December 2017

Integration of Scale-themed Instruction Across the General Chemistry Curriculum and Selected In-depth Studies

Jaclyn Trate
University of Wisconsin-Milwaukee

Recommended Citation

Trate, Jaclyn, "Integration of Scale-themed Instruction Across the General Chemistry Curriculum and Selected In-depth Studies" (2017). *Theses and Dissertations*. 1713.
https://dc.uwm.edu/etd/1713

INTEGRATING SCALE-THEMED INSTRUCTION ACROSS THE GENERAL CHEMISTRY CURRICULUM AND SELECTED IN-DEPTH STUDIES.

by

Jaclyn M. Trate

A Dissertation Submitted in
Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
in Chemistry

at

The University of Wisconsin-Milwaukee
December 2017

ABSTRACT

INTEGRATION OF SCALE-THEMED INSTRUCTION ACROSS THE GENEREAL CHEMISTRY CURRICULUM AND SELECTED IN-DEPTH STUDIES

by

Jaclyn Trate

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Kristen Murphy

In 1982, in response to a growing demand for a scientifically literate population, two organizations, the AAAS and NCISE published reports that proposed using themes to bridge scientific disciplines[1,2]. The NCISE report identified “9 explanatory concepts” which included organization, cause and effect, systems, scale, models, change, structure or function, discontinuous and continuous properties, and diversity. The AAAS report, as part of Project 2061, identified 4 themes that define science literacy which included systems, models, constancy and change, and scale. In 1993, the AAAS released the Benchmarks for Science Literacy[3] which outlined what all students should know or be able to do related to each common theme by the end of grades 2, 5, 8, and 12. However, prior to the release of the Framework for K-12 Science Education in 2012, and subsequent release of the Next Generation Science Standards in 2013, scale was not included in any national science education standards[4,5]. Now incorporated as one of seven crosscutting concepts, “scale, proportion, and quantity”, little is known regarding the degree to which scale is incorporated into instruction.

In disciplines like chemistry, undergraduate students are routinely confronted with concepts of scale and consistently demonstrate underdeveloped skills in understanding and applying concepts of scale. Previous research in this field led to the development of two assessments, the Scale Literacy Skills Test and Scale Concept Inventory[6], for measuring student scale literacy. Using these assessments, scale literacy was found to better predict student success in general chemistry than other traditional predictors of student success such as ACT and placement test scores. Expanding upon the work of Gerlach and co-workers, the work described here outlines the development and systematic integration of a scale-themed curriculum in both general chemistry I and II courses. Throughout 10 semesters of testing, supplemental instruction, laboratory experiments, and lecture instructional materials were developed and adapted to feature explicit themes of scale and implemented into both courses. When all three instructional methodologies are simultaneously administered, consistent positive conceptual learning gains are observed over repeated semesters of testing in general chemistry I.

References

1. National Center for Improving, Science Education; *Science and technology education for the elementary years: frameworks for curriculum and instruction;* Washington, D.C., 1989.

2. American Association for the Advancement of Science, Project 2061; *Science for all Americans: a project 2061 report on literacy goals in science, mathematics, and technology;* Washington, D.C., 1989.

3. American Association for the Advancement of Science, Project 2061; *Benchmarks for science literacy;* New York, New York: Oxford University Press, 1993.

4. National Research Council; *A framework for K-12 science education: practices, crosscutting concepts, and core ideas;* Washington, D.C., The National Academies Press: 2012.

5. National Research Council; *Next Generation Science Standards: for states, by states;* Washington, D.C., National Academies Press: Washington, D.C., 2013.

6. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Valid and Reliable Assessments to Measure Scale Literacy of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1538-1545.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Standing at the finish line and looking back on the journey that has brought me to this point, the truth is that I simply would not have accomplished this goal without the help and support of many people. Most importantly, I must thank my advisor, Kristen Murphy, for her role in my success. As both a mentor and role model, she has taught me how to conduct my research with great rigor and integrity, as well as, how to uphold the professional standards expected of a leading researcher at an R1 research university. Her passion for education and continual drive to improve the standards and expectations of both students and faculty within this department has been an inspiration over the last six years and I am beyond grateful for the opportunities I've had as her student. I also wish to thank the members of my committee, Peter Geissinger, Anja Blecking, Alan Schwabacher, and Don Wink, for their invaluable contributions to my research. Additional thanks are also given to the current and former members of my research group: Shalini Srinivasan, Lisa Kendhammer, Victoria Fisher, Erin O'Connell, and Karrie Gerlach, for their contributions to the success of the Chemistry Education Research discipline at UWM; the many undergraduates who contributed to this work, especially Brian Mohs and Ann Hackl; and the general chemistry instructors, teaching assistants, and students who made this work possible. Lastly, special thanks are due to my family, for their unwavering support in my pursuit of this goal. To Ken and Mary Jane, who I'm sure are relieved I finally have a "real" job; to Kim, Tim, and Lily, for "renting" me apartment Q (and a million other things I'd never be able to fit here); to Steve, Jaime, Riley, and Emma, for stepping up whenever help was needed; and most importantly, to Matt, Leo, and Jameson, for loving me and pushing me to finish this degree (even if it was just so that we would have more time to go to the zoo).

“Scale is a slippery concept, one that is sometimes easy to define but often difficult *to grasp. In the practice of archaeology, there is much equivocation about scale, as it is at the same time a concept, a lived experience, and an analytical framework.”* - *Gary Lock and Brian Molyneaux*

Chapter 1: Introduction

1.1 Introduction

In 1989 both the National Center for Improving Science Education (NCISE) and the American Association for the Advancement of Science (AAAS) published reports[1,2] in response to growing demand for a scientifically literate population. The NCISE report focused on outlining an elementary education curriculum framework built upon the idea that the world is changing at an accelerating pace and certain “explanatory” concepts could be used to organize students’ thoughts about the world. These concepts included organization, cause and effect, systems, scale, models, change, structure or function, discontinuous and continuous properties, and diversity. As part of Project 2061, the AAAS report identified four common themes that pervade science, mathematics, and technology that transcend disciplinary boundaries that included systems, models, constancy and change, and scale. In 1993, the AAAS followed Science for all Americans with Benchmarks for Science Literacy[3] which outlined what all students should know or be able to do in science, mathematics, and technology by the end of grades 2, 5, 8, and 12, with specific benchmarks aligned to each common theme.

If the goal of an educator in science is to increase science literacy[7], one must assume that instruction and assessment in science will align with those goals. As outlined in both the AAAS and the NCISE reports, the use of explanatory concepts or unifying themes in instruction are necessary to increase the effectiveness of science education and meet the desired outcomes.

However, of the four common themes identified by the AAAS, only scale had no supporting literature either upon initial publication or revision. Furthermore, it was not until the release of the National Research Council's Framework for K-12 Science Education[4] in 2012 and subsequent release of the Next Generation Science Standards[5] in 2013 that scale was explicitly included in national education standards as one of seven crosscutting concepts "scale, proportion, and quantity". Even more concerning is the fact that these curriculum guides specify only what a student needs to do to demonstrate proficiency and does not provide any guidelines for incorporation of any standard into instruction. Without detailed research pertaining to how students conceptualize scale and what concepts and ideas go in to understanding scale, an effective curriculum for teaching scale at any grade level cannot be developed. While prior research has attempted to answer these questions for science students at the K-12 and doctoral levels[8-10], pre- and in-service teachers[11-13], and experts in all domains of science, technology, engineering, and math[8], discipline based research on the importance of understanding scale, such as in chemistry, has been comparatively understudied in the post-secondary population[6,14,15].

In chemistry, students are immediately confronted with issues of scale as the entire discipline is rooted in a world far below the threshold of human sight. Given the lack of explicit scale instruction during primary and secondary education, it is no surprise that beginning college chemistry students demonstrate a profound deficiency in understanding and applying concepts of scale as it relates to an understanding of chemistry concepts. However, the work outlined in this dissertation demonstrates that performance on high-stakes final assessments in college chemistry courses can be predicted by how well a student understands scale and that the science literacy of undergraduate college chemistry students can be increased through targeted scale-themed instruction. Selected in-depth studies related to establishing the validity of this work are also presented in this dissertation.

1.2 References

1. National Center for Improving, Science Education; *Science and technology education for the elementary years: frameworks for curriculum and instruction.* Washington, D.C., 1989.

2. Project 2061 American Association for the Advancement of Science; *Science for all Americans : a project 2061 report on literacy goals in science, mathematics, and technology;* Washington, D.C. : American Association for the Advancement of Science: 1989; .

3. Project 2061 American Association for the Advancement of Science; *Benchmarks for science literacy;* New York, New York: Oxford University Press: 1993.

4. Taylor, A. R.; Jones, M. G.; Broadwell, B.; Oppewal, T. Creativity, inquiry, or accountability? Scientists' and teachers' perceptions of science education. *Science Education* **2008**, *92*, 1058-1075.

5. National Research Council *A framework for K-12 science education: practices, crosscutting concepts, and core ideas.* Washington, D.C.: The National Academies Press: 2012.

6. National Research Council *Next Generation Science Standards: for states, by states* Washington, D.C., 2013.

7. Tretter, T. R.; Jones, M. G.; Andre, T.; Negishi, A.; Minogue, J. Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching* **2006**, *43*, 282-319.

8. Tretter, T. R.; Jones, M. G.; Minogue, J. Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching* **2006**, *43*, 1061-1085.

9. Jones, M. G.; Taylor, A. R. Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching* **2009**, *46*, 460-475.

10. Trend, R. D. Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching* **2001**, *38*, 191-221.

11. Jones, M. G.; Tretter, T.; Taylor, A.; Oppewal, T. Experienced and Novice Teachers' Concepts of Spatial Scale. *International Journal of Science Education* **2008**, *30*, 409-429.

12. Jones, M. G.; Paechter, M.; Yen, C.; Gardner, G.; Taylor, A.; Tretter, T. Teachers' Concepts of Spatial Scale: An international comparison. *International Journal of Science Education* **2013**, *35*, 2462-2482.

13. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K. Valid and Reliable Assessments To Measure Scale Literacy of Students in Introductory College Chemistry Courses. *Journal of Chemical Education* **2014**, *91*, 1538-1545.

14. Karrie Gerlach; Jaclyn Trate; Anja Blecking; Peter Geissinger; Kristen Murphy Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry Courses. *Journal of Chemical Education* **2014**, *91*, 1526.

15. Swarat, S.; Light, G.; Park, E. J.; Drane, D. A typology of undergraduate students' conceptions of size and scale: Identifying and characterizing conceptual variation. *Journal of Research in Science Teaching* **2011**, *48*, 512-533.

Chapter 2: Literature Review

2.1 Defining Scale

In the dictionary scale is defined in terms such as “the proportion that a representation of an object bears to the object itself” and “a certain relative or proportionate size or extent.”[2] These definitions elicit a contextualization of scale in solely a “relational sense”, or the idea that scale refers only to a measure of proportion existing between something abstract and something more concrete. Numerous examples exist focusing on scale as an issue of “quantity” or on things that are in some way quantifiable[1] and scale has even been broadly defined as “any quantification of a property that is measured.”[3] However, Gary Lock and Brian Molyneaux explicitly detail why limiting scale to a quantifiable dimension ignores the fundamental element upon which scale differentiates itself from simply proportion or quantity.

> *“This understanding of scale as “analytical scale” is obviously important as it feeds into the process of archaeology’s basic tasks: collection, classification, and interpretation. Yet, there is much more to scale than this. Archaeology is not a remote laboratory pastime – it is a human task responding to a seemingly innate curiosity about history and a human construction of past events, meanings, and processes, from the traces that are left. Archaeologists deal implicitly with this qualitative and phenomenological aspect of scale every time they ponder the passing of time and the transformation of space.”*[1]

While written with a great deal of domain specificity, the underlying themes of this message are easily transferred to all disciplines rooted in human inquiry. The ability to not just understand that different phenomena occur on different scales but to be able to operate on the scale of which different phenomena occur becomes the foundation for an understanding of scale.

2.2 Prior research on Scale

2.2.1 Identifying conceptual boundaries

While one could argue that scale has appeared in literature prior to this[4], the first application of research into scale conceptions specifically within a science context didn't appear until 2001 and focused on K-5 teachers in the United Kingdom's perceptions of geologic time[5]. Participants in this study were asked to rank 20 "geo-events" using a 9-point Likert-type scale ranging from "more than approximately a million million years ago" to "less than a thousand years ago". Results of this study showed that these teachers demonstrated increased accuracy in ranking events closest to modern day and that these teachers held conceptions of historical time with distinct boundaries that could be categorized as "extremely ancient", "moderately ancient", and "less ancient". Interestingly, these categories shrunk in range as events moved from "extremely ancient" (a span of 10+ billion years) to "moderately ancient" (a span of 3+ billion years) to "less ancient" (a span of 50+ million years). This study concluded that conception of time becomes less well understood the farther back in history one goes, and that distinct "breaks" or "boundaries" existed in how these teachers conceptualized historical time across a continuum.

Expanding upon this work in 2006 to include students, Thomas Tretter, Gail Jones, and Amy Taylor[6] set out to measure the existing conceptualizations of of 5th, 7th, 9th, 12th, and doctoral students in science as it related to understanding linear distances. These students were picked to represent novice (5th-9th), gifted (12th) and expert (doctoral students) groups and were asked to complete activities or interviews that gave insight into how students in each group conceptualized scale. Students completed the Scale of Objects Questionnaire (SOQ) to assess the perceived size of 26 specified objects ranging from the size of an atomic nucleus to the distance between the Earth and the Sun. Students were given an object and a specified dimension (such as “width of a human hair”) and asked to indicate the size of each dimension using a Likert-type scale ranging from <1 nm to > 1 billion meters. After completing the SOQ, students were given 31 cards containing the name and picture of an object and asked to sort the cards by similarity of size. These cards also ranged in size from subatomic and galactic. Considering the results, moving across the expected trajectory of perceived scale knowledge (from novice through expert), the novice groups showed more variability in their relative ranking when compared to the gifted and expert groups and demonstrated the most difficulty in ranking the microscopic items. The gifted seniors exhibited less difficulty in ranking the microscopic items to within 1 place of the correct order and unsurprisingly, the experts placed all items correctly. Similar to the results found by Trend, the novice groups also consistently identified fewer categories as being distinctly different from one another when compared to the gifted senior and expert group.

Armed with the findings that students demonstrate different conceptions of scale depending on their age, Tretter, Jones, and colleagues[7] set out to measure how accuracy of spatial scale varies according to age and education and what strategies experts use to maneuver between different scales. In this study which utilized the same 5th, 7th, 9th, 12th, and doctoral students in science, participants were administered the Scale Anchoring Objects (SAO) assessment. Unlike the SOQ in which participants were given objects and asked to assign a dimension to them, the SAO gave participants a list of dimensions (first in units of meters, second in units of "body lengths") and asked them to identify an object typical of each size given. Results of this study showed that all groups were most accurate describing objects closest to their own size, between 1 decimeter and 10 meters in size. The novice students' accuracy dropped consistently between 1 decimeter and 1 millimeter before dropping drastically outside of 1 millimeter. Surprisingly, a similar result was not seen when considering measurements larger than 10 meters as novice students' accuracy continued to drop at a consistent pace between 10 meters and 1 billion meters. The expert and gifted senior groups showed comparable results to one another in accurately describing objects of a given size between 1 micrometer and 1000 meters. The accuracy of these groups outside of these dimensions followed the same pattern as the novice group, although not to the same degree, with accuracy dropping rapidly outside of 1 micrometer to the small end but steadily between 1000 meters and 1 billion meters to the large end. The novice groups also consistently reported feeling more confident when using their own body length as the unit as opposed to meters, while the gifted and expert groups favored the metric unit. In fact, when asked to use "body length" as the unit experts reported assigning the size of "1 meter" to their body and basing the rest of the comparisons from that unit. Lastly, when the experts were asked to describe how they thought about objects at the extreme small end of the scale used, the experts frequently mentioned the need to mentally jump to another scale in order to accurately think about the requested comparison, an observation not made with the novice students.

2.2.2 Scaling strategies of experts

Continuing to explore this work with experts, in 2008, Jones and Taylor[3] interviewed 50 experts from predetermined scale-laden professions and asked them to reflect upon both the importance of scale in their chosen careers and the educational experiences (both formal and informal) that contributed to their own developed sense of scale. The professions of those interviewed ranged from chemists, physicists, biologists, zoologists, neurologists, and engineers to pilots, sculptors, and auto body mechanics with all participants unequivocally stating that scale was integral to their understanding of and success in their chosen career. Looking across the self-reported experiences these experts used in the development of their sense of scale, many common themes emerged. Most notably, the experts frequently mentioned the use of body rulers and anchor points. For experts, using one's own body became a fast and reliable way to estimate distances such as the architect who commonly described using strides or arm lengths to estimate the functionality of a space, or the neurosurgeon who recalled using his thumb to identify a specific location on the brain that was "3 finger widths up and 2 over". The use of known size references, or anchor points, was often frequently referenced by the experts as well, such as the zoologist who used a red blood cell as the size reference for a micron or the materials scientist who used a virus as the size refence for a nanometer. These objects then become a useful standard for the comparison of other measurements. Culminating from the results of this research, Jones and Taylor proposed a Trajectory of Scale Concept Development (**Figure 2.1**) which outlines the 16 identified skills or concepts that contribute to an understanding of scale along with a relative timeline for development of each skill from novice to experienced. This trajectory along with a discussion of each included component follows in the next section.

Figure 2.1: Trajectory of Scale Concept Development

Novice

- Developing measurement estimation skills
- Conceptualizing relative sizes
- Using measurement tools skillfully
- Development of number sense

Developing

- Converting measurements and scales
- Surface area to volume relationships
- Being aware of changing scales
- Using body rules for measurement and estimation
- Visualizing scales
- Understanding different types of scales
- Development of proportional reasoning; Visual spatial skills

Experienced

- Automaticity and accuracy
- Creating reliable scales
- Relating one scale to another
- Developing accuracy in using scale
- Applying conceptual anchors when estimating scale

2.2.3 Trajectory of Scale Concept Development

One possible explanation for the apparent lack of literature referencing scale before the early 2000s could simply be that the term "scale" did not exist to mean what it does today. While several of the concepts and skills identified in the Trajectory of Scale Concept Development were in fact identified by Jones and Taylor through the research described in the previous section, many others find their roots in literature dating as far back as 1982. The trajectory outlined by Jones and Taylor proposes how one's sense of scale is developed over time beginning with concrete facts and moving into abstract conception.[8,9] Jones and Taylor provide evidence for this model using both their own research[10,11] as well as other key studies in how STEM curriculum in the United States is structured. For example, Jones and Taylor reference how, very early on in childhood education, students explore concepts of mathematical comparison and number sense[12,13] and elementary age students frequently work with measurement tools such as rulers and balances and explore ideas of estimation.[14] Jones and Taylor go on to reference that as students begin to mature, new skills such as proportional reasoning begin to emerge which allow students to begin to understand how changing scales can influence other variables such as surface area and volume.[15,16] Other skills such as converting measurements, increasing accuracy in making measurements or estimating, and learning to visually represent and manipulate scales[7] are also introduced and reinforced during this level of schooling. Finally, as was frequently observed during expert interviews and briefly described in the previous section, experts described accurately using body rulers and anchor points to maneuver between scales and were able to apply both strategies with increasing speed and accuracy as they gained experience both in school and on the job.

2.2.4 Deficiencies in scale

Based on the trajectory described in sections 2.2.2 and 2.2.3, and the attention paid to many of the identified concepts during primary and secondary education, one might wonder why students struggle when it comes to developing an understanding scale. One explanation for this observation could have to do with common reasoning patterns attributed to students. For example, students often assume a one-to-one correspondence between a model and the object or process being modeled.[17] Students may lack the understanding that interpretation of a model requires one to be able to fluently move between their world and the world in which the model exists on, which often requires the use of a new unit. This process, called "unitizing"[18], requires students to identify a new unit and mentally manipulate the new unit to make sense of numerical values.

Another possible explanation is that proportional reasoning skills for late elementary- and middle school-aged students don't emerge at the same time or rate for all students. Despite a heavy emphasis placed on proportional reasoning skills throughout middle school mathematics standards, [19-21] it is likely that students are in various stages of development of proportional reasoning skills during this time. As the use of proportional reasoning is required to move beyond a "developing" sense of scale in the trajectory outlined by Jones and Taylor, a lack of focus on these skills in post-secondary education could explain why some students demonstrate only a novice level of scale literacy.

2.3 Scale in chemistry

2.3.1 Scale as a theme in chemistry

"It's key. I mean in chemistry it is key. Again, because we are living in a macroscopic world and all the things that are composed are microscopic."- chemist on the importance of scale in their chosen career[8]

When considering the role of scale within the context of chemistry one can easily understand how maneuvering between the different representations used within chemistry would require a fluency with concepts of scale. Specifically, students need to be able to generate meaningful representations and use visual spatial and proportional reasoning skills to make meaning of representations to be successful in these courses. Alex Johnstone first described the 3 most used representations in chemistry as macroscopic, representational (symbolic), and sub-microscopic (particulate).[22] As students are most likely to have experienced chemistry only the macroscopic level it is not surprising that students would only demonstrate novice level ability to maneuver between these different levels of representation and demonstrate operational functionality within each dimension. Aligning with both the Trajectory of scale concept development and scale as defined by Lock and Molyneaux, the most relevant application of scaling concepts within the chemistry discipline were identified as falling into either "macroscopic/particle" or "number sense" categories (or combinations of both, call "scale").

These categories (**Figure 2.2**) feature predominantly within this work as an embodiment of both the quantitative and qualitative explanatory power scale brings to understanding chemistry concepts. For example, in beginning college chemistry courses students spend a great deal of time learning about states of matter and phase changes. Connecting the macroscopic observations made when ice is melted or water is boiled (disappearance of solid ice and appearance of liquid water) to the particle level properties (molecules gaining kinetic energy and overcoming intermolecucular forces) to the quantifiable aspects of this phenomenon (energy required to overcome these forces) requires students to use concepts of scale that fall into each category such as converting, relating scales, and applying conceptual anchors, among others. The connection of chemistry content through the use of these categorizations became the foundation upon which all scale-themed instructional materials were built.

Figure 2.2 Alignment of Trajectory of Scale Concept Development with chemistry content categories[a].

Macroscopic/Particle	Number Sense	Scale
• Relating one scale to another • Applying conceptual anchors when estimating scale	• Developing measurement estimation skills • Using measurement tools skillfully • Development of number sense • Converting measurements and scales • Surface area to volume relationships • Visualizing scales • Using body rules for measurement and estimation	• Development of proportional reasoning; Visual spatial skills • Creating reliable Scales • Understanding different types of scales • Conceptualizing relative sizes

[a] Three additional concepts: developing accuracy in using scale, automaticity and accuracy, and being aware of changing scales, were determined to reflect concepts related to expertise development and fall outside the scope of the work presented here.

2.3.2 Prior research on scale in chemistry

As previously stated, how undergraduate students in chemistry conceptualize scale has only recently been of interest in the literature. One study by Karrie Gerlach and colleagues[23] adapted the SOQ and SAO activities used by Tretter, Jones and Taylor to measure how beginning college chemistry students conceptualized scale. In this one-on-one interview activity participants were asked to create conceptual bins to encompass the entire spectrum of size (as perceived by the participant) before sorting 20 cards containing the name of an object into the previously identified bins. Results of this study showed consistency between the conceptual boundaries of scale held by beginning college chemistry students and the novice (5th, 7th, and 9th grade) students in Jones and Taylor's study. While Jones and Taylor had found that gifted seniors had begun to demonstrate a conception of scale closer to that of doctoral students, the undergraduate students in this study did not replicate that result.

In a separate publication related to the previously described work, Gerlach and co-workers[24] described the development and validation of two assessments, the Scale Concept Inventory and the Scale Literacy Skills Test, for use as class-wide assessments for measuring student ability in scale. These assessments were developed to assess student misconceptions about scale identified during preliminary student interviews (Scale Concept Inventory) and student conceptions of scale related to the content areas identified in the Trajectory of Scale Concept Development (Scale Literacy Skills Test). Both instruments were subjected to rigorous testing to ensure reliability and validity of these assessments for measuring conceptions of scale held by students through trial testing, expert content validation, and classical test theory. Comparison of performance on these assessments to final exam scores are shown in **Table 2.1** and surprisingly showed that scale literacy correlated as well or better to final exam scores than other traditional predictors of student success in general chemistry courses such as ACT composite and sub-scores or placement test scores.

Table 2.1: Common predictors of General Chemistry Performance[a]

	Final 1	Final 2
Math Placement	0.486	0.444
Chemistry Placement	0.513	0.493
Combined Placement	0.583	0.563
ACT Composite	0.514	0.509
ACT Mathematics	0.484	0.487
ACT Science Reasoning	0.430	0.437
Scale Literacy Skills Test (SLST)	0.550	0.606
Scale Concept Inventory (SCI)	0.401	0.466
Scale Literacy Score (SLS)[b]	0.583	0.650

[a]Pearson's product-moment correlation coefficient, r ($p < .001$ for all values); $n = 736$, [b]SLS calculated as average performance on both SLST and SCI

While it wasn't unexpected to find that student performance on these assessments could indicate a likelihood of success in general chemistry, the strength by which these association exists should not be understated. While this observation served not only as evidence that understanding scale plays a key role in understanding chemistry, but also that data collected from administration of these assessments could be used to develop, integrate, and assess meaningful instruction.

2.4 References

1. Lock, G.; Molyneaux, B. *Confronting Scale in Archaeology: Issues of Theory and Practice;* Springer Science & Business Media: New York, NY, 2007.

2. Definition of Scale*https://www.merriam-webster.com/dictionary/scale* (accessed Aug 29, 2017).

3. Jones, M. G.; Taylor, A. R., Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching.* **2009**, *46*, 460-475.

4. Golledge, R. G.; Gale, N.; Pelligrino, J. W.; Doherty, S., Spatial Knowledge Acquisition by Children: Route Learning and Relational Distances. *Annals of the Association of American Geographers.* **1992**, *82*, 223-244.

5. Trend, R. D. Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching.* **2001**, *38*, 191-221.

6. Tretter, T. R.; Jones, M. G.; Andre, T.; Negishi, A.; Minogue, J., Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching.* **2006**, *43*, 282-319.

7. Tretter, T. R.; Jones, M. G.; Minogue, J., Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching.* **2006**, *43*, 1061-1085.

8. Siegler, R. S. *Children's thinking;* Prentice Hall: Upper Saddle River, NJ, 1998.

9. Chen, Z.; Siegler, R. S. In *Intellectual Development in Childhood;* Sternberg, R. J., Ed.; Handbook of Intelligence; Cambridge University Press: Cambridge, UK, 2000; pp 92-116.

10. Jones, G.; Taylor, A.; Broadwell, B., Estimating Linear Size and Scale: Body rulers. *International Journal of Science Education.* **2009**, *31*, 1495-1509.

11. Taylor, A.; Jones, G., Proportional Reasoning Ability and Concepts of Scale: Surface area to volume relationships in science. *International Journal of Science Education.* **2009**, *31*, 1231-1247.

12. Greeno, J. G. Number Sense as Situated Knowing in a Conceptual Domain. *Journal for Research in Mathematics Education.* **1991**, *22*, 170-218.

13. Lamon, S. J. Ratio and Proportion: Connecting Content and Children's Thinking. *Journal for Research in Mathematics Education.* **1993**, *24*, 41-61.

14. Hofstein, A.; Lunetta, V. N., The Role of the Laboratory in Science Teaching: Neglected Aspects of Research. *Review of Educational Research.* **1982**, *52*, 201-217.

15. Tourniaire, F.; Pulos, S., Proportional Reasoning: A Review of the Literature. *Educational Studies in Mathematics.* **1985**, *16*, 181-204.

16. Vergnaud, G. In *Multiplicative structures;* Lesh, R., Landau, M., Eds.; Acquisition of mathematics concepts and processes; Academic Press, Inc.: New York, NY, 1983; pp 127-174.

17. Vicente Talanquer Minimizing misconceptions. *The Science Teacher.* **2002**, *69*, 46.

18. Lamon, S. In *Ratio and proportion: Cognitive foundations in unitizing and norming;* Harel, G., Confrey, J., Eds.; The development of multiplicative reasoning in the learning of mathematics; State University of New York Press: Albany, NY, 1994; pp 89-122.

19. Beaton, A. E.; Mullis, I. V.; Martin, M. O.; Gonzalez, E. J.; Kelly, D. L.; Smith, T. A. *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS);* International Association for the Evaluation or Educational Achievement. Center for the Study, Testing, Evaluation, and Education Policy: Chestnut Hill, NC, 1996.

20. Carpenter, T. P.; Coburn, T. G.; Reys, R. E.; Wilson, J. W., Notes from National Assessment: addition and multiplication with fractions. *The Arithmetic Teacher.* **1976**, *23*, 137-141.

21. Carpenter, T. P.; Corbitt, M. K.; Kepner, H. S.; Lindquist, M.; Reys, R. E. In *National assessment: prospective of students' mastery of basic skills;* Lindquist, M., Ed.; Selected issues in mathematics education; McCutchan: Berkeley, CA, 1980; .

22. Johnstone, A. H. The development of chemistry teaching. *Journal of Chemical Education.* **1993**, *70*, 701.

23. Gerlach, K; Trate, J; Blecking, A; Geissinger, P; Murphy, K., Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1526.

24. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Valid and Reliable Assessments to Measure Scale Literacy of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1538-1545.

Chapter 3: Methods

3.1 Introduction

The methods section is broken down into three main sections. The first section contains methods for data collection, course measures used, and data treatment over the entire data collection period. The second part contains experimental methods for the development of scale-themed laboratory experiments, and the third part contains experimental methods for the development of a scale-themed lecture curriculum.

3.2. General methods and courses of interest

This research was conducted over ten semesters at a large Midwestern, public, doctoral, R1 research university. The university has an undergraduate population of 21,000 with approximately one-third minority and first-generation students. The student population is 47% male and 53% female[1]. The research was conducted in both semesters of a two-semester general chemistry course with a course population majority of first and second year students. Data collection began during the Fall 2011 semester in general chemistry I and continued from the Fall 2012 semester through the Fall 2016 semester. In general chemistry II, data collection began during the Spring 2015 semester and continued through the Spring 2017 semester. All student data included in this dissertation were obtained via signed consent from all study eligible students to the University of Wisconsin-Milwaukee Institutional Review Board (IRB approval #s 09-047 and 14-404).

All statistical analyses described in this work were conducted using IBM® SPSS Statistics® unless otherwise noted.

3.2.1 General Chemistry I

General chemistry I is a 16-week traditional 5 credit laboratory/lecture/discussion course taken by science majors. The university prerequisite for this course includes a passing grade in intermediate algebra (or demonstrated algebra proficiency on a math placement test), or a passing grade in a preparatory chemistry course. Additionally, students are required to earn a passing score of 50% or above on a chemistry placement test (ACS Toledo Test) to maintain enrollment in the course. Students who do not score above that threshold are directed to a preparatory chemistry course.

The course typically covers 11 of the first 12 chapters of a traditional general chemistry textbook covering all content from classification of matter through intermolecular forces, while omitting the short introduction to organic chemistry found within this stream of content. Students are expected to attend 3 hours of lecture, 3 hours of laboratory, and 1 hour of teaching assistant-led discussion each week. Lecture assignments/assessments including 4 hourly exams, online homework, in class quizzes, and 2 nationally standardized final exams account for 75% of the student’s final grade in the course. Laboratory meets for 12 weeks of the semester completing 11 experiments or activities and 1 laboratory practical. Students complete weekly laboratory quizzes and experiment write-ups that account for 18.75% of their total grade in the course. The final 6.25% of the course is accounted for by discussion, in which students are expected to hand in answers to selected problems to earn weekly credit. At selected time points (described below) students may have received course credit or extra credit by completing selected assessments or surveys related to the research described in this dissertation. A detailed description of all course measures used in general chemistry I relevant to this research follows below. The full title of each assessment is followed by its more commonly referred to name in parenthesis.

Participant information (**Table 3.1**) including sex and ACT composite score (and sub scores) were collected from university institutional research data.

Table 3.1: General chemistry I descriptive statistics

	Male[a]	Female[b]	ACT Composite	ACT Reading	ACT English	ACT Math	ACT Science and Reasoning
n	*1133*	*1308*	*1981*	*1981*	*1981*	*1981*	*1981*
High			35	36	36	35	36
Low			11	7	10	13	11
Mean			23.37	23.63	22.71	23.14	23.46
Median			23	23	23	23	23
Mode			23	23	21	24	23
Standard Deviation			3.528	4.988	4.487	3.960	3.594
Skewness			-.089	.193	.071	.005	.261
Kurtosis			-.167	-.471	0.086	-.361	.546

[a,b] Includes total number of students who consented to participate in research via IRB protocol

Scale Supplemental Instruction (SI) consists of two self-paced adaptive activities that are “opened” to students during week 2 (activity 1) and week 14 (activity 2) of the semester. Each activity consists of 8 individual activities or assessments which are conditionally released based on performance in each of the eight segments of the activity. Developed and tested during the Fall 2010 and Spring 2011 semesters[2], these modules were designed to provide supplemental instruction in the development of both novice (activity 1) and developing (activity 2) concepts of scale and was officially launched into the general chemistry I curriculum during the fall 2011 semester. Beginning with this semester, supplemental instruction was offered to one section of the 2-section general chemistry I course each semester until full integration during the Fall 2016 semester (see **Tables 3.24 and 3.25** for cohort descriptions). Only students who completed all parts of both activities were eligible to remain in the data set for analyses in which supplemental instruction was considered. Beginning in the fall 2016 semester, the supplemental instruction portal was moved from the Desire2Learn course management system to a free-standing website[3].

The **ACS Exams Toledo Exam** (math placement, chemistry placement, total placement) is a 60-item placement test (20 math items and 40 chemistry items) administered during week 1 of the semester. Descriptive statistics related to this assessment are detailed in **Table 3.2**.

Table 3.2: General chemistry I placement test descriptive statistics

	Math Placement	Chemistry Placement	Total Placement
n	*2357*	*2357*	*2357*
High	20	37	57
Low	2	4	17
Mean	16.5	25.0	41.5
Median	17	25	42
Mode	17	26	42
Standard Deviation	2.4	4.2	5.6
Skewness	-.960	-.223	-.318
Kurtosis	1.986	.450	.429

The **Scale Literacy Skills Test** (SLST) is a 45-item multiple choice test that is administered via an online course management system (Desire2Learn) that is made active for one week (typically week 1, "SLST pre") in the beginning of the semester and one week (typically week 15, "SLST post") of the semester. Students receive their weekly lecture quiz points for completing the SLST pre and extra credit for completing the SLST post. The development and validation of this assessment is described comprehensively elsewhere[4]. Details related to the administration of this assessment along with selected descriptive statistics can be found in **Tables 3.3** and **3.4**. Complete item statistics for the Scale Literacy Skills Test can be found in **Appendix A**.

Table 3.3: General chemistry I Scale Literacy Skills Test administration.

Testing period		*n*
Fall 2011, Fall 2012-Fall 2016	Pre	141[a], 1893[b]
Fall 2011, Fall 2012-Fall 2016	Post	1419[b]

[a]administered via paper and pencil [b]administered via course management system

Table 3.4: General chemistry I Scale Literacy Skills Test descriptive statistics

		Pre	Post
n		*2034*	*1419*
Difficulty	High	0.916	0.942
	Low	0.060	0.178
	Mean	0.556	0.627
Discrimination	High	0.654	0.654
	Low	0.008	-0.011
	Mean	0.361	0.394
Overall (out of 45 possible)	High	42	43
	Low	6	8
	Mean	25.0	28.2
	Median	25	28
	Mode	25	29
	Standard deviation	6.3	6.9
	Skewness	.095	-.246
	Kurtosis	-.469	-.397

The **Scale Concept Inventory** (SCI) is a 40 item 5-point Likert scale survey (strongly agree to strongly disagree) that is administered via an online platform (Qualtrics™) to which students are emailed a link during week 1 (“SCI pre”) and week 15 (“SCI post”) of the semester. Students receive their week 1 discussion points for completing the SCI pre and extra credit for completing the SCI post. The development and validation of this survey is described comprehensively elsewhere[4]. Details related to the administration of this assessment along with selected descriptive statistics can be found in **Tables 3.5** and **3.6.** Complete item statistics for the objectively scored items of the Scale Concept Inventory can be found in **Appendix A**.

Table 3.5: General chemistry I Scale Concept Inventory administration.

Testing period		*n*
Fall 2011, Fall 2012-Fall 2016	Pre	472[a], 1187[b]
Fall 2011, Fall 2012-Fall 2016	Post	262[c], 906[b]

[a] administered via paper and pencil during the fall 2011, fall 2012, and one section of fall 2013 semesters. [b] administered via QualtricsTM. [c] administered via paper and pencil during the fall 2011 and fall 2012 semesters.

Table 3.6: General chemistry I Scale Concept Inventory descriptive statistics

		Pre	Post
n		*1659*	*1168*
Overall %	High	86	96
	Low	51	46
	Mean	66	68
	Median	65	67
	Mode	63	66
	Standard Deviation	6	7
	Skewness	.550	.756
	Kurtosis	.982	.671

Student performance on both the SLST and SCI are averaged to give the **Scale Literacy Score** “SLS pre” and “SLS post”.

The final measures used in general chemistry I are the **ACS Exams 2005 First Term General Chemistry Paired Questions Exam** ("paired final") and the **ACS Exams 2008 General Chemistry Conceptual Exam - First Term** ("conceptual final"). The paired final consists of 20 traditional/conceptual item pairings (40 total items) in which the traditional item always precedes the conceptual item. The conceptual final consists of 40 conceptual items. Selected descriptive statistics for these assessments can be found in **Table 3.7.**

Table 3.7: General chemistry I final exam descriptive statistics

		Paired Final	Conceptual Final
	n	*2036*	*2036*
Overall (out of 40)	High	39	40
	Low	8	7
	Mean	27.6	23.5
	Median	27	23
	Mode	27	22
	Standard Deviation	5.7	6
	Skewness	-.276	.112
	Kurtosis	-.445	-.464

3.2.2 General Chemistry II

General chemistry II is structured in the same way as general chemistry I as a 5-credit lecture/laboratory/discussion course with a university prerequisite of a grade of C or higher in general chemistry I or a score of 4 or higher on the AP® Chemistry exam. This course traditionally covers 8 chapters, beginning with a review of intermolecular forces and ending with electrochemistry. In general chemistry II, students are again expected to attend 3 hours of lecture, 3 hours of laboratory, and 1 hour of teaching assistant led discussion each week. Lecture assignments/assessments including 4 hourly exams, online homework, in class quizzes, and 2 nationally standardized final exams account for 75% of the student's final grade in the course. In laboratory, general chemistry II students meet for 11 weeks of the semester completing 10 experiments and 1 laboratory practical. Unlike general chemistry I, regular lab meetings do not being until the third week of instruction to allow for completion of an online nomenclature review activity. Students complete weekly laboratory quizzes and experiment write-ups that account for 18.75% of their total grade in the course. The final 6.25% of the course is accounted for by discussion, in which students are expected to hand in answers to selected problems to earn weekly credit. At selected time points (described below) students may have received extra credit or earned regular credit by completing selected assessments or surveys related to the research described in this dissertation. All measures used in general chemistry II remained consistent with those used in general chemistry I with the exception of the ACS Exams Toledo placement test and the ACS Exams 2008 General Chemistry Conceptual Exam - First Term, which were not used. The **ACS Exams 2005 First Term General Chemistry Paired Questions Exam** (placement test) was used as a low stakes placement test and the 40 item **ACS Exams 2008 General Chemistry Conceptual Exam – Second Term** (Conceptual final) was administered as the second final measure. Details related to the administration of these assessments along with selected descriptive statistics can be found in **Tables 3.8-3.13.** Scale **supplemental instruction** (SI) was developed and tested during the Spring 2017 semester (see **Appendix D.1** for a description of these activities). See **Tables 3.26 and 3.27** for complete cohort descriptions.

Table 3.8: General chemistry II selected course measure descriptive statistics

	Male[a]	Female[b]	ACT Composite	ACT Reading	ACT English	ACT Math	ACT Science and Reasoning
n	*371*	*477*	*675*	*675*	*675*	*675*	*675*
High			34	36	36	34	36
Low			14	12	9	15	11
Mean			23.83	23.98	23.29	23.60	23.85
Median			24	24	23	24	24
Mode			23	24	22	26	24
Standard Deviation			3.68	5.10	4.63	4.05	3.74
Skewness			.046	.078	.062	-.062	.277
Kurtosis			-.362	-.670	.078	-.439	.393

[a,b] Includes total number of students who consented to participate in research via IRB protocol

Table 3.9: General chemistry II selected course measure descriptive statistics

		Placement	Paired Final	Conceptual Final
n		*818*	*759*	*759*
Overall (out of 40)	High	40	40	36
	Low	7	15	6
	Mean	24.8	29.6	22
	Median	25	30	22
	Mode	24	29	21
	Standard Deviation	6.4	5.2	6
	Skewness	-.223	-.480	.159
	Kurtosis	-.422	-.184	-.637

Table 3.10: General chemistry II Scale Literacy Skills Test administration

Testing period		*n*
Spring 2015 - Spring 2017	Pre	740
Spring 2015 - Spring 2017	Post	540

Table 3.11: General chemistry II Scale Literacy Skills Test descriptive statistics

		Pre	Post
n		*740*	*540*
Difficulty	High	0.943	0.937
	Low	0.158	0.146
	Mean	0.633	0.624
Discrimination	High	0.681	0.733
	Low	0.059	0.111
	Mean	0.398	0.419
Overall (out of 45 possible)	High	43	42
	Low	9	8
	Mean	28.5	28.1
	Median	29	29
	Mode	30	33
	Standard deviation	6.9	7.4
	Skewness	-.333	-.336
	Kurtosis	-.276	-.599

Table 3.12: General chemistry II Scale Concept Inventory administration

Testing period		*n*
Spring 2015 - Spring 2017	Pre	647
Spring 2015 - Spring 2017	Post	470

Table 3.13: General chemistry II Scale Concept Inventory descriptive statistics

		Pre	Post
n		*647*	*470*
Overall %	High	93	95
	Low	55	53
	Mean	68.8	68.4
	Median	68	67
	Mode	67	65
	Standard Deviation	6.6	7
	Skewness	.862	.996
	Kurtosis	.705	1.038

3.2.3 Cleaning data

When appropriate, student data was cleaned prior to analysis. This consisted of removing student scores for those students who did not correctly answer verification items, those students with a response set variance equal to zero, or when system generated time stamps showed students completing an assessment in less time than required to read each item (a threshold of 4 minutes). This method resulted in a <5% removal rate of students in any particular data set. In only one case was post hoc removal of student data considered when it was determined that the student's residual score on a portion of the final exam was identified through statistical means as an extreme outlier after the distribution of residuals from that semester failed assumptions of normality. Removal of that one score did not alter the predictive model upon which his scores had previously contributed and the normality assumption upon his removal was reinstated.

