: $CDQ = (CFC - CFW)/SD_{item}$. Students were moderately confident about their understanding of acid strength in organic chemistry contexts. A negative CDQ (bolded) value indicates that students who answer incorrectly are more confident with their responses than students who answer correctly, suggesting that students do not know what they do not know.

that phenol was more acidic than *p*-methylphenol. The remaining students predicted that the methyl-substituted phenol was more acidic, with 63.2% of them justifying their decision based on the fact that the methyl group destabilized the conjugate base of *p*-methylphenol. In such a model of acid strength, ostensibly, destabilization is associated with increased acid strength. However, a closer look at students' responses for the entire concept inventory revealed that of the 24 students who selected this reason to justify why *p*-methylphenol was more acidic than phenol, 21 (87.5%) of them also chose a response on ACID I that contained a reference to stabilization of a conjugate base increasing acid strength. This inconsistency in responses—i.e., in some cases students believe that stabilization increases acid strength while in others stabilization decreases acid strength—is due to students heuristically selecting a response that related to conjugate base stability without truly understanding how the concept related to acid strength. It is important to note that Item 9 had a mean confidence of 45.7%; the conception that destabilization of the conjugate base increased acid strength was not held strongly in students' mental models.

Confidence Mean Quotients

ACID I identified two significant alternative conceptions that students held about acid strength and used to make inferences for seven specific cases involving three distinct sets of acids. The confidence tier provided a broad sense of how students understood acid strength after one semester of organic chemistry. A confidence mean quotient (CDQ) was calculated for each item to determine if students were actually aware of what they did and did not know.

The confidence mean quotient (CDQ) for each of the nine items (Table 3) suggested participants were generally unaware of their own understanding of acid strength as measured by ACID I. Students who answered the items *incorrectly* were more confident, on average, than students who answered the items *correctly*. Item 2 had the most negative CDQ of all nine items as students were particularly unaware of their strong association between benzene and stability had on their ability to predict whether phenol was more or less acidic than 2,4-pentadione. Furthermore, the A–R pair (i.e., Items 2 and 3) in Set 1 had a considerably lower CDQ ($CDQ_{Set1} = -0.42$) compared with the other A–R pairs of items, further supporting the claim that more students associated conjugate base stability with benzene than with two α-carbonyl groups.

Not all items on ACID I had a negative confidence mean quotient (CDQ). Item 4 and Item 8 both had positive quotients. To understand why students were largely unsuccessful at choosing the best answers while being generally confident, the item difficulty (p), item discrimination (D), and point biserial coefficients (r_{pbi}) were calculated for each item (Table 3). The difficulty of the nine items varied widely, though no item was easy (Popham, 2005). Less than one-quarter of participants answered at least five items correctly. No item was answered correctly by more than 70% of the students. Not surprisingly, less than half of the students chose the best response for all seven items on which significant alternative conceptions were detected. According to Popham (2005, p. 249), item difficulty is related to the instructional program. If students are not exposed to types of problems about acid strength that challenge their assumptions and mental models, then the items on ACID I ought to have low p values because the deep structure tasks upon which they were based required students to meaningfully consider the complex interrelatedness of the factors that affect acid strength.

Item Quality

Even though ACID I was difficult for participants ($M = 3.09 \pm 1.64$), the items discriminated between high and low scorers. Furthermore, the mean of the point biserial coefficients was 0.41, suggesting acceptable item quality (Ding & Beichner, 2009). The reliability of ACID I, calculated as a Cronbach alpha, was 0.41, which is below the standard value of 0.50 (Nunnally & Bernstein, 1994). A typical explanation for a lower Cronbach's α might suggest that the ACID I concept inventory was not consistently measuring students' understanding of acid strength. To address this possibility, ACID I was administered again later in the semester. While a full discussion of the results from a test–retest condition will be addressed elsewhere, the internal consistency remained low for 58 students who completed ACID I twice. Finding Cronbach's α to be low and yet consistent across repeated measures within the same population points not to unreliable functioning of the measure, but rather to the possibility that because most participants' conceptions of acid strength are not coherent, their conceptions seem to be fragmented (Adams & Wieman, 2011). For example, on the initial administration of ACID I, students chose the stability of the conjugate base as the best answer for four items on ACID I (i.e., Items 3, 4, 6, and 7) whereas only 6.7% of students consistently chose the best response option across all four of these items.