3.2.4 Missing data

Missing data were treated according to predetermined methods as deemed appropriate. At times, missing scores warranted the removal of other associated scores, such as the removal of ACT sub-scores when no ACT composite score was reported (1 case), the removal of placement test total scores when either the math (4 cases) or chemistry portion (0 cases) was not completed, or removal of final exam scores if the paired final (3 cases) and conceptual final (13 cases) were not both completed. For assessments first completed on paper and later completed electronically, such as the SLST or SCI, scores were removed for students completing the assessment on paper who did not answer every question. These removals often caused complete removal from analysis in instances where only consenting students who completed all necessary measures could be included (i.e. paired sample t-tests and multiple regression and residual analysis). Additionally, students receiving an overall grade in the course but who did not complete the final exam(s), were not eligible to be included in analysis (217 total cases across both courses).

3.2.5 Building a predictive model for general chemistry I:

A thorough evaluation of all collected data from the fall 2011 and fall 2012 semesters revealed the three most significant predictors of student success in general chemistry I to be the ACT composite score, the combined math and chemistry placement test score, and the scale literacy pre score (**Table 3.14**). These variables were chosen for multiple regression analysis to predict student performance on each final exam. The decision to use final exams as performance measures as opposed to final course percent was made to account for other course aspects (laboratory, homework, extra credit) that are included in that calculation but are less indicative of true student ability on targeted chemistry concepts. That is not to say however, that these other measures were ignored in totality but were rather analyzed within the context of which that data was collected. The full correlation matrix for all analyzed course measures for both the 142 students included in the control group semesters as well as for all semesters of testing can be found in **Appendix A**.

Table 3.14: Predictors of General Chemistry I Final Exam Performance[a]

	Final 1	Final 2
Math Placement	0.508	0.419
Chemistry Placement	0.587	0.527
Combined Placement	0.680	0.593
ACT Mathematics	0.468	0.451
ACT Science Reasoning	0.424	0.443
ACT Composite	0.578	0.571
Scale Literacy Skills Test (SLST)	0.582	0.651
Scale Concept Inventory (SCI)	0.368	0.383
Scale Literacy Score (SLS)	0.587	0.646

[a] Pearson's product-moment correlation coefficient, r ($p < .001$ for all values); n = 142.

3.2.5 Building a predictive model for general chemistry II:

A similar evaluation of all collected data from the spring 2015 semester (**Table 3.15**) revealed the three most significant predictors of student success in general chemistry II to also be the ACT composite score, the placement test score, and the scale literacy pre score. These variables were again chosen for multiple regression analysis to predict student performance on the conceptual final exam. The decision to only use the conceptual final exam was made to account the paired question final being used as both a final exam and as a placement test in this course. The full correlation matrix for all analyzed course measures for both the 93 students included in the control group as well as for all semesters of testing can be found in **Appendix A**.

Table 3.15: Predictors of General Chemistry II Final Exam Performance[a]

	Final 1	Final 2
Placement Test[b]	0.785	0.646
ACT Composite	0.642	0.606
ACT Mathematics	0.586	0.499
ACT Science Reasoning	0.584	0.586
Scale Literacy Skills Test (SLST)	0.547	0.515
Scale Concept Inventory (SCI)	0.451	0.545
Scale Literacy Score (SLS)	0.577	0.585

[a]Pearson's product-moment correlation coefficient, r ($p < .001$ for all values); $n = 93$. [b]Placement Test and Final 1 are both the ACS Exams Paired Question Exam.

3.3 Development of scale-themed laboratory experiments

3.3.1 Development of laboratory experiments for general chemistry I:

Following the implementation of scale supplemental instruction, laboratory was selected as the next aspect of the general chemistry I curriculum to feature explicit scale themes. However, before any work could be done to this end, a thorough evaluation of the current laboratory curriculum was needed. The laboratory manual previously used in General chemistry I was examined for both content coverage and explicit scale themes. As seen in **Table 3.16**, no experiment covered concepts in material beyond an introduction to thermodynamics (7th week of lecture instruction) and a heavy emphasis was placed on aqueous solutions (4th/5th week of lecture instruction). Furthermore, no explicit use of scale themes was evident in any of the existing laboratory experiments. Instead a technical focus existed in which students were frequently expected to provide only a surface level understanding of the system being studied in that experiment. Expanding both the content coverage and depth of information covered in each experiment was of primary interest to the research team. This meant not only adapting the objectives, pre-laboratory questions, and results and discussion questions of the existing laboratory activities but also adapting published activities or writing new experiments to better fit the desired content coverage. Of the ten experiments conducted over 11 weeks of the course, only the "Safety and Skill Inventory" was retained in its same format for use in the new laboratory sequence. To make room for experiments that expanded the content coverage of laboratory, the "Nomenclature" activity (completed online) was moved into the grading structure for lecture, the "Find the Relationship" activity and "Standardization of Solutions" experiment were eliminated, and the two week "Qualitative Analysis" experiment was adjusted to be completed during a single 3-hour laboratory period. The seven experiments that remained were altered from their current formats to fit the new scale themed curriculum objectives. As these experiments would be taught by teaching assistants an accompanying "TA Manual" was developed for each experiment in which answers and grading schemes for all pre-lab, results and calculation questions could be found, as well as helpful hints and descriptions of things to include during laboratory instruction.

Table 3.16: General chemistry I laboratory experiments

Content Area	Ch.	Experiment
Skills, Safety	1	Safety and Skill Inventory
Measurements, Physical changes, and using the equipment	1	Physical Properties of Water
Using the software and Nomenclature	2	Find the Relationship and Nomenclature
Aqueous reactions	4	Qualitative Analysis, Week 1
Aqueous reactions	4	Qualitative Analysis, Week 2
Aqueous reactions	4	Standardization of Solutions
Aqueous reactions	4	Stoichiometry and Acid/Base Titrations
Gases	5	Gas Laws
Enthalpy	6	Enthalpy
Reactions	3	Copper Cycle
Concentration	4	Beer's Law

3.3.2 Development of laboratory experiments for general chemistry II:

Following a similar method as to what was done in general chemistry I, the laboratory manual used in general chemistry II was evaluated for both content coverage and use of explicit scale themes. While these experiments were also found to lack any use of explicit scale themes, unlike in general chemistry I, the content coverage of these experiments was unexpectedly broad. As seen in **Table 3.17**, these experiments covered the entire breadth of the content covered in general chemistry II.

Table 3.17: General chemistry II laboratory experiments

Content Area	Ch.	Experiment
Gas laws/Properties of liquids	5/12	The Molar Mass of a Volatile Liquid
Physical properties of solutions	13	Freezing Point Depression
Chemical kinetics	14	Rate and Order of a Chemical Reaction
Chemical kinetics	14	Rate determination and Activation Energy
Chemical Equilibrium	15	Determination of an equilibrium constant
Acids and Bases/Acid-Base Equilibria	16/17	Buffers
Solubility Equilibria	17	Determine the Ksp of Calcium Hydroxide
Thermodynamics	6/18	Hess's Law – Heat of Combustion
Complex Ion Equilibria/Electrochemistry	17/19	What's in a Penny
Electrochemistry	19	Electrochemistry – Voltaic Cells

However, while the content coverage was sufficient, the experiments themselves were written in such a way that a very heavy emphasis was placed on error determination and very little information about the system of study was given in the introduction. Given the heavy quantitative nature of the topics covered in general chemistry II, it was suggested that this in and of itself makes the content more difficult for students to comprehend. It was therefore determined that these labs would also be altered to give students information related to an analogous system of study to the one covered by the laboratory experiment with the ultimate goal of students being able to focus on the chemistry of the experiment as opposed to trying to simultaneously understand the methods employed in the experiment as well as the chemistry that is happening. A detailed explanation of how this was accomplished follows in section 4.2.4. Similar to general chemistry I, an accompanying "TA Manual" was also produced for this set of laboratory experiments.

3.3.3 Development of a laboratory survey

Keeping in mind the importance of student feedback when instituting curricular change, it was decided that a laboratory survey would be developed and administered during both the first laboratory class meeting and following the practical exam during the last laboratory class meeting. Details related to the administration of this survey and selected descriptive statistics for both pre and post administrations in both general chemistry I and II are shown in **Tables 3.18 and 3.19**. The survey contained 13 objective items, 6 subjective items, and 1 verification item and was scored on a 5-poing Likert scale (strongly agree (1) to strongly disagree (5)). The 13 objective items centered around specific ideas and misconceptions that have appeared in laboratory related to measurement, error, and number sense. These items specifically addressed ideas such the accuracy of common measurements taken in the lab and determining the reasonableness of commonly calculated values. The subjective items gathered information related to the role of lab (as perceived by students) and the desired outcomes of the enhanced curriculum (as planned by the researchers). The verification item stated, "Of lab, lecture, and discussion, lab gives the most hands-on approach to understanding chemistry concepts".

Students who did not respond “strongly agree” or “agree” to the verification item were removed from analysis, as well as, students whose responses did not have a variance greater than zero. A list of the survey items as well as the percent chosen for all objectively scored items can be found in **Appendix B**.

Table 3.18: General chemistry I and II laboratory survey administration

Semester Included		General Chemistry I, n	Semester Included		General chemistry II, n
Spring 2013 - Fall 2016	Pre	1724	Spring 2015-Spring 2017	Pre	732
Fall 2012 - Fall 2016	Post	1613	Spring 2015-Spring 2017	Post	661

Table 3.19: General chemistry I and II laboratory survey descriptive statistics

Overall %	GC I pre	GC I post	GC II pre	GC II post
n	*1724*	*1613*	*732*	*661*
High	98	100	97	100
Low	50	50	53	53
Mean	70	77	73.3	75
Median	70	77	73	75
Mode	70	77	73	75
Standard Deviation	7	8	7	8
Skewness	.319	.132	.091	.179
Kurtosis	.339	.061	-.122	-.071

3.3.5 Development of pre-laboratory quizzes

Prior to the integration of the newly developed curriculum, students took pre-laboratory quizzes that centered solely around answering the question “did the student read the experiment prior to entering the laboratory?" Given the idea that each laboratory experiment can be viewed independently from an assessment standpoint, the researchers felt that the pre-laboratory quizzes offered a unique opportunity to gauge student understanding of chemistry concepts both before completing an experiment and after completing an experiment centered around those concepts. The quizzes were written so that a student would need to engage with specific scale concepts as they related to the chemistry concept tested in the quiz item. The quizzes consisted of three questions and followed the same format every week. The first question, with the exception of the first quiz of the semester, followed the format of "From your experiment last week…" and asked the student to answer a follow up question to the previously conducted experiment (the "post" question). The second question, "what is one concept from your textbook that you are applying in this week's experiment?" remained the same on every quiz, was graded on completion, and student responses to this item were not analyzed. The third question followed the format of "For your experiment this week…" and asked the student to answer a question related to the experiment they would be completing that day in laboratory (the "pre" question). Quizzes were given within the first 10 minutes of the lab period (prior to being given any information regarding the experiment they would be conducting) and were collected and graded by a member of the research team or by their teaching assistant using an established set of quiz keys (**Table 3.20**). All questions of the quiz were graded on a 4-point scale initially using the rubrics depicted in **Figures 3.1-3.2** while simultaneously cataloguing student responses for the purposes of developing the more descriptive quiz keys used later. The rubrics were used by four raters initially, revised and tested again with six raters. The reliability was 0.898 (as measured by Cronbach's alpha). Scores for each quiz question were used to calculate a single quiz score which was then added to each student's lab report as a possible five extra credit points. A list of the quiz questions for both general chemistry I and general chemistry II can be found in **Appendix B**.

Table 3.20: General chemistry I and II pre-laboratory quiz grading

Course	Graded by research team	Graded by teaching assistant
GC I	Spring 2013 Spring 2014	Fall 2014 – present
GC II	Spring 2015 Fall 2015	Spring 2016 - present

Figure 3.1: Pre-laboratory quiz grading rubric for items not requiring reasoning

Numerical rating	Answer correctness
4	The student answered all parts correctly
3	The student answered the majority of the question(s) correctly
2	The student answered half of the question(s) correctly or was partially correct
1	The student recorded an answer however, that answer was totally incorrect
0	The student did not answer the question

Figure 3.2: Pre-laboratory quiz grading rubric for items requiring the student to provide reasoning for their answer.

Numerical rating	Answer correctness	
4	The student provided a correct response	The student's reasoning is correct and supports their answer
3	The student provided a correct answer	The student provided a reason for their answer, however the reasoning is not correct
2	The student did not answer correctly	The student provided a reason that supported their answer, but the answer was incorrect
1	The student did not answer correctly	The student either did not provide a reason, or their reason did not support their answer.
0	The student did not answer the question	

3.3.6 Complexity analysis of pre-laboratory quiz items

As described previously, the researchers desired to compare student performance from the pre quiz question to the post quiz question. However, given the inherent (and purposeful) more difficult nature of the post question, making a straight comparison between performance on the two items would make it appear as though students actually performed lower on the post question as compared to performance on the pre question[5]. For this reason, the complexity of each quiz question needed to be considered. Following an established protocol for rating the complexity of general chemistry items[6] independent ratings were made for each of the pre-laboratory quiz items used in both general chemistry I and II in order to create weighted performance scores based on complexity.

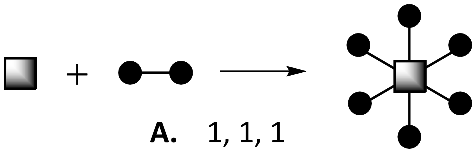
3.4 Development of scale-themed lecture slides and activities

3.4.1 Conceptual versus Algorithmic analysis

A variety of methods were employed to develop a lecture curriculum built on empirical data. The first method that was employed attempted to compare student performance on hourly course exams with their scale literacy score. To do this, the multiple-choice items from four hourly exams given during the Fall 2011 and Fall 2012 semesters in general chemistry I (n = 116) and the Spring 2015 semester in general chemistry II (n = 103) were rated as either testing conceptual or algorithmic content knowledge. Algorithmic items were identified as anything a student could solve using a defined process (see **Figures.3.3-3.4** for example items). The ratings were completed by 4 raters (2 faculty and 2 graduate student) and discrepancies were discussed until agreement was met. Performance sub-scores were calculated for total performance in each category as well as for performance in each category for each chapter in which the content the items tested were from. The items fell into categories as described in **Tables 3.21 and 3.22**. Categories containing fewer than 3 items were excluded from analysis.

Figure 3.3: General chemistry I item rated as algorithmic

What are the simplest whole number coefficients in this equation when balanced?



A. 1, 1, 1
B. 1, 3, 1
C. 1, 6, 1
D. 2, 1, 2

Figure 3.4: General chemistry I item rated as conceptual

Which is true?

A. Salt is an element.
B. CO_2 is a compound.
C. C_2H_6O is a homogeneous mixture.
D. Chlorine gas is a heterogeneous mixture.

Table 3.21: General chemistry I items by chapter

Chapter	1	2	3	4	5	6	7	8	9	10	11	12	Total
Conceptual Items	5	5	5	11	7	12	2	1	4	3	--	--	55
Algorithmic Items	-	8	12	11	8	5	5	6	2	3	1		61

Table 3.22: General chemistry II items by chapter

Chapter	13	14	15	16	17	18	19	Total
Conceptual Items	8	10	8	7	11	12	1	66
Algorithmic Items	5	7	3	7	3	5	3	47

3.4.2 Development of a scale concept learning progression

The second method that was employed was to create a learning progression for the general chemistry I students based on their performance on the Scale Literacy Skills Test as it related to the scale concept trajectory published by Jones and Taylor[7]. Using a weighted average based on complexity, a performance sub-score in each of the 12 scale concept areas tested on the exam were calculated for the 1750 general chemistry I students who took the Scale Literacy Skills Test between the fall semester of 2009 and the spring semester of 2013. Based on these averages, shown in **Table 3.23**, a learning progression was generated to visualize the order in which these students appear to develop ability in these areas.

Table 3.23: Average weighted performance of general chemistry I students on content areas tested on the Scale Literacy Skills Test.

Scaling concept	Trajectory assignment	Average weighted difficulty
Measurement and Estimation	Novice	1.151
Relative sizes	Novice	1.093
Making measurements	Novice	1.061
Number sense	Novice	.792
Converting	Developing	.701
Unitizing to self	Developing	.713
Visualizing scales	Developing	.811
Understanding different scales	Developing	.948
Visual spatial skills	Developing	1.352
Creating reliable scales	Experienced	1.225
Relating scales	Experienced	1.438
Applying conceptual anchors	Experienced	.626

3.4.3 Development of a general chemistry I content map

The third method employed by the research team was to identify and map the entire curriculum of a traditional general chemistry I lecture course. There were two primary objectives to this practice. The first was done to ensure that the research team was satisfied with the order in which the content was presented to the students. If any content seemed to be out of place or should logically be moved to another place within the course, it would need to be done at this stage of the project. Additionally, this practice allowed the team to identify both where concepts of scale were inherently used (at least in theory) or could be explicitly used to connect aspects of the curriculum. The mapping was completed by two faculty members (veteran instructors of record for the course) and a graduate student. The mapping occurred using a magnetic white board and magnets containing the concepts or ideas contained in each chapter.

Each chapter's ideas were given their own color coding scheme and the pieces for the entire curriculum were laid out and manipulated on the board.

3.4.3 Development of scale-themed lecture slides and notes for general chemistry I:

During the fall 2013 semester, the existing slides and notes used during instruction in general chemistry I were transformed to include explicit scale themes. With the exception of the addition of several slides explaining what scale is and how it related to the course of study, there were relatively few changes to the slides themselves. The most obvious change to the lecture slides was the addition of a "scale symbol" to the bottom left hand corner of each slide (pictured in **Figure 3.6**) meant to serve as a reminder to students to draw upon scale when thinking about the concept, problem, or idea presented on that slide. Class notes, however, drew upon scale concepts frequently, although these connections were often made verbally using the idea or concept presented in the slide. Additionally, a rubric was developed for classroom observations in order to capture the students' responses to the inclusion of scale in the lecture materials. Aiding these observations, all lectures were also audio and video recorded.by video and audio. These were used as references when evaluating the inclusion of scale into instruction. Following the first use of the scale-themed lectures, there was also a discussion between the observer and the lecturer following every lecture on the inclusion of scale. These slides were also used during the Spring 2014 semester combined treatment in conjunction with the scale-themed laboratory experiments.

Figure 3.5: symbol used to denote presence of scale on lecture slides

3.4.4 Development of active learning lecture activities for general chemistry I and II:

The scale themed lecture slides and notes used during the Fall 2013 and Spring 2014 semesters of general chemistry I were used to create scale-themed active learning lecture activities. These activities were designed to act as an outline for each lecture in which students would follow along and fill in the activities as content was covered. Additionally, these activities provided opportunities for students to collaborate with other students to predict and build explanations for ideas contained in the activity. These activities built upon the scale-themed lecture slides and notes by transforming what was previously only a verbal connection to scale into more formal instruction. Simultaneous to the development of the scale-themed active learning lecture activities, a second set of non-scale active learning lecture activities in which scale was not present was developed for the purposes of elucidating the impact of both the scale-themed content itself and of the content delivery method. During the fall 2015 semester, an analogous set of scale-themed active learning lectures activities as well as non-scale active learning lecture activities were developed and implemented in general chemistry II. Examples of how these activities were presented to students in these courses can be found section 4.2.2. The active learning lecture activities developed for general chemistry I were implemented during the fall semesters of 2014, 2015, and 2016. The non-scale active learning lecture activities were implemented during the spring semesters of 2015 and 2016. In general chemistry II, the active learning lecture activities were implemented in both sections of the spring 2016 semester and one section of the spring 2017 semester. The non-scale active learning lecture activities were implemented during the fall 2016 semester and one section of the spring 2017 semester.

3.5 Scale-themed instruction integration schedule

Table 3.24: General chemistry I Scale-themed instruction cohort assignments

	Fall 2011 Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014 Fall 2015	Spring 2015 Spring 2016	Fall 2016
Lec I	Control	Laboratory Experiments	Lecture	Lecture + Laboratory Experiments	Active learning + Laboratory Experiments	Non-scale Active learning + Laboratory Experiments	Active learning + Laboratory Experiments + Supplemental Instruction
Lec II	Control + Supplemental Instruction	Laboratory Experiments + Supplemental Instruction	Lecture + Supplemental Instruction	Lecture + Laboratory Experiments + Supplemental Instruction	Active learning + Laboratory Experiments + Supplemental Instruction	Non-scale Active learning + Laboratory Experiments + Supplemental Instruction	Active learning + Laboratory Experiments + Supplemental Instruction

Table 3.25: General chemistry I cohort sample sizes

	Fall 2011 Fall 2012	Spring 2013	Fall 2013	Spring 2014	Fall 2014 Fall 2015	Spring 2015 Spring 2016	Fall 2016
Lec I	73 146	117	87	118	104 109	90 113	117
Lec II	78 152	152	109	141	151 148	132 150	157

Table 3.26: General chemistry II scale-themed instruction cohort assignments

	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017
Lec I	Control	Laboratory Experiments	Active learning + Laboratory Experments	Non-scale Active learning + Laboratory Experiments	Active learning + Laboratory Experiments + Supplemental Instruction
Lec II	Control	Not offered	Active learning + Laboratory Experments	Not offered	Non-scale Active learning + Laboratory Experiments + Supplemental Instruction

Table 3.27: General chemistry II cohort sample sizes.

	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017
Lec I	78	145	86	174	85
Lec II	104	[a]	91	[a]	86

[a] section not offered

3.6 References

1. Facts and Impacts *http://uwm.edu/facts/* (accessed Sep 1, 2017).

2. Gerlach, K. L. The Measure and Instruction of Scale in Introductory Chemistry, Thesis (M.S. in Chemistry), University of Wisconsin-Milwaukee, 2012.

3. The Scale Project *www.scale.chem.uwm.edu* (accessed September 14, 2017).

4. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Valid and Reliable Assessments to Measure Scale Literacy of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1538-1545.

5. Murphy, K. L.; Holme, T. A., Improving Instructional Design with Better Analysis of Assessment Data. *Journal of Learning Design.* **2014**, *7*, 29-46.

6. Knaus, K.; Murphy, K.; Blecking, A.; Holme, T., A valid and reliable instrument for cognitive complexity rating assignment of chemistry exam items. *Journal of Chemical Education.* **2011**, *88*, 554.

7. Jones, M. G.; Tretter, T.; Taylor, A.; Oppewal, T., Experienced and Novice Teachers' Concepts of Spatial Scale. *International Journal of Science Education.* **2008**, *30*, 409-429.

Chapter 4: Results and Discussion

4.1 Introduction

The results section is broken down into three main sections. The first section contains the final details and examples of the scale-themed laboratory curriculum developed for general chemistry I and II. The second section contains both the results that guided the development of and examples of the scale-themed lecture curriculum developed for both general chemistry I and II. The final section contains results and discussion related to the integration of all aspects of the developed scale-themed curriculum including supplemental instruction, laboratory, and lecture and all statistical evidence for the efficacy of each type of intervention for both general chemistry I and II.

4.2 Scale-themed laboratory curriculum

4.2.1 Scale-themed laboratory experiments for general chemistry I:

The finalized experiment list developed for general chemistry I is depicted in **Table 4.1**. In total, four new experiments were added to the laboratory sequence of which "color my nanoworld[1]" and the "scale activity[2,3]" were adapted from existing literature resources, and "classification of matter" and "intermolecular forces" were written by the research team. In all 11 laboratory experiments, the scale concepts outlined by Jones and Taylor[4] were explicitly incorporated in as many ways as possible. The specific way in which this was done is outlined in section 4.2.2.

Table 4.1: General chemistry I scale-themed laboratory experiment list[a]

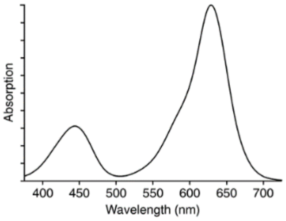
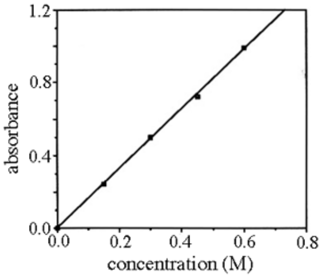
Content Area	Ch.	Experiment	Scaling concept
Skills, safety	1	Safety and Skill Inventory	Measurements
Scale, proportion, and measurement	1	Scale Activity	Scale, proportion and measurements; linear vs. logarithmic measurements
Physical and chemical changes and properties	1	Classification of Matter	Macroscopic Observations and Particle representations
Aqueous reactions (precipitation and complex ion formation)	4	Qualitative Analysis	
Aqueous reactions (acid/base)	4	Stoichiometry and Acid/Base Titrations	
Concentration and spectroscopy	4	Beer's law[5]	Scale, Proportion, Measurements – specifically sizes between particles and Macroscopic Observations and Particle Representations
Concentrations and spectroscopy	4	Color My Nanoworld	
Gases	5	Gas Laws	
Enthalpy	6/12	Physical Properties of Water	
Enthalpy	6	Enthalpy	
Reactions	4	Copper Cycle	
Intermolecular Forces	9/12	Intermolecular Forces	

[a] The complete General Chemistry I Laboratory Manual (110 pages) and General Chemistry I Laboratory Teaching Manual (115 pages) developed as a product of this work are available upon request.

4.2.2 Example of scale-themed general chemistry I laboratory experiment

Specific changes that were made to each section of the "Beer's Law" experiment to highlight explicit themes of scale are shown below in **Figures 4.1 and 4.2**. **Figure 4.1** shows a comparison between the pre lab questions of the initial experiment and of the scale-themed experiment. Students are asked to draw upon key scale concepts as they work through the problems such as relating one scale to another (weight to % weight and volume to % volume) and using number sense, converting, and visualizing different scales to envision a cube and calculate the volume of a fraction of that cube.

Figure 4.1: Comparison of pre lab questions between non-scale and scale-themed laboratory experiment for general chemistry I.[a]

Non-scale experiment	Scale-theme experiment
• What is one real world or practical application for this experiment or portion of this experiment? • In your own words, define: absorbance, absorbance spectrum, electromagnetic radiation, (lamda) max, molar absorptivity, path length, ultraviolet (UV) radiation, and visible radiation. • The dye you will use in this experiment is malachite green, search the internet or other source and find lamda max for this dye that you will use in this experiment. • Identify and potentially hazardous steps in your procedure. In your own words, explain what safety procedures should be followed and why. • A solution of dye in a 0.80 cm cuvette of concentration 1.25×10^{-3} M had an absorbance of 0.115 at a particular wavelength. What was the molar absorptivity of the dye?	• Envision building a cube with an edge length of 1 m, calculate the volume of 1 one millionth (1 ppm) of this cube. • How would you make a 5% sugar solution that has a total mass of 100 g? • What volume of water and dye would you need to make a 10% solution? • A solution of dye in a 0.80 cm cuvette of concentration 1.25×10^{-3} M had an absorbance of 0.115 at a particular wavelength. What was the molar absorptivity of the dye? • Below is the absorbance spectrum for a malachite green dye solution, what is lamda max?  • Given the following calibration curve, what is the concentration of a solution that has an absorbance of 0.800?  • Describe how you would prepare 50.00 mL of 0.100 M solution of NaOH using: a. solid NaOH b. a 1.00 M solution of NaOH • Identify any potentially hazardous steps in your procedure. In your own words, explain what safety procedures should be followed and why.

---

[a] The complete non-scale “Beer’s Law” experiment as well as its scale-themed counterpart can be found in **Appendix C.**

Figure 4.2 shows the comparison between the results and discussion questions of the initial experiment and of the scale-themed experiment. In the results and discussion questions of the scale-themed experiment students are asked to draw upon several more concepts of scale including conceptualizing relative sizes (determining how far apart the dye particles in the diluted solution are), number sense (determining a ratio of solvent particles to dye particles), estimation (approximating the number of dye particles and solvent particles in the solution), and converting (using concentration to determine number of dye particles and solvent particles.

Figure 4.2: Comparison of results and calculations questions of non-scale and scale-themed laboratory experiment for general chemistry I.

Non-scale experiment	Scale-themed experiment
• If your dye was copper(II) sulfate, describe an alternate method for determining the molar concentration of your unknown sample. • Make a Beer's Law plot for your dye. Plot the absorbance versus concentration. Make sure your plot includes the point (0,0), determine the molar absorptivity of your known, and the concentration of your unknown. • Using your plot, estimate the error in your molar absorptivity. Comment on its value. • Identify at least one random and at least one systematic error in this experiment. How would each change your results?	• Explain the plot you made. a. Does your line of best fit go through 0? b. Should it? c. Using your plot, find the equation of your line. d. What is the molar absorptivity of the dye? • In which cup did the solution first appear colorless? What is the concentration of dye in this cup? • What is the concentration of the unknown dye solution? • Considering the solutions you made in this experiment: a. If you were to continue to dilute your original solution down to 1.0 part per billion, what would be the molar concentration of dye in this solution? b. How do you know that there is still dye present in the solution even though the solution appears colorless? c. Approximately how many dye particles would be in 1.0 mL of this solution? d. Approximately how many water molecules would be in 1.0 mL of this solution?

	e. What is the ratio of water molecules to one dye particle? f. Using the ratio calculated above, what is a real world comparison you can make to help you understand the number of solute particles to solvent molecules. g. Thinking about the comparison you made above, approximately how far apart are the dye particles in this solution?

4.2.3 Scale-themed laboratory experiments for general chemistry II

The finalized scale-themed experiment list for general chemistry II is given in **Table 4.2**. In all, 10 scale-themed laboratory experiments were created by altering the introduction, objectives, pre lab questions, and results and calculations questions of each experiment. The specific ways in which this was done for general chemistry II are outlined in section 4.2.4.

Table 4.2: General chemistry II scale-themed laboratory experiment list

Content Area	Ch.	Experiment	Scaling concept
Gas laws/properties of liquids	5/12	The Molar Mass of a Volatile Liquid	Number sense, converting, relating different scales
Physical properties of solutions	13	Freezing Point Depression	Macroscopic Observations and Particle representations
Chemical Kinetics	14	Rate and Order of a Chemical Reaction	
Chemical Kinetics	14	Rate Determination and Activation Energy	
Chemical Equilibrium	15	Determination of an Equilibrium Constant	
Acids and Bases/Acid-Base equilibria	16/17	Buffers	
Solubility equilibria	17	Determining the Ksp of Calcium Hydroxide	
Thermodynamics	6/18	Hess's Law – Heat of Combustion	Measurements – Specifically sizes of and distances between particles and Macroscopic Observations and Particle Representations
Complex Ion Equilibria/Electrochemistry	7/19	What's in a Penny	Macroscopic Observations and Particle representations
Electrochemistry	19	Electrochemistry – Voltaic Cells	

[a] The full General Chemistry II Laboratory Manual (100 pages) and General Chemistry I Laboratory Teaching Manual (88 pages) developed as a product of this work are available upon request.

4.2.4 Example of scale-themed general chemistry II laboratory experiment

For general chemistry II, the objectives, pre lab questions, and results and calculation questions of each experiment were altered in the same manner as for general chemistry I. These changes asked students to think about and draw upon specific scale concepts and skills as they worked through the experiment or answered questions about the experiment. One change that was made to the experiments in general chemistry II that was not made in general chemistry I was the inclusion of a much more detailed introduction to the experiment. This change gave students an overview of all the relevant ideas and concepts needed to understand the chemistry happening in the experiment so that the pre lab and results and calculation questions could focus on extending student understanding to specific systems of study. For example, the introduction in **Figure 4.3** was written specifically to help students understand the presence and relative amount of gas particles present at the liquid-vapor interface of a pure solvent or of a solution on the particle level. One pre lab questions asks students to choose two different liquids and diagram them on the particle level and a results and calculation question follows up to ask them to diagram the solvent and solution they used in the experiment on the particle level. In this example, the student is given a generic system of study to reference with all necessary information to answer both of these questions, but the added complexity of applying this information to a specific system of study requires the student to engage with concepts such as relative sizing in order to accurately complete the questions.

Figure 4.3: Introduction to “Molar Mass of a Volatile Liquid” experiment both before and after adaptation.[a]

Before Adaptation:

One of the properties that helps characterize a substance is its molar mass. If the substance in question is a volatile liquid, a common method to determine its molar mass is to use the ideal gas law, $PV = nRT$. Because the liquid is volatile, it can easily be converted to a gas. While the substance is in the gas phase, you can measure its volume, pressure, and temperature. You can then use the ideal gas law to calculate the number of moles of the substance. Finally, you can use the number of moles of the gas to calculate molar mass.

After Adaptation:

To the unaided eye the surface of a liquid may seem of little interest. However, as shown in **Figure 1** there is a lot of chemistry occurring at what is frequently referred to as the liquid/vapor interface. Many observations about a substance can be explained by modeling the interface of that substance (both pure substances and solutions will have unique interfaces, see **Figure 2** for a solution/gas interface). If a substance has a high vapor pressure, that is, the pressure exerted on the surface of a liquid by evaporated molecules of that liquid is high, it is said to also be a *volatile liquid*. Volatility is a measure of the ease in which liquid molecules gain sufficient kinetic energy to escape into the gas phase. These gas molecules will exert a pressure and this pressure is called the *vapor pressure*. Given this definition, it can be determined that a solution with a *high vapor pressure* and high volatility would contain *many gas molecules* at the liquid/vapor interface while conversely, a substance with *low vapor pressure* and low volatility would represent a solution in which *fewer liquid molecules* are able to escape into the gas phase.

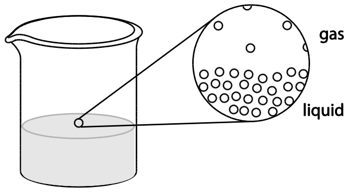


Figure 1: Macroscopic and particle diagram of the interface of a pure liquid

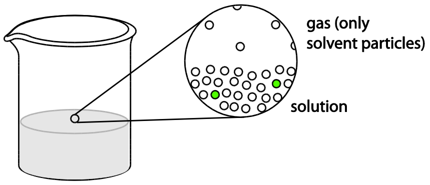


Figure 2: Macroscopic and particle diagram of the interface of a solution and the gas (of the solvent).

You have already learned several chemical methods to determine the identity of an unknown substance such as melting point and density. Another intensive property that can be used to identify an unknown substance is its molar mass. If the substance in question is a volatile liquid, a common method to determine its molar mass is to use the ideal gas law, $PV = nRT$. Because the liquid is volatile, it can easily be converted to a gas. While the substance is in the gas phase, you can measure its volume, pressure, and temperature. You can then use the ideal gas law to calculate the number of moles of the substance. Finally, you can use the number of moles of the gas to calculate molar mass.

---

[a] The complete “unchanged” “Molar Mass of a Volatile Liquid” experiment as well as its “scale-themed” counterpart are attached in **Appendix C**).

4.3 Scale-themed lecture curriculum

4.3.1 Conceptual versus algorithmic analysis

For both general chemistry I and II, comparison of student performance on both conceptual and algorithmic hourly exam items with scale literacy scores showed significant correlations (see **Table 4.3**) for all groups. Further analysis of student performance by both chapter and item type also showed strong positive correlations for most of the content areas analyzed. For general chemistry I, performance on algorithmic items from the chapters on classification of matter, stoichiometry, aqueous solutions, gases, periodic trends, and advanced bonding showed statistically significant correlations to scale literacy performance. Similarly, performance on conceptual items from chapters on measurement, classification of matter, stoichiometry, aqueous solutions, gases, energy, electronic structure, bonding, and advanced bonding showed significant correlations to scale literacy performance. In general chemistry II, performance on algorithmic items from chapters on physical properties of solutions, acids and bases, acid-base and solubility equilibria, and thermodynamics showed significant correlation to scale literacy performance. Performance on conceptual items from all chapters (physical properties of solutions, kinetics, equilibrium, acids and bases, acid-base and solubility equilibria, and thermodynamics) showed significant correlation to scale literacy performance. In all, 15 of the 16 (94%) conceptual item sub-scores had significant correlations to scale literacy while only 11 of 16 (69%) algorithmic item sub-scores had significant correlations. This result is consistent with the stronger correlation seen between scale literacy and performance on the conceptual final than between scale literacy and performance on the paired final for general chemistry I students (.646 versus .587).

Table 4.3: General chemistry content areas showing significant correlations to scale literacy.

		Algorithmic			Conceptual		
Chapter (GC I)	Content area tested	r	p	n	r	p	n
	All areas	.409	<.001	320	.573	<.001	320
1	Measurement	.183	.054	111[a]	.330	<.001	320
2	Classification of Matter	.255	<.001	320	.277	<.001	320
3	Stoichiometry	.462	<.001	320	.419	<.001	209[b]
4	Aqueous Solutions	.316	<.001	320	.381	<.001	320
5	Gases	.221	<.001	320	.335	<.001	320
6	Energy	.244	<.001	302	.460	<.001	320
7	Electronic Structure	.079	.160	320	.348	<.001	111[a]
8	Periodic Trends	.119	.033	320	.160	.093	111[a]
9	Bonding	--[c]			.296	<.001	320
10	Bonding II	.116	.038	320	.293	<.001	209[b]
12	Intermolecular forces	--[c]			--[c]		
Chapter (GC II)	Content area tested	r	p	n	r	p	n
	All areas	.420	<.001	113	.461	<.001	113
13	Physical properties of solutions	.465	<.001	113	.200	.034	113
14	Kinetics	.041	.665	113	.385	<.001	113
15	Equilibrium	-.169	.074	113	.198	.036	113
16	Acids and Bases	.308	.001	113	.221	.019	113
17	Acid/Base/solubility equilibria	.361	<.001	113	.253	.007	113
18	Thermodynamics	.208	.027	113	.201	.033	113
19	Electrochemistry	.118	.212	113	--[c]		

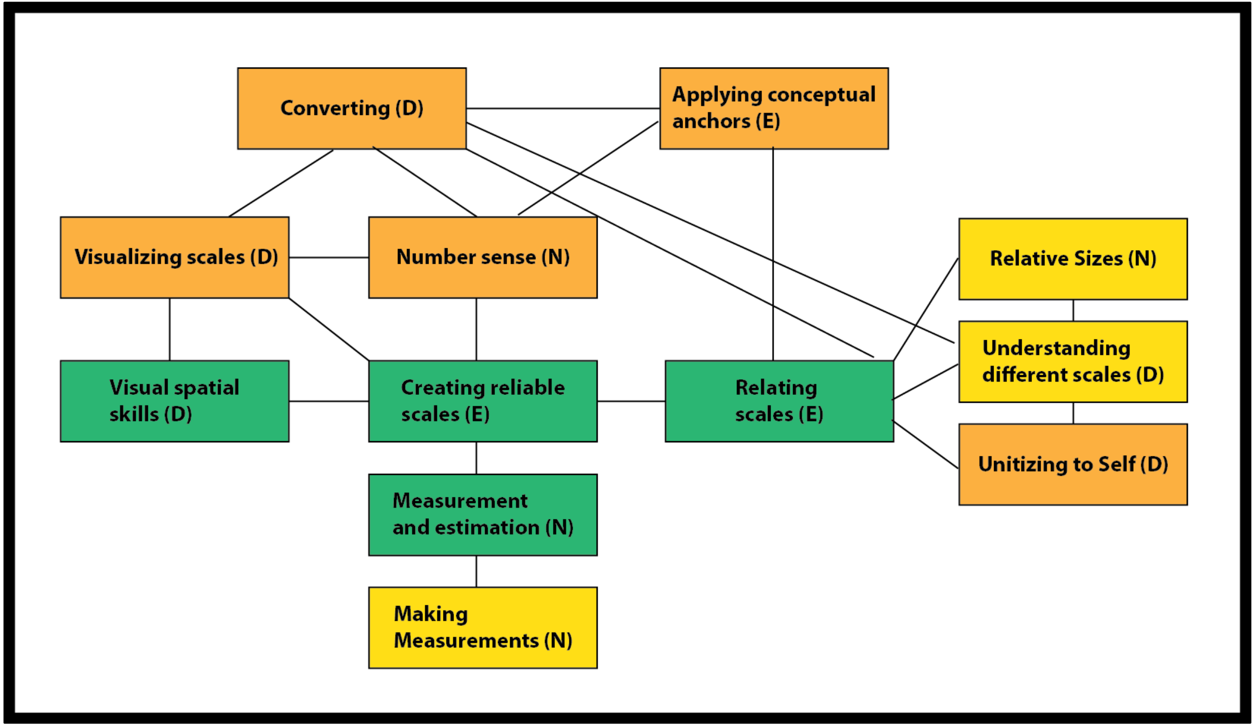
[a]only enough items tested during the fall 2011 semester. [b]only enough items tested during the fall 2012 semester [c]not enough items tested

4.3.2 General Chemistry I Scale learning progression

The general chemistry I scale learning progression was constructed by comparing the scale concept trajectory proposed by Gail Jones[4] to the performance of general chemistry I students on the scale content areas tested on the Scale Literacy Skills Test. The scale learning progression was built upon the premise of beginning scale instruction with the concepts students are most comfortable with and using that prior knowledge to construct new knowledge about less familiar concepts. In some places, even though the data suggested otherwise, an intuitive progression was maintained to protect the natural progression a student would follow in the development of certain skills such as at the bottom of the learning progression shown in **Figure 4.4** where "making measurements" appears before "measurement and estimation" despite student performance related to making a measurement being lower than performance related to measurement and estimation skills. The general chemistry I scale learning progression shows where general chemistry I students compare to those used to construct the scale concept trajectory by color coating each content area as it corresponds to the general chemistry I student's performance (green = highest performance, yellow = average performance, orange = lowest performance) and by denoting how each scaling concept fell in the original trajectory of scale concept development, denoted by the letter in parenthesis next to the concept ("E" = experienced, "D" = developing, "N" = novice). As the Trajectory of Scale Concept Development was built largely upon the retrospective perceptions of experts on how they developed an understanding of scale, not surprisingly, several key differences exist between the Trajectory of Scale Concept Development proposed by Jones and co-workers and the General

Chemistry I Scale Learning Progression proposed here. As the General Chemistry I Scale Learning Progression is built upon empirical evidence derived from general chemistry I students' performance on items testing concepts of scale, it is likely this proposed progression more accurately represents the understanding of students at this level.