Conclusions and Implications for Instruction

A nine-item concept inventory was developed to validly and reliably identify alternative conceptions regarding acid strength, the frequency with which those alternative conceptions appear, and the strength with which the alternative conceptions are held in the mental models of organic chemistry students. Students hold two significant alternative conceptions (Table 2). The prevalence and strength of *functional group determines acid strength* and *stability determines acid strength* varied depending on the trio of acids. Participants chose distracters that were primarily based on structure/composition features of the acid(s), even when better choices that involved electronic properties of the acid(s) were available, consistent with previously published studies (McClary & Talanquer, 2011a).

Finding that alternative conceptions related to structure/composition and stability are prevalent and held strongly enough to repeatedly bias students' reasoning is not unexpected in the domain of chemistry (McClary & Talanquer, 2011a; Maeyer & Talanquer, 2010; Taber, 2009). Chemical and physical properties of substances and reactions *greatly depend on molecular structure and composition*, and instructors explicitly teach the importance of molecular structure/composition to students. The challenge for science education researchers and practitioners is to help students learn to separate intuitive knowledge from explicit prior knowledge that has been particularly emphasized in chemistry classrooms. First, researchers and practitioners must elicit students' understandings and discern how such understandings impact decision-making and learning.

While a body of literature on alternative conceptions exists in many domains (Duit, 2009), a complementary corpus of work regarding science students' intuitive ideas is now emerging (cf. McClary & Talanquer, 2011a; Cheng & Brown, 2010; Hammer, 1996; Maeyer & Talanquer, 2010; diSessa, 1993; Stavy & Tirosh, 2000; Taber & García-Franco, 2010; Talanquer, 2009). Within chemistry, studies have explored students' intuitive ideas on topics including chemical and physical properties (McClary & Talanquer, 2011a; Maeyer & Talanquer, 2010) and particulate nature of matter (Taber & García-Franco, 2010). Taber and García-Franco (2010) recently reported a methodology to distinguish intuitive knowledge, specifically p-prims, from explicit knowledge like alternative conceptions. Assuming that students' intuitive ideas could be elicited and consciously expressed, the authors analyzed transcripts from secondary school students in England who participated in a study about particulate nature of matter. Five themes emerged, including one relevant to findings in this paper: *component gives property*, which was described as 'properties of substances derive from *components* that have an inherent property' (p. 112).

Functional groups within chemical substances certainly influence chemical and physical properties. Students' conception that *functional group determines acid strength* is an alternative conception because acid strength is an emergent property that does not depend on isolated functional groups. Rather, acid strength in the Brønsted–Lowry and Lewis models depends on the structure or composition of the entire molecule (ion), as well as on the structure/composition of solvent molecules (ions), reaction product(s), and other solutes in solution. Furthermore, structure/composition also helps one predict likely electronic properties of molecules (ions) that contribute to lowering the Gibbs free energy of the system, in particular, the conjugate base of the acid. To claim that *functional group determines acid strength* and *stability determines acid strength* are cognitive resources that operate at the intuitive level, which Taber and García-Franco (2010) argued p-prims originate, is beyond that which can be argued from the data collected using ACID I. The qualitative study on which the items and distracters for ACID I were developed, however, does provide compelling evidence that for at least some participants in the current study, the two significant alternative conceptions constitute functional intuitive knowledge and may be conflated for certain tasks. In general, most students' knowledge related to acids and acid strength seemed to bias their heuristic reasoning toward distracters that were based on the structure or composition of chemical substances rather than on implicit, electronic properties.