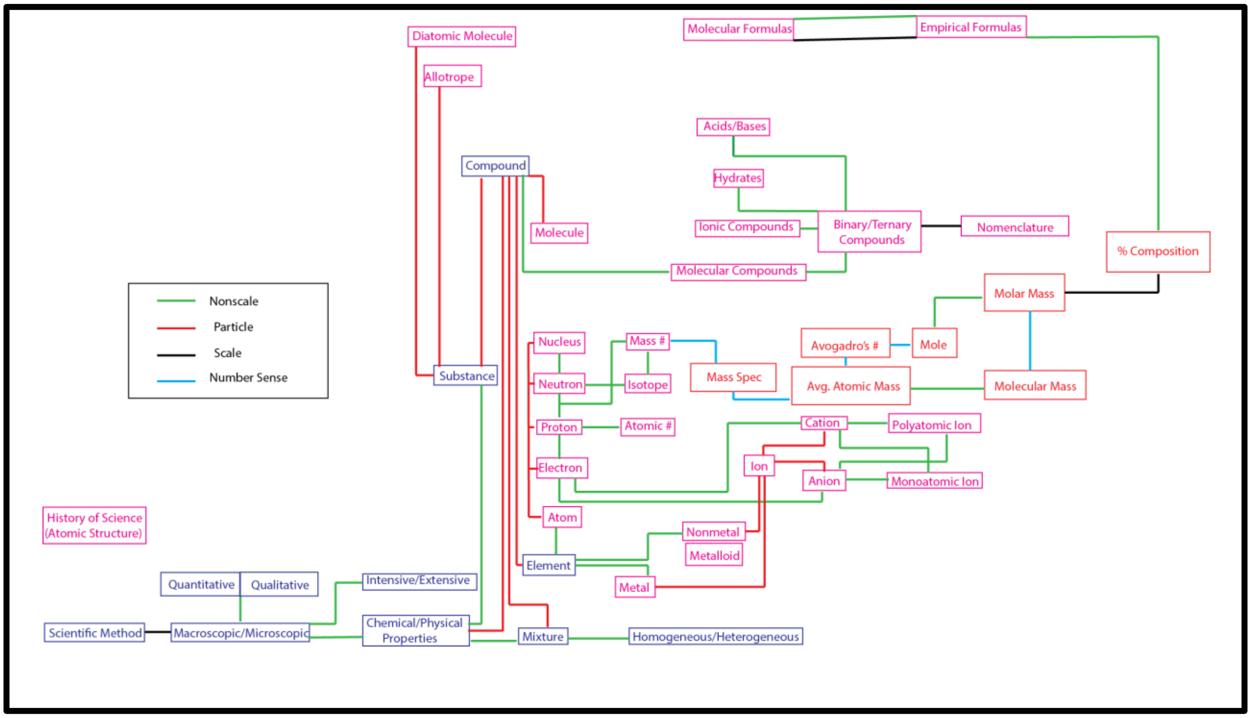
Figure 4.4: General chemistry I proposed scale learning progression



4.3.3 General chemistry I content map

Separate from the General Chemistry I Scale Learning Progression, a General Chemistry I Content Map was also constructed by identifying each element of chemistry content covered in the course and arranging each element according to both the order of presentation in the course and its connectivity to other elements in the course. Content areas identified by chapters were each given different color text while the connectivity of the elements was given a second color coding scheme relating to how the connectivity was made in the context of understanding scale. In many cases an initial intuitive connectivity was followed as a “foundational element” was required to precede learning of new elements or content areas before being analyzed for connections to understanding scale. Keeping in mind both the concrete (number sense) and abstract (macroscopic/particle) components of scale in chemistry, the connections between each piece of chemistry curriculum that could be made using these distinctions were identified. These links were color-coded by number sense (blue), macroscopic/particle (red), or those utilizing both (black). Those elements in which no inherent connection to scale was determined were color-coded in green. A small piece of the generated general chemistry I content map for the introductory concepts of the scientific method, atomic structure, and nomenclature are shown in **Figure 4.5**. The content map is arranged from bottom to top by the order in which content areas are introduced and by increasing difficulty.

Figure 4.5: General chemistry I content map



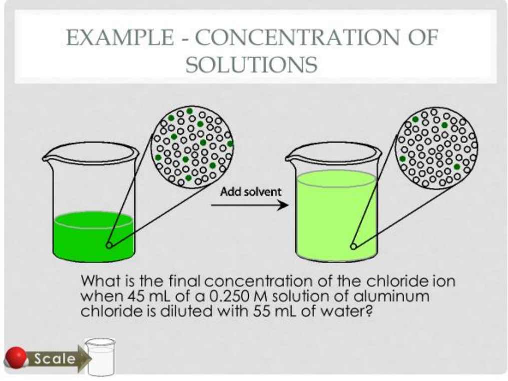
4.3.4. Scale-themed lecture slides and notes for general chemistry I

Using the General Chemistry I Scale Learning Progression and content mapping described in sections 4.3.2 and 4.3.3, the general chemistry I lectures notes and slides were arranged and adapted to feature explicit instruction in scale. The instructional materials and methods were specifically designed to help students learn to use and engage with concepts of scale as they work to understand chemistry concepts. The inclusion of scale into instruction in this way was explicitly explained to students at the start of the course and included an explanation of both why these themes were being incorporated and all of the different concepts and skills that would be emphasized throughout the semester. Students were made aware of the presence of the "scale symbol" (**Figure 3.5**) on each slide in which scale was featured with the expectation that after enough explicit instruction to do so, students might see the symbol and be cued to think about scale. The inclusion of scale was the only change made to the instructional materials of the course at this time and no other additional scale-themed activities were presented to students.

An example of how the lecture slides and notes were updated to include explicit themes of scale is shown in **Figure 4.6**. In this example where students are learning about dilution they are asked to complete a very common practice problem related to calculating a new concentration following dilution. Students are also shown a very common depiction that helps explain what the dilution process looks like on the particle level. The difference between the scale-themed instruction and the non-scale instruction is in the questions that follow the calculation of the new concentration. Asking students to think about the ratio of solvent particles to solute particles in not just a relative sense but also in an absolute sense forces students to engage with multiple themes of scale as they relate to understanding both the particle-level models of the undiluted and diluted solutions and also the calculated numerical value. While the lecture slides don't explicitly ask students to calculate the number of solvent or solute particles, as is also shown in **Figure 4.6**, the lecture notes do actually go in to this level of detail.

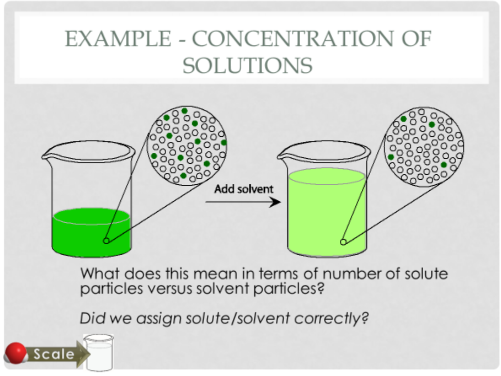
Classroom observations made during the initial implementation of this lecture material yielded several results. Most importantly, in some cases where the lecturer thought the inclusion of scale was obvious and clear, classroom observation was not consistent with that perception. Discussions with the lecturer following the class period provided a mechanism to both improve the lecture materials for the next implementation as well as possibly reveal the need for a clarification to students at the next lecture. Additionally, classroom observation revealed that while students did not appear to disengage with material during explicit scale instruction, students also did not seem to engage with it. This was evident as students continued to take notes throughout the instruction but could only be seen writing when the instructor wrote something on the board. As much of the explicit scale instruction was made through verbal connections it was not surprising to see students not write anything down during verbal instruction. These observations strengthened the argument for the development of the active learning lecture activities described in sections 3.4.4 and 4.3.5.

Figure 4.6: Scale-themed lecture slides and accompanying lecture notes on dilution.



EXAMPLE - CONCENTRATION OF SOLUTIONS

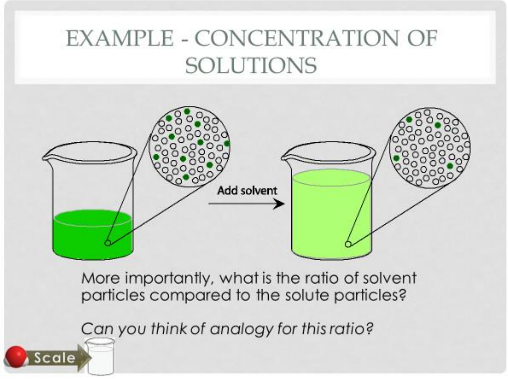
What is the final concentration of the chloride ion when 45 mL of a 0.250 M solution of aluminum chloride is diluted with 55 mL of water?



EXAMPLE - CONCENTRATION OF SOLUTIONS

What does this mean in terms of number of solute particles versus solvent particles?

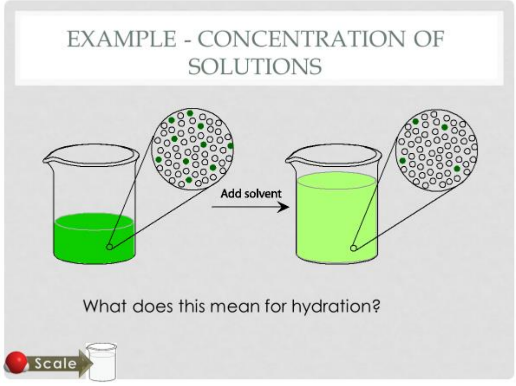
Did we assign solute/solvent correctly?



EXAMPLE - CONCENTRATION OF SOLUTIONS

More importantly, what is the ratio of solvent particles compared to the solute particles?

Can you think of analogy for this ratio?



EXAMPLE - CONCENTRATION OF SOLUTIONS

What does this mean for hydration?

$$\left(\frac{0.375\text{ mol Cl}^-}{\text{L}}\right)(0.090\text{ L})\left(\frac{6.022\times10^{23}\text{ Cl}^-\text{ ions}}{\text{mol Cl}^-}\right)=2.0\times10^{22}\text{ Cl}^-\text{ ions}$$

$$(90\text{ mL water})\left(\frac{1\text{ g}}{\text{mL}}\right)\left(\frac{\text{mol}}{18.02\text{ g}}\right)\left(\frac{6.022\times10^{23}\text{molecules H}_2\text{O}}{\text{mol H}_2\text{O}}\right)=3.0\times10^{24}\text{ H}_2\text{O molecules}$$

There are approximately 150 to 1 water molecules to chloride ions.

An analogy would be similar to one professor to 150 students. The number of spheres of hydration would be more than one around the ions.

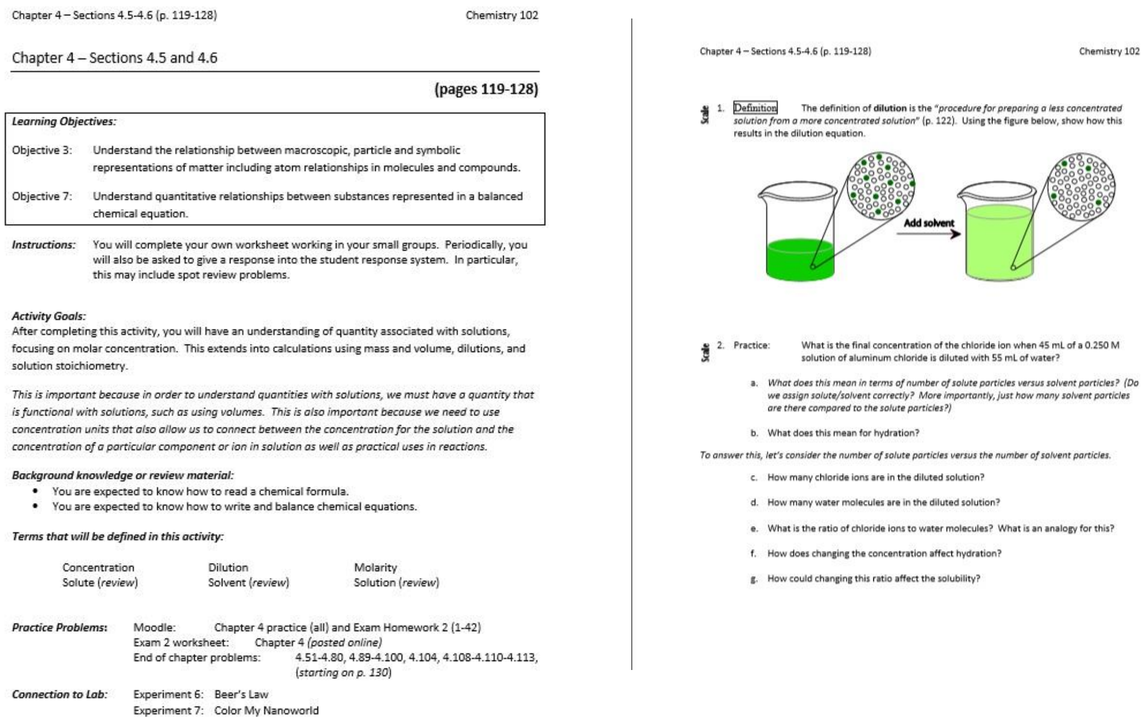
(Then the representations used throughout this chapter, the omission of water molecules may allow us to (erroneously) think the solute particles are closer than they are.)

It is important to remember that there are also cations present (aluminum) at 1/3 of the ratio to that of the chloride ion (and water molecules are needed for the hydration of aluminum as well).

4.3.5 Active learning lecture activities for general chemistry I and II

As the lecture materials for general chemistry I were adapted into active learning activities for immediate implementation, minor adjustments were made throughout the semester to improve formatting and inclusion of content. For example, at the beginning of the semester, the activities largely did not replicate the material presented on the slides (which generally remained unchanged for comparison on teaching methodologies only). However, it was found that many students were attempting to copy all of the slide content into the active learning exercises including longer, textbook definitions. Because this was time consuming, these were included into the activities with a notation (and textbook location) for a definition. Consistent with the lecture slides, a notation was included in the activities to flag “scale” when a portion of the activity incorporated the theme of scale. An example of how the lecture material displayed in **Figure 4.6** was adapted into the active learning lecture activities can be seen in **Figure 4.7**. The format of the activities was explained to the students on the first day of lecture with an expectation that they would follow along with the active learning activities and actively participate in all discussions. Additionally, selected items from the activities that were not completed during lecture were assigned as “lecture assignments”. These lecture assignments became the foundation for the discussion content for the week (including the subset that was assigned for weekly discussion credit). In spring, 2015, the lecture activities were altered to remove all scale-themed components and any reference to scale but with the same format and expectations of the students as the scale-themed active learning activities. The process by which the adaptation and development of active learning lecture materials for general chemistry II mirrored what was done in general chemistry I with the creation of both scale-themed and non-scale active learning lecture activities. The complete General Chemistry I Lecture Activity book both with explicit scale themes (331 pages) and without explicit scale themes (319 pages), as well as the complete General Chemistry II Lecture Activity book both with explicit scale themes (258 pages) and without scale themes (239 pages) developed as a product of this work are available upon request.

Figure 4.7: General chemistry I active learning lecture activity example. The cover page (left) for each activity summarizes what information can be found in the activity's pages (right).



Chapter 4 – Sections 4.5-4.6 (p. 119-128) Chemistry 102

Chapter 4 – Sections 4.5 and 4.6

(pages 119-128)

Learning Objectives:

Objective 3: Understand the relationship between macroscopic, particle and symbolic representations of matter including atom relationships in molecules and compounds.

Objective 7: Understand quantitative relationships between substances represented in a balanced chemical equation.

Instructions: You will complete your own worksheet working in your small groups. Periodically, you will also be asked to give a response into the student response system. In particular, this may include spot review problems.

Activity Goals:
After completing this activity, you will have an understanding of quantity associated with solutions, focusing on molar concentration. This extends into calculations using mass and volume, dilutions, and solution stoichiometry.

This is important because in order to understand quantities with solutions, we must have a quantity that is functional with solutions, such as using volumes. This is also important because we need to use concentration units that also allow us to connect between the concentration for the solution and the concentration of a particular component or ion in solution as well as practical uses in reactions.

Background knowledge or review material:

- You are expected to know how to read a chemical formula.
- You are expected to know how to write and balance chemical equations.

Terms that will be defined in this activity:

Concentration
Solute (*review*)
Dilution
Solvent (*review*)
Molarity
Solution (*review*)

Practice Problems: Moodle: Chapter 4 practice (all) and Exam Homework 2 (1-42)
Exam 2 worksheet: Chapter 4 *(posted online)*
End of chapter problems: 4.51-4.80, 4.89-4.100, 4.104, 4.108-4.110-4.113, (*starting on p. 130*)

Connection to Lab: Experiment 6: Beer's Law
Experiment 7: Color My Nanoworld

Chapter 4 – Sections 4.5-4.6 (p. 119-128) Chemistry 102

Scale 1. Definition The definition of **dilution** is the "*procedure for preparing a less concentrated solution from a more concentrated solution*" (p. 122). Using the figure below, show how this results in the dilution equation.

Scale 2. Practice: What is the final concentration of the chloride ion when 45 mL of a 0.250 M solution of aluminum chloride is diluted with 55 mL of water?

a. *What does this mean in terms of number of solute particles versus solvent particles? (Do we assign solute/solvent correctly? More importantly, just how many solvent particles are there compared to the solute particles?)*

b. What does this mean for hydration?

To answer this, let's consider the number of solute particles versus the number of solvent particles.

c. How many chloride ions are in the diluted solution?

d. How many water molecules are in the diluted solution?

e. What is the ratio of chloride ions to water molecules? What is an analogy for this?

f. How does changing the concentration affect hydration?

g. How could changing this ratio affect the solubility?

4.4 Efficacy of instructional approaches in general chemistry I

4.4.1 Predicting efficacy

For students in general chemistry I, multiple regression analysis was conducted on the fall 2011 and fall 2012 control data set (n = 224). For the paired final, all combined predictors (total placement test score ($\beta = 0.444$, $p < 0.001$), ACT composite score ($\beta = 0.241$, $p = 0.001$), and scale literacy ($\beta = 0.210$, $p = 0.005$), had significant standardized coefficients. All three predictors accounted for 56% of the variance with $R^2 = 0.56$, $F(3, 138) = 59.85$, $p < 0.001$. For the conceptual final, all combined predictors (total placement test score ($\beta = 0.273$, $p < 0.001$), ACT composite score ($\beta = 0.237$, $p = 0.001$), and scale literacy ($\beta = 0.363$, $p < 0.001$), had significant standardized coefficients. All three predictors accounted for 52% of the variance with $R^2 = 0.52$, $F(3, 138) = 52.20$, $p < 0.001$. Both models were checked for assumptions of homoscedascity and normality of residuals.

4.4.2 Pre laboratory quizzes

Analysis of student performance on pre-laboratory quizzes, **Table 4.4**, showed on average positive changes from pre question to post question. On average the greatest change was seen for the “Physical Properties of Water” experiment in which scores ranged from 0.40 to 1.72 (on a 4 point scale) and the most consistent average change score was seen for the “Scale Activity” in which all 7 semesters of analyzed data showed a positive change score of between .56 and .90. Unexpected data points that might warrant further investigation include the negative (or close to zero) change scores that emerge for the “Gas Laws” and “Enthalpy” experiments when non-scale active learning lecture activities are used. These same experiments exhibited positive change scores when other instructional methodologies were utilized suggesting the possibility that either this instructional methodology hinders student performance within this content area as it relates to this experiment or some other instructional/instructor effect is affecting performance on one or both of these quiz questions.

Table 4.4: General chemistry I pre-laboratory quiz performance

	Lab	Lecture	Non-scale AL	Non-scale AL	AL	AL	AL
n	231	226	206	194	204	217	216
Scale Activity	0.71	0.77	0.56	0.81	0.80	0.83	0.90
Classification of Matter	1.05	0.22	1.11	0.21	0.42	0.79	0.69
Qualitative Analysis	0.95	0.09	-0.11	0.46	0.10	0.20	0.59
Acid/Base Titrations	0.77	0.06	0.52	1.82	1.21	1.73	0.24
Beer’s Law and Dilutions	0.50	1.31	0.78	1.40	0.77	1.09	0.85
Color My Nanoworld	0.16	0.82	-0.01	0.41	0.40	0.26	0.30
Gas Laws	0.20	0.73	-0.38	0.04	0.44	0.65	0.66
Enthalpy	0.66	-0.16	-0.23	0.01	0.61	0.63	0.87
Physical Properties of water	1.08	0.40	0.50	0.93	1.72	0.87	1.12
Intermolecular Forces	0.26	0.16	0.56	0.59	0.51	0.12	-0.06
Overall performance	0.66	0.62	0.60	0.69	0.69	0.72	0.75

4.4.3 Laboratory survey

Analysis of student performance on the laboratory survey in general chemistry I, **Table 4.5**, show significant increases in student performance from pre to post on the 13 objectively scored items of the survey. These increases also show medium to large effect sizes for all groups. With the exception of the initial semester of using scale-themed active learning in conjunction with scale-themed laboratory experiments, larger effect sizes are seen for the section of general chemistry I that also completed scale-themed supplemental instruction.

Table 4.5: General chemistry I laboratory survey group comparisons

	n	p	effect size
Control		Not given	
Supplemental Instruction			
Laboratory Experiments	71	0.001	0.442
*Laboratory Experiments + Supplemental Instruction**	*18*	*0.023*	*0.755*
Lecture	65	<.001	1.4
Lecture + Supplemental Instruction	23	<.001	1.61
Lecture + Laboratory Experiments	65	<.001	0.57
Lecture + Laboratory Experiments + Supplemental Instruction	38	<.001	1.36
Non-scale Active learning + Laboratory Experiments (initial)	55	0.005	0.461
Non-scale Active learning + Laboratory Experiments (repeat)	72	<.001	0.716
Non-scale Active learning + Laboratory Experiments + Supplemental Instruction (initial)	22	0.002	0.938
Non-scale Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	20	0.001	1.02
Active learning + Laboratory Experiments (initial)	69	<.001	0.968
Active learning + Laboratory Experiments (repeat)	73	<.001	0.556
Active learning + Laboratory Experiments + Supplemental Instruction (initial)	35	<.001	0.70
Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	26	<.001	0.847
Active learning + Laboratory Experiments + Supplemental Instruction (2 section repeat)	86	<.001	0.891

** n less than 20*

4.4.4 Scale literacy

Comparisons between student scale literacy scores from pre to post in general chemistry I, **Table 4.6**, showed consistently significant increases. With the exception of several semesters in which too few students remained for meaningful analyses, only two additional treatments (scale-themed laboratory experiments and active learning lecture activities with scale-themed laboratory experiments) did not show significant increases from pre to post. The combined treatments of active learning lecture activities, laboratory experiments, and supplemental instruction consistently showed significant increases from pre to post with large effect sizes.

Table 4.6: General chemistry I scale literacy group comparisons

	n	p	effect size
Control	118	<.001	0.425
Supplemental Instruction	46	<.001	1.30
Laboratory Experiments	22	0.355	
*Laboratory Experiments + Supplemental Instruction**	*10*	*0.288*	
Lecture	29	0.001	0.413
*Lecture + Supplemental Instruction**	*16*	*0.001*	*0.767*
Lecture + Laboratory Experiments	32	<.001	0.491
Lecture + Laboratory Experiments + Supplemental Instruction	21	0.004	0.584
Non scale Active learning + Laboratory Experiments (initial)	25	0.012	0.411
*Non scale Active learning + Laboratory Experiments (repeat)**	*17*	*0.012*	*0.642*
*Non scale Active learning + Laboratory Experiments + Supplemental Instruction (initial)**	*14*	*<.001*	*1.315*
*Non scale Active learning + Laboratory Experiments + Supplemental Instruction (repeat)**	*11*	*<.001*	*1.11*
Active learning + Laboratory Experiments (initial)	27	0.671	
Active learning + Laboratory Experiments (repeat)	29	0.001	0.661
Active learning + Laboratory Experiments + Supplemental Instruction (initial)	25	<.001	1.15
Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	30	<.001	0.879
Active learning + Laboratory Experiments + Supplemental Instruction (2 section repeat)	65	<.001	0.836

** n of less than 20*

4.4.5 Residual analyses

In general chemistry I, multiple regression residual analysis of the ACS First Term Paired Questions Exam followed a consistent trend over 10 semesters of testing (**Table 4.7**). Supplemental instruction alone accounted for a 2% increase in student performance on the final exam, an effect that was replicated in all semesters in which supplemental instruction was used in conjunction with other treatments. For this exam, the combined treatments of active learning (both scale-themed and non-scale), laboratory experiments, and supplemental instruction accounted for a consistent positive residual average of 5.1%-6.7% over five semesters. When supplemental instruction was not included, a 2% increase in student performance was observed for one semester of students receiving scale-themed active learning versus those receiving the non-scale active learning. This effect was not consistent, however, with a previous semester of testing.

For the ACS General Chemistry Conceptual Exam (first term), multiple regression residual analysis revealed no significant increase in student performance from any treatment until the combined treatments of active learning, laboratory experiments, and supplemental instruction were implemented **(Table 4.8)**. This combined treatment yielded consistently significant increases of 4.9%-6.7% over three repeated semesters of testing. No other observable trends exist in the data although further analysis of the semesters in which negative residuals were found when non-scale active learning lecture activities were used is warranted.

Table 4.7: General chemistry I Paired Final residual averages and group comparisons

	n	Residual average	p	effect size
Control	143	<1%		
Supplemental Instruction	49	2.0%	0.092	
Laboratory Experiments	41	3.7%	0.045	0.346
Laboratory Experiments + Supplemental Instruction	15	5.6%	0.008	0.52
Lecture	40	1.8%	0.179	
Lecture + Supplemental Instruction	21	3.1%	0.142	
Lecture + Laboratory Experiments	40	2.0%	0.206	
Lecture + Laboratory Experiments + Supplemental Instruction	26	3.4%	0.104	
Non scale Active learning + Laboratory Experiments (initial)	35	1.90%	0.356	
Non scale Active learning + Laboratory Experiments (repeat)	29	2.10%	0.417	
Non scale Active learning+ Laboratory Experiments + Supplemental Instruction (initial)	15	6.20%	0.004	0.564
Non scale Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	12	5.10%	0.18	
Active learning + Laboratory Experiments (initial)	42	<1%	0.901	
Active learning + Laboratory Experiments (repeat)	50	4.20%	0.009	0.305
Active learning + Laboratory Experiments + Supplemental Instruction (initial)	27	6.70%	0.003	0.522
Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	30	4.90%	0.016	0.420
Active learning + Laboratory Experiments + Supplemental Instruction (2 section repeat)	63	5.80%	<.001	0.504

** n of less than 20*

Table 4.8: General chemistry I Conceptual final residual averages and group comparisons

	n	Residual average	p	effect size
Control	143	<1%		
Supplemental Instruction	49	<1%	0.819	
Laboratory Experiments	41	<1%	0.683	
Laboratory Experiments + Supplemental Instruction	15	<1%	0.973	
Lecture	40	<1%	0.604	
Lecture + Supplemental Instruction	21	2.90%	0.166	
Lecture + Laboratory Experiments	40	<1%	0.946	
Lecture + Laboratory Experiments + Supplemental Instruction	26	<1%	0.988	
Non scale Active learning + Laboratory Experiments (initial)	35	1%	0.565	
Non scale Active learning + Laboratory Experiments (repeat)	29	-2.60%	0.161	
Non scale Active learning + Laboratory Experiments + Supplemental Instruction (initial)	15	-1.50%	0.42	
Non scale Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	12	-1.70%	0.423	
Active learning + Laboratory Experiments (initial)	42	2%	0.187	
Active learning + Laboratory Experiments (repeat)	50	1.70%	0.297	
Active learning + Laboratory Experiments + Supplemental Instruction (initial)	27	4.20%	0.036	0.34
Active learning + Laboratory Experiments + Supplemental Instruction (repeat)	30	5.10%	0.005	0.441
Active learning + Laboratory Experiments + Supplemental Instruction (2 section repeat)	63	4.90%	0.001	0.393

* *n of less than 20*

4.5 Efficacy of instructional approaches in general chemistry II

4.5.1 Predicting efficacy

For students in general chemistry II, multiple regression analysis was conducted on the spring 2015 control data set (n = 182). For the conceptual final, all combined predictors (placement test score ($\beta = 0.363$, $p < 0.001$), ACT composite score ($\beta = 0.263$, $p = 0.008$), and scale literacy ($\beta = 0.221$, $p = 0.025$), had significant standardized coefficients. All three predictors accounted for 51% of the variance with $R^2 = 0.51$, $F(3, 89) = 32.553$, $p < 0.001$. This model was also checked for assumptions of homoscedascity and normality of residuals.

4.5.2 Pre laboratory quizzes

Analysis of student performance on pre-laboratory quizzes in general chemistry II, **Table 4.9**, was not consistent with the trends observed in general chemistry I. While no experiment in general chemistry I showed negative average change values across all groups, three quizzes ("Buffers", "Hess's Law", and "What's in a Penny") consistently had negative average change scores for multiple groups. "Freezing point depression", "Rate and Order of a Reaction", and "Rate and Activation Energy" had the highest average change scores across all semesters (.60 to

2.06) and “Determination of Ksp” had the most consistent average score in which all analyzed groups had an average change score between .13 and .46. Further investigation of the experiments and quiz questions in which negative average change scores exist for all groups is warranted.

Table 4.9: General chemistry II pre-laboratory quiz performance

	control	Lab	AL	Non scale AL	AL + SI	Non scale AL + SI
n	163	129	157	155	78	80
Molar mass of a volatile liquid	0.05	0.86	1.13	0.58	0.65	0.43
Freezing point depression	0.98	2.06	0.92	1.80	1 .55	1.58
Rate and order of a reaction	0.64	1.03	0.99	1.28	0.97	0.77
Rate and Activation Energy	0.89	1.59	1.20	0.60	0.72	0.77
Determination of K	0.33	1.15	0.52	0.87	0.23	0.63
Buffers	-0.07	-0.11	-0.19	-0.23	-0.26	-0.10
Determination of Ksp	0.19	0.28	0.22	0.46	0.13	0.22
Hess's Law	0.10	-0.20	-0.14	0.06	-0.32	-0.19
What’s in a penny	--[a]	-0.15	-0.59	-0.06	0.04	-0.04
Electrochemistry	--[b]	--[a]	--[a]	--[a]	--[a]	--[a]
Overal performance	0.56	0.60	0.54	0.69	0.61	0.65

[a]No quiz given for this experiment [b]quiz given for this semester only

4.5.3 Laboratory survey

Analysis of student performance on the laboratory survey in general chemistry II, **Table 4.10**, showed significant increases in student performance from pre to post for only the groups completing laboratory experiments alone and in conjunction with non-scale active learning lecture activities. These increases show small to medium effect sizes. While this data suggests that students in general chemistry II are underperforming on this survey as compared to students in general chemistry I, the students in these groups have higher initial scores which could explain the smaller observed increases. Most surprisingly, is the appearance of data which suggests that the incorporation of supplemental instruction hinders student performance on this survey. Repeated testing with larger samples for the combined treatments including both scale-themed and non-scale themed active learning lecture activities, laboratory experiments, and supplemental instruction is warranted to verify this observed data.

Table 4.10: General chemistry II laboratory survey group comparisons

	n	p	effect size
Control	126	.321	
Laboratory Experiments	93	< .001	.538
Non scale Active learning + Laboratory Experiments	121	.002	.266
Non scale Active learning + Laboratory Experiments + Supplemental Instruction	27	.359	
Active learning + Laboratory Experiments	126	.079	
Active learning + Laboratory Experiments + Supplemental Instruction	24	.653	

4.5.4 Paired question final

Analysis of student performance on the ACS General Chemistry Paired Questions exam from pre (placement test) to post (final exam) in general chemistry II, **Table 4.11**, showed significant increases in student performance across all groups with medium to large effect sizes.

This result is not surprising given the low stakes testing environment in which the placement test is given and the repeated use of this instrument throughout the general chemistry curriculum at this institution.

Table 4.11: General chemistry II Paired final group comparisons.

	n	p	effect size
Control	160	< .001	.684
Laboratory Experiments	125	< .001	1.05
Non scale Active learning + Laboratory Experiments	146	< .001	1.02
Non scale Active learning + Laboratory Experiments + Supplemental Instruction	34	.004	.474
Active learning + Laboratory Experiments	156	< .001	.696
Active learning + Laboratory Experiments + Supplemental Instruction	27	< .001	.722

4.5.5 Scale literacy

Comparisons between student scale literacy scores from pre to post in general chemistry II, **Table 4.12**, showed no significant increases for any group. While general chemistry I students' scale literacy increased regardless of type of instruction, the same result was not replicated in general chemistry II. While general chemistry II students began the course with a higher average scale literacy pre score (66%) than general chemistry I students (61%), it is possible that students in general chemistry II have encountered a ceiling effect in which scale literacy is not increased further. Another possible explanation for this observation could be due to the scale-themed instruction integration cohort testing occurring in general chemistry II. As students frequently do not enroll in general chemistry II for the semester immediately following the completion of general chemistry I, students in any one section of general chemistry II likely have a large variation in both their exposure to scale instruction and the recentness of that exposure. This variation could account for the lack of observed growth in scale literacy. Once again, repeated testing with larger samples is recommended for the combined treatments of supplemental instruction, laboratory experiments, and active learning lecture activities in order to further investigate the seemingly positive trend in the data for incorporation of scale-themed instruction in both laboratory and lecture.

Table 4.12: General chemistry II scale literacy group comparisons

	n	p	effect size
Control	75	.667	
Laboratory Experiments	56	.290	
Non scale Active learning + Laboratory Experiments	76	.718	
*Non scale Active learning + Laboratory Experiments + Supplemental Instruction**	*17*	*.533*	
Active learning + Laboratory Experiments	55	.125	
*Active learning + Laboratory Experiments + Supplemental Instruction**	*16*	*.835*	

** n of less than 20*

4.5.6 Residual analysis

In general chemistry II, multiple regression residual analysis of the ACS Exams General Chemistry Conceptual exam (second term), **Table 4.13**, revealed significant increases in student performance for all treatments until supplemental instruction was implemented into the course. While the trend in positive residual averages (3.1%-4.6%) is not consistent with what is expected for the given treatments (i.e. laboratory experiments yielding a more positive residual average that laboratory experiments in conjunction with active learning), it is highly inconsistent for residual averages for the combined treatments of both scale-themed and non-scale active learning, laboratory experiments, and supplemental instruction to yield -1.4% to 1% residual averages. While these results once again suggest that incorporation of supplemental instruction in general chemistry II actually hinders development of scale literacy, the residual averages are higher (or less negative) for those students in each section who completed the supplemental instruction than for those who did not. More concerning is the fact that those students not completing supplemental instruction have residual averages of -2.5% and -1.9% for those students receiving scale-themed active learning lecture instruction and non-scale-themed active learning lecture instruction, respectively. This observation can likely be attributed to an instructional effect stemming from the split cohort during the Spring 2017 semester of testing, although, repeated testing of both combined treatments with larger samples should be considered before investigating the existence of these effects.

Table 4.13: General chemistry II Conceptual final residual averages and group comparisons.

	n	Residual average	p	effect size
Control	87	<1%	.930	
Laboratory Experiments	67	4.6%	<.001	.390
Non scale Active learning + Laboratory Experiments	90	3.6%	.006	.303
Non scale Active learning + Laboratory Experiments + Supplemental Instruction	26	-1.4%	.507	
Active learning + Laboratory Experiments	86	3.1%	.014	.238
Active learning + Laboratory Experiments + Supplemental Instruction	20	1.0%	.644	

4.6 References

1. McFarland, A. D.; Haynes, C. L.; Mirkin, C. A.; Van Duyne, R. P.; Godwin, H. A., Color My Nanoworld. *Journal of Chemical Education.* **2004**, *81*, 544A.

2. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1526.

3. Tretter, T. R.; Jones, M. G.; Minogue, J., Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching.* **2006**, *43*, 1061-1085.

4. Jones, M. G.; Tretter, T.; Taylor, A.; Oppewal, T., Experienced and Novice Teachers' Concepts of Spatial Scale. *International Journal of Science Education.* **2008**, *30*, 409-429.

5. Jones, M. G. *Extreme science : from nano to galactic : investigations for grades 6-12 / M. Gail Jones, Amy R. Taylor, Michael R. Falvo;* Arlington, Va. : NSTA Press: 2009.

Chapter 5: Conclusions and Implications for Practice

This chapter presents conclusions based on the results described in this dissertation and is broken down into three sections. The first section focuses on conclusions related to the integration of scale themed instruction into general chemistry I and II courses. The second section focuses on the limitations of the research presented and the third section focuses on the implications these results have for instruction in both the domain of chemistry and the discipline of science. The final section briefly describes the on-going continuation of this work and possible future directions of this project.

5.1 Conclusions

Integration of explicit scale-themed instruction in an undergraduate general chemistry course has been accomplished through rigorous control/treatment cohort testing by incorporating supplemental instruction and adapting laboratory experiments and lecture instructional materials. In general chemistry I, several important results demonstrate that explicit incorporation of this theme into instruction directly led to increased demonstrated proficiency by students on chemistry assessments. This proficiency was measured by positive changes in student performance on pre-laboratory quizzes, a laboratory survey, and final exams. Most importantly, trends in the observed positive changes were consistent with expected trends based on control testing and combined treatment effects. Notably, residual averages for the combined treatments of supplemental instruction, laboratory experiments, and active learning lecture activities over three repeated semesters of testing showed consistency in improving student final exam scores 4.9%-6.7% on the ACS Exams Paired Questions Exam and 4.2%-5.1% on the ACS Exams First Term Conceptual Exam.

5.2 Limitations

Limitations to this work center around the many variables that exist when instituting curricular change such as the instructor of record and teaching assistant turn over. During the time period in which this data was collected it is possibly that any one of three rotating instructors of record taught general chemistry I, general chemistry II, or both courses during a single semester. While the rotation of these instructors provides validity to the observed consistency in results, it is unlikely that any one instructor was able to completely separate themselves from incorporating concepts of scale during lecture control semesters once implemented. In a similar vein, over the course of the 10 semesters in which this research was conducted, countless teaching assistants were responsible for overseeing student completion of the scale themed laboratory experiments and completion of the active learning lecture assignments in discussion. Teaching assistants in these positions were often unaware of the great deal of influence their perceptions of these research objectives had over student perceptions and it is possible that some teaching assistants' experiences influenced the experience of the student.

5.3 Implications for teaching

While instructional standards such as the AAAS Benchmarks for Science Literacy and more recently, the Next Generation Science Standards provide a strong argument for the incorporation of themes into instruction, at this time only 18 states plus the District of Columbia have adopted the standards. Furthermore, as the standards only specify what a student needs to know or do to demonstrate proficiency little is known as to the extent by which the standards in those states are incorporated into teaching. Compounding factors, such as incorporating of a single cross-cutting concept which draws together scale, proportion, and quantity greatly diminished the extent by which scale is conceptualized as distinctly different from proportion or quantity. The implication of this being that upon assessment, an instructor might feel as though they adequately address the cross-cutting concept in their classroom while actually only addressing two thirds of the standard.

As noted throughout the body of this dissertation, all of the instructional materials developed as a part of this work are available to any instructor who may wish to integrate them into their curriculum. While it is the goal of the research that an instructor would be able to drop these materials into their chemistry curriculum and observe positive changes in student performance, more importantly, what this research does is provide an instructional guide for any instructor in science who wishes to incorporate themes into instruction. Although the efficacy of scale-themed instruction in increasing student performance in chemistry has been demonstrated in this work, it is possible that incorporation of other themes, possibly in conjunction with scale, could further improve science literacy.

5.4 Future directions

As found in this work, the role understanding scale plays in understanding chemistry and the effectiveness of scale-themed instruction in general chemistry II is less well understood than in general chemistry I. A longitudinal study into the conceptions of scale held by students at the start of general chemistry II or for those continuing into organic chemistry courses could provide insight into instructional targets related to helping students further develop skills related to scale. As evidenced in the work of Jones and Taylor, scale is a theme that pervades any science course, not just chemistry. Knowing the degree to which scale conception impacts student performance in chemistry, it is logical to assume that undergraduate students in other science disciplines demonstrate the same deficiencies. As many of the themes of scale that are present in chemistry are also mirrored in biology courses, most noticeably the connection between the macroscopic and microscopic realms, the extension of this work into a biology course is currently underway.