When teaching students about acid strength and factors that affect it, instructors should de-emphasize explicit, structure/composition features of acids and instead focus students' learning on implicit, electronic properties that are critical for meaningful learning of chemistry. Instructors, particularly in prerequisite courses for organic chemistry, should be especially mindful of students' reliance on structure/composition features—not only when teaching acid strength, but also when preparing assessments. Assessment items must be valid; arriving at a correct solution to a problem meant to assess students' understanding of acid strength ought to include the nuances of how both structure *and electronic effects* influence acid strength.

Instructors should also consider incorporating into class content the intuitive assumptions, p-prims, and heuristics that secondary students and college students have been found to use when thinking and reasoning about chemistry concepts (McClary & Talanquer, 2011a; Maeyer & Talanquer, 2010; Taber & Garcia-Franco, 2010; Talanquer, 2009). Contexts provided in these studies elicit students' prior knowledge, which can provide helpful feedback to initiate in-class discussions or to design assessments. Knowledge of intuitive ideas that students are likely to bring into science classrooms benefits not only instructors but also any educator who facilitates learning (e.g., teaching assistants, in-service teachers, curriculum developers).

ACID I has been shown to be a useful diagnostic tool for assessing conceptions about acid strength that organic chemistry undergraduates hold after completing one semester of a spiral curriculum. While our concept inventory can be administered following instruction on acid strength in the first semester, it does not explicitly assess prior knowledge of acid strength related to general chemistry. General chemistry students' conceptions of acid strength are quite different from what organic chemistry students are expected to know (Ferguson & Bodner, 2008). Therefore, ACID I is not recommended as a pre-assessment of students' understanding of acid strength prior to formal instruction within the context of organic chemistry. ACID I may be useful as a pre-assessment in second-semester organic chemistry and in courses like biochemistry where conceptions of acids and acid strength of organic molecules are important prior knowledge.

The inclusion of confidence tiers on two-tier multiple-choice instruments is uncommon in science education literature (Caleon & Subramaniam, 2010a). Moreover, inclusion of confidence tiers has rarely been reported in the domain of chemistry (cf. Potgieter & Davidowitz, 2011). Confidence tiers proved to be an important psychometric for ACID I. By adding confidence tiers to the items, evidence emerged that students were generally unaware of what they do not know as indicated by the negative confidence mean quotient of most items (Table 3). Quantifying the strength of students' alternative conceptions is important for instructors so that they can better allocate their instructional time to address alternative conceptions about which many students are relatively confident. Such ingrained alternative conceptions may adversely affect meaningful learning of subsequent, related content, e.g., the success of students when proposing reaction mechanisms (Bhattacharyya & Bodner, 2005; Ferguson & Bodner, 2008).

Instructors can only do so much to address students' alternative conceptions and intuitive knowledge. Ways to better facilitate students' awareness of their own knowledge must be found and disseminated to the science education community. Confidence tiers offer a promising opportunity to enable students to explicitly consider their own knowledge about particular concepts; they can be included on formative assessments where feedback is provided. For example, instructors could use clicker questions to determine students' confidence about particular items or concepts. The results can then be the basis to initiate dialogue where students actively engage in constructing and deconstructing their own understanding of key concepts throughout a science curriculum. Flynn (2012) recently reported on integrating complementary active learning methods to help students make meaning of their own learning. Without sacrificing the breadth of course content, Flynn used post-class online questions, clicker questions, and think-pair-share, and organic chemistry students retained knowledge topics such as reaction mechanisms and spectroscopy. Other active learning methods (e.g., peer-led team learning, process-oriented guided inquiry learning, and supplemental instruction) as well as self-initiated study groups (Christian & Talanquer, 2012) are also excellent means to encourage dialogue outside of formal lecture time.