Chapter 6: Class-wide Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry

6.1 Introduction

As described in sections 3.3.1 and 4.2.1, as part of an on-going research study into the efficacy of instructional approaches in teaching scale to undergraduate general chemistry I students, a sequence of laboratory experiments was developed that highlight specific concepts of scale. One activity, the "scale activity" was adapted from the work of Thomas Tretter and Gail Jones[8] and from a one-on-one interview acitivity[14] previously published by this research team. Of particular interest was the fact that this activity not only sought to help students increase their knowledge of scale concepts but also gave feedback into the conceptual boundaries of scale held by students. Provided that the activity continues to function as it was intended, adapting this activity into a class-wide activity could provide valuable insight into how a larger proportion of this population of students think of and conceptualize scale.

In this activity students created "bins" to sort objects spanning a wide range of sizes and then given 20 cards containing the names of objects to sort into their bins. The preliminary data collected from this activity shows consistency between the class-wide activity and the previously published data in which students frequently operate within a very narrow range of scale, typically centered around the height of an adult. Additionally, students often lumped all nonvisible items into a single bin, ignoring the many orders of magnitude separating these objects. Finally, when asked to place the items in order within their bins, students struggled to correctly order the nonvisible items.

6.2 Methods

6.2.1 General methods and activity description

Unlike the work of Tretter and Jones and the one-on-one interview activity, the “scale activity” laboratory activity not only set out to measure the current conceptions of scale held by students but also to allow students to become familiar with and practice several of the concepts they’d be working with throughout the semester of scale-themed instruction. This was accomplished through several distinct portions of the activity which included a card sorting activity, a worksheet to familiarize oneself and practice using logarithms, and an absolute scaling activity in which students worked to develop new anchor points for scaling from the size of a human down to the size of an atom. Of most interest to the research team were how the results of the card sorting task compared to those of the one-on-one interviews.

Similar to one-on-one interviews, the card sorting task had three parts:

1) Bin Creation and Item Sort
2) Item Ordering within Bins
3) Item Ordering with Measurements

The card sorting task was designed to be completed within the first 60 minutes of a 3-hour laboratory period. Students chose their own groups, and were permitted to work in pairs, or groups of three. Instructions for completing the activity were given to students by either verbal (teaching assistant) or written (laboratory manual) instruction in a way that intended to not prompt, cue, or guide them in their bin description creation or item placement. Students were also told that they could change their bin descriptions or quantity of bins at any point throughout the activity.

In *Part 1 (Bin Creation and Item Sort)*, students were instructed to create bin descriptions to organize 20 items of varying size in a single dimension such as length. Students were encouraged to create their smallest and largest bins on an open interval so that items that fell outside the boundaries of their bin descriptions would still be encompassed. For example, an open interval bin designation may be represented as, “less than an ant” or “taller than a 30 story building”. All other bins were to be created using a closed interval range such as, “from the size of a mouse to the size of a dog”. This portion of the study differed from Tretter and Jones’ study in that students were asked to create their bins prior to receiving the object cards so as to prevent any cueing in the bin description creation process that could occur from seeing the items in the activity. Participants were advised that they may add, delete, or edit their bin descriptions at any time throughout the activity. Students were asked to fill out bin cards with the descriptions that they created and to record their bin descriptions in their laboratory notebooks. Both the bin cards and student laboratory reports (collected the following week) were used for analysis.

Following the creation of their initial bin descriptions, students were asked to record their bin descriptions in their notebook before being given 20 cards containing the name of an object such as “atom” or “diameter of the earth” (see **Table 6.1** for the full list of objects). These items covered a vast range of sizes (from femtometer to terrameter), although students were not given the numerical size for any object during this portion of the activity. Without regard to any order, students were told to sort these objects into the bins using the descriptions they had created. As specific objects were used in creation of the object cards (such as red blood cell for “cell”) clarification regarding the actual object or dimension of the object described by the card was given to the student upon asking. Students were asked to record which bin each item was initially placed.

Table 6.1: Object cards with measurements*

Object	Dimension
Atomic nucleus	10 fm
Atom	100 pm
Virus	100 nm
Bacterium	1 μm
Cell	7 μm
Hair Width	100 μm
Ant	2 mm
Postage stamp	1.5 cm
Finger	8 cm
New pencil length	21 cm
Textbook	28 cm
Adult height	2 m
Semi-truck	20 m
Football field	91 m
Cruising altitude of a 747	11 km
Width of Wisconsin	450 km
New York to Los Angeles	4800 km
Diameter of Earth	13 Mm
Distance from Earth to Moon	384 Mm
Distance from Earth to Sun	146 Tm

*Measurements for objects are reported in the most commonly used unit to describe the object.

In *Part 2 (Item Ordering within Bins)*, students were asked to order the object cards by size (from smallest to largest), within their bins. In this portion of the activity, students continued to use the object cards only containing the object name or dimension. Students were advised that they could move items into different bins from where originally sorted and were again informed that they could change their bins at any point. When the groups were finished ordering the cards, they were asked to once again record their bin descriptions along with the placement of each object within their bins.

In *Part 3 (Item Ordering with Measurements)*, students were given cards containing the same 20 objects that they had been given in the previous activities. However, in this portion of the activity, each object card contained not only the object name, but the numerical size of the item with associated metric units. Students were again asked to sort these cards (from smallest to largest) within their bins and reminded that they could change their bins at any time. Following the final ordering activity, students were asked to one last time record their bin descriptions and item placements.

6.2.2 Spring 2013

The scale activity laboratory activity was implemented beginning in the spring 2013 semester. Teaching assistants for this semester were required to attend an additional training session for this activity in which they were given the opportunity to view the materials and gain a better understanding of what was expected of both themselves and of the students during the different tasks of the activity. During this training, teaching assistants were asked to introduce the activity without cueing students to possible bin descriptions although they were not given specific instruction as to how to do this. Additionally, based on the results of the one-on-one interviews in which students most commonly created 6 bins, the bin cards for the scale activity were dispersed to students in groups of 6. Students were advised that there were additional cards available should they choose to change bins or create new bins during the activity. Lastly, a member of the research team observed a portion of each of the 21 individual sections of the course in which the activity was being trialed to identify areas in need of refinement.

6.2.3 Spring 2014 – Fall 2015

Based on the results from the initial class-wide implementation of the scale activity, a more comprehensive training was offered to all teaching assistants leading students in this activity for the first time. In this training, the research team instructed the teaching assistants to join into groups and complete all three parts of the card sorting task using the mindset of an undergraduate student. This allowed for the teaching assistants to gain a better understanding of what was being asked of the students which in turn made them better equipped to instruct students and answer questions related to the different tasks of the activity while still maintaining the integrity of the data collected by the research team. Analogous examples to describe the activities objectives, such as categorizing events by length of time, were developed and distributed to teaching assistants to help them introduce the activity without using any reference to linear distances. Additionally, quantities of bin cards were distributed around the activity room to allow students to determine the number and quantity of bin cards they wished to use. Students were explicitly asked to record whether they had changed their bins and asked to record their bin descriptions at the end of each part of the card sorting task in distinct tables.

6.3 Results and Discussion

6.3.1 Spring 2013

Data collected from the spring 2013 implementation of the scale activity revealed several aspects in need of refinement. Beginning with qualitative observations made by the research team during the laboratory periods, it was clear that teaching assistants did not know how to introduce the idea of a “bin card” or how to instruct students in how to “correctly” label their bins (“bin descriptions”). In 10 of the observed sections, a teaching assistant gave an example bin description to the entire group of students that included analogous objects to those included in the actual activity. In two of those same sections, a complete example set of bin descriptions were given to students by a teaching assistant during the activities introduction. The same bin descriptions then frequently reappeared in student laboratory reports used for analysis in those sections. In all sections, including those in which obvious cueing was observed from the teaching assistant, the teaching assistants struggled to properly instruct students in how to create bin descriptions that encompassed an inclusive range. Of the 145 initial groups completing all aspects of the activity, 98 groups were excluded from analysis for creating bin descriptions that either were not a range or whose bins featured descriptions that were not inclusive. An example set of each type of excluded bin description is shown in **Figure 6.1**. Five additional groups were excluded for creating generic bin names that did not convey any meaningful information such as "bin 1" or "really small"

Figure 6.1: Example of student created bin descriptions excluded during analysis

Not a range	Not inclusive
Atom	< molecule
Bee	dust-cell
Truck	man to ant
Mountain	mountain to man
Moon	state to country
Sun	> continent

Of the 145 initial groups and 42 analyzed groups, 101 (70%) and 27 (64%) respectively, used the exact amount of bin cards (6) that were distributed to them. Despite repeated reminders to students that they could change their bin names or quantity of bin cards at any time, no group formally reported changing either at any point during the activity or in their laboratory report. Despite the small number of student groups remaining in the sample, preliminary analysis demonstrated consistent results to those found during the one-on-one interviews in terms of bin description selection and item placement and ordering. Given the increased likelihood of bias in this data, however, these results are not presented here.

6.3.2 Spring 2014-Fall 2015

Following the fall 2013 laboratory control experiment, data collection for the scale activity resumed during the spring 2014 semester. Upon implementation of more detailed teaching assistant training with regards to bin description creation, with the exception of one section of data from the spring 2015 semester, no evidence of cueing from teaching assistants exists in the data. Furthermore, the occurrence rate of exclusion for groups making bin names not encompassing either ranges or non-inclusive ranges dropped from 68% to an average of 23% (range of 15%-37%). Similarly, the percentage of students creating exactly 6 bins dropped from 64% during the initial activity to no more than 23% (range of 19-23%) when students were not given a pre-determined number of blank bin cards.

Part 1: Bin Creation and Item Sort

For each of the 4 semesters in which data was analyzed, the student bin descriptions were recorded and analyzed in the same manner as those from the one-on-one interviews. As the students created bins that encompassed a wide variety of descriptions (from everyday objects to metric sizes), the bin descriptions used by students were categorized based on how the object cards were sorted into the bins. Using the lower boundary of the bin description, a bin boundary was identified based on the object that either exactly aligned with the boundary or the first smallest item to fall outside of the bin. For example, “atom” would have been assigned as the boundary for a bin description of “atom-amoeba” as well as for a bin description of “molecule-amoeba”. As the goal of this activity was to explore student conceptions of the “small” end of the size spectrum, the lower boundary of each bin description was aligned with the objects.

Given the inclusive nature of each set of bins, each object or dimension used as a bin description was only counted once. For example, if the first two bin descriptions were "less than the size of an atom" and "from the size of an atom to the size of a virus" the bin boundaries would have been assigned as "atomic nucleus" and "atom" respectively. If the first two bin descriptions were "less than the diameter of DNA" and "from DNA to a virus" only one bin boundary of "atom" would have been assigned. The greatest amount of judgement in assigning bin boundaries was used when students chose generic items such as "building" or "lake" as boundaries for their bin descriptions. While students were encouraged to create bin descriptions specific enough that another student would interpret the description in the same manner, the level of specificity required by this was sometimes overlooked. A full set of student bin descriptions and their assigned bin boundaries with rationale are presented in **Figure 6.2**.

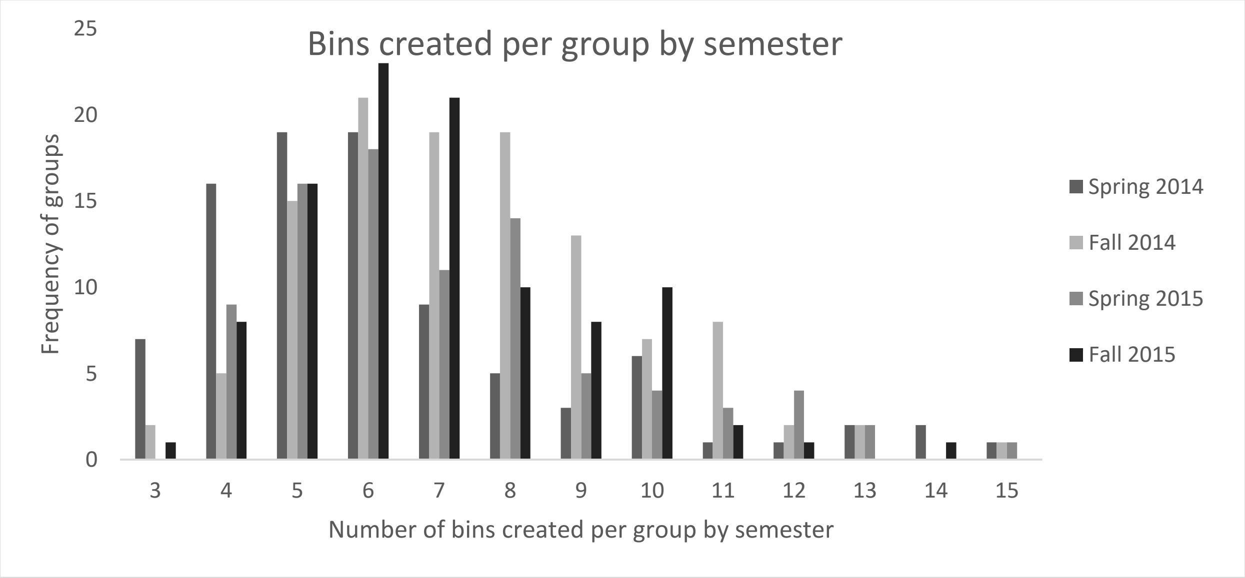
Figure 6.2: Example of student created bin descriptions and their assigned bin boundary with rationale

Bin description	Categorization and rationale
Smaller than a sugar cube	Ant – because an ant is smaller than a sugar cube
The size of a sugar cube to the height of a telephone booth	Ant – because an ant is smaller than a sugar cube
The height of a telephone booth to the width of Texas	Adult height – because an adult is smaller than a telephone booth
The width of Texas and larger	Width of the state of WI – because the width of WI is smaller than the width of Texas

Another important distinction was made in determining the threshold for different categorizations of student bin descriptions. Based on results of the one-on-one interview activity, identification of those bin descriptions fitting the categories of nonvisible, outside 3 orders of magnitude of the height of an adult to the small, and outside 3 orders of the magnitude of the height of an adult to the large were of most interest. Maintaining the most rigid of standards, it was determined that only those bin descriptions in which an actual nonvisible object or dimension were used would fall into the nonvisible category. In most cases this was very straightforward as groups commonly used descriptions such as “less than the size of a cell” or “cannot be seen with the unaided eye”. However, students using bin descriptions of “less than the width of a human hair” or “smaller than what we can see” were not counted as having a bin boundary in the nonvisible realm as the item to which their description was anchored, was actually a visible item. Hair width and football field were selected as the threshold items for groups creating bins falling outside of 3 orders of magnitude to the height of an adult on either end.

Examination of the bin descriptions used by students showed that on average, and consistent with one-on-one interviews, students created 6-7 bins per group (**Table 6.2**). Further analysis shows the range in number of bin descriptions created extends from as few as 3 to as many as 15 (**Figure 6.3**).

Figure 6.3: Distribution of number of bins created per group by semester



When considering the creation of bin descriptions which fell into the nonvisible region, **Table 6.2**, unlike the one-on-one interviews in which only 37% of students used a nonvisible item in their bin description, in the class-wide data a larger percentage of students (48%-72%) used a bin description that anchored to an object in the nonvisible realm. This increase likely can be attributed to both the completion of the pre-laboratory assignment and the collaborative nature of working in a group. These numbers drop significantly to an average of only 24% (range 16% to 37%) when considering groups using 2 nonvisible objects within their bin descriptions and 5% (range 2% to 11%) when considering groups using 3 nonvisible objects within their bin descriptions. This steep decline when considering groups creating multiple nonvisible bin descriptions is once again consistent with the one-on-one interviews.

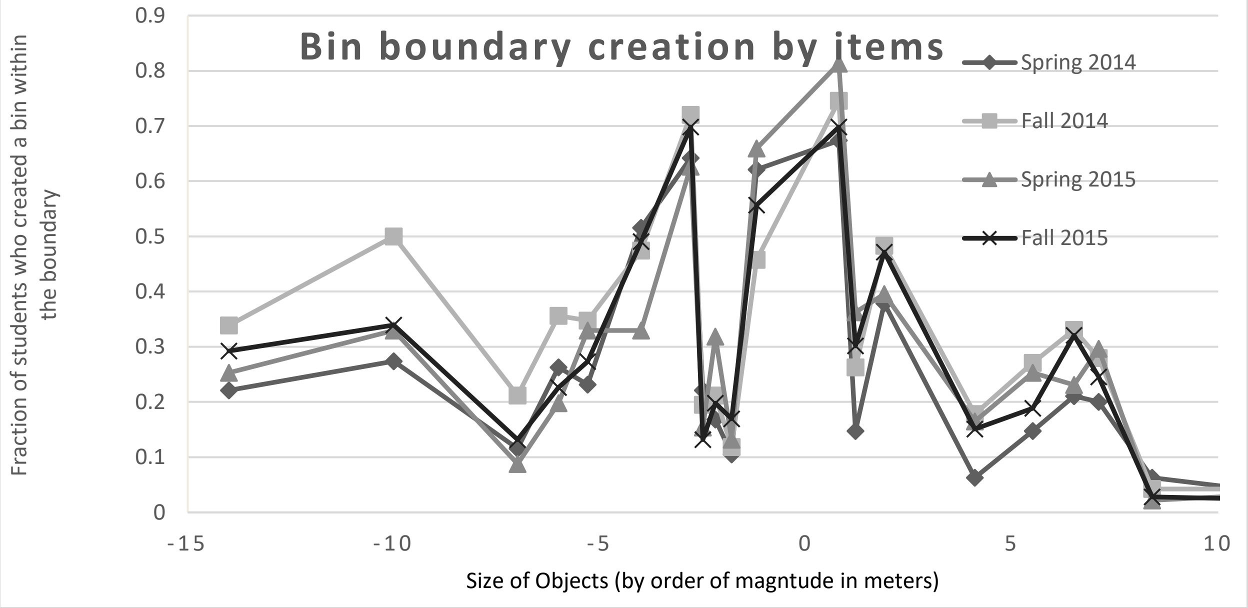
Table 6.2: Class-wide student bin description creation and bin boundary analysis

Parameters	Spring 2014	Fall 2014	Spring 2015	Fall 2015
Total # of groups (students)	95 (186)	118 (228)	91 (179)	106 (216)
Total number of bins created	628	883	653	740
Average number of bins per group	6.5	7.4	7.2	6.9
Total number of nonvisible bins created	68 (48%)	142 (72%)	64 (50%)	85 (55%)
Total number of bins larger than a football field created	110 (58%)	202 (86%)	136 (71%)	164 (84%)
Groups making 1 bin in each extreme	49 (52%)	99 (84%)	53 (58%)	54 (51%)
Groups making 2 bins in each extreme	21 (22%)	51 (43%)	26 (29%)	19 (18%)
Groups making 3 bins in each extreme	4 (4%)	7 (6%)	6 (7%)	4 (4%)
Groups using measurements instead of objects	29 (31%)	24 (20%)	11 (12%)	20 (19%)
Groups using measurements who also created a nonvisible bin	14 (48%)	20 (83%)	1 (9%)	13 (65%)
Groups who changed their bins	35 (37%)	17 (14%)	31 (34%)	30 (28%)
Groups who didn't initially make a nonvisible bin but changed bins to include a nonvisible bin	6 (17%)	2 (12%)	8 (26%)	1 (3%)

Of all the objects, the greatest number of bin descriptions (70%) aligned to the height of an adult and to the objects that fell within 3 orders of magnitude on either side of the height of an adult.

Given the familiarity of students with objects of this dimension, it is not surprising that a much larger percentage of groups (average of 75%) used a bin description that fell into this category as depicted in **Figure 6.4**. Interestingly, as seen in table 3, 62% of all groups created one bin on each end of this spectrum. This number drops drastically when considering groups using 2 (28%) or 3 (5%) bin descriptions in each of these size extremes.

Figure 6.4: Bin boundaries of group-created bins reported using the fraction of groups who created a bin within the boundary of the objects.



When considering the use of objects versus the use of measurements in the creation of bin descriptions, it was found that no more than 31% of groups in any semester (range 12%-31%) used measurements as opposed to objects to describe their bins (**Table 6.2**). This percentage is much smaller than the average 40% of students who used measurements to describe their bins in the one-on-one interviews. Given that the use of bin descriptions centered on measurements as opposed to objects implies a better understanding of scale, it was expected that those students using measurements may have been more likely to use a bin description anchored to a nonvisible dimension. However, data from repeated use of this activity suggests that this is not the case and students were equally likely to use a nonvisible bin description regardless of the type of bin descriptions used. Of concern, when considering the small fraction of groups who reported changing their bins upon seeing the object cards, only an average of 14% (3% to 26%) of those groups who did not initially use a nonvisible bin description changed their first bin to include a nonvisible object or dimension. However, given that no student in the one-on-one interview sample changed their bin quantity or description at any point, the observation of any group reporting a change in their bin descriptions should not be discounted.

Lastly, when considering the results of the initial object card sorting task (**Figure 6.5** and **Table 6.3**), students often only created a single bin to hold the entirety of the nonvisible spectrum encompassed by the items in the activity and an additional 1-2 bins to hold every item used in the activity that was larger than a football field. The remaining 5 orders of magnitude falling between the width of a human hair and the length of a football field were split over the remaining 4-5 bins. Once again consistent with the one-on-one interviews, adult height was frequently placed in a bin by itself.

Figure 6.5: Results of initial sort

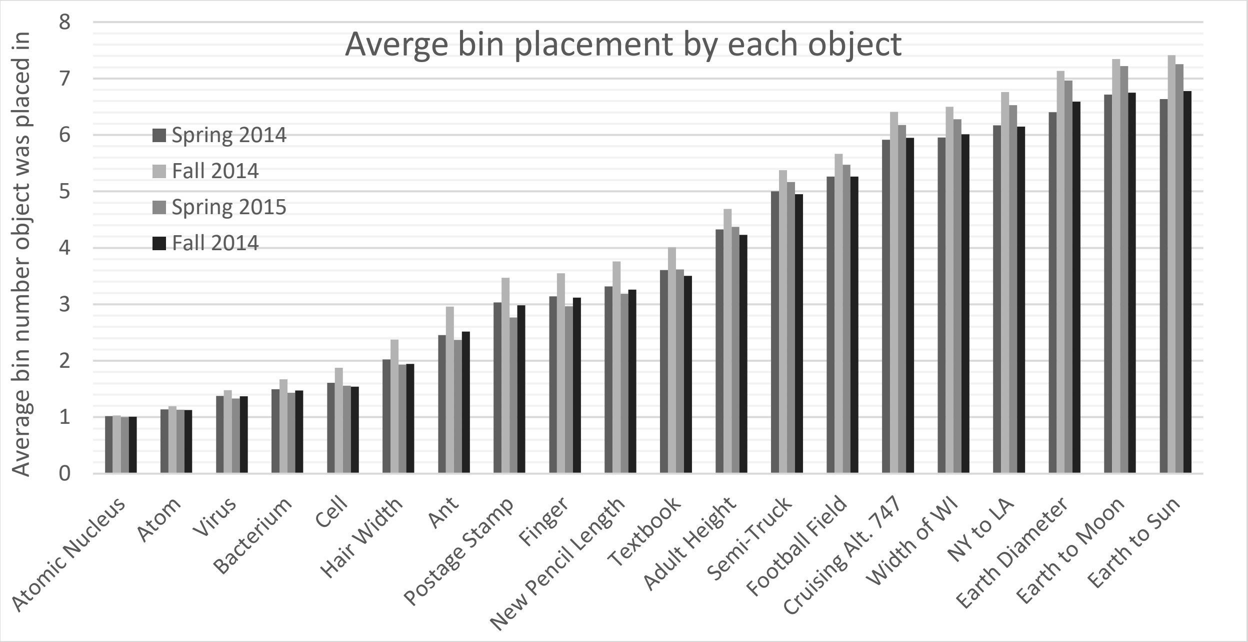


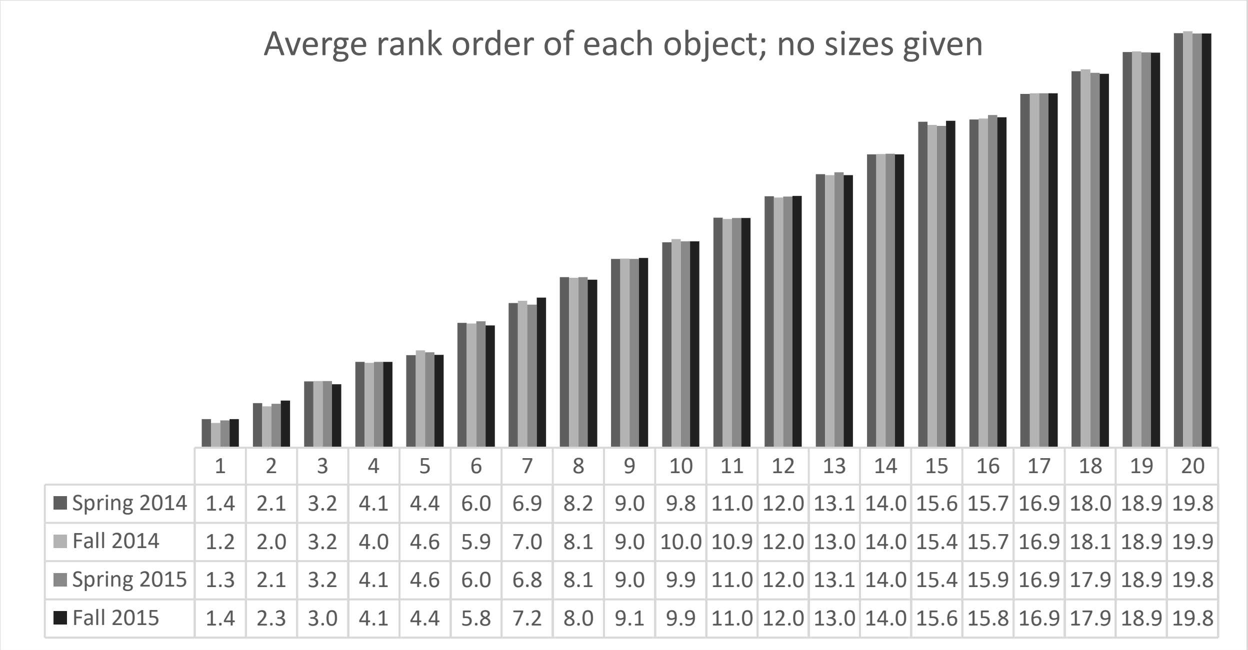
Table 6.3: Average span of each bin in orders of magnitude

Bin number	1	2	3	4	5	6	7
Objects placed in bin	Nucleus Atom Virus Bacterium Cell	Hair Ant	Stamp Finger Pencil Textbook	Adult	Semi Football field	747 WI NY to LA Earth	Earth to Moon Earth to Sun
Range of bin (in orders of magntitude)	8	1	1	--	<1	4	5

Part 2: Item ordering within Bins

In Part 2 of the activity, students were asked to order the objects within each bin from smallest to largest relative to the other objects (rank order). **Figure 6.6** displays the average rank order (1-20) of each object as recorded in the student's laboratory notebook as compared to the true rank order of each object. As can be seen in the figure, students were able to place many of the items in the correct rank order. Consistent with the one-on-one interviews students had the greatest difficulty in correctly placing “bacterium” and “cell” on the small end and “the cruising altitude of a 747” and “the width of the state of WI” on the large end. While the size of a bacterium was often placed correctly relative to the size of a virus, the size of a cell was often placed lower than its expected position as evidenced by its average placement value being less than its expected value. This opposes the results of the one-on-one interviews in which on average, virus, bacterium, and cell, were frequently placed as too large and too small respectively. On the large end, and consistent with the one-on-one interviews, the cruising altitude of a 747 was often overestimated and placed higher than expected while the width of the state of WI was often underestimated and placed lower than expected. One explanation for this phenomenon comes from Tretter and Jones who suggest that students face more difficulty ordering objects they conceive of to be similar. For example, in one study students were reported as saying that although the difference in distance between the span of a bridge and the distance between Miami and Boston is large, in comparison to a planetary distance, the difference is less noticeable[1]. This could explain how when comparing the cruising altitude of an airplane and the width of the state of WI to much larger distances such as the distance between New York and Los Angeles or the diameter of the earth, the two objects become grouped together as being similar in size. Another compounding factor that was frequently mentioned by students in that study was a lack of direct experience with either of the objects in question.

Figure 6.6: Average rank order of each object (no sizes given). Reported as average item placement (1-20) as compared to correct item placement (1-20).

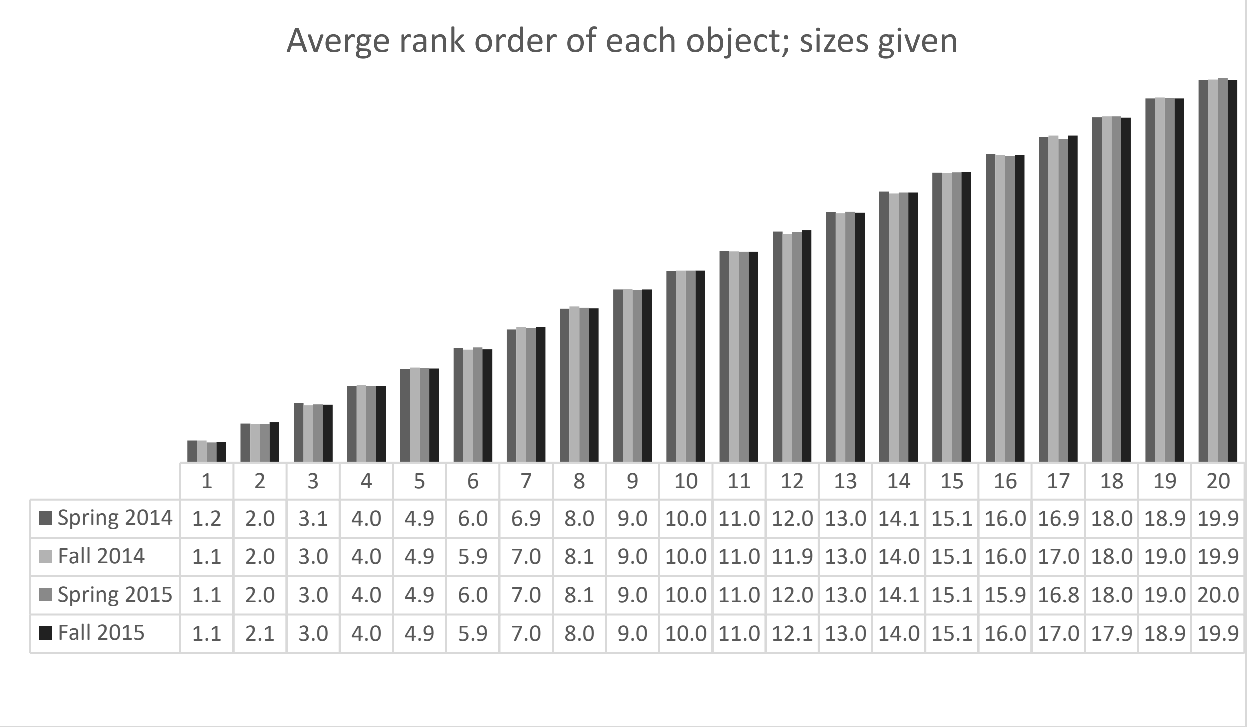


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
■ Spring 2014	1.4	2.1	3.2	4.1	4.4	6.0	6.9	8.2	9.0	9.8	11.0	12.0	13.1	14.0	15.6	15.7	16.9	18.0	18.9	19.8
■ Fall 2014	1.2	2.0	3.2	4.0	4.6	5.9	7.0	8.1	9.0	10.0	10.9	12.0	13.0	14.0	15.4	15.7	16.9	18.1	18.9	19.9
■ Spring 2015	1.3	2.1	3.2	4.1	4.6	6.0	6.8	8.1	9.0	9.9	11.0	12.0	13.1	14.0	15.4	15.9	16.9	17.9	18.9	19.8
■ Fall 2015	1.4	2.3	3.0	4.1	4.4	5.8	7.2	8.0	9.1	9.9	11.0	12.0	13.0	14.0	15.6	15.8	16.9	17.9	18.9	19.8

Part 3: Item ordering with measurements

In part 3, students were asked to once again order the objects from part 2 within each bin from smallest to largest, but this time were given the measurement and unit associated with the object. Unlike the one-on-one interviews in which performance was only slightly improved for the objects on the small end of the spectrum, when given the actual measurements, all groups made dramatic improvements in the placement of all objects (**Figure 6.7**).

Figure 6.7: Average rank order of each object (with sizes given). Reported as average item placement (1-20) as compared to correct item placement (1-20).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
■ Spring 2014	1.2	2.0	3.1	4.0	4.9	6.0	6.9	8.0	9.0	10.0	11.0	12.0	13.0	14.1	15.1	16.0	16.9	18.0	18.9	19.9
■ Fall 2014	1.1	2.0	3.0	4.0	4.9	5.9	7.0	8.1	9.0	10.0	11.0	11.9	13.0	14.0	15.1	16.0	17.0	18.0	19.0	19.9
■ Spring 2015	1.1	2.0	3.0	4.0	4.9	6.0	7.0	8.1	9.0	10.0	11.0	12.0	13.0	14.1	15.1	15.9	16.8	18.0	19.0	20.0
■ Fall 2015	1.1	2.1	3.0	4.0	4.9	5.9	7.0	8.0	9.0	10.0	11.0	12.1	13.0	14.0	15.1	16.0	17.0	17.9	18.9	19.9

6.4 Limitations

An important note of distinction between the one-on-one interviews and the different semesters of the scale activity laboratory activity is in both the preparation for the activity and the on-going integration of scale-themed instruction simultaneously occurring while this data was being collected. Unlike the one-on-one interviews, students enrolled in these laboratory courses were expected to complete a pre-laboratory assignment. As part of this assignment students were given a list of several common nonvisible objects (atoms, viruses, bacteria, molecules, etc.) and were asked to list other nonvisible objects which likely cued them to think about and use these objects as they worked on this activity in the laboratory. Furthermore, depending on the semester in which the student was enrolled (see **Table 24**), the students may have already had a brief introduction to scale prior to completing the scale activity laboratory activity. As the scale activity laboratory activity occurs very early in the laboratory sequence, it was not expected that these instructional effects would impact the data significantly, however, it is for this reason that all results of this activity are not presented in aggregate, but rather by the semester in which the data was collected.

6.5 Conclusions and Implications for Practice

In comparing the results of the one-on-one interviews with the results of class-wide data collection, consistent findings were found for both groups. When considering the bin descriptions used by students the data once again shows that students are most comfortable operating in a narrow range of scale (centered around their own size) and when working with familiar objects. Noting the frequency with which adult height and ant were selected as bin boundary descriptions suggests that students in the class-wide group are comfortable scaling down to the size of 1 mm and beginning to use 1 mm as an anchor point for scaling to sizes smaller than 1 mm. However, the infrequency with which any object or dimension falling outside ± 3 orders of magnitude to the size of a human was selected suggests that students are only in the very initial stages of developing this skill. When considering the relative scaling of the specific objects used in this activity, objects falling outside ± 3 orders of magnitude to the size of a human often represented a threshold for which accuracy in placing these objects decreased markedly. Although the students in the class-wide studies seemed to recognize a distinct size difference between an "atomic" and "microscopic" scale, students often failed to correctly differentiate between objects falling into those categories. The results of both the one-on-one interviews and the class-wide studies indicate that novice college chemistry students do not demonstrate a high level of scale literacy upon entering a beginning chemistry course and highlight the need for explicit scale instruction to be included in instruction.

6.6 References

1. Tretter, T. R.; Jones, M. G.; Andre, T.; Negishi, A.; Minogue, J., Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching.* **2006**, *43*, 282-319.

2. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1526.

Chapter 7: Response Process Validity Studies of the Scale Literacy Skills Test

7.1 Introduction

As described in section 3.2.1, as part of an on-going research study into the efficacy of instructional approaches in teaching scale to undergraduate general chemistry I and II students, two assessments were used to measure student ability in scale. The scale literacy skills test is a 45-item multiple choice assessment developed to measure student ability in several distinct content areas related to scale and the scale concept inventory is a 40-item Likert scale survey that measures the degree to which students agree or disagree with several common misconceptions related to scale. The preliminary reliability and validity studies of both of these instruments have been published previously[1]. Despite the established validity of the assessments using domain experts, item statistics, and trial testing, it was always the intention of the research team to enhance the validity through a response process study[2]. As many claims upon which the foundation of this research has been built around student performance on these assessments, of critical importance was the identification of any items that pose a threat to the validity of the assessment. Following the response process study published by Jack Barbera and co-workers[3], the response processes of general chemistry I and anatomy and physiology I students on the scale literacy skills test and scale concept inventory were collected and analyzed.

7.2 Methods

7.2.1 General Chemistry I

General chemistry I interview participants were solicited during the spring 2014, fall 2014, and spring 2015 semesters. The spring 2015 solicitation was only opened to male students to ensure equal sampling of each gender. Students received a copy of the ACS General Chemistry Study Guide or a gift card for their participation in the study. Interview participants digitally presented with each item of both assessments and instructed to verbalize their problem-solving process as they worked to arrive at an answer and select their chosen answer. Students were provided a calculator and scratch paper but were given the caveat that if used, the student must verbalize everything they are writing down or entering into the calculator. The interviewer asked follow-up questions as needed to ensure the process used by the student was thoroughly captured. Example items were presented to the student in which the interviewer first demonstrated the expected process to the interviewee and the interviewee then completed an example of their own with feedback from the interviewer. Each interview was video recorded and the audio transcribed for further analysis.

7.2.2 Anatomy and Physiology I

Based on the results of the response process validity study in chemistry and the results of the initial implementation of the scale assessments in anatomy and physiology I, it was determined that changes to both assessments would be made. The scale literacy skills test would be adjusted to address the threats found through response process which included changing the stem (2), changing the distractors (2), or adding clarifying language (2) to 6 of the items featured on the assessment. Additionally, 4 items of the scale concept inventory were updated to include clarifying language that was revealed to be necessary to account for the domain specific content knowledge of the anatomy and physiology students. These changes were made to the assessments prior to the start of the fall 2016 semester of data collection. Students enrolled in this semester of anatomy of physiology were solicited to participate in the response process study of the scale literacy skills test. The scale concept inventory was not selected to be included in the response process study as it was revealed during the chemistry study that students often did not have a justification for the selection of their level of agreement. Students were seated at a table and given a copy of the entire 45 item test and an answer sheet with workspace. Each item was given the same amount of work space regardless of the type of item so as not to cue students to the anticipated problem-solving process. Students in this interview set were also instructed to verbalize their problem-solving process as they worked to arrive at an answer and were instructed to verbalize anything written down in the work space along with their selected answer. The interviewer again asked follow-up questions as needed to clarify the student’s process.

7.2.3 Coding

Following the work of Jack Barbera, each interview response was coded based on the process used by the student. Those response codes could fall under the category of "intended process", "potential threat", or "other". Items coded under "intended processes" included **Totally Correct (TC)** in which a student who selected the correct multiple-choice option also provided a correct reasoning for selection of that response and **Totally Incorrect (TI)** in which a student who selected an incorrect multiple-choice option also provided an incorrect reasoning that supported selection of that response. For those items testing misconceptions, an additional code for **Supported Misconception Response (SMR)** was added in which a student selected a response related to a misconception and also demonstrated in their response that they held that misconception.

Any process demonstrated by a student falling outside of the intended processes listed above would be flagged as a potential threat and further evaluated. Potential threats were flagged as **Correct for the Wrong Reason (CW)** in which a student selects a correct response but does not provide reasoning that supports selection of that choice, **Incorrect for the Wrong Reason (WR)** in which a student selects an incorrect response but does not provide reasoning that supports the selection of that choice (i.e. the student was actually correct but for some reason found a different distractor appealing), and **Used a Test-Taking Strategy (TTS-E, TTS-TE, TTS-NM)** in which a student was able to apply a test taking strategy to increase their odds of answering correctly (further broken down by type of test-taking strategy applied – elimination, trial and error, or number matching). For those items testing misconceptions, the category of **Did not**

Support Misconception Response (DSMR) was applied for those students who selected a misconception response but did not demonstrate having the misconception tested by the distractor. A final category of "**other**" was applied for those students whose process could either not be elucidated from interview transcripts or for students who explicitly stated they were guessing.

Items were rated by independent raters (3 for the chemistry study and 5 for the anatomy and physiology study) and any discrepancies in coding were discussed until an agreement was met. Upon final assignment of the codes, any item in which 2 or more students reported using the same discrepant reasoning process were determined to pose threats to the validity of the assessment and considered for removal from analysis.

7.3 Results and Discussion

7.3.1 Response process support for content validity

In total, 38 students participated in the interviews from general chemistry I and completed all 85 items of both assessments during the interview period. As seen in **Table 7.1**, for the scale literacy skills test, intended processes were used by students at least 50% of the time on all items. This number only drops to 42 when considering students using an intended process at least 70% of the time and 33 when considering those in which an intended process was used at least 80% of the time. In anatomy and physiology, 20 students participated in the interviews where 43 items demonstrated an intended process use rate of greater than or equal to 50%. This number drops to 33 and 25, respectively, when considering those items in which intended processes were used greater than or equal to 70% and 80%.

Table 7.1: Number of items in which intended processes were used by students

Threshold	General Chemistry	Anatomy & Physiology
n	38	20
≥50%	45	43
≥70%	42	33
≥80%	33	25
≥90%	11	14

7.3.2 Response process support for threats

As described in **Table 7.2**, the response assignments in general chemistry I were first used to determine those items in which potential threats existed. Using a 2.5% occurrence rate 19 items fell into the category of having at least 1 student use a discrepant process in responding to the item. Upping the threshold to 5% (2 students using 1 or more discrepant reasoning processes) led to the identification of 6 items as potential threats to the validity of the assessment. Further inspection of each item and the type of processes used by the students narrowed the number of threats existing on the assessment to 5. The remaining item was determined not to pose a threat to the validity as no two students reported choosing the same incorrect answer while using the same process.

Table 7.2: Suspect items identified through response process in general chemistry I.

Gen Chem I	2.5%	Item(s)	5%	Item(s)	Threats	Item(s)
CW	12	2,3,9,11,17,19,21,23,27,29,31,39	2	3,9	2	3,9
TTS-E	2	10, 16	0	--	--	--
TTS-TE	2	12,14	2	12,14	2	12,14
TTS-NM	1	30	1	30	1	30
WR	2	5,24	1	24	--	--
DSMR	0	--	--	--	--	--

Using the same method as with the chemistry students, a 5% occurrence rate (1 student using 1 discrepant process) was chosen for the initial assessments of items posing a potential threat to the assessment in anatomy and physiology (displayed in **Table 7.3**). Of the 16 items initially identified, upping the threshold to 10% (2 students using 1 or more discrepant reasoning processes) led to the reduction of potential threats to 6 items. Further inspection of each item and the type of processes used by the students narrowed the number of threats existing on the assessment to 4. The remaining two items were determined not to pose a threat to the validity as while more than 2 students did use a discrepant process to select a correct response, no 2 students chose that response while using a similar process. While no common threats between courses were found on any items, several of the items were found to function very differently in each of the disciplines sampled.

Table 7.3: Suspect items identified through response process in anatomy and physiology I.

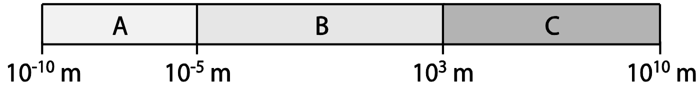
A&P I	5% (16)	Item(s)	10%(6)	Item(s)	Threats (4)	Item(s)
CW	10	1,2,3,11,19,20,21,29,31,39	4	2,20,21,29	2	2,20
TTS-E	4	7,8,20,31	1	8	1	8
TTS-TE	2	8,13	1	8	1	8
TTS-NM	2	16,31	1	31	1	31
WR	1	12	0	--	--	--
DSMR	1	5	0	--	--	--

7.3.3 Items displaying Correct for the wrong reason threats

Items 2 and 3 (**Figure 7.1**) were developed purposefully with several common student errors in mind and demonstrated very interesting results in both response process studies. The regions of the number line (specifically the boundary between region A and region B) were selected to identify those students who only use the given metric unit and not the value and metric unit together. Those students who correctly take into account both parts of the value in item 2 should correctly arrive at answer B, while those who do not, will choose answer A. Item 3 allowed many students to catch the missed value in item 2 and to go back and correctly identify choice A. These results were verified from the percent chosen of each distractor in both the class-wide and interview data sets.

Figure 7.1: Items 2/3 of the Scale Literacy Skills Test administered in chemistry

Use the figure for items 2 and 3 by selecting the region (by letter) for each value. If the region is not shown, respond (D).



2. The value 100 µm is in which region?

(A) A **(B)** B **(C)** C

(D) Region not shown on figure.

3. The value 1 µm is in which region?

(A) A **(B)** B **(C)** C

(D) Region not shown on figure.

Table 7.4: Item statistics and response frequency by percentage for items 2/3 as administered in chemistry

	Item 2 interview (n = 38)	Item 2 Class-wide (n = 2034)	Item 3 interview (n = 38)	Item 3 Class-wide (n = 2034)
Difficulty	.342	.544	.789	.688
Discrimination	.100	.464	.500	.454
Percent chosen A	**47.4**	**32.3**	78.9	68.8
Percent chosen B	**34.2**	**54.4**	2.6	21.8
Percent chosen C	13.2	8.2	10.5	5.9
Percent chosen D	5.3	5.1	7.9	3.5

Based on the item statistics of the interview set, shown in **Table 7.4**, item 2 appears to warrant further investigation as the both item difficulty and discrimination fall out of the recommended range. In the class-wide data set, item 2 appears to function as expected. However, during the interviews it became apparent that if a student confused µm to equal $1\text{x}10^{-9}$ m, while the student would correctly choose an incorrect response ($1\text{x}10^{-7}$, option A) to item 2, the student would incorrectly ($1\text{x}10^{-9}$ m, option A) choose the correct response to item 3.

Chemistry student responding to item 2: ***"I was looking for the range that micrometers was in and that is 10^{-9} meters so I would have to pick [region] A."***

Given that the threat to item 2 only existed due to the chosen unit, it was determined that changing the item to include a more familiar unit like millimeter could eliminate the threat posed by this item as students would be less likely to confuse its conversion for another unit. The number line and value in question were altered (**Figure 7.2**) to account for this change prior to presentation of the item to anatomy and physiology students.

Figure 7.2: Items 2/3 of the Scale Literacy Skills Test administered in anatomy and physiology. For clarity, these items are marked with an *.

Use the figure for items 2 and 3 by selecting the region (by letter) for each value. If the region is not shown, respond (D).

2*. The value 100 mm is in which region?

(A) A **(B)** B **(C)** C

(D) Region not shown on figure.

3*. The value 1 mm is in which region?

(A) A **(B)** B **(C)** C

(D) Region not shown on figure.

Table 7.5: Item statistics and response frequency by percentage for items 2*/3* as administered in anatomy and physiology.

	Item 2* interview (n = 20)	Item 2* Class-wide (n = 538)	Item 3* interview (n = 20)	Item 3* Class-wide (n = 538)
Difficulty	.850	.591	.500	.385
Discrimination	0.00	.341	.500	.348
Percent chosen A	5	13.8	50	38.5
Percent chosen B	85	59.1	35	42.6
Percent chosen C	5	15.8	0	5.2
Percent chosen D	5	11.3	15	13.8

Looking at the item statistics for items 2* and 3* in **Table 7.5**, it appears as though the modifications to items 2 and 3 have increased the students' likelihood of getting 2* correct, while decreasing the likelihood of getting item 3* correct (the desired outcome). However, through response process it was revealed that as opposed to the chemistry students who only paid attention to the unit, the anatomy and physiology students only paid attention to the value itself. As the region between 10^{-2} m and 10^{-3} m encompasses both the correct value of $1x10^{-1}$ m and the original value of 100 (or 10^2), 5 (25%) students in the interview set reported a process that resulted in the selection of the correct answer for item 2* while using a process other than that intended by the item.

A&P student responding to item 2*: ***"I'd say it's in [region] B because it's between 10…I think it's a -2 and 10^3 and 100 is 10^2 so it would be in that [region]."***

7.3.4 Items displaying Test Taking Strategies Threats

In each course, items were identified in which the distractors chosen allowed students to increase the odds of selecting a correct answer through use of a test taking strategy. One test taking strategy used by both groups was trial and error. This test taking strategy appeared in items in which students were able to use the distractors to attempt to back calculate an original value in order to determine the correct answer choice. In one particularly interesting example (**Figure 7.3**) that emerged as a threat in anatomy and physiology, students were able to correctly select "one order of magnitude" while demonstrating the use of strategy that should have lent to the selection of an incorrect answer choice.

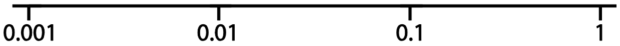
Figure 7.3: Items 7/8 of the Scale Literacy Skills Test

7. For the scale shown below, what is the size for each increment?

(A) 10^1 **(B)** 10 **(C)** 1

(D) 1 order of magnitude

8. For the scale shown below, what is the size for each increment?



(A) 10^{-1} **(B)** 0.1 **(C)** 1

(D) 1 order of magnitude

Table 7.6: Item statistics and response frequency by percentage for items 7/8 as administered in anatomy and physiology.

	Item 7 interview (n = 20)	Item 7 Class-wide (n = 538)	Item 8 interview (n = 20)	Item 8 Class-wide (n = 538)
Difficulty	.45	.414	.5	.377
Discrimination	.667	.548	.667	.622
Percent chosen A	30	38.8	30	35.7
Percent chosen B	20	16.7	15	22.9
Percent chosen C	5	3	5	3.7
Percent chosen D	45	41.4	50	37.7

Looking at the item statistics for each of these items in **Table 7.6**, it appears performance on both of these items is similar and might lead to the assumption that students were consistent in their answer selections (i.e. students selecting 10^1 in item 7 choosing 10^{-1} in item 8). However, upon further analysis, 40% of the students (class-wide) who were incorrect on the first question changed to a correct response on the second question. Response process interviews provide a possible explanation for this observation based on the common strategies used by students who answered item 7 incorrectly.

Anatomy and physiology student responding to item 7: ***"10^0 is 1, 10^1 is 10, 10^2 is 100 and to get from 10^1 to 10^2 you have to times your answer by 10 so I'm going to say it's 10"***

Anatomy and physiology student responding to item 7: ***"10^0 is just I think that's just 10. I don't remember. Okay 10^1 should just be um, should just be 10 and then this is 100 and this is 1000. So um, to like check the order of magnitude I think you just take the biggest one and divide whatever is before that one. So the answer is 10."***

If the student uses the same strategy in item 8 and attempts to find the distractor which relates to the common multiplier between the values on the number line, the student does not find the "correct" value. By failing to include "10" as a distractor for the item, the student using this strategy is inadvertently given an advantage over other students not using this strategy. This is further evidenced by the fact that of the 60 students who changed to selecting "1 order of magnitude" in item 8, all but 4 of them chose "10" or "10^1" in item 7.

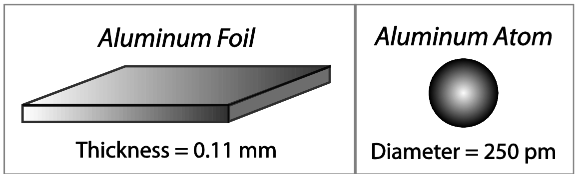
Anatomy and physiology student responding to item 8: ***"That would be 10. No, that would be 0.1. Cause if you were to take 0.001 and you were to multiply that by 0.1 [doing math on calculator] No, it would have to be 1. If you were to take 0.01 and multiply that by 1 [doing math on calculator] No, what am I talking about? 10^{-1} [converting 10^{-1} to decimal notation] so the decimal would be here. Well it would be…so it would be an order of magnitude then cause A [10^{-1}] and B [0.1] would be equivalent to each other and that doesn't give you the right answer. C [1] doesn't give you a right answer so it would have to be D."***

Two additional items in each course (**Figure 7.4**) were found to exhibit a threat to the validity of the assessment due to number matching. In each case, a student was able to discern the correct multiple-choice response through selection of the only distractor containing a certain number sequence. For example, in chemistry, item 30 featured only one distractor containing two “4”s in succession to one another. As demonstrated by the student excerpt below, If a student did not convert or incorrectly converted either value given in the problem but still correctly divided the thickness of the foil by the diameter of the atom, the student was able to identify the correct choice simply by matching the “44” of their calculated answer with the answer choices.

Chemistry student responding to item 30: ***“So the thickness in atoms. You’d have to do conversion of 250 picometers to millimeters and that’s well 100 picometers is*** 10^{-9} ***millimeters so*** $2.5\text{x}10^{-10}$ ***I think and then I would have to calculate .11 millimeters divided by*** $2.5\text{x}10^{-10}$ ***and I get 44 hundred million, but that’s not up there so I did something wrong but since it is 44 and then a string of zeros I will just do 440,000.”***

Figure 7.4: Items 30/31 of the Scale Literacy Skills Test

30. A sheet of aluminum foil has the thickness as shown in the figure. What is the thickness of the sheet in aluminum atoms?



(A) 28 atoms **(B)** 2300 atoms

(C) 440 000 atoms **(D)** 2 300 000 atoms

31. A book has a thickness of 0.0235 m (not counting the cover) with paper which has an individual thickness of 57 µm. How many sheets of paper are in the book?

(A) 1.3 sheets **(B)** 410 sheets

(C) 2400 sheets **(D)** 750 000 sheets

Comparable results emerged for item 31 in anatomy and physiology where students were able to use a similar strategy to correctly select choice B through number matching. As demonstrated by the student interview excerpt for this example, the student was able to correctly identify choice B as the correct answer through matching the “41” from their calculated value with the response choices.

Anatomy and Physiology student responding to item 30: ***“I’m gonna divide um .0235…yeah gonna divide .0235 by 57 [doing work on calculator] which gives the answer of .0041 approximately and the only one that’s close to that is [choice] B, 410.”***

7.3.3 Treatment of Identified Threats

As student performance on the scale literacy skills test has been used to predict student performance on the final exams used in both general chemistry I and II, upholding the validity of the claims made based on student performance on this assessment is of critical importance. Therefore, performance sub-scores with the five identified threats removed were calculated for the scale literacy skills test for the fall 2011 and fall 2012 general chemistry I control groups. Using a procedure developed by Meng, Rosenthal, and Rubin[4], comparison of student performance on the assessment to performance on other course proficiency measures showed a non-significant difference between the computed correlations (**Table 7.7**). These results provide evidence that exclusion of these 5 items from analysis of the assessment is not necessary and that the predictive model from which these scores were built is still valid. As this research is only in its preliminary stage in anatomy and physiology, the four items identified through response process will be considered further as necessary.

Table 7.7: Meng's Test of Correlated Correlation Coefficients

Correlated items	r (Pearson product moment coefficient)	Z	*p*
SLST (all items), paired final	0.582	0.644	0.520
SLST (threats removed), paired final	0.569		
SLST (all items), conceptual final	0.651	0.277	0.782
SLST (threats removed), conceptual final	0.645		

7.4 Limitations

The response process data collected for the scale literary skills test demonstrated that response validity exists for 80% of the items in chemistry and for 64% of the items in anatomy and physiology. While this difference could be accounted for by the relatively small sample of students interviewed from the same semester of anatomy and physiology, another explanation for this observation could be due to differing levels of domain specific content knowledge held by students within the chemistry and biological science disciplines. As these items were written for an assessment with chemistry students in mind, these items might hold an inherent bias towards knowledge thought to be held by chemistry students.

7.5 Conclusions

The response process data collected during these studies identified five items in chemistry and four items in anatomy and physiology that are threats to the validity of the scale literacy skills test. Two out of the five items identified in chemistry that were changed prior to the response process study in anatomy and physiology were revealed to no longer pose a threat in the new discipline. Two of the remaining three items were written in such a way that changing the stem or distractors was not likely to eliminate the threat and were anticipated to remain a threat in the new discipline. Interestingly, the four items identified to pose a threat to the assessment in anatomy and physiology are unique to those identified in chemistry.

7.5 References

1. Gerlach, K.; Trate, J.; Blecking, A.; Geissinger, P.; Murphy, K., Valid and Reliable Assessments to Measure Scale Literacy of Students in Introductory College Chemistry Courses. *Journal of Chemical Education.* **2014**, *91*, 1538-1545.

2. Arjoon, J. A.; Xu, X.; Lewis, J. E., Understanding the state of the art for measurement in chemistry education research: examining the psychometric evidence. *Journal of Chemical Education.* **2013**, *90*, 536-545.

3. Schwartz, P.; Barbera, J., Evaluating the Content and Response Process Validity of Data from the Chemical Concepts Inventory. *Journal of Chemical Education.* **2014**, *91*, 630-640.

4. Meng, X.; Rosenthal, R.; Rubin, D. B., Comparing Correlated Correlation Coefficients. *Psychological bulletin.* **1992**, *111*, 172-175.

Chapter 8: Usability studies of scale themed active learning lecture activities

8.1 Introduction

As described in sections 3.4.4 and 4.3.5, as part of an on-going research study into the efficacy of instructional approaches in teaching scale to undergraduate general chemistry I and II students, a sequence of active learning lecture activities were developed. The intended use of these activities was to provide students with a comprehensive outline of chemistry content covered throughout each lecture period and to provide more opportunities for engagement with course material than what is available during a didactic lecture. However, as the research team was intimately aware, the intended use and perceived benefit of the activities to students by the research team and the intended use and perceived benefit of the activities to students by the students themselves were not always in alignment. For that reason, a series of usability studies were designed to judge both the effectiveness and reception of the active learning lecture activities in both courses in which the activities were being used.

8.2 Methods

8.2.1 General methods

Students enrolled in general chemistry II were solicited for participation in the active learning lecture activity usability studies during the spring 2016 semester and students enrolled in general chemistry I were solicited during the fall 2016 semester. As described in **Table 8.1**, students were either selected to complete the usability study on paper (half of those selected for general chemistry II and all of those selected for general chemistry I) or electronically using an eye-tracking instrument (half of those selected for general chemistry II).

Table 8.1: Usability Study Participants

Course	General chemistry II	General Chemistry I
Semester	Spring 2016	Fall 2016
Electronically	9	--
On paper	8	17

Students selected to complete the study on paper were seated at a table with a copy of the entire activity and a member of the research team and students selected to complete the study electronically were seated at an eye-tracking instrument and presented with each page of the usability study packet electronically. The electronic stimuli were identical to the pages of the paper usability study packet. These students were also given a packet of "work space" containing unlabeled, uniformly sized blank spaces so as not to cue students to an expected written response or an expected type or length of response.

Students in both groups were instructed that while completion of the activity during the interview was not meant to mimic the exact environment of how a student would complete the activity during a lecture period, that completion of the packet was also not meant to feel like a test or a quiz. As the intent of the usability study was not to discern the level of content knowledge of the student, students were informed they were free to use any resource they wished, including asking questions of the researcher. All interviews were audio and video recorded for later analysis.

8.2.2 Design of usability study

As the goal of the usability study interviews were to discern areas of the lecture activities in need of refinement, the design of the usability study packets was guided by the following research questions:

1. Are the prompts enough? Do students understand what they are supposed to be doing?
2. Do students have enough expertise to make the requested drawings and make meaning of them?

To answer these questions, the study packets were designed to include several types of items requiring students to interpret information, make drawings, and apply knowledge between related items. The content areas chosen to be covered on the study packets were specifically selected to overlap between courses and included aqueous solutions and intermolecular forces for general chemistry I and intermolecular forces, colligative properties, and properties of solutions for general chemistry II. The timing of the interviews was scheduled to align with the end of the semester (but prior to taking the final exam) so that no new content was presented to the student through the activity and that participants would be familiar with the structure and features of the lecture activities. As the content covered in the activity spanned a large portion of content covered in the course, the usability studies also served as a useful review leading up to the final exam. The member of the research team present in the room took detailed notes regarding the questions, comments, and actions of the student as they worked through the activity.

8.3 Results and Discussion

8.3.1 Results of electronic usability study interviews

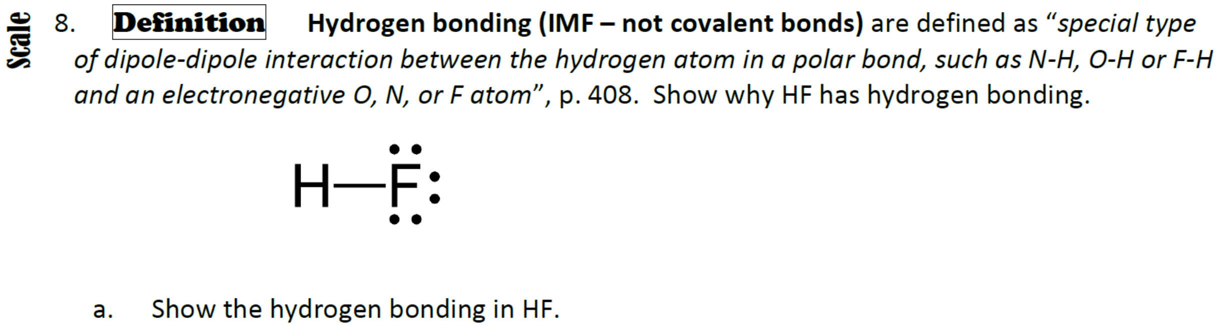
While completion of the usability study using an eye-tracking instrument could have provided valuable information regarding how students interact with the features of the lecture activities, results of the eye-tracking study showed that electronic presentation of the active learning lecture activities created too much variation in how students approached the activity. Most noticeably, the electronic stimulus did not require students to hold themselves accountable to answering each question and students often only responded verbally with little detail. For example, the first item presented to all students from both courses asked students to model a solid, liquid, and gas on the particle level. Of the 9 students who completed the activity electronically, only 5 drew a picture on the provided answer sheet, with the other 4 opting instead to verbally describe what each picture would look like. Of the remaining 25 participants from both courses who completed the activity on paper, all but 1 made a drawing. Furthermore, of the 30 items and 30 provided blank work spaces given, the most any one student used during electronic completion of the activity was 13 with a mean of 6.5 blanks used. By comparison, no student who completed the activity on paper left more than 3 items without a written answer. As this observation more closely replicates the completion of these activities in a lecture setting, it was decided that no usability study interviews with an electronic stimulus would be conducted for general chemistry I students.

8.3.2. Results of paper usability study interviews

Student interpretation of prompts and cues embedded in the activities

In response to research question 1, the usability study interviews demonstrated that a disconnect often existed between what the student thought an item was asking them to do and what the author of the item intended for the item to ask students to do. For example, items 8 and 8a (**Figure 8.1**) first asked students to show why HF has hydrogen bonding before being asked to show the hydrogen bonding in HF. For this item, students often interpreted item 8 to mean "show the hydrogen bonding" and drew a single molecule of H-F and identified the covalent bond between the H atom and F atom as a hydrogen bond. Students would comment "*HF has hydrogen bonding because it has a bond to hydrogen*". While the finding of this common misconception regarding the difference between an inter-and intramolecular force is not surprising, these items were written purposefully to help lead students to the identification of hydrogen bonding as an intermolecular force through the interaction of the "positive end" of one molecule with the "negative end" of another. Ignoring the fact that item 8a was asking students to draw hydrogen bonding and that item 8 must have been asking a different question, students would simply read item 8a and comment "*oh, that's what I just did*" and continue to the next item.

Figure 8.1: Items 8 and 8a of the usability study packet



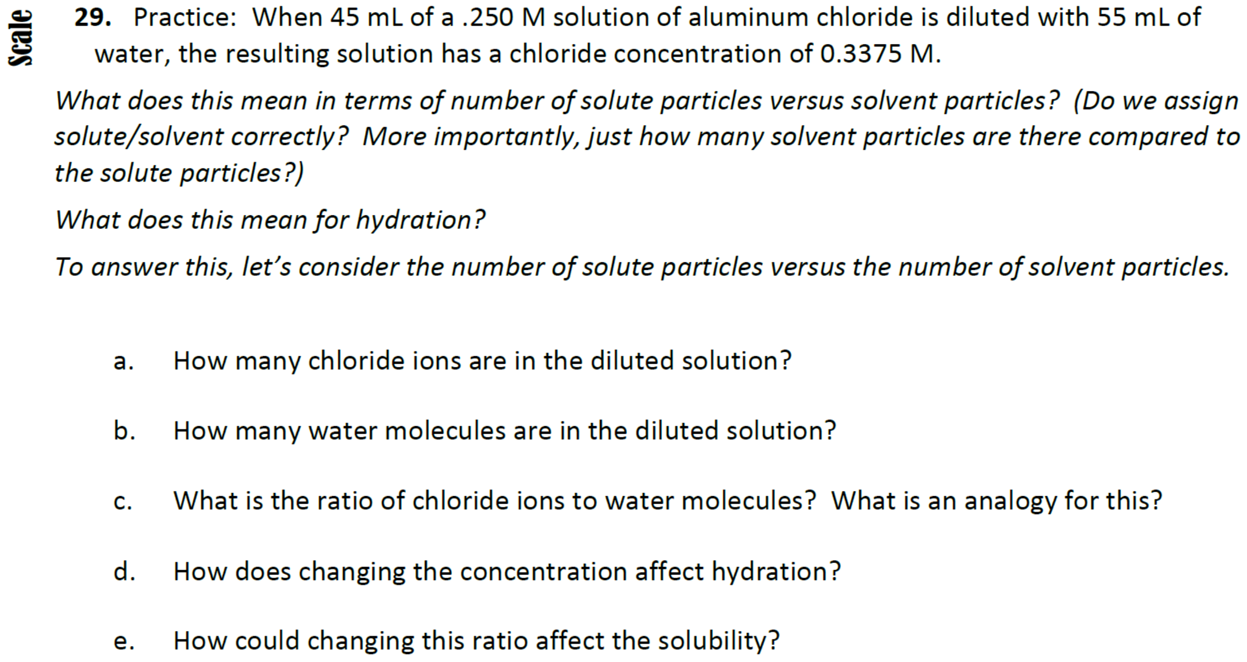
Scale

8. **Definition** **Hydrogen bonding (IMF – not covalent bonds)** are defined as "*special type of dipole-dipole interaction between the hydrogen atom in a polar bond, such as N-H, O-H or F-H and an electronegative O, N, or F atom*", p. 408. Show why HF has hydrogen bonding.

a. Show the hydrogen bonding in HF.

This observation was replicated throughout the activity as students frequently did not use the other items in the activity as contextual references. In one particularly interesting example, shown in **Figure 8.2**, in which students were given a worked dilution problem and asked a series of follow up questions regarding the solute and solvent particles, all students required confirmation that the worked portion of the problem did not require any additional work on their part, with several completing calculations to verify that the given information was correct. It was only after expressing this confusion and receiving verbal confirmation that the information given in this problem would be used in the upcoming items that students would continue in the activity. After reading the italic print in which the intent of the item was revealed, students would frequently attempt to answer these questions with generic statements such as the identification of a solvent being in larger quantity than a solute. Only after attempting to answer the questions given in italic print would the student continue down the page and see that these questions had a clear intention.

Figure 8.2: Item 29 of the usability study packet



29. Practice: When 45 mL of a .250 M solution of aluminum chloride is diluted with 55 mL of water, the resulting solution has a chloride concentration of 0.3375 M.

What does this mean in terms of number of solute particles versus solvent particles? (Do we assign solute/solvent correctly? More importantly, just how many solvent particles are there compared to the solute particles?)

What does this mean for hydration?

To answer this, let's consider the number of solute particles versus the number of solvent particles.

a. How many chloride ions are in the diluted solution?

b. How many water molecules are in the diluted solution?

c. What is the ratio of chloride ions to water molecules? What is an analogy for this?

d. How does changing the concentration affect hydration?

e. How could changing this ratio affect the solubility?

These observations suggest that integrity of items such as those in **Figure 8.2** may be lost if students are not explicitly taught to look for cues that additional information that can guide them to understanding given tasks might appear in other places within the activity book. This is especially evident given that this particular item of the activity was purposefully designed so that the entirety of this item would appear on a single page and that italic print was consistently used throughout the activities as a way to link related items or give context to an item. As the active learning lecture activities are purchased by students in a bound, double-sided activity book it is possible that other similar types of multi-part items span multiple pages and important information linking those items is missed by students.

Demonstrated ability on tasks

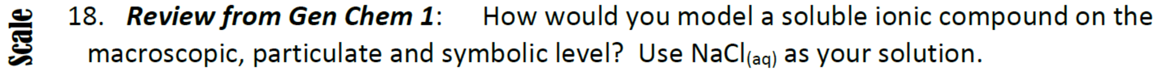
In response to research question 2, students demonstrated only a very low level of ability when asked to make drawings or interpret given information. For example, the only item in which almost all (96%) students correctly made a drawing was a review item in which students were asked to model a solid, liquid, and gas on a particle level. As this type of generic drawing was seen and referenced numerous times throughout both courses, students were very quickly able to recall and recreate this when asked. In comparison, when asked to model a solution of sodium chloride on the macroscopic and particle levels, only 7 out of 16 (44%) general chemistry I students and 2 out 8 (25%) general chemistry II students did so correctly. When asked to do the same for an aqueous methanol solution, no general chemistry I student and only 3 out of 8 (38%) general chemistry II students did so correctly. It cannot be determined from this work if these students actually lack the required knowledge to make these drawings or because it was prompted to them in a low stakes environment and students were less motivated to fully engage with the task.

8.3.3 Identifying Misconceptions

While the primary objective of the interview set was to determine the usability of the active learning lecture activities, a secondary objective emerged during the interviews to identify student misconceptions related to scaling concepts and the content embedded in the activities.

As a lot of emphasis was placed on connecting the macroscopic and particulate levels of representations it was both surprising and discouraging to find that while students could clearly distinguish between a macroscopic and particulate representation, a great deal of confusion surrounded the features of a particle level representation itself. For example, in item 18, shown in **Figure 8.3**, general chemistry II students were asked to model an aqueous NaCl solution on the macroscopic, particulate, and symbolic levels. Ignoring the accuracy of the chemistry shown in the drawn representations themselves, while all students were able to correctly make a macroscopic level representation of an aqueous solution (i.e. all students showed a container with liquid inside with no visible particles), no student was able to correctly make both symbolic or particle level representations of an aqueous solution (i.e.no student wrote a balanced equation for the symbolic and represented the solution using particles). All students demonstrated varying levels of confusion regarding the features that would be shown in each of the symbolic and particle level representations such as the common response shown in **Figure 8.4**. This confusion was even explicitly expressed by students when prompted to explain the drawings they had made.

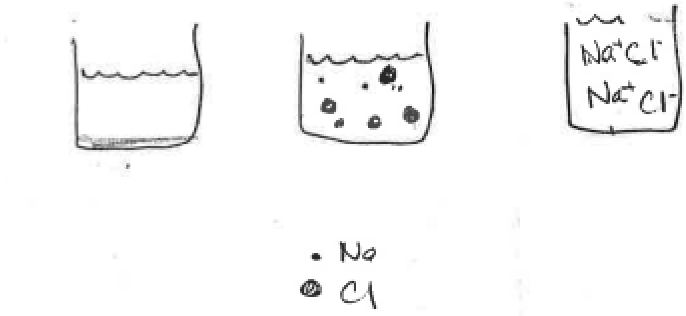
Figure 3: Item 18 of the general chemistry II usability study packet



Scale

18. ***Review from Gen Chem 1***: How would you model a soluble ionic compound on the macroscopic, particulate and symbolic level? Use $NaCl_{(aq)}$ as your solution.

Figure 4: Student macroscopic, particulate, and symbolic drawing for an aqueous solution of NaCl and explanation of drawing.



Student: ***"When we do these in lecture right now it's usually just the macroscopic and particle [representations] so it's nice that the symbolic is on there because we usually interchange particle with symbolic."***

Interviewer: ***"Based on the drawings that you've made, what is the difference between the symbolic and the particle level?"***

Student: ***"When you do the particle level there's usually more room to put more dots, um otherwise there's not much difference"***

More specifically within the chemistry content, another misconception that emerged during the interviews with both classes of students was related to students' understanding of how the observed physical properties (function) of an element or molecule, such as boiling point, can be explained by the structure of that element or molecule on the particle level. More specifically, students in this study demonstrated a belief that while the boiling point of a group of nonpolar molecules could be explained by the number of electrons, that the boiling point of a group of polar molecules could be explained by relative differences in electronegativity. **Figure 8.5** shows two related items of the usability study packet which ask students to identify the relationship between both dispersion forces and boiling point and dipole forces and boiling point and then to use tables of observed data to confirm these relationships.

Figure 8.5: Items 5/5a and 7/7a of the lecture activity usability study

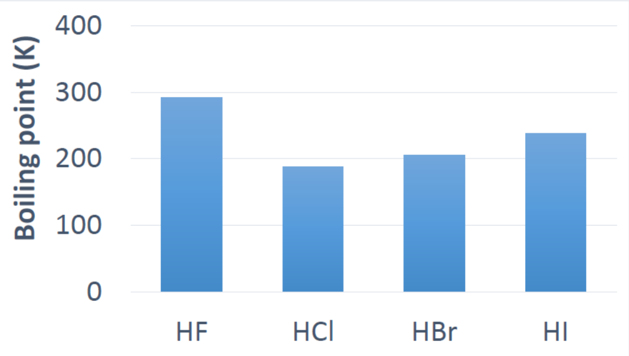
Scale 5. What is the relationship between dispersion forces and boiling point?

a. Is this relationship confirmed with the states of the halogens under standard conditions?

Halogen	State	Total # of electrons
F_2	g	18
Cl_2	g	34
Br_2	l	70
I_2	s	108

Scale 7. What is the relationship between dipole forces and boiling point?

a. Is this consistent with the boiling points of the hydrogen halides? Why or why not?

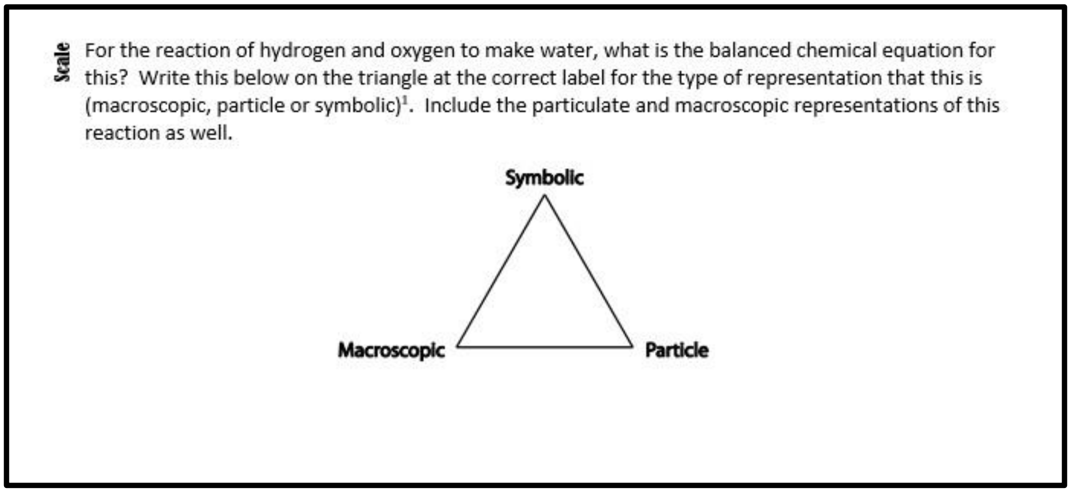


For items 5 and 5a, all students would correctly state that increasing dispersion forces would increase the observed boiling point of a molecule and that this trend was confirmed by the increasing number of electrons in each molecule. However, when asked how boiling point was represented in the table, several students were unable to identify how the molecule's state of matter predicted boiling point. For items 7 and 7a, students once again correctly identified that as dipole forces increase, the observed boiling point of a compound would also increase. However, when asked to interpret the observed boiling points of the hydrogen halides in the context of this relationship, students stated that while HF followed an expected trend having the highest observed boiling point, the other hydrogen halides did not. Students explained that as polarity was determined based upon differences in electronegativity, increasing electronegativity from iodine to chlorine should increase the observed boiling point, contradicting the observed boiling points for HCl, HBr, and HI. Even more puzzling is the fact that when questioned about how an increasing number of electrons affects dispersion versus dipole forces, students reiterated the idea that "*everything has dispersion forces*" but that at a certain point the effects of other factors such as electronegativity, take precedence over those seen from increasing number of electrons. The confusion exhibited by students here could perhaps be explained by incomplete knowledge related to periodic trends, such as the fact that reported electronegativity values are relative values while the number of electrons is absolute. This observation might make students more likely to interpret differences in electronegativity along an interval scale that would indicate a linear progression of "strength of dipole forces", regardless of whether the difference manifests itself in a deviation from expected behavior.

8.4 Guiding Instruction

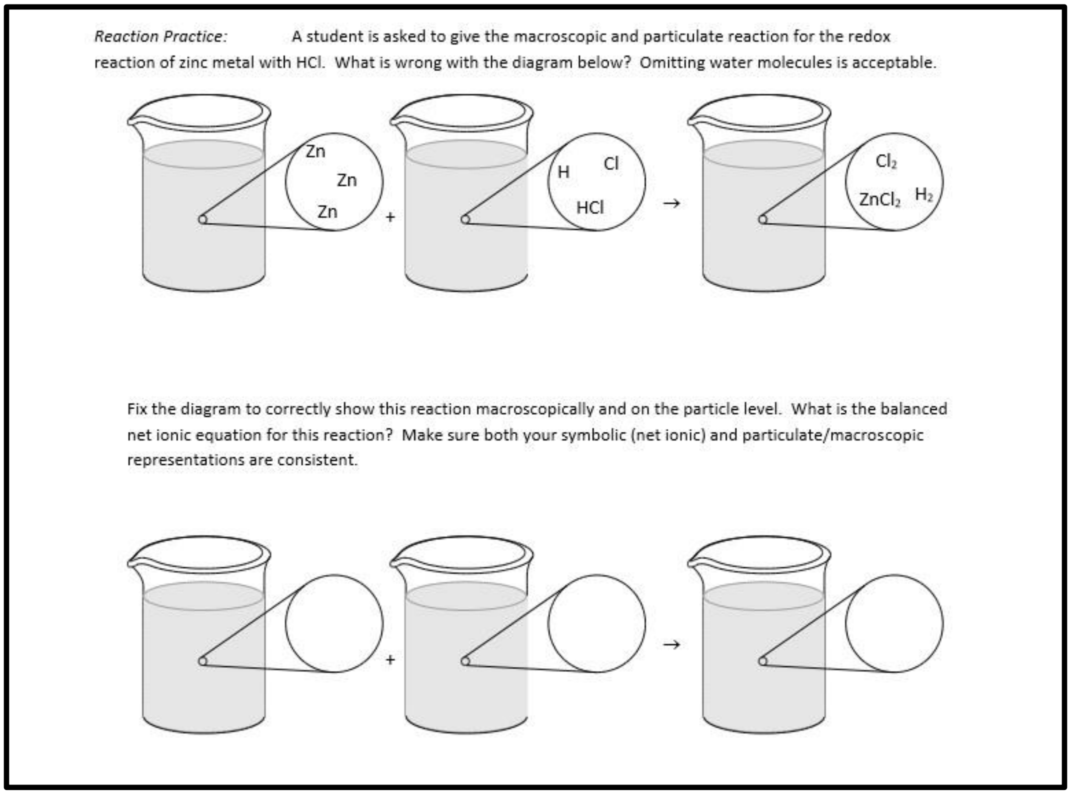
The usability study interviews described in the previous sections provided key insight into the implementation of scale-themed curriculum into lecture. Most importantly, the interviews identified areas of the lecture activities in need of refinement, identified instructional targets, and guided the development of a supplemental instruction module for general chemistry II. Within the lecture activities themselves, inclusion of items such as those used in general chemistry I and shown in **Figure 8.6**, could help alleviate some of the confusion students demonstrate regarding what constitutes a symbolic and particulate representation and what information is conveyed at each level of representation.

Figure 8.6: Items from general chemistry I active learning lecture activities



Scale

For the reaction of hydrogen and oxygen to make water, what is the balanced chemical equation for this? Write this below on the triangle at the correct label for the type of representation that this is (macroscopic, particle or symbolic)[1]. Include the particulate and macroscopic representations of this reaction as well.



Reaction Practice: A student is asked to give the macroscopic and particulate reaction for the redox reaction of zinc metal with HCl. What is wrong with the diagram below? Omitting water molecules is acceptable.

Fix the diagram to correctly show this reaction macroscopically and on the particle level. What is the balanced net ionic equation for this reaction? Make sure both your symbolic (net ionic) and particulate/macroscopic representations are consistent.

Relating to instructional targets, the interviews provided valuable information related to how students interpret the common language used to teach chemistry concepts and how the use of this language can lead students to develop incomplete knowledge about concepts and subsequently develop misconceptions. For example, in item 8a the definition given for hydrogen bonding should more clearly specify that two or more molecules are required to model hydrogen bonding.

Even though both items 8 and 8a refer only to “HF”, item 8 can be answered showing a single molecule while item 8a cannot. While experienced chemistry students might understand the nuances of the language used within these items, a beginning student likely does not. Lastly, based on the low level of proficiency demonstrated by students related to maneuvering between different levels of representation, and connecting macroscopic observations with particle level properties (specifically modeling solutions and understanding the role of intermolecular forces), supplemental instruction modules that mirror those offered in general chemistry I featuring these concepts (see Appendix D.1 for an outline of these activities) were designed and integrated into the general chemistry II curriculum.