Finally, while this paper represents a quantitative approach to evaluating a diagnostic tool, the role of qualitative methods such as think-aloud interviews with students and fine-grained analyses of individual distracters must be emphasized in the development of multiple-choice assessments, including concept inventories. Without a rich, qualitative foundation to support the development of diagnostic tools, the analysis would have been restricted to merely reporting alternative conceptions (Hammer, 1996) without a basis for more deeply discussing the cognitive origins of the alternative conceptions (Smith et al., 1993). The use of psychometrics in tandem with qualitative methods to create and modify questions offers the most viable methodology to deeply explore students' understanding of science concepts.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. 0733642 and 0728614. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Adams, W.K., & Wieman, C.E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289–1312.

Anslyn, E.V., & Dougherty, D.A. (2006). *Modern physical organic chemistry*. Sausalito, CA: University Science Books.

Bauer, C.F. (2005). Beyond 'Student Attitudes': Chemistry self-concept inventory for assessment of the affective component of student learning. *Journal of Chemical Education*, *82,* 1864–1870.

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

Bhattacharyya, G., & Bodner, G.M. (2005). 'It gets me to the product': How students propose organic mechanisms. *Journal of Chemical Education*, *82,* 1402–1407.
Bodner, G.M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, *63,* 873–878.
Brandreit, A.R., Xu, X., Bretz, S.L., & Lewis, J.E. (2011). Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential. *Chemistry Education Research and Practice, 12,* 271–278.
Brown, W.H., Foote, C.S., Iverson, B.L., & Anslyn, E.V. (2009). *Organic chemistry* (5th ed.). Belmont, CA: Brooks/Cole Cengage Learning.
Caleon, I., & Subramaniam, R. (2010a). Do students know what they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, *40,* 313–337.
Caleon, I., & Subramaniam, R. (2010b). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, *32,* 939–961.
Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
Carr, M. (1984). Model confusion in chemistry. *Research in Science Education*, *14,* 97–103.
Cartrette, D.P., & Mayo, P.M. (2011). Students' understanding of acids/bases in organic chemistry contexts. *Chemistry Education Research and Practice*, *12,* 29–39.
Cheng, M.F., & Brown, D. (2010). Conceptual resources in self-developed explanatory models: The importance of integrating conscious and intuitive knowledge. *International Journal of Science Education, 32*, 2367–2392.
Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5,* 121–152.
Christian, K., & Talanquer, V. (2012). Modes of reasoning in self-initiated study groups in chemistry. *Chemistry Education Research and Practice*. Retrieved March 31, 2012, http://www.chem.arizona.edu/tpp/Christian_Tal_CERP12r.pdf on 03/31/2012
Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50,* 66–71.
Cloonan, C.A., & Hutchinson, J.S. (2011). A chemistry concept reasoning test. *Chemistry Education Research and Practice*, *12,* 205–209.
Cros, D., Maurin, M., Amouroux, R., Chastrette, M., Leber, J., & Fayol, M. (1986). Conceptions of first-year university students of the constituents of matter and the notions of acids and bases. *International Journal of Science Education*, *8,* 305–313.
Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics – Physics Education Research*, *5,* 1–17.
Drechsler, M., & van Driel, J. (2008). Experienced teachers' pedagogical content knowledge of teaching acid–base chemistry. *Research in Science Education*, *38,* 611–631.
Drechsler, M., & Schmidt, H.-J. (2005). Textbooks' and teachers' understanding of acid–base models used in chemistry teaching. *Chemistry Education Research and Practice*, *6,* 19–35.
Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education*, *5,* 61–84.
Driver, R., & Erickson, G. (1983). Theories-in-action: Some theoretical and empirical issues in the study of students' conceptual frameworks in science. *Studies in Science Education*, *10,* 37–60.
Duis, J.M. (2011). Organic chemistry educators' perspectives on fundamental concepts and misconceptions: An exploratory study. *Journal of Chemical Education*, *88,* 346–350.
Duit, R. (2009). Bibliography STCSE: Students' and teachers' conceptions and science education. Retrieved July 3, 2011, http://www.ipn.uni-kiel.ed/aktuell/stcse/download_stcse.html
Eidinoff, M.L. (1945). Dissociation constants of acetylacetone, ethyl acetoacetate and benzoylacetone. *Journal of the American Chemical Society*, *67,* 2072–2073.