Appendix A: Scale Assessments

- Full correlation matrix of course measures used in general chemistry I
- General Chemistry I Scale Literacy Skills Test pre-administration item statistics
- General Chemistry I Scale Literacy Skills Test post-administration item statistics
- General Chemistry I Scale Concept Inventory pre-administration item statistics
- General Chemistry I Scale Concept Inventory test post-administration item statistics
- General Chemistry I Laboratory Survey pre-administration item statistics
- General Chemistry I Laboratory Survey post-administration item statistics
- General Chemistry I Laboratory quiz items

- Full correlation matrix of course measures used in general chemistry II
- General Chemistry II Scale Literacy Skills Test pre-administration item statistics
- General Chemistry II Scale Literacy Skills Test post-administration item statistics
- General Chemistry II Scale Concept Inventory pre-administration item statistics
- General Chemistry II Scale Concept Inventory test post-administration item statistics
- General Chemistry II Laboratory Survey pre-administration item statistics
- General Chemistry II Laboratory Survey post-administration item statistics
- General Chemistry II Laboratory quiz items

Figure A.1 Correlation matrix of all general chemistry I course measures

n = 1092		ACT COMP	ACT MATH	ACT SCI&R	Math Placement	Chemistry Placement	Combined Placement	SLST Pre	SCI Pre	Scale Literacy	Paired Final	Conceptual Final	Course percent
ACT COMP	Pearson Correlation	1	.767**	.820**	.457**	.423**	.513**	.489**	.299**	.498**	.526**	.533**	.418**
ACT MATH	Pearson Correlation	.767**	1	.625**	.538**	.399**	.529**	.505**	.240**	.492**	.516**	.508**	.435**
ACT SCI&R	Pearson Correlation	.820**	.625**	1	.392**	.377**	.450**	.431**	.291**	.448**	.454**	.469**	.359**
Math Placement	Pearson Correlation	.457**	.538**	.392**	1	.402**	.730**	.474**	.236**	.465**	.431**	.446**	.406**
Chemistry Placement	Pearson Correlation	.423**	.399**	.377**	.402**	1	.919**	.443**	.290**	.458**	.486**	.507**	.447**
Combined Placement	Pearson Correlation	.513**	.529**	.450**	.730**	.919**	1	.535**	.317**	.542**	.548**	.570**	.508**
SLST Pre	Pearson Correlation	.489**	.505**	.431**	.474**	.443**	.535**	1	.414**	.954**	.514**	.588**	.430**
SCI Pre	Pearson Correlation	.299**	.240**	.291**	.236**	.290**	.317**	.414**	1	.667**	.332**	.394**	.236**
Scale Literacy	Pearson Correlation	.498**	.492**	.448**	.465**	.458**	.542**	.954**	.667**	1	.530**	.611**	.429**
Paired Final	Pearson Correlation	.526**	.516**	.454**	.431**	.486**	.548**	.514**	.332**	.530**	1	.767**	.803**
Conceptual Final	Pearson Correlation	.533**	.508**	.469**	.446**	.507**	.570**	.588**	.394**	.611**	.767**	1	.747**
Course Percent	Pearson Correlation	.418**	.435**	.359**	.406**	.447**	.508**	.430**	.236**	.429**	.803**	.747**	1

** All correlations significant at <.001

Table A.1 General chemistry I Scale Literacy Skills Test (pre-administration) Item Statistics n = 2034

Item	Key	DIF	Discr	%A	%B	%C	%D	Attr A	Attr B	Attr C	Attr D
1	**C**	0.916	0.191	2.6	4.3	91.6	1.5	-0.06	-0.10	0.19	-0.03
2	**B**	0.544	0.464	32.3	54.4	8.2	5.1	-0.22	0.46	-0.17	-0.07
3	**A**	0.688	0.454	68.8	21.8	5.9	3.5	0.45	-0.29	-0.11	-0.06
4	**A**	0.357	0.505	35.7	57.4	3.2	3.7	0.50	-0.39	-0.06	-0.06
5	**A**	0.219	0.470	21.9	17.7	4.0	56.5	0.47	0.04	0.01	-0.51
6	**A**	0.505	0.582	50.5	38.9	4.7	5.8	0.58	-0.44	-0.03	-0.11
7	**D**	0.492	0.627	1.4	15.8	33.7	49.2	-0.04	-0.20	-0.38	0.63
8	**D**	0.476	0.654	38.2	12.8	1.4	47.6	-0.36	-0.27	-0.03	0.65
9	**D**	0.690	0.446	5.3	24.7	1.1	69.0	-0.08	-0.33	-0.03	0.45
10	**C**	0.487	0.585	35.0	2.8	48.7	13.4	-0.47	-0.06	0.59	-0.05
11	**A**	0.708	0.391	70.8	4.0	22.5	2.6	0.39	-0.10	-0.26	-0.04
12	**B**	0.774	0.458	1.3	77.4	11.4	9.8	-0.04	0.46	-0.23	-0.19
13	**B**	0.769	0.334	12.5	76.9	5.8	4.8	-0.13	0.33	-0.11	-0.10
14	**B**	0.416	0.432	41.8	41.6	2.9	13.7	-0.14	0.43	-0.07	-0.22
15	**D**	0.565	0.454	16.1	8.0	19.4	56.5	-0.18	-0.17	-0.10	0.45
16	**C**	0.507	0.477	9.6	17.7	50.7	22.1	-0.15	-0.28	0.48	-0.05
17	**B**	0.256	0.369	17.2	25.6	52.0	5.2	-0.18	0.37	-0.20	0.02
18	**B**	0.306	0.409	8.9	30.6	34.2	26.2	-0.05	0.41	-0.22	-0.14
19	**A**	0.452	0.428	45.2	20.2	23.6	11.0	0.43	-0.09	-0.24	-0.10
20	**B**	0.641	0.420	17.4	64.1	17.0	1.5	-0.18	0.42	-0.22	-0.02
21	**D**	0.507	0.405	17.2	12.1	19.9	50.7	-0.12	-0.10	-0.18	0.40
22	**C**	0.672	0.367	9.6	16.8	67.2	6.3	-0.12	-0.16	0.37	-0.09
23	**B**	0.507	0.456	33.0	50.7	9.0	7.4	-0.21	0.46	-0.11	-0.13
24	**A**	0.893	0.104	89.3	1.8	2.8	6.1	0.10	-0.02	-0.06	-0.02
25	**C**	0.065	0.014	4.6	80.9	6.5	8.0	-0.04	0.01	0.01	0.01
26	**B**	0.060	0.043	3.0	6.0	65.4	25.5	-0.02	0.04	0.04	-0.06
27	**D**	0.736	0.204	1.2	12.5	12.7	73.6	-0.03	-0.03	-0.15	0.20
28	**C**	0.568	0.509	38.0	3.2	56.8	1.9	-0.38	-0.08	0.51	-0.04
29	**C**	0.554	0.525	13.3	27.4	55.4	4.0	-0.25	-0.24	0.52	-0.04
30	**C**	0.511	0.544	4.4	20.1	51.1	24.4	-0.09	-0.40	0.54	-0.05
31	**B**	0.768	0.383	4.1	76.8	16.6	2.6	-0.10	0.38	-0.25	-0.04
32	**A**	0.539	0.096	53.9	43.3	1.9	0.9	0.10	-0.03	-0.05	-0.02
33	**C**	0.478	0.415	0.9	18.2	47.8	33.1	-0.02	-0.24	0.41	-0.16
34	**C**	0.201	0.126	9.1	24.8	20.1	46.0	-0.15	0.00	0.13	0.02
35	**A**	0.205	0.008	20.5	30.2	23.1	26.2	0.01	0.14	-0.02	-0.13
36	**C**	0.513	0.460	4.8	24.3	51.3	19.6	-0.08	-0.39	0.46	0.01
37	**C**	0.910	0.185	1.7	4.3	91.0	3.0	-0.05	-0.11	0.18	-0.02
38	**B**	0.899	0.208	1.3	89.9	3.8	5.0	-0.03	0.21	-0.09	-0.08
39	**A**	0.767	0.346	76.7	12.3	6.6	4.4	0.35	-0.12	-0.14	-0.09
40	**B**	0.714	0.424	4.1	71.4	18.3	6.2	-0.07	0.42	-0.26	-0.09
41	**B**	0.870	0.279	0.5	87.0	2.1	10.4	-0.02	0.28	-0.06	-0.20
42	**A**	0.641	0.320	64.1	12.3	17.9	5.7	0.32	-0.22	-0.05	-0.06
43	**A**	0.480	0.269	48.0	47.9	3.1	1.0	0.27	-0.19	-0.06	-0.02
44	**A**	0.721	0.234	72.1	5.4	13.0	9.5	0.23	-0.05	-0.14	-0.05
45	**A**	0.464	0.194	46.4	7.7	25.2	20.6	0.19	-0.10	0.00	-0.10

Table A.2 General chemistry I Scale Literacy Skills Test (post-administration) Item Statistics n= 1419

Item	Key	DIF	Discr	%A	%B	%C	%D	AttrA	AttrB	AttrC	AttrD
1	**C**	0.923	0.189	2.7	3.5	92.3	1.6	-0.08	-0.08	0.19	-0.03
2	**B**	0.654	0.465	28.5	65.4	3.9	2.3	-0.31	0.46	-0.09	-0.06
3	**A**	0.846	0.369	84.6	8.9	4.3	2.2	0.37	-0.22	-0.10	-0.05
4	**A**	0.436	0.563	43.6	49.3	3.9	3.2	0.56	-0.44	-0.08	-0.04
5	**A**	0.311	0.586	31.1	17.5	3.6	47.8	0.59	-0.02	-0.03	-0.54
6	**A**	0.551	0.592	55.1	35.3	3.5	6.1	0.59	-0.41	-0.08	-0.11
7	**D**	0.677	0.552	1.3	12.5	18.6	67.7	-0.04	-0.21	-0.30	0.55
8	**D**	0.634	0.654	24.4	10.9	1.3	63.4	-0.34	-0.28	-0.04	0.65
9	**D**	0.704	0.485	5.1	23.6	0.8	70.4	-0.11	-0.35	-0.02	0.48
10	**C**	0.530	0.639	31.6	3.3	53.0	12.1	-0.48	-0.09	0.64	-0.08
11	**A**	0.797	0.423	79.7	4.2	14.3	1.8	0.42	-0.13	-0.26	-0.04
12	**B**	0.789	0.468	2.1	78.9	12.9	6.1	-0.08	0.47	-0.27	-0.12
13	**B**	0.819	0.349	12.5	81.9	3.8	1.8	-0.20	0.35	-0.10	-0.06
14	**B**	0.455	0.414	43.8	45.5	4.1	6.6	-0.14	0.41	-0.11	-0.16
15	**D**	0.553	0.538	11.1	8.2	25.4	55.3	-0.15	-0.19	-0.19	0.54
16	**C**	0.540	0.532	7.8	18.7	54.0	19.6	-0.17	-0.30	0.53	-0.07
17	**B**	0.374	0.541	13.5	37.4	43.6	5.5	-0.19	0.54	-0.33	-0.02
18	**B**	0.459	0.555	7.5	45.9	30.8	15.9	-0.07	0.55	-0.33	-0.15
19	**A**	0.488	0.456	48.8	20.9	21.1	9.2	0.46	-0.11	-0.26	-0.08
20	**B**	0.709	0.445	17.8	70.9	11.1	0.3	-0.25	0.45	-0.19	-0.01
21	**D**	0.565	0.406	21.9	7.3	14.3	56.5	-0.14	-0.08	-0.18	0.41
22	**C**	0.789	0.310	9.3	10.4	78.9	1.3	-0.13	-0.15	0.31	-0.03
23	**B**	0.638	0.549	27.1	63.8	6.6	2.5	-0.34	0.55	-0.14	-0.06
24	**A**	0.903	0.132	90.3	2.1	2.7	4.9	0.13	-0.07	-0.06	0.00
25	**C**	0.178	0.268	3.9	70.3	17.8	8.1	-0.07	-0.19	0.27	-0.01
26	**B**	0.283	0.361	3.2	28.3	47.4	21.1	-0.07	0.36	-0.16	-0.12
27	**D**	0.693	0.273	1.8	19.3	9.5	69.3	-0.06	-0.10	-0.12	0.27
28	**C**	0.635	0.507	29.7	4.4	63.5	2.5	-0.35	-0.09	0.51	-0.06
29	**C**	0.588	0.490	9.2	26.1	58.8	5.9	-0.19	-0.26	0.49	-0.04
30	**C**	0.592	0.592	3.2	18.0	59.2	19.6	-0.08	-0.38	0.59	-0.13
31	**B**	0.784	0.425	3.6	78.4	16.1	1.9	-0.11	0.43	-0.28	-0.03
32	**A**	0.670	0.158	67.0	30.1	1.5	1.4	0.16	-0.09	-0.05	-0.01
33	**C**	0.363	0.327	1.8	9.7	36.3	52.1	-0.06	-0.14	0.33	-0.13
34	**C**	0.297	0.279	6.5	26.6	29.7	37.2	-0.12	-0.06	0.28	-0.10
35	**A**	0.254	-0.011	25.4	29.2	26.4	19.1	-0.01	0.15	-0.06	-0.08
36	**C**	0.576	0.487	5.2	24.9	57.6	12.3	-0.12	-0.36	0.49	0.00
37	**C**	0.942	0.104	1.6	1.8	94.2	2.5	-0.05	-0.04	0.10	-0.01
38	**B**	0.897	0.177	2.3	89.7	3.2	4.9	-0.07	0.18	-0.06	-0.05
39	**A**	0.815	0.330	81.5	11.0	4.9	2.6	0.33	-0.17	-0.11	-0.05
40	**B**	0.760	0.389	3.5	76.0	16.9	3.6	-0.09	0.39	-0.23	-0.07
41	**B**	0.918	0.228	0.8	91.8	2.0	5.4	-0.03	0.23	-0.06	-0.14
42	**A**	0.714	0.355	71.4	12.5	12.0	4.1	0.35	-0.22	-0.08	-0.05
43	**A**	0.738	0.301	73.8	23.6	2.0	0.6	0.30	-0.24	-0.05	-0.01
44	**A**	0.792	0.273	79.2	4.0	8.0	8.8	0.27	-0.08	-0.10	-0.10
45	**A**	0.589	0.206	58.9	5.1	20.6	15.4	0.21	-0.10	-0.02	-0.09

Table A.3 General chemistry I Scale Concept Inventory (pre-administration) Item Statistics n = 1659

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1	+	35.7	38.3	3.3	15.7	6.9	0.06	74.0	22.6
2	-	1.2	7.3	19.3	42.5	29.7	0.00	8.5	72.2
3	-	31.3	32.1	6.9	20.0	9.7	0.00	63.4	29.7
4	+	12.9	31.0	21.3	23.9	10.7	0.12	43.9	34.7
5	-	2.1	11.5	7.9	43.0	35.4	0.06	13.6	78.5
6	+	28.5	40.4	8.4	16.9	5.7	0.06	68.9	22.6
7	+	14.6	29.3	28.5	18.1	9.5	0.06	43.9	27.5
8	+	33.3	37.1	18.4	8.3	2.8	0.06	70.4	11.2
9		5.4	23.6	27.9	32.3	10.8	0.00	29.0	43.1
10	+	14.1	36.8	26.2	17.0	6.0	0.00	50.9	23.0
11	+	11.4	34.5	24.7	23.5	5.9	0.00	45.9	29.4
12	+	8.1	24.9	8.7	41.1	17.0	0.18	33.0	58.1
13	+	18.3	42.6	22.8	13.4	2.7	0.18	60.9	16.1
14	-	1.4	9.6	13.9	39.4	35.6	0.00	11.1	75.0
15	+	29.7	46.9	5.3	14.6	3.3	0.18	76.6	18.0
16	+	16.2	37.7	24.4	18.0	3.7	0.06	53.8	21.8
17	-	30.5	43.3	10.2	12.6	3.4	0.00	73.8	16.0
18		5.2	19.7	21.9	39.3	14.0	0.00	24.8	53.3
19	-	6.7	20.3	19.9	35.4	17.5	0.18	26.9	53.0
20	+	20.1	45.1	18.0	13.7	2.8	0.30	65.2	16.5
21	+	11.5	33.4	21.3	28.8	4.9	0.06	44.9	33.7
22	+	56.1	38.2	3.0	2.2	0.6	0.00	94.3	2.8
23	-	4.5	17.4	24.8	37.0	16.2	0.12	21.9	53.2
24		3.9	16.2	21.9	39.1	18.9	0.00	20.0	58.0
25	+	19.1	40.2	16.4	19.5	4.8	0.00	59.3	24.3
26	+	10.1	25.1	24.7	33.0	7.1	0.00	35.2	40.1
27	**V**	62.7	37.3	0.0	0.0	0.0	0.00	100.0	0.0
28	+	25.8	39.0	22.5	9.9	2.7	0.06	64.8	12.7
29	+	7.3	18.2	20.9	42.2	11.3	0.12	25.5	53.5
30	+	4.0	12.6	31.4	34.3	17.7	0.06	16.6	52.0
31	-	5.6	13.6	11.2	45.9	23.7	0.06	19.2	69.6
32	-	10.8	46.1	21.3	16.0	5.8	0.00	56.9	21.8
33	+	9.8	39.3	16.3	26.9	7.6	0.12	49.1	34.5
34	-	16.5	41.0	25.1	13.4	4.0	0.06	57.5	17.4
35	-	8.7	39.5	21.3	21.7	8.7	0.06	48.3	30.4
36	+	11.8	42.6	31.2	10.8	3.6	0.00	54.4	14.5
37	-	3.2	17.1	12.2	48.5	19.0	0.00	20.3	67.5
38	+	3.1	21.2	20.0	43.5	12.2	0.06	24.3	55.6
39	+	13.4	34.7	33.0	16.6	2.2	0.00	48.2	18.8
40	-	14.8	43.3	19.7	16.4	5.7	0.18	58.0	22.1

Table A.4 General chemistry I Scale Concept Inventory (post-administration) Item Statistics n = 1168

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1	+	25.6	40.6	5.6	20.2	7.8	0.26	66.2	28.0
2	-	4.5	9.8	11.3	45.1	29.1	0.26	14.2	74.2
3	-	17.9	40.7	12.1	21.5	7.7	0.17	58.6	29.2
4	+	15.0	46.2	17.6	16.0	5.0	0.17	61.2	21.0
5	-	4.1	13.5	4.8	44.2	33.3	0.09	17.6	77.5
6	+	22.0	42.9	9.0	18.8	7.2	0.17	64.9	25.9
7	+	19.6	43.1	10.4	19.0	7.7	0.26	62.7	26.7
8	+	29.0	43.0	16.5	8.0	3.3	0.17	72.0	11.3
9		3.8	21.1	22.3	42.7	9.9	0.09	24.9	52.7
10	+	13.3	43.0	22.9	16.6	4.2	0.00	56.3	20.8
11	+	12.0	34.8	25.5	23.1	4.5	0.17	46.7	27.6
12	+	5.4	26.5	12.2	44.3	11.6	0.09	31.8	55.9
13	+	13.8	54.0	18.7	11.8	1.7	0.00	67.8	13.5
14	-	1.4	9.6	11.6	40.5	36.8	0.17	11.0	77.3
15	+	32.0	52.8	4.5	8.3	2.3	0.00	84.8	10.6
16	+	15.8	46.5	19.5	15.3	2.8	0.09	62.2	18.2
17	-	25.6	48.6	8.9	12.6	4.2	0.09	74.2	16.8
18		8.8	36.4	21.7	27.6	5.6	0.00	45.2	33.1
19	-	5.1	22.7	14.5	39.4	18.2	0.26	27.7	57.5
20	+	18.8	53.9	13.1	12.8	1.5	0.09	72.6	14.2
21	+	17.6	46.9	17.4	15.6	2.5	0.09	64.5	18.1
22	+	50.1	45.0	2.7	2.1	0.2	0.00	95.1	2.2
23	-	4.9	19.5	23.7	37.2	14.7	0.00	24.4	51.9
24		5.1	15.6	16.8	46.2	16.4	0.00	20.6	62.6
25	+	18.3	42.3	15.8	20.5	3.0	0.00	60.6	23.5
26	+	12.2	36.8	22.9	24.4	3.6	0.00	49.1	28.0
27	**V**	54.4	45.6	0.0	0.0	0.0	0.00	100.0	0.0
28	+	29.1	43.6	15.1	10.1	2.1	0.00	72.7	12.2
29	+	21.4	33.4	12.8	26.5	5.8	0.09	54.8	32.4
30	+	7.1	24.7	23.2	31.0	13.8	0.17	31.8	44.8
31	-	3.6	13.7	11.9	48.5	22.2	0.09	17.3	70.7
32	-	6.9	39.8	19.1	24.2	9.9	0.00	46.7	34.2
33	+	9.8	43.0	17.3	24.5	5.5	0.00	52.7	30.0
34	-	14.6	46.1	12.2	21.5	5.6	0.09	60.7	27.1
35	-	6.6	40.2	22.1	21.7	9.5	0.00	46.7	31.2
36	+	10.4	44.9	24.3	15.7	4.6	0.00	55.4	20.3
37	-	2.4	11.5	10.4	52.6	23.1	0.00	13.9	75.7
38	+	2.8	20.9	20.5	46.7	9.1	0.00	23.7	55.8
39	+	16.2	44.0	21.4	15.8	2.7	0.00	60.2	18.4
40	-	6.5	32.6	18.7	28.8	13.2	0.26	39.1	42.0

Figure A.2 Correlation matrix of all general chemistry II course measures

N = 442		ACT COMP	ACT MATH	ACT SCI&R	Placement Test	SLST Pre	SCI Pre	Scale Literacy	Paired Final	Conceptual Final	Course percent
ACT COMP	Pearson Correlation	1	.785**	.811**	.458**	.478**	.366**	.496**	.535**	.497**	.327**
ACT MATH	Pearson Correlation	.785**	1	.626**	.485**	.480**	.324**	.482**	.544**	.444**	.365**
ACT SCI&R	Pearson Correlation	.811**	.626**	1	.428**	.441**	.336**	.457**	.476**	.437**	.307**
Placement Test	Pearson Correlation	.458**	.485**	.428**	1	.546**	.518**	.601**	.698**	.553**	.486**
SLST Pre	Pearson Correlation	.478**	.480**	.441**	.546**	1	.533**	.954**	.601**	.516**	.401**
SCI Pre	Pearson Correlation	.366**	.324**	.336**	.518**	.533**	1	.761**	.457**	.506**	.295**
Scale Literacy	Pearson Correlation	.496**	.482**	.457**	.601**	.954**	.761**	1	.622**	.574**	.411**
Paired Final	Pearson Correlation	.535**	.544**	.476**	.698**	.601**	.457**	.622**	1	.662**	.643**
Conceptual Final	Pearson Correlation	.497**	.444**	.437**	.553**	.516**	.506**	.574**	.662**	1	.678**
Course Percent	Pearson Correlation	.327**	.365**	.307**	.486**	.401**	.295**	.411**	.643**	.678**	1

** All correlations significant at <.001

Table A.5 General chemistry II Scale Literacy Skills Test (pre-administration) Item Statistics n = 740

Item	**Key**	DIF	Discr	%A	%B	%C	%D	AttrA	AttrB	AttrC	AttrD
1	**C**	0.936	0.178	2.2	3.1	93.6	1.1	-0.06	-0.09	0.18	-0.02
2	**B**	0.665	0.443	26.5	66.5	5.8	1.2	-0.30	0.44	-0.13	-0.02
3	**A**	0.803	0.432	80.3	13.2	4.5	2.0	0.43	-0.28	-0.10	-0.05
4	**A**	0.485	0.681	48.5	47.0	1.9	2.6	0.68	-0.61	-0.04	-0.03
5	**A**	0.361	0.681	36.1	17.8	3.8	42.3	0.68	-0.06	-0.03	-0.59
6	**A**	0.597	0.676	59.7	31.2	3.0	6.1	0.68	-0.49	-0.05	-0.14
7	**D**	0.703	0.514	1.2	9.9	18.6	70.3	-0.03	-0.21	-0.28	0.51
8	**D**	0.653	0.632	23.8	9.9	1.1	65.3	-0.34	-0.27	-0.03	0.63
9	**D**	0.730	0.508	4.2	21.8	1.1	73.0	-0.11	-0.36	-0.04	0.51
10	**C**	0.592	0.665	23.4	4.1	59.2	13.4	-0.45	-0.12	0.66	-0.09
11	**A**	0.785	0.465	78.5	3.8	15.3	2.4	0.46	-0.12	-0.27	-0.07
12	**B**	0.786	0.449	1.5	78.6	12.7	7.2	-0.06	0.45	-0.24	-0.15
13	**B**	0.838	0.292	10.8	83.8	3.6	1.8	-0.16	0.29	-0.08	-0.05
14	**B**	0.468	0.438	44.5	46.8	2.7	6.1	-0.20	0.44	-0.05	-0.19
15	**D**	0.584	0.530	10.8	7.7	23.1	58.4	-0.08	-0.19	-0.25	0.53
16	**C**	0.576	0.519	7.4	17.4	57.6	17.6	-0.11	-0.37	0.52	-0.04
17	**B**	0.361	0.557	13.2	36.1	43.9	6.8	-0.22	0.56	-0.37	0.03
18	**B**	0.461	0.568	8.0	46.1	31.9	14.1	-0.03	0.57	-0.34	-0.20
19	**A**	0.499	0.551	49.9	21.8	20.4	8.0	0.55	-0.23	-0.21	-0.11
20	**B**	0.715	0.443	17.3	71.5	10.8	0.4	-0.22	0.44	-0.21	-0.01
21	**D**	0.572	0.519	18.5	11.8	12.6	57.2	-0.12	-0.19	-0.21	0.52
22	**C**	0.799	0.270	8.5	8.9	79.9	2.7	-0.10	-0.10	0.27	-0.07
23	**B**	0.628	0.562	27.6	62.8	6.8	2.8	-0.37	0.56	-0.10	-0.09
24	**A**	0.897	0.059	89.7	1.8	2.7	5.8	0.06	-0.04	-0.04	0.02
25	**C**	0.158	0.243	2.7	73.1	15.8	8.4	-0.05	-0.17	0.24	-0.02
26	**B**	0.204	0.292	1.9	20.4	60.4	17.3	-0.05	0.29	-0.18	-0.06
27	**D**	0.673	0.330	1.5	20.0	11.2	67.3	-0.02	-0.14	-0.17	0.33
28	**C**	0.664	0.492	28.1	3.5	66.4	2.0	-0.38	-0.07	0.49	-0.04
29	**C**	0.562	0.578	9.2	28.5	56.2	6.1	-0.17	-0.34	0.58	-0.07
30	**C**	0.631	0.492	1.6	14.9	63.1	20.4	-0.03	-0.33	0.49	-0.13
31	**B**	0.811	0.351	2.6	81.1	14.2	2.2	-0.06	0.35	-0.24	-0.05
32	**A**	0.591	0.141	59.1	38.1	1.8	1.1	0.14	-0.09	-0.03	-0.02
33	**C**	0.428	0.351	0.7	14.3	42.8	42.2	-0.01	-0.23	0.35	-0.11
34	**C**	0.251	0.249	7.7	32.8	25.1	34.3	-0.15	-0.06	0.25	-0.04
35	**A**	0.220	0.114	22.0	27.4	34.1	16.5	0.11	0.06	-0.09	-0.09
36	**C**	0.607	0.443	4.3	22.8	60.7	12.2	-0.06	-0.35	0.44	-0.03
37	**C**	0.943	0.103	1.2	2.2	94.3	2.3	-0.03	-0.05	0.10	-0.02
38	**B**	0.903	0.189	1.4	90.3	3.8	4.6	-0.02	0.19	-0.10	-0.08
39	**A**	0.836	0.351	83.6	9.2	3.8	3.4	0.35	-0.13	-0.12	-0.10
40	**B**	0.746	0.405	3.4	74.6	17.6	4.5	-0.09	0.41	-0.24	-0.08
41	**B**	0.935	0.162	0.5	93.5	1.4	4.6	-0.02	0.16	-0.04	-0.10
42	**A**	0.714	0.351	71.4	12.6	12.0	4.1	0.35	-0.21	-0.07	-0.07
43	**A**	0.738	0.259	73.8	24.3	1.5	0.4	0.26	-0.23	-0.02	-0.01
44	**A**	0.799	0.216	79.9	3.9	8.8	7.4	0.22	-0.03	-0.14	-0.04
45	**A**	0.565	0.151	56.5	4.5	23.2	15.8	0.15	-0.08	0.03	-0.10

Table A.6 General chemistry II Scale Literacy Skills Test (post-administration) Item Statistics n = 540

Item	**Key**	DIF	Discr	%A	%B	%C	%D	AttrA	AttrB	AttrC	AttrD
1	**C**	0.937	0.185	2.6	2.8	93.7	0.9	-0.08	-0.07	0.19	-0.03
2	**B**	0.693	0.533	23.7	69.3	5.6	1.5	-0.33	0.53	-0.17	-0.03
3	**A**	0.796	0.444	79.6	14.4	4.8	1.1	0.44	-0.30	-0.13	-0.01
4	**A**	0.531	0.578	53.1	41.1	3.7	2.0	0.58	-0.47	-0.07	-0.03
5	**A**	0.404	0.733	40.4	15.9	4.3	39.4	0.73	-0.05	-0.05	-0.63
6	**A**	0.631	0.593	63.1	26.9	3.7	6.3	0.59	-0.38	-0.07	-0.15
7	**D**	0.711	0.519	1.3	12.4	15.2	71.1	-0.04	-0.27	-0.21	0.52
8	**D**	0.678	0.593	18.5	11.7	2.0	67.8	-0.26	-0.28	-0.05	0.59
9	**D**	0.746	0.556	7.2	16.9	1.3	74.6	-0.18	-0.34	-0.04	0.56
10	**C**	0.604	0.637	23.3	3.5	60.4	12.8	-0.44	-0.10	0.64	-0.10
11	**A**	0.752	0.481	75.2	7.8	15.4	1.7	0.48	-0.24	-0.22	-0.01
12	**B**	0.748	0.511	2.6	74.8	16.3	6.3	-0.07	0.51	-0.31	-0.13
13	**B**	0.837	0.333	12.2	83.7	3.5	0.6	-0.24	0.33	-0.09	-0.01
14	**B**	0.457	0.437	43.7	45.7	5.7	4.8	-0.19	0.44	-0.13	-0.12
15	**D**	0.522	0.548	9.6	11.3	26.9	52.2	-0.07	-0.24	-0.24	0.55
16	**C**	0.533	0.511	7.2	20.7	53.3	18.7	-0.14	-0.30	0.51	-0.07
17	**B**	0.409	0.430	15.0	40.9	35.6	8.5	-0.33	0.43	-0.10	0.01
18	**B**	0.430	0.578	9.6	43.0	33.1	14.3	-0.04	0.58	-0.38	-0.16
19	**A**	0.498	0.489	49.8	18.1	21.9	10.2	0.49	-0.10	-0.29	-0.10
20	**B**	0.711	0.481	18.7	71.1	10.2	0.0	-0.28	0.48	-0.20	0.00
21	**D**	0.456	0.452	23.9	12.2	18.3	45.6	-0.21	-0.09	-0.16	0.45
22	**C**	0.794	0.341	9.3	9.1	79.4	2.2	-0.13	-0.16	0.34	-0.06
23	**B**	0.617	0.600	30.7	61.7	4.4	3.1	-0.37	0.60	-0.12	-0.11
24	**A**	0.904	0.178	90.4	2.8	2.8	4.1	0.18	-0.09	-0.11	0.02
25	**C**	0.146	0.252	4.8	71.5	14.6	9.1	-0.11	-0.13	0.25	-0.01
26	**B**	0.248	0.193	3.0	24.8	55.2	17.0	-0.08	0.19	-0.05	-0.06
27	**D**	0.620	0.533	1.9	23.0	13.1	62.0	-0.06	-0.24	-0.23	0.53
28	**C**	0.663	0.519	26.3	6.3	66.3	1.1	-0.36	-0.13	0.52	-0.04
29	**C**	0.587	0.585	8.7	28.3	58.7	4.3	-0.22	-0.32	0.59	-0.04
30	**C**	0.646	0.467	4.1	16.5	64.6	14.8	-0.10	-0.25	0.47	-0.11
31	**B**	0.728	0.459	3.1	72.8	19.6	4.4	-0.10	0.46	-0.27	-0.10
32	**A**	0.670	0.222	67.0	29.4	2.2	1.3	0.22	-0.10	-0.09	-0.04
33	**C**	0.383	0.511	3.9	10.6	38.3	47.2	-0.12	-0.24	0.51	-0.16
34	**C**	0.183	0.244	9.1	38.1	18.3	34.4	-0.14	-0.13	0.24	0.02
35	**A**	0.265	0.111	26.5	28.0	33.1	12.4	0.11	0.05	-0.12	-0.04
36	**C**	0.622	0.467	5.2	23.0	62.2	9.6	-0.16	-0.29	0.47	-0.02
37	**C**	0.920	0.156	1.9	3.1	92.0	3.0	-0.07	-0.09	0.16	0.00
38	**B**	0.874	0.296	3.0	87.4	3.7	5.9	-0.10	0.30	-0.12	-0.08
39	**A**	0.824	0.378	82.4	8.5	6.5	2.6	0.38	-0.13	-0.18	-0.07
40	**B**	0.735	0.378	3.3	73.5	19.8	3.3	-0.06	0.38	-0.28	-0.04
41	**B**	0.913	0.267	1.7	91.3	2.6	4.4	-0.07	0.27	-0.09	-0.11
42	**A**	0.693	0.356	69.3	15.9	10.7	4.1	0.36	-0.26	-0.06	-0.04
43	**A**	0.687	0.304	68.7	26.3	3.1	1.9	0.30	-0.16	-0.08	-0.06
44	**A**	0.756	0.311	75.6	5.2	9.4	9.8	0.31	-0.12	-0.16	-0.04
45	**A**	0.522	0.126	52.2	6.1	24.3	17.4	0.13	-0.13	0.11	-0.10

Table A.7 General chemistry II Scale Concept Inventory (pre-administration) Item Statistics n = 647

Item	Key	%A	%B	%C	%D	%E	Positive (%)	Negative (%)
1	+	32.6	43.1	3.9	15.6	4.8	75.7	20.4
2	-	0.8	4.5	9.4	48.1	37.2	5.3	85.3
3	-	22.4	36.6	10.0	24.3	6.6	59.0	30.9
4	+	15.3	43.9	13.8	23.0	4.0	59.2	27.0
5	-	0.2	7.6	4.6	46.7	41.0	7.7	87.6
6	+	26.4	45.0	9.1	15.8	3.7	71.4	19.5
7	+	18.7	40.0	14.5	19.8	7.0	58.7	26.7
8	+	31.7	42.5	16.5	6.8	2.5	74.2	9.3
9		4.0	19.6	25.7	40.3	10.4	23.6	50.7
10	+	13.4	48.4	20.1	14.7	3.4	61.8	18.1
11	+	10.8	40.0	25.2	20.7	3.2	50.9	24.0
12	+	5.4	31.4	10.7	38.5	14.1	36.8	52.6
13	+	15.9	51.6	19.5	11.7	1.2	67.5	13.0
14	-	0.5	8.5	9.9	43.3	37.9	9.0	81.1
15	+	29.8	57.2	4.3	6.8	1.9	87.0	8.7
16	+	15.9	46.2	18.4	16.4	3.1	62.1	19.5
17	-	25.8	48.1	9.6	12.4	4.2	73.9	16.5
18		7.9	37.2	23.6	24.4	6.8	45.1	31.2
19	-	4.9	22.1	16.8	38.3	17.8	27.0	56.1
20	+	21.9	53.0	11.4	11.6	2.0	75.0	13.6
21	+	15.5	48.5	15.8	17.3	2.9	64.0	20.2
22	+	48.2	45.9	3.4	1.5	0.9	94.1	2.5
23	-	4.0	19.2	22.4	42.0	12.4	23.2	54.4
24		3.4	15.5	19.8	44.8	16.5	18.9	61.4
25	+	18.4	46.7	14.8	17.5	2.6	65.1	20.1
26	+	12.2	34.0	23.5	26.0	4.3	46.2	30.3
27	**V**	51.2	48.8	0.0	0.0	0.0	100.0	0.0
28	+	27.7	47.6	15.6	7.0	2.2	75.3	9.1
29	+	14.7	35.2	17.0	28.0	5.1	49.9	33.1
30	+	8.5	25.5	24.0	30.1	11.9	34.0	42.0
31	-	4.9	15.8	12.2	46.4	20.7	20.7	67.1
32	-	7.4	43.4	16.7	24.7	7.7	50.9	32.5
33	+	13.0	43.9	17.9	22.3	2.9	56.9	25.2
34	-	14.1	42.8	16.7	20.4	6.0	56.9	26.4
35	-	7.1	42.0	18.1	24.4	8.3	49.1	32.8
36	+	11.3	41.9	27.4	14.7	4.8	53.2	19.5
37	-	1.5	13.3	8.8	55.5	20.9	14.8	76.4
38	+	3.1	24.4	23.3	40.5	8.7	27.5	49.1
39	+	16.1	42.8	25.7	12.7	2.8	58.9	15.5
40	-	8.8	34.9	19.5	25.8	11.0	43.7	36.8

Table A.8 General chemistry II Scale Concept Inventory (post-administration) Item Statistics n = 470

Item	Key	%A	%B	%C	%D	%E	Positive (%)	Negative (%)
1	+	28.1	51.1	2.6	15.5	2.8	79.1	18.3
2	-	1.3	6.6	10.0	49.8	32.3	7.9	82.1
3	-	21.3	46.4	9.1	19.8	3.4	67.7	23.2
4	+	13.6	50.0	15.7	17.4	3.2	63.6	20.6
5	-	1.5	7.9	4.0	49.8	36.8	9.4	86.6
6	+	23.4	47.7	10.6	16.0	2.3	71.1	18.3
7	+	15.1	46.0	15.7	18.9	4.3	61.1	23.2
8	+	26.4	47.2	18.1	7.0	1.3	73.6	8.3
9		2.6	20.2	20.9	47.2	9.1	22.8	56.4
10	+	12.8	47.0	20.6	15.5	4.0	59.8	19.6
11	+	10.9	40.6	22.6	22.6	3.4	51.5	26.0
12	+	5.7	26.4	13.6	43.8	10.4	32.1	54.3
13	+	13.6	54.5	21.5	9.1	1.3	68.1	10.4
14	-	1.1	12.8	13.0	41.3	31.9	13.8	73.2
15	+	25.3	59.4	5.5	8.3	1.5	84.7	9.8
16	+	14.7	48.7	17.4	17.2	1.9	63.4	19.1
17	-	20.6	53.2	10.9	12.1	3.2	73.8	15.3
18		9.1	38.5	20.0	27.4	4.9	47.7	32.3
19	-	4.5	22.8	21.1	38.5	13.2	27.2	51.7
20	+	19.1	58.7	10.4	10.6	1.1	77.9	11.7
21	+	16.2	54.9	14.3	13.2	1.5	71.1	14.7
22	+	38.3	55.3	4.0	1.7	0.6	93.6	2.3
23	-	5.3	22.8	22.1	39.8	10.0	28.1	49.8
24		3.0	20.6	17.2	46.4	12.8	23.6	59.1
25	+	16.8	48.3	16.2	16.0	2.8	65.1	18.7
26	+	14.3	38.9	25.7	19.8	1.3	53.2	21.1
27	**V**	43.0	57.0	0.0	0.0	0.0	100.0	0.0
28	+	25.5	45.5	17.9	9.8	1.3	71.1	11.1
29	+	13.8	37.7	14.0	29.8	4.7	51.5	34.5
30	+	6.6	30.4	23.4	28.5	11.1	37.0	39.6
31	-	4.7	19.4	10.4	49.1	16.4	24.0	65.5
32	-	7.4	42.3	18.9	23.0	8.3	49.8	31.3
33	+	10.6	46.4	17.4	21.3	4.3	57.0	25.5
34	-	10.4	49.4	16.6	18.7	4.9	59.8	23.6
35	-	8.5	44.5	19.4	18.7	8.9	53.0	27.7
36	+	10.4	45.1	26.0	14.7	3.8	55.5	18.5
37	-	2.6	14.7	10.4	53.6	18.7	17.2	72.3
38	+	4.0	23.6	21.7	43.6	7.0	27.7	50.6
39	+	17.7	47.4	17.9	14.9	2.1	65.1	17.0
40	-	8.3	39.8	15.3	26.8	9.8	48.1	36.6

Appendix B: Laboratory Assessments

- Laboratory Survey Items

- General Chemistry I Laboratory Survey pre-administration item statistics
- General Chemistry I Laboratory Survey post-administration item statistics
- General Chemistry I Laboratory quiz items

- General Chemistry II Laboratory Survey pre-administration item statistics
- General Chemistry II Laboratory Survey post-administration item statistics
- General Chemistry II Laboratory quiz items

Table B.1 Laboratory Survey Items[a]

Objective items

- If the balance reads 0.153 grams, you should record 0.11 grams in your notebook.
- Volume is most accurately measured using a beaker.
- Precision of a measurement can be estimated by calculating a standard deviation.
- A reasonable percent yield for your experiment is 103.2%.
- Laboratory balances can measure 1 gram to 3 places with certainty.
- A percent error calculation reveals an error of ~50%, this tells you that your experiment value is off by a factor of 2 from the accepted value.
- Lab will be a helpful component to this chemistry course for demonstrating chemical concepts.
- You are asked to calculate the molar concentration of a sulfuric acid solution and your answer is 110 M. This is a reasonable concentration for this solution.
- You are asked to measure out 2.50 grams of a material, anything between 2.47g – 2.53g is an acceptable value.
- Increasing the number of measurements decreases the amount of error associated with that measurement.
- Percent error calculations tell you the degree to which your experimental value differs from an accepted value.
- Overfilling a volumetric flask while making a solution would result in a higher calculated concentration.
- Using a volumetric flask instead of an Erlenmeyer flask to make a solution will make the measurement more precise.

Subjective items

- I expect the lab will help reinforce the chemistry concepts taught in lecture.
- I expect to understand things better on the molecular level because of lab.
- I don’t think I will learn anything in lab.
- The laboratory activities will help me learn lecture concepts that are unable to be demonstrated in a classroom setting.
- I expect my understanding of the particulate nature of matter will be increased by the laboratory activities.
- I don’t expect the laboratory activities to match well with the lecture topics.
- Of lab, lecture, and discussion, lab gives the most hands on approach to understanding chemistry concepts.