Evans, J.B.T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Ferguson, R., & Bodner, G.M. (2008). Making sense of the arrow-pushing formalism among chemistry majors enrolled in organic chemistry. *Chemistry Education Research and Practice*, *9,* 102–113.

Flynn, A.B. (2012). Development of an online, postclass question method and its integration with teaching strategies. *Journal of Chemical Education*, *89,* 456–464.

Furió-Más, C., Calatayud, M.L., Guisasola, J., & Furió-Gómez, C. (2005). How are concepts and theories of acid–base reactions presented? Chemistry in textbooks and presented by teachers. *International Journal of Science Education*, *27*(11), 1337–1358.

Gentner, D. (2002). Psychology of mental models. In N.J. Smelser & P.B. Bates (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 9683–9687). Amsterdam: Elsevier Science.

Greca, I.M., & Moreira, M.A. (2000). Mental models, conceptual models, and modelling. *International Journal of Science Education, 22,* 1–11.

Grove, N.P., Hershberger, J.W., & Bretz, S.L. (2008). Impact of a spiral organic curriculum on student attrition and learning. *Chemical Education Research and Practice, 9,* 157–162.

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perception and intentions? *The Journal of the Learning Sciences*, *5,* 97–127.

Hand, B. (1989). Student understanding of acids and bases: A two year study. *Research in Science Education*, *19,* 133–144.

Hand, B.M., & Treagust, D.F. (1988). Application of a conceptual conflict teaching strategy to enhance student learning of acids and bases. *Research in Science Education*, *18,* 53–63.

Hewson, P.W., & Hewson, M.G.A. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. *Instructional Science*, *13,* 1–13.

Holum, J.R. (1969). *Introduction to organic and biological chemistry.* New York, NY: John Wiley & Sons.

Johnstone, A.H. (1993). The development of chemistry teaching: a changing response to changing demand. *Journal of Chemical Education*, *70,* 701–705.

Lin, J.-W., & Chiu, M.-H. (2007). Exploring the characteristics and diverse sources of students' mental models of acids and bases. *International Journal of Science Education*, *29,* 771–803.

Liptak, M.D., Gross, K.C., Seybold, P.G., Feldgus, S., & Shields, G.C. (2002). Absolute pK_a determinations for substituted phenols. *Journal of the American Chemical Society*, *124,* 6421–6427.

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

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

McClary, L., & Talanquer, V. (2011b). Heuristic reasoning in chemistry: Making decisions about acid strength. *International Journal of Science Education, 33*, 1433–1454.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science*, *210,* 1139–1141.

Mulford, D.R., & Robinson, W.R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education*, *79,* 739–744.

Nakhleh, M.B. (1994). Students' models of matter in the context of acid–base chemistry. *Journal of Chemical Education*, *71,* 496–499.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Nyachwaya, J.M., Mohamed, A.-R., Roehrig, G.H., Wood, N.B., Kern, A.L., & Schneider, J.L. (2011). The development of an open-ended drawing tool: An alternative diagnostic tool for assessing students' understanding of the particulate nature of matter. *Chemistry Education Research and Practice*, *12,* 121–132.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*, 988–1010.

Pavelich, M., Jenkins, B., Birk, J., Bauer, R., & Krause, S. (2004). *Development of a chemistry concept inventory for use in chemistry, materials and other engineering courses.* Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition (Paper no. 2004–1907), Salt Lake City, UT.

Popham, W.G. (2005). *Classroom assessment: What teachers need to know* (4th ed.). Boston, MA: Pearson Education.

Potgieter, M., & Davidowitz, B. (2011). Preparedness for tertiary chemistry: Multiple applications of the Chemistry Competence Test for diagnostic and prediction purposes. *Chemistry Education Research and Practice, 12,* 193–204.

Rahayu, S., Chandrasegaran, A.L., Treagust, D.F., Kita, M., & Ibnu, S. (2011). Understanding acid–base concepts: Evaluating the efficacy of a senior high school student-centered instructional program in Indonesia. *International Journal of Science and Mathematics Education, 9,* 1439–1458.

Redish, E.F. (2004). A theoretical framework for physics education research: Modeling student thinking. In E.F. Redish & M. Vicentini (Eds.), *Proceedings of the International School of Physics, 'Enrico Fermi' Course CLVI* (pp. 1–64). Amsterdam: IOS Press.

Ross, B., & Munby, H. (1991). Concept mapping and misconceptions: A study of high-school students' understandings of acids and bases. *International Journal of Science Education, 13,* 11–23.

Schmidt, H.-J. (1997). Students' misconceptions—looking for a pattern. *Science Education, 81,* 123–135.

diSessa, A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10,* 105–225.

Shah, A.K., & Oppenheimer, D.M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134,* 207–222.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.

Smith, J.P., diSessa, A.A., & Roschelle, J. (1993–1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115–163.

Stavy, R., & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive rules.* New York, NY: Teachers College Press.

Taber, K.S. (2008). Conceptual resources for learning science: Issues of transience and grain-size in cognition and cognitive structure. *International Journal of Science Education, 30,* 1027–1053.

Taber, K.S. (2009). College students' conceptions of chemical stability: The widespread adoption of a heuristic rule out of context and beyond its range of application. *International Journal of Science Education, 31,* 1333–1358.

Taber, K.S., & Garcia-Franco, A. (2010). Learning processes in chemistry: Drawing upon cognitive resources to learn about the particulate structure of matter. *The Journal of the Learning Sciences, 19,* 99–142.

Talanquer, V. (2006). Commonsense chemistry: A model for understanding students' alternative conceptions. *Journal of Chemical Education, 83,* 811–816.

Talanquer, V. (2009). On cognitive constraints and learning progressions: The case of structure of matter. *International Journal of Science Education, 31,* 2123–2136.

Talanquer, V. (2010). Exploring dominant types of explanations built by general chemistry students. *International Journal of Science Education, 32,* 2393–2412.

Todd, P.M., & Gigerenzer, G. (2000). Précis of *Simple heuristics that make us smart*. *Behavioral and Brain Sciences, 23,* 727–780.

Treagust, D. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education, 73,* 233–235.

Treagust, D.F., Chandrasegaran, A.L., Crowley, J., Yung, B.H., Cheong, I.P.-A., & Othman, J. (2010). Evaluating students' understanding of kinetic particle theory concepts relating to the states of matter, changes of state and diffusion: A cross-national study. *International Journal of Science and Mathematics Education, 8,* 141–164.

Turner, M., VanderHeide, K., & Fynewever, H. (2011). Motivations for and barriers to the implementation of diagnostic assessment practices – a case study. *Chemistry Education Research and Practice*, *12,* 142–157.

Villafañe, S.M., Bailey, C.P., Loertscher, J., Minderhout, V., & Lewis, J.E. (2011). Development and analysis of an instrument to assess student understanding of foundational concepts before biochemistry coursework. *Biochemistry and Molecular Biology Education*, *39,* 102–109.

Villafañe, S.M., Loertscher, J., Minderhout, V., & Lewis, J.E. (2011). Uncovering students' incorrect ideas about foundational concepts for biochemistry. *Chemistry Education Research and Practice*, *12,* 210–218.

de Vos, W., & Pilot, A. (2001). Acids and bases in layers: The strata structure of an ancient topic. *Journal of Chemical Education*, *78*(4), 494–499.

Voska, K.W., & Heikkinen, H.W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, *37,* 160–176.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45–69.

Vosniadou, S. (2002). Mental models in conceptual development. In L. Magnani & N. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp. 353–368). New York: Kluwer Academic Press.

Vosniadou, S., & Brewer, W. (1987). Theories of knowledge restructuring development. *Review of Educational Research*, *57,* 51–67.