[a] Items for “pre” survey shown. “Post” items are identical except in cases where past tense language was added – for example changing “I don’t think I will learning anything in lab” to “I didn’t learn anything in lab”

Table B.2 General chemistry I Laboratory Survey (pre-administration) Item Statistics
n = 1724

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1		51.9	45.5	1.7	0.7	0.2	0.06	97.4	0.9
2		40.4	51.2	6.9	1.0	0.5	0.06	91.6	1.5
3	-	4.4	12.5	15.8	39.2	27.8	0.35	16.8	67.0
4	-	3.9	18.0	17.4	34.9	25.1	0.58	22.0	60.0
5	+	7.9	40.5	39.3	8.7	2.5	1.10	48.4	11.2
6		0.8	0.9	2.4	35.3	60.0	0.70	1.7	95.2
7	-	1.5	8.5	31.2	37.6	20.6	0.46	10.0	58.3
8	+	13.1	51.8	26.7	6.8	0.9	0.81	64.8	7.7
9	+	3.0	26.7	51.9	15.3	2.4	0.70	29.6	17.7
10		44.1	52.4	2.8	0.4	0.2	0.06	96.5	0.6
11	-	2.7	17.2	42.6	25.7	11.3	0.52	19.9	36.9
12	+	10.6	40.7	12.6	27.7	8.3	0.12	51.3	36.0
13		40.2	55.0	3.9	0.5	0.3	0.06	95.2	0.8
14	+	21.1	46.1	17.5	12.8	2.4	0.12	67.2	15.2
15	V	53.4	46.6	0.0	0.0	0.0	0.00	100.0	0.0
16	+	27.1	61.1	8.7	2.3	0.2	0.58	88.3	2.4
17	-	5.2	23.8	19.3	36.9	14.6	0.17	29.1	51.5
18		31.3	59.1	7.7	1.6	0.1	0.23	90.4	1.7
19		1.0	4.6	13.1	56.6	24.6	0.06	5.7	81.1
20	+	15.1	37.6	34.6	9.7	2.8	0.12	52.7	12.5

Table B.3 General chemistry I Laboratory Survey (post-administration) Item Statistics n = 1613

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1		28.5	58.2	8.6	4.2	0.6	0.06	86.6	4.7
2		28.3	54.6	12.8	3.5	0.6	0.25	83.0	4.0
3	-	3.2	11.4	9.4	39.5	36.1	0.31	14.6	75.6
4	-	1.7	5.4	3.5	36.8	52.0	0.68	7.1	88.7
5	+	8.2	39.6	38.7	10.2	2.9	0.37	47.8	13.1
6		0.7	2.0	4.8	40.4	51.5	0.56	2.7	91.9
7	-	1.8	9.3	11.0	34.3	43.0	0.56	11.1	77.4
8	+	17.7	61.9	12.6	5.8	1.1	0.81	79.7	6.9
9	+	4.6	34.7	36.9	19.7	3.7	0.25	39.4	23.4
10		29.4	58.8	7.9	3.3	0.3	0.25	88.2	3.7
11	-	2.1	10.2	12.7	41.4	33.5	0.12	12.3	74.9
12	+	19.3	61.1	6.9	10.0	2.5	0.12	80.4	12.5
13		26.0	58.5	10.6	4.3	0.6	0.12	84.4	4.8
14	+	15.7	48.7	15.9	16.2	3.4	0.12	64.4	19.6
15	V	41.5	58.5	0.0	0.0	0.0	0.00	100.0	0.0
16	+	31.0	58.5	6.9	3.0	0.5	0.12	89.5	3.5
17	-	4.8	13.8	9.7	43.6	28.0	0.06	18.6	71.6
18		22.4	59.0	13.9	3.9	0.7	0.06	81.4	4.6
19		5.0	13.7	24.5	42.7	14.1	0.06	18.7	56.8
20	+	29.8	45.6	12.8	9.7	2.0	0.06	75.4	11.8

Table B.4 General Chemistry I pre-laboratory quiz items and complexity ratings

Exp	Pre/ Post	Item	Complexity rating
2	Pre	What are two things that are smaller than we can see with our unaided eyes? Are the two things the same size? Which is smaller?	6
	Post	A diagram of a plant cell is shown. Diagram a water molecule in relation to this.	9
3	Pre	On the particle level, show the reaction of hydrogen and oxygen to form water; how does this diagram help explain the Law of Conversation of Matter?	8
	Post	What is the difference between macroscopic observations and particle-level diagrams? How does this difference make it clear that we do not make observations on the same scale as the particle-level diagrams we use to describe them?	9
4	Pre	On the particle level, show sodium chloride in aqueous solution. You must include at least 4 water molecules in your drawing.	8
	Post	On the particle level, show the reaction of sodium chloride with silver nitrate. How does this differ from what you would observe on the particle level? You must include at least 4 water molecules in your drawing.	11
5	Pre	Draw the particulate representation for the reaction of HCl (aq) and NaOH (aq).	7
	Post	Using the reaction of HCl with NaOH, describe how using particle level diagrams or pictures helps our understanding of the difference between acids and bases and the process of neutralization.	11
6	Pre	What volume of water and dye would you need to make 10 mL of a 10% dye solution? If the original dye solution is light orange, what would you expect to observe for the diluted solution? What does this mean on the particle level?	8
	Post	If you continued to dilute your original solution down to 1.0 ppb (1 ppb), how would you know that there is still dye present in the solution even though it will appear colorless? Draw a picture of a particle diagram of a ppb solution using an “x” as a dye particle and a circle as a water molecule.	11
7	Pre	Using the 13 nm gold nanoparticle that you will make in lab this week, which is larger, a gold nanoparticle or a water molecule? Draw a particle diagram of a gold nanoparticle and a water molecule – make sure to label both.	7
	Post	Draw a picture or diagram what happens on the particle level as the electrolyte is added to the solution. Make sure to include at least 2 nanoparticles (labeled), 2 ions (labeled), and 4 water molecules (labeled).	11
8	Pre	Draw a picture or diagram what happens to gas particles (on the particle level) as they are heated. Include how this relates to a change in pressure if the volume and quantity are held constant.	8

	Post	Draw a particle level picture or diagram of neon under standard conditions. To do this correctly, consider how far apart particles are under these conditions. Using your diagram, show that the gas particle(s) are moving and explain how this movement gives us the macroscopic measurement of pressure (what is pressure?).	10
9	Pre	Draw a picture or diagram showing a liquid and a gas on the particle level. Add a labeled arrow for boiling and a labeled arrow for condensing between the liquid and gas. Note which is exothermic and which is endothermic.	7
	Post	Draw a picture of a particle level diagram of the transition of boiling for water. To do this correctly you must show why this requires energy input. Do you think a different substance (like carbon dioxide) requires the same amount of energy input to boil as water? Why or why not?	11
10	Pre	Draw a picture or diagram showing the hydration of sodium hydroxide. Include an energy diagram correctly showing the sign on enthalpy of solution for sodium hydroxide.	8
	Post	Draw a picture of a particle level diagram of the hydration of ammonium chloride. To do this correctly you must include water molecules hydrating the ammonium chloride. The solution cooled as you made it. Draw an energy diagram to the right showing the cooling for this hydration. Include the reactants and products as you have drawn above on the energy diagram. Why do you think it got colder?	11.5
11	Pre	Are all molecules the same size? How do dispersion forces change as size changes? How does size affect the macroscopic property of boiling point?	7.5
	Post	What was your reasoning for the assignment of size for the molecules in part 3 (pure molecular substances) and how did this contribute to your decision of the relative boiling points of these substances?	9

Table B.5 General chemistry II Laboratory Survey (pre-administration) Item Statistics n = 732

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1		42.9	55.3	1.8	0.0	0.0	0.00	98.2	0.0
2		35.0	59.3	5.5	0.3	0.0	0.00	94.3	0.3
3	-	3.1	11.5	12.3	50.1	22.7	0.27	14.6	72.8
4	-	1.9	8.6	11.1	47.4	31.0	0.00	10.5	78.4
5	+	5.1	46.0	37.4	10.0	1.1	0.41	51.1	11.1
6		0.5	1.1	3.0	50.5	44.5	0.27	1.6	95.1
7	-	1.9	10.0	15.3	41.9	30.7	0.14	11.9	72.7
8	+	9.4	63.9	19.3	5.9	0.5	0.96	73.4	6.4
9	+	4.1	36.1	38.1	18.9	2.6	0.27	40.2	21.4
10		32.0	64.8	3.0	0.3	0.0	0.00	96.7	0.3
11	-	1.2	11.3	16.5	48.4	22.5	0.00	12.6	70.9
12	+	13.1	57.5	11.5	16.0	1.9	0.00	70.6	17.9
13		29.6	66.8	2.7	0.5	0.1	0.14	96.4	0.7
14	+	15.2	54.9	13.5	13.5	2.7	0.14	70.1	16.3
15	V	44.3	55.7	0.0	0.0	0.0	0.00	100.0	0.0
16	+	20.2	57.9	8.6	10.5	2.3	0.41	78.1	12.8
17	-	4.4	22.8	14.6	43.6	14.5	0.14	27.2	58.1
18		23.1	68.2	7.8	0.7	0.1	0.14	91.3	0.8
19		0.8	11.1	20.1	51.4	16.7	0.00	11.9	68.0
20	+	13.1	52.3	22.7	10.5	1.2	0.14	65.4	11.7

Table B.6 General chemistry II Laboratory Survey (post-administration) Item Statistics n = 661

Item	Key	%A	%B	%C	%D	%E	%Omit	Positive (%)	Negative (%)
1		29.5	62.3	5.4	2.7	0.0	0.00	91.8	2.7
2		24.1	58.1	12.6	4.8	0.3	0.15	82.1	5.1
3	-	4.5	12.6	7.9	48.3	26.6	0.15	17.1	74.9
4	-	3.3	7.0	4.8	46.6	38.1	0.15	10.3	84.7
5	+	7.3	40.4	39.6	10.3	2.1	0.30	47.7	12.4
6		1.7	2.9	4.5	46.7	43.6	0.61	4.5	90.3
7	-	3.8	12.9	11.5	39.8	30.9	1.21	16.6	70.7
8	+	15.9	64.6	12.9	5.0	1.1	0.61	80.5	6.1
9	+	5.1	34.9	37.1	20.1	2.6	0.15	40.1	22.7
10		27.4	62.5	6.2	3.5	0.2	0.30	89.9	3.6
11	-	2.1	8.5	10.6	45.8	32.8	0.15	10.6	78.7
12	+	20.9	60.1	7.3	9.5	2.3	0.00	80.9	11.8
13		24.1	64.4	5.9	5.3	0.3	0.00	88.5	5.6
14	+	15.4	49.3	17.1	16.2	1.8	0.15	64.8	18.0
15	V	36.5	63.5	0.0	0.0	0.0	0.00	100.0	0.0
16	+	32.7	63.4	2.6	1.1	0.0	0.30	96.1	1.1
17	-	5.6	22.4	12.6	43.6	15.6	0.30	28.0	59.2
18		19.1	59.5	15.1	5.6	0.8	0.00	78.5	6.4
19		4.7	13.2	19.5	49.9	12.4	0.30	17.9	62.3
20	+	18.8	49.6	18.3	11.3	2.0	0.00	68.4	13.3

Table B.7 General Chemistry II pre-laboratory quiz items and complexity ratings

<table>
<tr><th>Exp</th><th>Pre/ Post</th><th>Item</th><th>Complexity rating</th></tr>
<tr><td rowspan="2">1</td><td>Pre</td><td>Diagram the surface of a liquid on the particle level. Identify the source of vapor pressure.</td><td>7</td></tr>
<tr><td>Post</td><td>Diagram two different liquids on the particle level. Describe how they can have different vapor pressures based on intermolecular forces.</td><td>9</td></tr>
<tr><td rowspan="2">2</td><td>Pre</td><td>Diagram a mixture on the particle level and show the difference in the assignment of solute versus solvent using your diagram.</td><td>5</td></tr>
<tr><td>Post</td><td>Diagram a solution on the particle level. Using your diagram, describe how the vapor pressure of the solution differs from the vapor pressure of the pure solvent and how this relates to boiling point elevation.</td><td>12</td></tr>
<tr><td rowspan="2">3</td><td>Pre</td><td>Describe the difference between the concentration of the iron (III) ion in trial 1 versus trial 5.
<table>
<tr><th>Trial</th><th>$FeCl_3$ (mL)</th><th>KI (mL)</th><th>H_2O (mL)</th></tr>
<tr><td>1</td><td>20.0</td><td>20.0</td><td>0.0</td></tr>
<tr><td>2</td><td>20.0</td><td>10.0</td><td>10.0</td></tr>
<tr><td>3</td><td>10.0</td><td>20.0</td><td>10.0</td></tr>
<tr><td>4</td><td>15.0</td><td>10.0</td><td>15.0</td></tr>
<tr><td>5</td><td>10.0</td><td>15.0</td><td>15.0</td></tr>
</table></td><td>6</td></tr>
<tr><td>Post</td><td>Describe how varying the concentration(s) of reactions can vary the rate of a reaction.</td><td>9</td></tr>
<tr><td rowspan="2">4</td><td>Pre</td><td>Draw an energy diagram, identifying the activation energy. Using your diagram, predict the sign on your activation energy and give your justification for this prediction.</td><td>7</td></tr>
<tr><td>Post</td><td>Draw two different energy diagrams for two different reactions, one fast and one slow. Justify your answer using activation energy.</td><td>10</td></tr>
<tr><td rowspan="2">5</td><td>Pre</td><td>Describe and/or diagram three different systems: one with very large K, one with very small K, and one with a K about equal to one.</td><td>6</td></tr>
<tr><td>Post</td><td>Diagram the reaction $2\ NO_2(g) \leftrightarrow N_2O_4\ (g)$ with two different systems: one with K > 1 and one with K < 1; you must include your actual K value in both of the diagrams.</td><td>10</td></tr>
<tr><td rowspan="2">6</td><td>Pre</td><td>Give the masses you calculated for pre-lab number 4, buffer A and buffer B. In Buffer A, What are all species present in solution?</td><td>9</td></tr>
<tr><td>Post</td><td>Diagram how your acetic acid-acetate buffer reacts with an added acid. Show how this affects the pH.</td><td>9</td></tr>
<tr><td>7</td><td>Pre</td><td>Diagram a solution of calcium hydroxide for which you can measure the Ksp.</td><td>7</td></tr>
</table>

	Post	For your solution of calcium hydroxide, diagram what happened when you added HCl to measure the Ksp (do not show the experimental setup).	9
8	Pre	In addition to reaction 3 listed below, which other reaction must also be exothermic and why? (1) $MgO (s) + 2HCl(aq) \rightarrow MgCl_2(aq) + H_2O(l)$ (2) $Mg(s) + 2HCl(aq) \rightarrow MgCl_2(aq) + H_2(g)$ (3) $H_2(g) + ½O_2(g) \rightarrow H_2O(l)$ (4) $Mg(s) + ½O_2(g) \rightarrow MgO(s)$	6
	Post	Describe why you cannot measure the enthalpy of formation of magnesium oxide directly using your calorimetric setup.	7
9	Pre	On the particle level, describe a redox reaction. Please include in your diagram, what happens when something is oxidized (showing this) and what happens when something is reduced (showing this).	8
	Post	Diagram the reaction of copper with nitric acid on the particle level using the boxes below. Include enough detail that you can infer that the mass of copper metal decreases while the concentration of the copper ions increases as the reaction proceeds.	10

Appendix C: Laboratory Experiments

- Non-scale General Chemistry I Beer’s Law laboratory experiment
- Scale-themed General Chemistry I Beer’s Law laboratory experiments
- Non-scale General Chemistry II Molar Mass of a Volatile Liquid experiment
- Scale-themed General Chemistry II Molar Mass of a Volatile Liquid experiment

Figure C.1: Non-Scale

Determining the Concentration of a Solution: Beer's Law

The primary objective of this experiment is to determine the concentration of an unknown dye solution. Use of a spectrometer will allow for the determination an appropriate wavelength based on the absorbance spectrum of the solution. A higher concentration of the colored solution absorbs more light (and transmits less) than a solution of lower concentration.

You will prepare five dye solutions of known concentration (standard solutions). Each solution is transferred to a small, rectangular cuvette that is placed into the Spectrometer. The amount of light that penetrates the solution and strikes the photocell is used to compute the absorbance of each solution. When you graph absorbance *vs*. concentration for the standard solutions, a direct relationship should result. The direct relationship between absorbance and concentration for a solution is known as *Beer's law*.

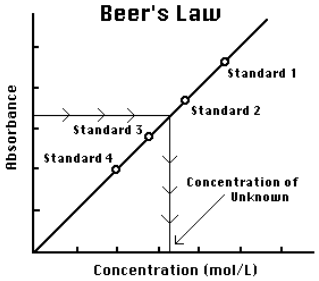


Figure 1

You will determine the concentration of an unknown dye solution by measuring its absorbance. By locating the absorbance of the unknown on the vertical axis of the graph, the corresponding concentration can be found on the horizontal axis. The concentration of the unknown can also be found using the slope of the Beer's law curve.

OBJECTIVES

In this experiment, you will

- Prepare and test the absorbance of five standard dye solutions.
- Calculate a standard curve from the test results of the standard solutions.
- Test the absorbance of a dye solution of unknown molar concentration.
- Calculate the molar concentration of the unknown dye solution.

MATERIALS

LabQuest
LabQuest App
Vernier Spectrometer
one cuvette
five 20 × 150 mm test tubes
two 10 mL pipets or graduated cylinders
two 100 mL beakers
dye solution of known concentration
unknown dye solution
pipet pump or pipet bulb
distilled water
test tube rack
stirring rod
tissues (preferably lint-free)

PROCEDURE

Both Colorimeter and Spectrometer Users

1. Obtain and wear goggles.

2. Obtain small volumes of known dye solution and distilled water in separate beakers. Record the concentration of the dye solution.

3. Label five clean, dry, test tubes 1–5. Use pipets to prepare five standard solutions according to the chart below. Thoroughly mix each solution with a stirring rod. Clean and dry the stirring rod between uses.

Test Tube	Known dye solution (mL)	Distilled H_2O (mL)	Concentration (M)
1	2	8	
2	4	6	
3	6	4	
4	8	2	
5	~10	0	

4. Prepare a *blank* by filling a cuvette 3/4 full with distilled water. To correctly use cuvettes, remember:
   - Wipe the outside of each cuvette with a lint-free tissue.
   - Handle cuvettes only by the top edge of the ribbed sides.
   - Dislodge any bubbles by gently tapping the cuvette on a hard surface.
   - Always position the cuvette so the light passes through the clear sides.

5. Connect the Spectrometer to LabQuest and choose New from the File menu.

6. Calibrate the Spectrometer.
   a. Place the blank cuvette in the Spectrometer.
   b. Choose Calibrate from the Sensors menu. The following message is displayed: “Waiting 60 seconds for lamp to warm up.” After 60 seconds, the message will change to “Warmup complete.”
   c. Select Finish Calibration. When the message “Calibration completed” appears, select OK.

7. Determine the optimal wavelength for creating this standard curve and set up the data-collection mode.
   a. Remove the blank cuvette, and place the known dye standard (highest concentration, test tube 5) into the cuvette slot.
   b. Start data collection. A full spectrum graph of the solution will be displayed. Stop data collection. The wavelength of maximum absorbance (λ max) is automatically identified.
   c. Tap the Meter tab. On the Meter screen, tap Mode. Change the mode to Events with Entry.
   d. Enter the Name (Concentration) and Units (mol/L). Select OK.

8. You are now ready to collect absorbance-concentration data for the five standard solutions.

   a. Start data collection.
   b. Using the solution in Test Tube 1, rinse the cuvette twice with ~1 mL amounts and then fill it 3/4 full. Wipe the outside with a tissue and place it in the device.
   c. When the value displayed on the screen has stabilized, tap Keep and enter the concentration in mol/L. Select OK. The absorbance and concentration values have now been saved for the first solution.
   d. Discard the cuvette contents as directed. Using the solution in Test Tube 2, rinse and fill the cuvette 3/4 full. Wipe the outside and place the cuvette in the device (close the lid of the Colorimeter). Wait for the value displayed on the screen to stabilize, and tap Keep. Enter the concentration in mol/L.
   e. Repeat the procedure for Test Tubes 3 and 4. Trial 5 is the original known dye solution. **Note:** Do not test the unknown solution until Step 11.
   f. When you have finished testing the standard solutions, stop data collection.
   g. To examine the data pairs on the displayed graph, tap any data point. As you tap each data point, the absorbance and concentration values are displayed to the right of the graph.

9. Write down the absorbance values, for each of the five trials, in your data table.

10. Display a graph of absorbance *vs.* concentration with a linear regression curve.
    a. Choose Graph Options from the Graph menu.
    b. Select Autoscale from 0 and select OK.
    c. Choose Curve Fit from the Analyze menu.
    d. Select Linear as the Fit Equation. The linear-regression statistics for these two data columns are displayed for the equation in the form: $y = mx + b$ where x is concentration, y is absorbance, a is the slope, and b is the y-intercept. **Note:** One indicator of the quality of your data is the size of b. It is a very small value if the regression line passes through or near the origin. The correlation coefficient, r, indicates how closely the data points match up with (or *fit*) the regression line. A value of 1.00 indicates a nearly perfect fit.
    e. Select OK. The graph should indicate a direct relationship between absorbance and concentration, a relationship known as Beer's law. The regression line should closely fit the five data points *and* pass through (or near) the origin of the graph.
11. Determine the absorbance and concentration values of the unknown dye solution.
    a. Tap the Meter tab.
    b. Obtain about 5 mL of the *unknown* dye solution in another clean, dry, test tube. Record the number of the unknown in your data table.
    c. Rinse the cuvette twice with the unknown solution and fill it about 3/4 full. Wipe the outside of the cuvette and place it into the device (close the lid of the Colorimeter).
    d. Monitor the absorbance value. When this value has stabilized, record it in your data table.
    e. Tap the Graph tab.
    f. On the Graph screen, choose Interpolate from the Analyze menu. Tap any point on the regression curve (or use the ◄ or ► keys on LabQuest) to find the absorbance value that is closest to the absorbance reading you obtained in Step 11d. Determine the concentration of your unknown dye solution and record the concentration in your data table.
12. Record the path length of your cuvette in centimeters.
13. Clean up.

DATA TABLE

Trial	Concentration (mol/L)	Absorbance
1		
2		
3		
4		
5		
6	Unknown number ____	

Path length of your cuvette: _______________

RESULTS AND CALCULATIONS

1. If your dye was copper(II) sulfate, describe an alternate method for determining the molar concentration of your unknown sample.

2. Make a Beer's Law plot for your dye. Plot the absorbance (ordinate) versus concentration (abscissa). Make sure your plot includes the point (0,0). Draw the best fit line through your data that includes the point (0,0), determine the molar absorptivity of your known, and the concentration of your unknown.

3. Using your plot, estimate the error in your molar absorptivity. Comment on its value.

4. Identify at least one random and at least one systematic error in this experiment. How would each change your results?

PRE-LABORATORY ASSIGNMENT *Before you come to class:*

1. What is one real-world or practical application for this experiment or portion of this experiment?
2. In your own words, define: absorbance, absorbance spectrum, electromagnetic radiation, λ_{max}, molar absorptivity, path length, ultraviolet (UV) radiation, and visible radiation.
3. The dye you will use in this experiment is malachite green. Search the internet or other source and find λ_{max} for this dye that you will use in this experiment.
4. Identify any potentially hazardous steps in your procedure. In your own words, explain what safety procedures should be followed and why.
5. A solution of a dye in a 0.80 cm cuvette of concentration 1.25×10^{-3} M had an absorbance of 0.115 at a particular wavelength. What was the molar absorptivity of the dye?

Figure C.2: Scale-Themed

Determining the Concentration of a Solution: Beer's Law

The primary objective of this experiment is to gain a better understanding of solution chemistry by determining the concentration of an unknown dye solution. You will begin by making a solution of known concentration (a 10% solution of dye and water) and dilute it down to the part per million (ppm) level. Although this solution will be clear and colorless you will be able to measure that it still contains dye.

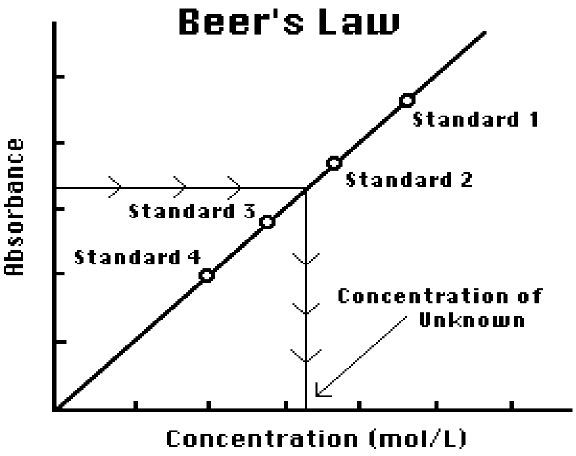


Figure 1

You will prepare these 6 dye solutions of known concentration (standard solutions) by performing a serial dilution. Each solution is transferred to a small, rectangular cuvette that is placed into the Spectrometer. The amount of light that penetrates the solution and strikes the photocell is used to compute the absorbance of each solution. When you graph absorbance *vs*. concentration for the standard solutions, a direct relationship should result. The direct relationship between absorbance and concentration for a solution is known as *Beer's law* and obeys the equation

$$a = \varepsilon bc$$

(a = absorbance, ε = the molar absorptivity constant with units $M^{-1}cm^{-1}$, b = the path length of the cuvette in cm, and c = concentration of the solution. From this "calibration curve" you will be able to calculate the molar absorptivity of the dye.

You will determine the concentration of an unknown dye solution by measuring its absorbance. By locating the absorbance of the unknown on the vertical axis of the graph, the corresponding concentration can be found on the horizontal axis. The concentration of the unknown can also be found using the above equation by plugging in the molar absorptivity value you find and the absorbance of the unknown solution.

OBJECTIVES

In this experiment, you will

- Prepare a 10% dye solution and serial dilute it to the ppm level.
- Create a standard curve from the absorbances and concentrations of the standard solutions.

- Test the absorbance of a dye solution of unknown molar concentration.
- Calculate the molar concentration of the unknown dye solution.

MATERIALS

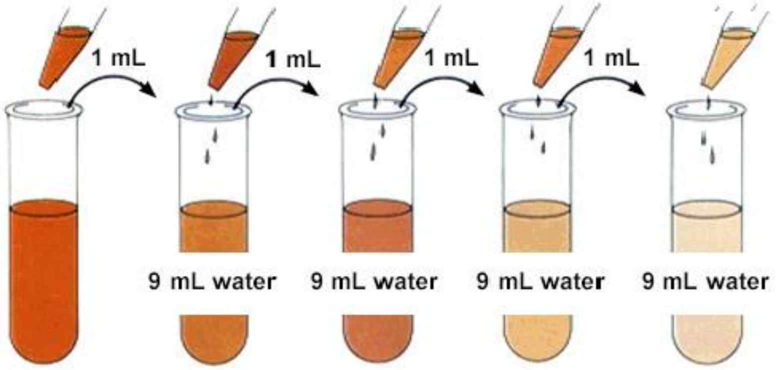
LabQuest
LabQuest App
Vernier Spectrometer
one cuvette
six 20 × 150 mm test tubes
two 10 mL graduated pipets
one 100 mL volumetric flask
one 10 mL graduated cylinder

1 M crystal violet dye solution
unknown dye solution
pipet pump or pipet bulb
distilled water
test tube rack
stirring rod
tissues (preferably lint-free)

PROCEDURE

1. Obtain and wear goggles.
2. Prepare a 10 mL 10% (by volume) dye solution of 1 M crystal violetdye and water in a 100 mL volumetric flask.
3. Label 5 clean, dry, test tubes 1–5 and use pipets to prepare 5 standard solutions according to the chart below. Thoroughly mix each solution with a stirring rod. Clean and dry the stirring rod between uses.

Test Tube	Known dye solution (mL)	Distilled H_2O (mL)	Concentration (M)
1	1 mL 10% solution	9	
2	1 mL from TT #1	9	
3	1 mL from TT #2	9	
4	1 mL from TT #3	9	
5	1 mL from TT #4	9	



4. Prepare a *blank* by filling a cuvette 3/4 full with distilled water. To correctly use cuvettes, remember:
   - Wipe the outside of each cuvette with a lint-free tissue.
   - Handle cuvettes only by the top edge of the ribbed sides.
   - Dislodge any bubbles by gently tapping the cuvette on a hard surface.
   - Always position the cuvette so the light passes through the clear sides.
5. Connect the Spectrometer to LabQuest and choose New from the File menu.
6. Calibrate the Spectrometer.
   d. Place the blank cuvette in the Spectrometer.
   e. Choose Calibrate from the Sensors menu. The following message is displayed: “Waiting 60 seconds for lamp to warm up.” After 60 seconds, the message will change to “Warmup complete.”
   f. Select Finish Calibration. When the message “Calibration completed” appears, select OK.
7. Determine the optimal wavelength for creating this standard curve and set up the data-collection mode.
   e. Remove the blank cuvette, and place a cuvette filled 3/4 full with the solution from Test Tube 1 into the cuvette slot.
   f. Start data collection. A full spectrum graph of the solution will be displayed. Stop data collection. The wavelength of maximum absorbance (λ max) is automatically identified.
   g. Tap the Meter tab. On the Meter screen, tap Mode. Change the mode to Events with Entry.
   h. Enter the Name (Concentration) and Units (mol/L). Select OK.
8. You are now ready to collect absorbance-concentration data for the five standard solutions.
   h. Start data collection.
   i. Begin with the solution in Test Tube 1 (should already be in your cuvette from the standardization in number 7).
   j. When the value displayed on the screen has stabilized, tap Keep and enter the concentration using decimal notation in mol/L. Select OK. The absorbance and concentration values have now been saved for the first solution.
   k. Discard the cuvette contents as directed. Using the solution in Test Tube 2, rinse and fill the cuvette 3/4 full. Wipe the outside and place the cuvette in the device. Wait for the value displayed on the screen to stabilize, and tap Keep. Enter the concentration in mol/L.
   l. Repeat the procedure for Test Tubes 3, 4 and 5. **Note:** Do not test the unknown solution until Step 11.
   m. When you have finished testing the standard solutions, stop data collection.

n. To examine the data pairs on the displayed graph, tap any data point. As you tap each data point, the absorbance and concentration values are displayed to the right of the graph.

9. Write down the absorbance values, for each of the five trials, in your data table.
10. Display a graph of absorbance *vs.* concentration with a linear regression curve.
    f. Choose Graph Options from the Graph menu.
    g. Select Autoscale from 0 and select OK.
    h. Choose Curve Fit from the Analyze menu.
    i. Select Linear as the Fit Equation. The linear-regression statistics for these two data columns are displayed for the equation in the form: $y = mx + b$ where x is concentration, y is absorbance, a is the slope, and b is the y-intercept. **Note:** One indicator of the quality of your data is the size of b. It is a very small value if the regression line passes through or near the origin. The correlation coefficient, r, indicates how closely the data points match up with (or *fit*) the regression line. A value of 1.00 indicates a nearly perfect fit.
    j. Select OK. The graph should indicate a direct relationship between absorbance and concentration, a relationship known as Beer's law. The regression line should closely fit the five data points *and* pass through (or near) the origin of the graph.
11. Determine the absorbance and concentration values of the unknown dye solution.
    g. Tap the Meter tab.
    h. Obtain about 5 mL of the *unknown* dye solution in another clean, dry, test tube.
    i. Rinse the cuvette twice with the unknown solution and fill it about 3/4 full. Wipe the outside of the cuvette and place it into the device (close the lid of the Colorimeter).
    j. Monitor the absorbance value. When this value has stabilized, record it in your data table.
    k. Tap the Graph tab.
    l. On the Graph screen, choose Interpolate from the Analyze menu. Tap any point on the regression curve (or use the ◄ or ► keys on LabQuest) to find the absorbance value that is closest to the absorbance reading you obtained in Step 10d. Determine the concentration of your unknown dye solution and record the concentration in your data table.
14. Save a copy of your calibration curve to print off and hand in with your lab report.
15. Record the path length of your cuvette in centimeters.
16. Clean up.

DATA AND OBSERVATIONS

1. Suggestion for your data table:

Trial	Concentration (mol/L)	Absorbance
1		
2		
3		
Trial	Concentration (mol/L)	Absorbance
4		
5		
6	Unknown	

Path length of your cuvette: _______________

2. Make sure to record all observations.

RESULTS AND CALCULATIONS

5. Explain the plot you made.
    a. Does your line of best fit go through 0?
    b. Should it?
    c. Using your plot, find the equation of your line.
    d. What is the molar absorptivity of the dye?
6. In which cup did the solution first appear colorless? What is the concentration of dye in this cup?
7. What was the concentration of the unknown dye solution?
8. Considering the solutions you made in this experiment:
    a. If you were to continue to dilute your original solution down to 1.0 part per billion (ppb), what would be the molar concentration (M) of dye in this solution?
    b. How do you know that there is still dye present in the solution even though the solution appears colorless?
    c. Approximately how many dye particles would be in 1.0 mL of this solution?
    d. Approximately how many water molecules would be in 1.0 mL of this solution (use the density of water as 1.0 g·mL^{-1} and the volume of the solution equal to the volume of water)?
    e. What is the ratio of water molecules to one dye particle?
    f. Using the ratio you calculated in part (e), what is a real-world comparison you can make to help you understand the number of solute (dye) particles to solvent (water) molecules.

g. Thinking about the comparison you made in (f), approximately how far apart are the dye particles in this solution?

PRE-LABORATORY ASSIGNMENT *Before you come to class:*

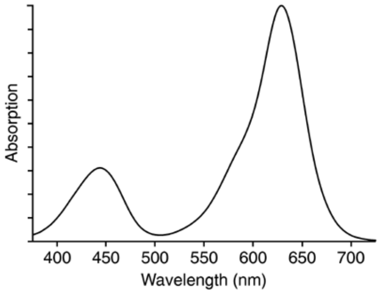
6. Envision building a cube with an edge length of 1m, calculate the volume of 1 one millionth (1 ppm) of this cube.
7. How would you make a 5% sugar solution that has a total mass of 100g? Hint: This is a weight percent calculation so use the equation:

   weight percent = (mass solute)/(mass solution) x 100

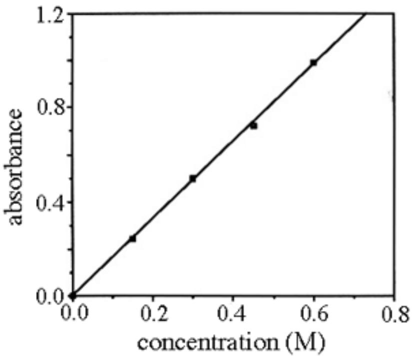
8. What volume of water and dye would you need to make a 10% solution? Hint: This is a volume/volume % calculation so use the equation:

   v/v% = (volume solute)/(volume solution) x 100

9. A solution of a dye in a 0.80 cm cuvette of concentration 1.25×10^{-3} M had an absorbance of 0.115 at a particular wavelength. What was the molar absorptivity of the dye?
10. Below is the absorbance spectrum for a malachite green dye solution, what is lamda max?



11. Given the following calibration curve, what is the concentration of a solution that has an absorbance of 0.800?



12. Describe how would you prepare a 50.00 mL of a 0.100 M solution of NaOH using:

a. solid NaOH

b. a 1.00 M solution of NaOH

You must include all calculations and describe the process you would follow to do this.

13. Identify any potentially hazardous steps in your procedure. In your own words, explain what safety procedures should be followed and why.

WHAT TO DISCUSS IN YOUR CONCLUSION

When writing your conclusion for this activity, make sure to include discussing:

- the intent of the activities;
- one real-world or practical application for this experiment or portion of this experiment (must include references);
- list at least one random and at least one systematic error in this experiment. How would each change your results;
- the molar absorptivity value you calculated? (i.e. Does it make sense?) Comment on the concentration value you found for the unknown. Does this make sense based on the color of the standard solutions?

C.3 Non-Scale:

The Molar Mass of a Volatile Liquid

One of the properties that helps characterize a substance is its molar mass. If the substance in question is a volatile liquid, a common method to determine its molar mass is to use the ideal gas law, $PV = nRT$. Because the liquid is volatile, it can easily be converted to a gas. While the substance is in the gas phase, you can measure its volume, pressure, and temperature. You can then use the ideal gas law to calculate the number of moles of the substance. Finally, you can use the number of moles of the gas to calculate molar mass.

OBJECTIVES

In this experiment, you will

- Evaporate a sample of a liquid substance and measure certain physical properties of the substance as it condenses.
- Determine the molar mass of an unknown liquid.

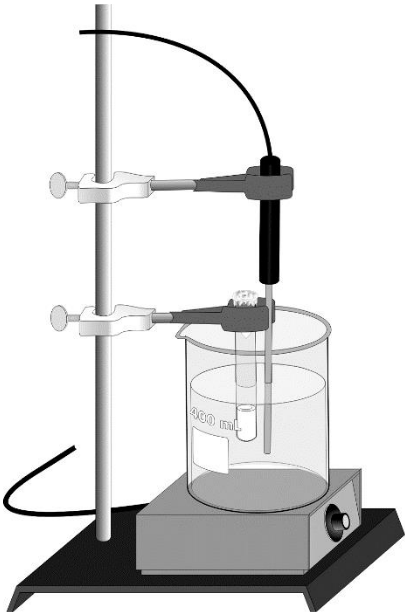


Figure 1

MATERIALS

Vernier computer interface
computer
Temperature Probe
(optional) Vernier Gas Pressure Sensor
ring stand
two utility clamps
aluminum foil
Ice

unknown volatile liquid (see Table 1.1)
fume hood
test tube, 13 × 100 mm, and holder
needle
hot plate
analytical balance
two 400 mL beakers
tissues or paper towels

PRE-LABORATORY ASSIGNMENT *To be completed before attending lab:*

1. What is one real-world or practical application for this experiment or portion of this experiment?

2. Identify all potentially hazardous steps in your procedure. In your own words, explain what safety precautions should be taken and why?

3. Draw the structure for each of the potential unknowns listed in Table 1.1.

Unknown	Formula	Unknown	Formula
butanone	C_4H_8O	ethanol	C_2H_6O
cyclohexane	C_6H_{12}	propanone	C_3H_6O

Table 1.1 - Potential unknowns and their formula

4. Review your procedures and identify as many potential sources for error as possible. Determine if your errors are systematic or random. Attempt to estimate how much error and which direction would each introduce into your calculated value for molar mass.

5. Use your textbook or other sources and find any constants or conversion factors you might need to evaluate for the molar mass. Assume that the pressure could be given in any one of the following units: inHg, cmHg, mmHg, torr, atm, or bar.

6. A flask with a total volume of 289.33 mL was found to contain 0.3546 g of vapor at 99.88°C on a day when the barometric pressure was 29.00 inHg. What was the molar mass of the unknown?

PROCEDURE

1. Obtain and wear goggles. Conduct this experiment in a fume hood or well-ventilated area.

2. Trim a piece of aluminum foil so that it just covers the top of a small, 13 × 100 mm, test tube. Use a needle to make a small hole in the middle of the foil. Measure the mass of the test tube and foil.

3. Prepare a hot-water bath by warming about 300 mL of tap water in a 400 mL beaker. Keep the beaker on a hot plate once the water is warm.

4. Use a second 400 mL beaker to prepare an ice-water bath.

5. Connect the Temperature Probe to LabQuest and choose New from the File menu. If you have an older sensor that does not auto-ID, manually set up the sensor.

6. Obtain a liquid sample of an unknown volatile compound. Pour about 0.5 mL of the liquid into the test tube and quickly cover the test tube with the aluminum foil. Place the test tube in the hot-water bath. Make sure that the foil is above the water level (see Figure 1).

7. Immerse the Temperature Probe in the hot water bath (see Figure 1). Do not allow the tip of the probe to touch the beaker. This will give you a more accurate reading of the water bath temperature. You will monitor the temperature readings during the experiment. There is no need to store and graph data.

8. Heat the beaker of water to boiling and maintain the boiling as your sample of liquid vaporizes. Note that some of your sample will escape the test tube through the needle hole in the foil. This process also serves to flush the air out of the test tube.

9. Keep the test tube in the boiling-water bath for at least three minutes *after* all of the liquid in the test tube has vaporized. Watch the temperature readings and record the temperature of the boiling-water bath, which will be used in the ideal gas law calculations.

10. Use a test-tube holder to *quickly* transfer the test tube to the ice water bath. Cool the test tube for about one minute, then remove it and dry it completely. Measure the mass of the test tube and the aluminum foil top.

11. Record the barometric pressure in the room.

12. Rinse out the test tube and fill it to the top with tap water. Cover the test tube with aluminum foil. Measure and record the mass of the test tube, water, and foil.

DATA TABLE

	Trial 1	Trial 2
Mass of test tube and foil cover (g)		
Temperature of water bath (ºC)		
Mass of test tube and foil and gas sample (g)		
Barometric pressure (kPa)		
Mass of test tube and foil and water (g)		

DATA ANALYSIS

1. Determine the mass of the condensed portion of the unknown that you placed in the test tube.
2. Use the mass of the water in the test tube from Step 12 of the procedure and its density to calculate the volume of the test tube.
3. Use the calculations from Questions 1 and 2 above, along with the temperature of the boiling water bath and the barometric pressure of the room, to calculate the molar mass of your unknown compound.
4. Identify the unknown liquid substance that you tested.
5. Calculate the error in your calculated molar mass.

RESULTS ANALYSIS

1. What are the possible sources of your error in this experiment? Discuss the possible reasons for your error.

2. Identify what you to consider to be the largest single source of error. Did you identify a systematic or random error? Explain.

3. How did you use the ideal gas law in your calculations?

4. Was the vapor really "ideal"? If not, how were your calculations affected? Explain.

5. If all of the vapor had not condensed to a liquid when you cooled the test tube, how would your calculations have been affected?

6. How would your experiment have been affected if you had used a different initial amount of the unknown compound?

7. Identify any changes you would make in your procedure.

C.4 Scale-themed:

The Molar Mass of a Volatile Liquid

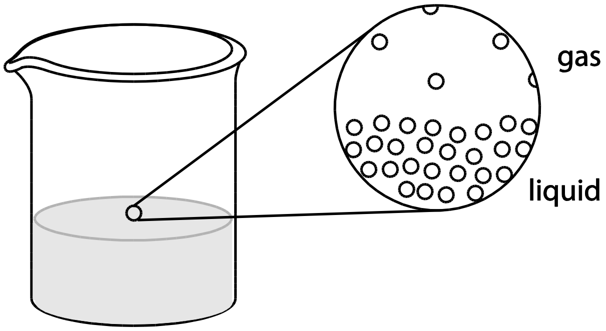


Figure 1: Macroscopic and particle diagram of the interface of a pure liquid and a gas.

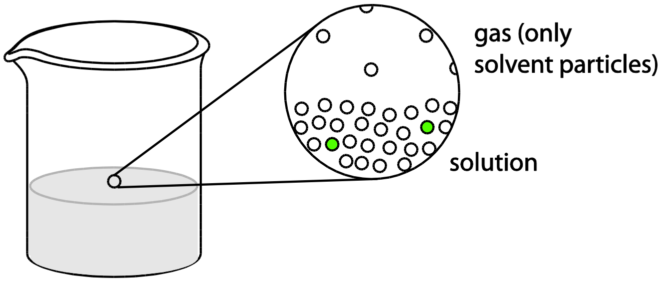


Figure 2: Macroscopic and particle diagram of the interface of a solution and the gas (of the solvent).

To the unaided eye the surface of a liquid may seem of little interest. However, as shown in **Figure 1** there is a lot of chemistry occurring at what is frequently referred to as the liquid/vapor interface. Many observations about a substance can be explained by modeling the interface of that substance (both pure substances and solutions will have unique interfaces, see **Figure 2** for a solution/gas interface). If a substance has a high vapor pressure, that is, the pressure exerted on the surface of a liquid by evaporated molecules of that liquid is high, it is said to also be a *volatile liquid*. Volatility is a measure of the ease in which liquid molecules gain sufficient kinetic energy to escape into the gas phase. These gas molecules will exert a pressure and this pressure is called the *vapor pressure*. Given this definition, it can be determined that a solution with a *high vapor pressure* and high volatility would contain *many gas molecules* at the liquid/vapor interface while conversely, a substance with *low vapor pressure* and low volatility would represent a solution in which *fewer liquid molecules* are able to escape into the gas phase.

You have already learned several chemical methods to determine the identity of an unknown substance such as melting point and density. Another intensive property that can be used to identify an unknown substance is its molar mass. If the substance in question is a volatile liquid, a common method to determine its molar mass is to use the ideal gas law, $PV = nRT$. Because the liquid is volatile, it can easily be converted to a gas. While the substance is in the gas phase, you can measure its volume, pressure, and temperature. You can then use the ideal gas law to calculate the number of moles of the substance. Finally, you can use the number of moles of the gas to calculate molar mass.

OBJECTIVES

In this experiment, you will

- Evaporate a sample of a liquid substance and measure certain physical properties of the substance as it condenses.
- Determine the molar mass of an unknown liquid.

MATERIALS

Vernier computer interface
computer
Temperature Probe
(optional) Vernier Gas Pressure Sensor
ring stand
two utility clamps
aluminum foil
Ice

unknown volatile liquid (see Table 1.1)
fume hood
test tube, 13 × 100 mm, and holder
needle
hot plate
analytical balance
two 400 mL beakers
tissues or paper towels

PROCEDURE

1. Obtain and wear goggles. Conduct this experiment in a fume hood or well-ventilated area.

2. Trim a piece of aluminum foil so that it just covers the top of a small, 13 × 100 mm, test tube. Use a needle to make a small hole in the middle of the foil. Measure the mass of the test tube and foil.

3. Prepare a hot-water bath by warming about 300 mL of tap water in a 400 mL beaker as shown in **Figure 3** (<u>make sure to use a ring around the beaker as shown</u>). Keep the beaker on a hot plate once the water is warm.

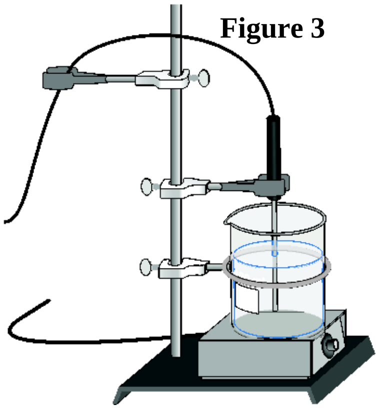


Figure 3

4. Use a second 400 mL beaker to prepare an ice-water bath.
5. Connect the Temperature Probe to LabQuest and choose New from the File menu. If you have an older sensor that does not auto-ID, manually set up the sensor.
6. Obtain a liquid sample of an unknown volatile compound. Pour about 0.5 mL of the liquid into the test tube and quickly cover the test tube with the aluminum foil. Place the test tube in the hot-water bath. Make sure that the foil is above the water level (see **Figure 4**).
7. Immerse the Temperature Probe in the hot water bath (see **Figure 4**). Do not allow the tip of the probe to touch the beaker. This will give you a more accurate reading of the water bath temperature. You will monitor the temperature readings during the experiment. There is no need to store and graph data.
8. Heat the beaker of water to boiling and maintain the boiling as your sample of liquid vaporizes. Note that some of your sample will escape the test tube through the needle hole in the foil. This process also serves to flush the air out of the test tube.
9. Keep the test tube in the boiling-water bath for at least three minutes *after* all of the liquid in the test tube has vaporized. Watch the temperature readings and record the temperature of the boiling-water bath, which will be used in the ideal gas law calculations.
10. Use a test-tube holder to *quickly* transfer the test tube to the ice water bath. Cool the test tube for about one minute, then remove it and dry it completely. Measure the mass of the test tube and the aluminum foil top.
11. Record the barometric pressure in the room.
12. Rinse out the test tube and fill it to the top with tap water. Cover the test tube with aluminum foil. Measure and record the mass of the test tube, water, and foil.

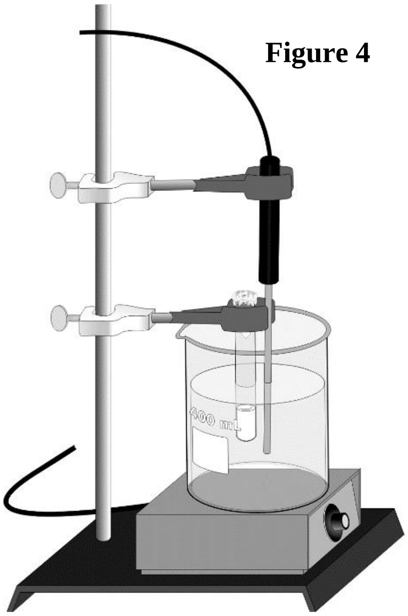


Figure 4

DATA AND OBSERVATIONS

6. Suggestion for your data table:

	Trial 1	Trial 2
Mass of test tube and foil cover (g)		
Temperature of water bath (ºC)		
Mass of test tube and foil and gas sample (g)		

Barometric pressure (kPa)		
Mass of test tube and foil and water (g)		

7. Make sure to record all observations.

RESULTS AND CALCULATIONS

8. Determine the mass of the condensed portion of the unknown you placed in the test tube.

9. Use the mass of the water in the test tube from Step 12 of the procedure and its density to calculate the volume of the test tube.

10. Use the calculations from Questions 1 and 2 above, along with the temperature of the boiling water bath and the barometric pressure of the room to calculate the molar mass of your unknown compound.

11. Identify the unknown liquid substance that you tested.

12. Calculate the error in your calculated molar mass.

13. A student calculates a molar mass of 8.6×10^{2} grams/mole. Is this reasonable? What error could have led to this? Another student calculates a molar mass of 1.2×10^{-2}. Is this reasonable?

14. Using the density and molar mass, calculate the number of gas particles contained in the test tube . Calculate the number of nitrogen molecules that would be contained in the same volume at STP. Was the vapor really “ideal”? Why or why not?

15. Make a particle level drawing of the substance in your test tube before and after you vaporized the liquid and after.

16. Using your drawing speculate as to how your experiment would have been affected if you

    a. had used a different initial amount of the unknown compound.

    b. not all of the vapor had condensed to a liquid when you cooled the test tube.

17. Identify any changes you would make in your procedure.

PRE-LABORATORY ASSIGNMENT

To be completed before attending lab:

7. Identify all potentially hazardous steps in your procedure. In your own words, explain what safety precautions should be taken and why?

8. Draw a picture of 2 different solutions showing the difference between a volatile liquid and non-volatile liquid. On your drawing show how intermolecular forces affect the two different liquids and indicate which would have high and low vapor pressure.

9. Draw the structure for each of the potential unknowns listed in Table 1.1.

Unknown	Formula	Unknown	Formula
butanone	C_4H_8O	ethanol	C_2H_6O
cyclohexane	C_6H_{12}	propanone	C_3H_6O

Table 2.1 - Potential unknowns and their formula

10. Use your textbook or other sources and find any constants or conversion factors you might need to evaluate for the molar mass. Assume that the pressure could be given in any one of the following units: inHg, cmHg, mmHg, torr, atm, or bar.

11. A flask with a total volume of 289.33 mL was found to contain 0.3546 g of vapor at 99.88°C on a day when the barometric pressure was 29.00 inHg. What was the molar mass of the unknown?

    a. A student completes the above calculation and determines the unknown to have a molar mass of 1206 g/mol. Is this reasonable? Why or why not?

WHAT TO DISCUSS IN YOUR CONCLUSION

When writing your conclusion for this activity, make sure to consider discussing:

- the intent of the experiment;
- one real-world or practical application for this experiment or portion of this experiment (must include references);
- Why you needed to convert the liquid to a gas to be able to determine the molar mass;
- The particle level properties of a liquid, specifically the intermolecular forces and vapor pressure of the liquid that allow you to determine the molar mass;
- Your resulting value(s). Is (are) they reasonable? How reliable was the method you used in this experiment?

Appendix D: Additional Projects

- D1: Development of scale-themed Supplemental Instruction for General Chemistry II
- D2: Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items
- D3: Assessment of NMR teaching and learning strategies in organic undergraduate labs

D.1 Supplemental Instruction for General Chemistry II

Solutions Activity

Initial Activity Questions:

1. Heating and Cooling Curves
2. Intermolecular Forces
3. Phase Changes
4. Phase Diagrams
5. Solutions Amounts
6. Intermolecular forces in solution
7. Vapor Pressure Lowering
8. Boiling point elevation
9. Phases Diagrams of solutions

Scoring (*1 point each*): 7-9 (>75%) – Scenario 3
5-6 (50-75%) – Scenario 2
0-4 (<50%) – Scenario 1

Scenario 1: Introduction

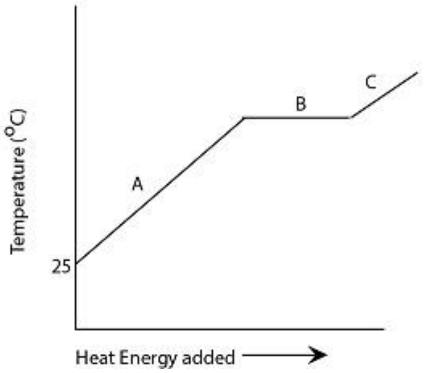
You go into your kitchen planning to make rice. You find your roommate left a measuring cup of a clear, colorless liquid (unknown liquid) right next to your measuring cup of water. You decide to boil both (in separate pots) to observe if there are differences.

Unknown liquid Water

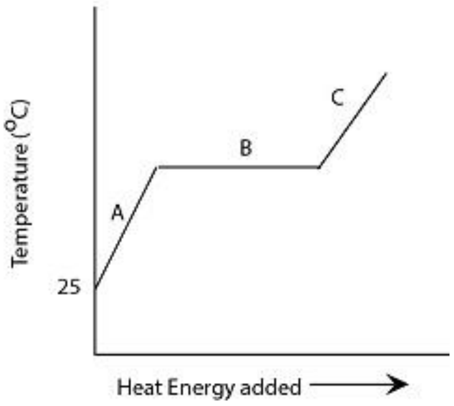
Please use the hints provided as they are designed to help you with answering the questions. Any time you see a definition, you will find the definition and related information in the hint. Good luck!

You slowly heat both liquids while plotting temperature of the liquid over time and generate a heating curve for each substance. You notice these graphs look very similar to ones you've seen in your chemistry class and remember that you can get a lot of information about a substance from a plot such as this.

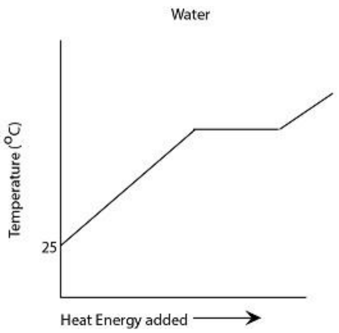
1. Identify where boiling is occurring on the heating curve generated for water.

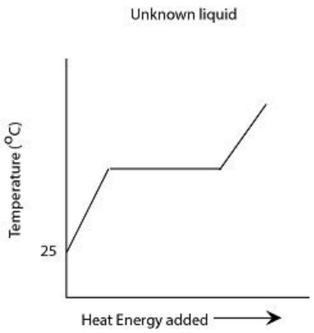


2. Identify where boiling is occurring on the heating curve generated for the unknown liquid.



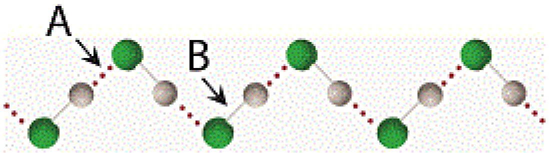
3. Which substance has a higher boiling point?





Now that you know the unknown liquid has a lower boiling point than water, you start to think about what particle level properties both of these liquids exhibit and how those properties relate to their relative boiling points.

4. Using this particle level diagram, which letter designates what is overcome to boil a substance?



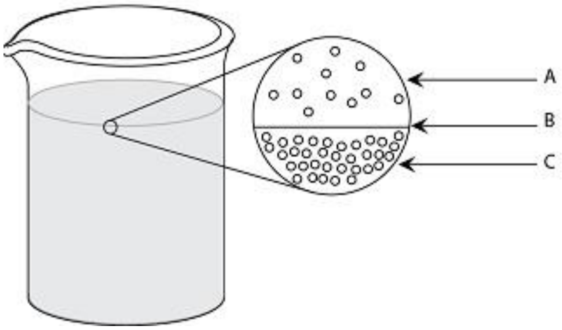
5. Is the strength of intermolecular forces of water equal to that of the unknown liquid?

a. Yes
b. No

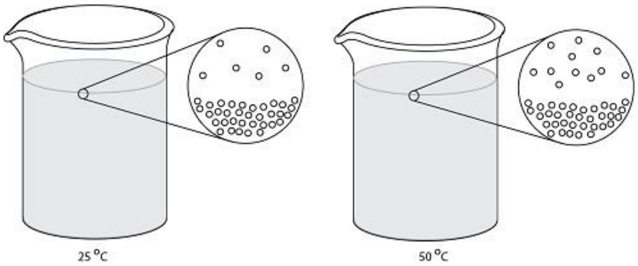
6. Explain why your answer for #5 is correct.
7. Which substance has stronger intermolecular forces: water or the unknown liquid?
8. Explain why your answer for #7 is correct.

Since water has a higher boiling point than the unknown liquid, you are certain that means water has stronger intermolecular forces than the unknown liquid. You also remember from chemistry class that all liquids have vapor pressure, but start to wonder how intermolecular forces affect the quantity of vapor particles that exist above your two liquids.

9. On the diagram, select the letter corresponding to where vapor pressure is measured.

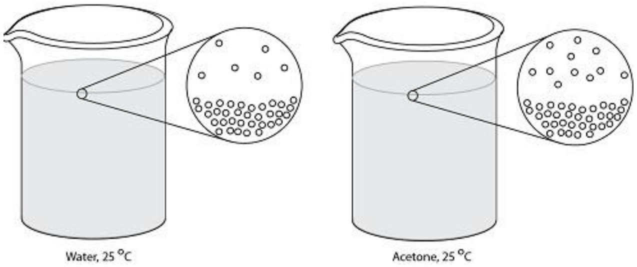


10. The figure above shows the same liquid on the particle level at different temperatures. Based on the figure, as the temperature of a liquid increases, the vapor pressure:



A. Increases
B. Decreases

11. The figure above shows different liquids on the particle level at the same temperature. Based on the figure, as intermolecular forces of pure substances increase, the vapor pressure:

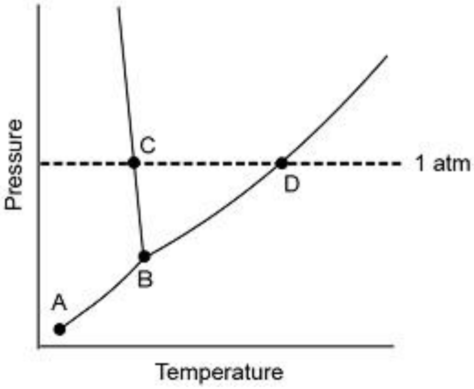


A. Increases
B. Decreases

12 Using your answers to numbers 10 and 11 explain the relationship between temperature, vapor pressure, and intermolecular forces.

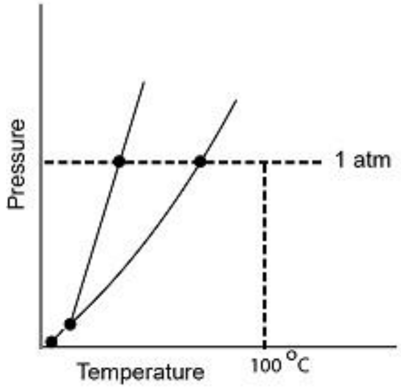
Knowing now that water has a lower vapor pressure than the unknown liquid, you want to understand how having a lower vapor pressure means water requires more energy (i.e. a higher temperature) than the unknown liquid to boil.

13. What must be true of the vapor pressure and the external pressure before a liquid will boil?
14. Explain your answer to number 13. Make sure to include why this must happen before boiling can be observed.
15. Which letter on the phase diagram corresponds to the normal boiling point of a liquid?

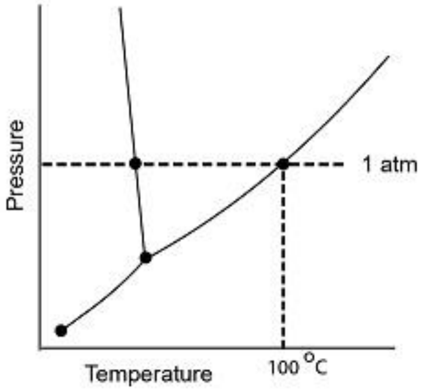


a. A
b. B
c. C
d. D

16. Which phase diagram corresponds to the unknown liquid?

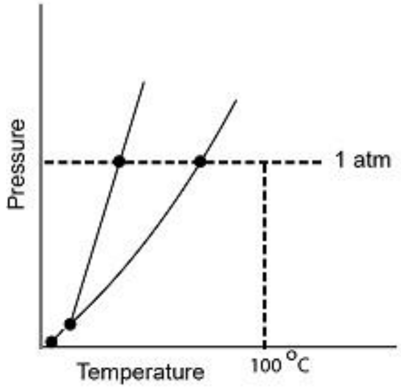


Phase Diagram 1

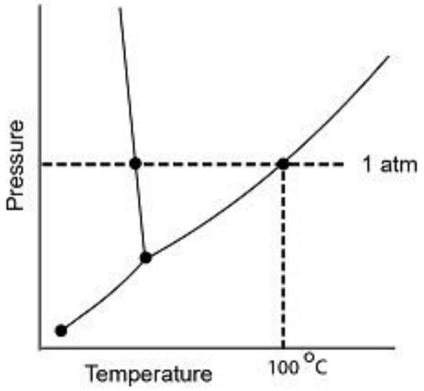


Phase Diagram 2

17. Which phase diagram corresponds to water?



Phase Diagram 1



Phase Diagram 2

18. Which liquid are you going to use to make your rice?

a. Unknown liquid
b. Water

Scoring *(1 point each)*: 11-18 (>=60%) – Scenario 1 Questions
0-10 (<60%) – repeat with a note to make sure to check the hints.

Scenario 1 Questions (5):

1. Heating and Cooling Curves
2. Intermolecular Forces (2)
3. Phase Changes
4. Phase Diagrams

Scoring *(1 point each)*: 4-5 (>=80%) Scenario 2 0-3 (<=60%) repeat

---

Scenario 2: Introduction

You are planning to make rice using a recipe that calls for a 2:1 ratio of water to rice. You measure out 2 cups of water and pour it in the pot. As you add a teaspoon of salt to the water and start the heat, you think about the ways solutions are different than pure substances, like water.

Please use the hints provided as they are designed to help you with answering the questions. Any time you see a definition you will find the definition and any other relevant information in the hint. Good luck!

1. If a teaspoon of salt weights 5 g and a metric cup is equal to 250 mL, what is the molar concentration of the salt solution in the pot? Report your answer to 5 significant figures. MW NaCl = 58.44 g/mol (2+2, fill in the blank)

2. If the molality (*m*) of the solution is actually 0.16974 *m*, what is the density of the solution (in g•mL^{-1})?

3. You go to the fridge looking for something to drink while you are cooking and see your roommate's container of juice. Select all of the possible concentration units for the container of juice you found.

a. ppm
b. %v/v
c. g
d. g/mol
e. mL
f. g/mL

After adding the salt you notice that your new solution doesn't appear to look any differently than it did before you added the salt. You can no longer see grains of salt in your pot of water so you know that on the symbolic and particle levels your solution would have to be represented differently to show what has happened. You've been studying for an upcoming chemistry exam and decide to test yourself first on symbolic representations.

4. What is the best symbolic representation for your salt solution?

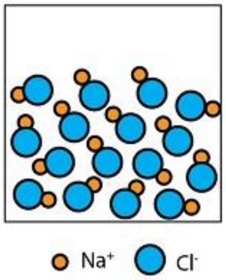
a. $Na^+(aq)$ and $Cl^-(aq)$
b. $Na(s)$ and $Cl_2(g)$
c. $Na(aq)$ and $Cl_2(aq)$
d. $NaCl(l)$

5. Methanol $(CH_3OH)(l)$ is also soluble in water. What is the best symbolic representation of an aqueous solution of methanol?

a. $CH_4(aq)$ and $H_2O(l)$
b. $CH_3^+(aq)$ and $OH^-(aq)$
c. $CH_3(aq)$ and $OH(aq)$
d. $CH_3OH(aq)$

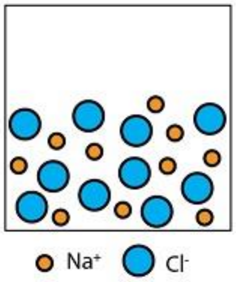
Feeling confident you understand how to represent solutions symbolically, you decide to test yourself on representing solutions on the particle level.

6. Which particle level diagram corresponds to the pure salt before it is added to the pot of water?

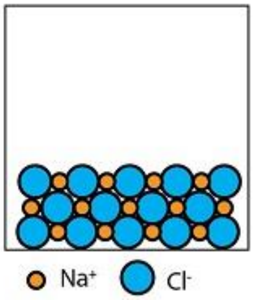


a.

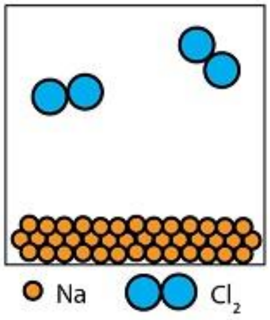
b.



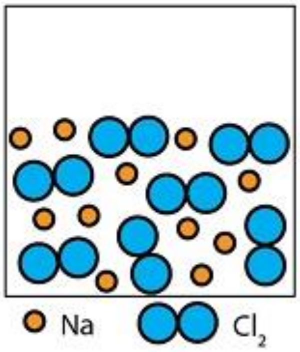
c.



d.

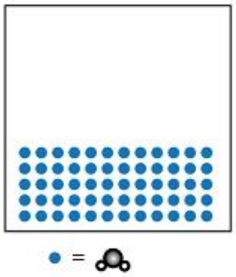


e.

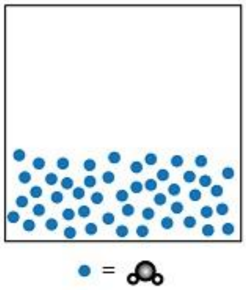


7. Which diagram corresponds to liquid water?

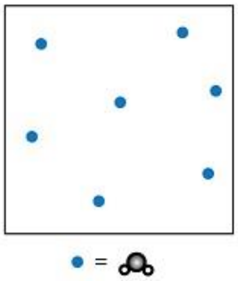
a.



b.

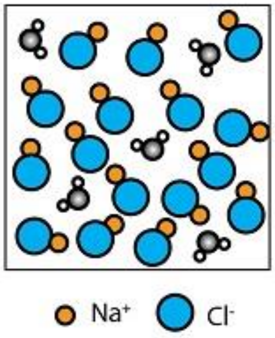


c.

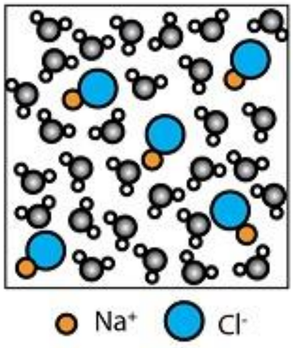


8. Which diagram corresponds to your salt solution?

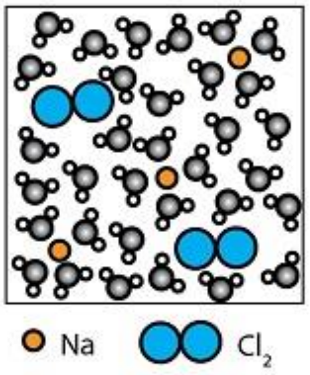
a.



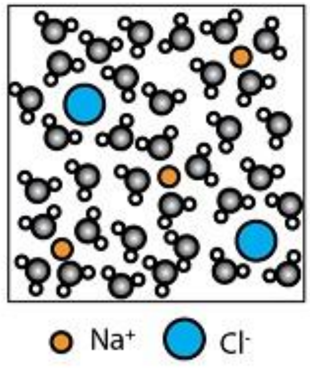
b.



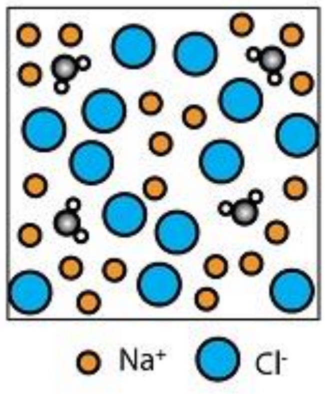
c.



d.



e.

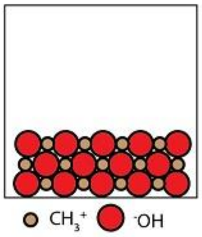


9. What happened to the distance between the sodium ions and the chloride ions from the solid to the solution?

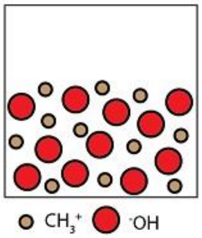
a. Increase
b. Decrease

10. Which diagram corresponds to pure methanol (CH_3OH)? (The normal boiling point of methanol is 64.70°C).

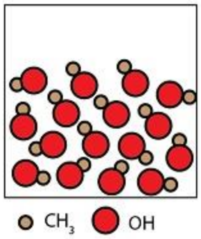
a.



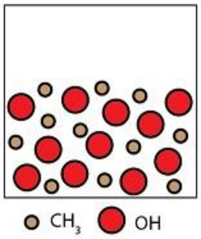
b.



c.

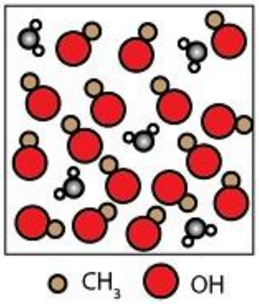


d.

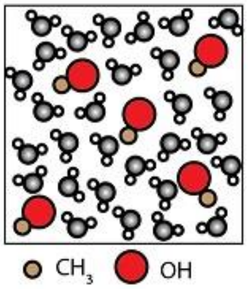


11. Which diagram corresponds to methanol in solution?

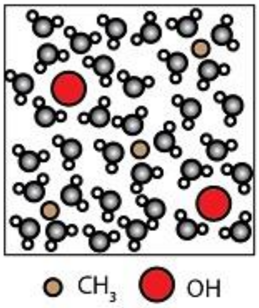
a.



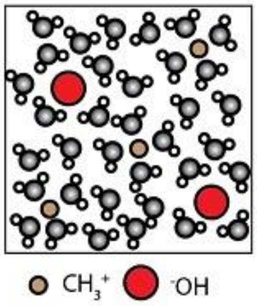
b.



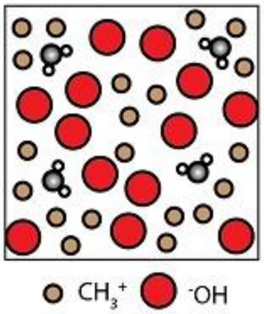
c.



d.



e.



12. What happened to the distance between the methanol molecules from the pure liquid to the solution?

   a. Increase
   b. Decrease

Based on your particle level drawings you can see that a solution is much different than a pure substance and start to think about how those differences affect the properties of a solution.

13. In addition to dispersion forces, what are the intermolecular forces present in your salt solution?
14. Qualitatively explain the forces present in a salt solution.
15. In addition to dispersion forces, what are the intermolecular forces present in your methanol solution?
16. Will methanol hydrogen bond with water?

   a. Yes
   b. No

17. Qualitatively explain the forces present in a methanol solution.

18. Based on the explanations you gave in numbers 14 and 17, do you think adding salt will make any difference in the time it takes to cook your rice?

Scoring *(1 point each)*: 11-18 (>=60%) – Scenario 2 questions
0-10 (<60%) – repeat with a note to make sure to check the hints.

Scenario 2 Questions (5):

1. Solution amounts (2)
2. Intermolecular forces in solution (3)

Scoring *(1 point each)*: 4-5 (>=80%) Scenario 3 0-3 (<=60%) repeat

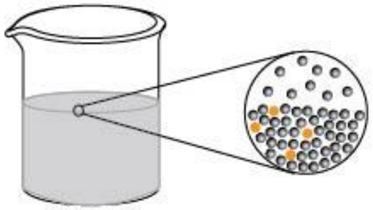
---

Scenario 3: Introduction

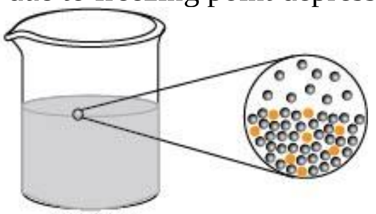
You are making rice using a boiling salt water solution. You relate this back to the chapter on freezing point depression that you just finished reading for your chemistry class. In lecture you learned that the freezing point of a solution is lower than the freezing point of the pure solvent used to make the solution. You remember that this is called freezing point depression and that it belongs to a group of phenomenon that are independent of the identity of the solute but are dependent on the quantity of solute in solution. You know that boiling point elevation and vapor pressure lowering also belong to this group and you start thinking about how you might be observing the effects of these properties as you cook.

Please use the hints provided as they are designed to help you with answering the questions. Any time you see a definition you will find the definition and any other relevant information in the hint. Good luck!

1. Which solution would have a lower freezing point due to freezing point depression?



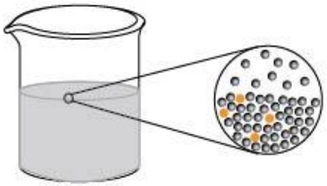
1.0 m NaCl

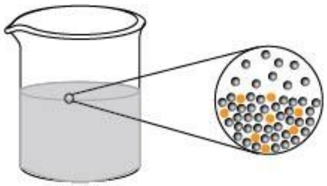


2.0 m NaCl

a. 1.0 *m* NaCl
b. 2.0 *m* NaCl

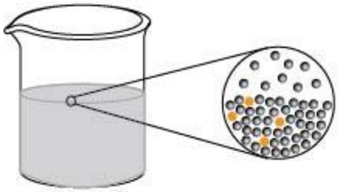
2. Which solution would have a higher boiling point due to boiling point elevation?

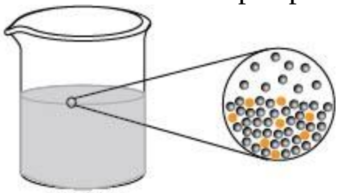

1.0 m NaCl


2.0 m NaCl

a. 1.0 *m* NaCl
b. 2.0 *m* NaCl

3. Which solution has the greatest number of water molecules in the vapor phase?

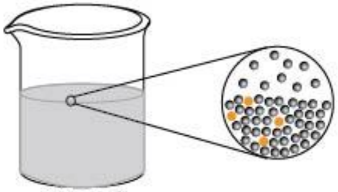

1.0 m NaCl

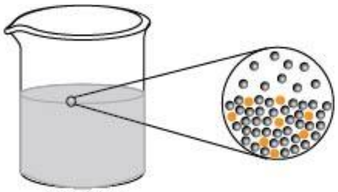

2.0 m NaCl

a. 1.0 *m* NaCl
b. 2.0 *m* NaCl
c. Both solutions have an equal number of water molecules in the vapor phase.

4. Explain why your answer to number 3 is correct.

5. Which solution has the highest vapor pressure?

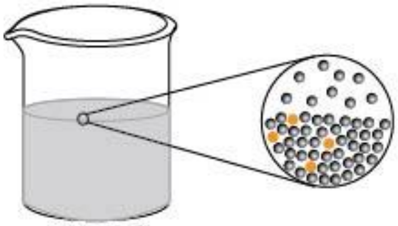

1.0 m NaCl

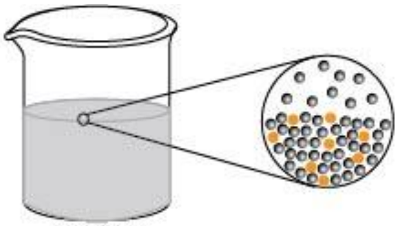

2.0 m NaCl

a. 1.0 *m* NaCl
b. 2.0 *m* NaCl
c. Both solutions have the same vapor pressure.

6. Explain why your answer to number 5 is correct.

7. Which solution has the highest boiling point?

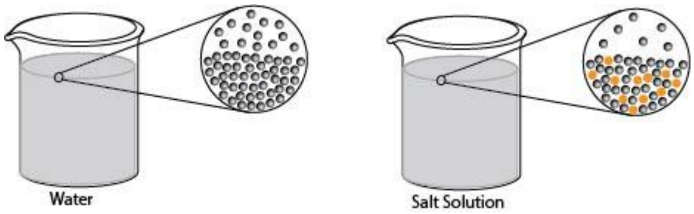

1.0 m NaCl


2.0 m NaCl

a. 1.0 *m* NaCl
b. 2.0 *m* NaCl
c. Both solutions have the same boiling point.

8. Explain why your answer to number 7 is correct.

9. Which substance has a higher vapor pressure?

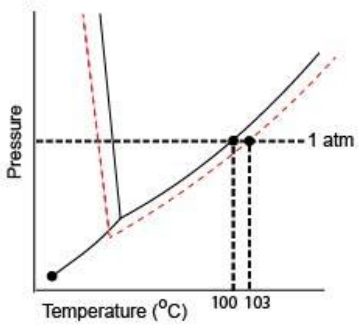


a. The salt solution has a higher vapor pressure.
b. Water has a higher vapor pressure.
c. Water and a salt solution have equal vapor pressures.

10. Explain why your answer to number 9 is correct.

Vapor pressure lowering and boiling point elevation are two examples of colligative properties. Because both are related to how much solute is present in a solution recall how vapor pressure relates to boiling point.

11. What happens when the vapor pressure equals the external pressure?
12. Using the phase diagram, what is the normal boiling point for the solvent?



13. Using the phase diagram, what is the boiling point for the solution?
14. Using the phase diagram, what is the change in temperature (ΔT_b) for the solution?
15. Did the addition of salt to the pot affect the cooking time of the rice?

a. Yes
b. No

16. Based on your answers to this activity, why do you think salt is added to water when cooking rice?

Scoring *(1 point each)*: 10-16 (>=60%) – move on
0-9 (<60%) – repeat with a note to make sure to check the hints.

Scenario 3 Questions (5):

1. Vapor pressure lowering (2)
2. Boiling point elevation (2)
3. Phase Diagrams of Solutions (1)

Scoring *(1 point each)*: 4-5 (>=80%) Final questions 0-3 (<=60%) repeat

---

Final Activity Questions (10):

1. Heating and Cooling Curves
2. Intermolecular Forces
3. Phase Changes
4. Phase Diagrams
5. Solutions Amounts
6. Intermolecular forces in solution (2)
7. Vapor Pressure Lowering
8. Boiling point elevation
9. Phases Diagrams of solutions

Macroscopic/symbolic/particulate activity

Initial Activity Questions (11):

1. Galvanic Cells
2. Cell Potential
3. System/Surroundings
4. Macroscopic – Gases
5. Symbolic – Gases
6. Particulate - Gases
7. Symbolic - Reactions
8. Nernst Equation
9. Spontaneity
10. Energy Diagrams
11. Reaction Mechanisms

Scoring (*1 point each*): 8-11 (>75%) – Scenario 3
6-7 (50-75%) – Scenario 2
0-5 (<50%) – Scenario 1

Scenario 1: Introduction

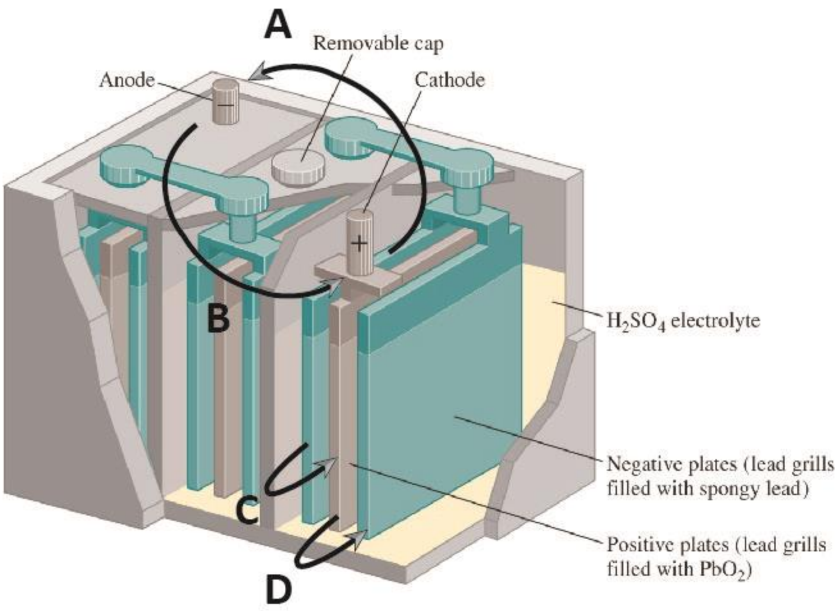
You have been chosen to test drive a hydrogen fuel cell car (referred to as fuel cell vehicle or FCV). You may have heard that these cars are more efficient and better for the environment than a car that runs on gasoline as the fuel (referred to as a standard vehicle or SV). As you walk to the new car you start to think about how this car is different than your car.

Please use the hints provided as they are designed to help you with answering the questions. Good luck!

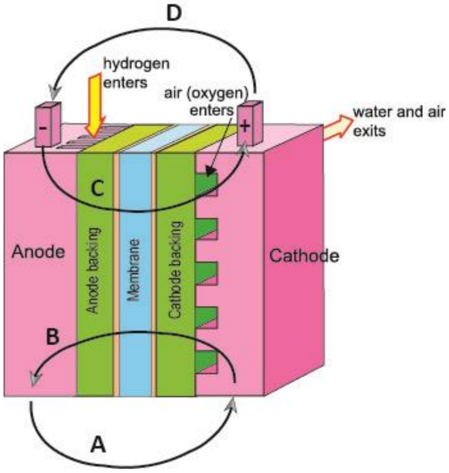
You go outside on a cold winter day to drive to school. You have recently been chosen to test drive a hydrogen fuel cell car. As you start your hydrogen fuel cell car you wonder if the temperature will affect how the car warms up compared to a summer day.

It is a particularly cold day today, and you know that sometimes a standard vehicle after sitting overnight in cold temperatures may not start. While this occurrence is actually a result of several factors, the most important is that the battery has failed. You know that the fuel cell in a fuel cell vehicle is analogous to the battery in a standard vehicle in that it produces energy, but you start to wonder what makes the fuel cell vehicle different.

1. Even though you hook up your car at the terminals which part (letter) of this image shows where a car battery produces electricity?



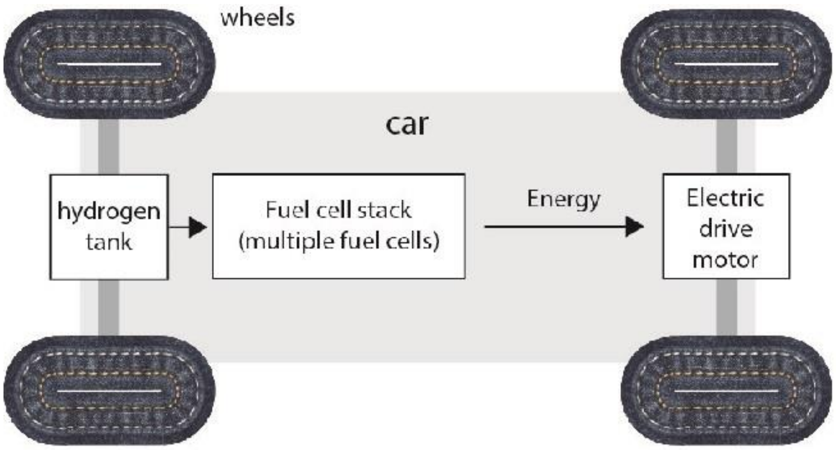
2. Which part (letter) of this image shows how a fuel cell produces electricity?



In theory, a single hydrogen fuel cell can produce 1.23 V of electricity, but in reality the output is closer to only 0.7 V of electricity.

3. Which object(s) could be powered by 0.7 V of electricity? Select all that apply. (MS)
    a. A small flash light
    b. A laptop
    c. A cell phone
    d. A house
4. Do you think 0.7 V is enough to power a car?
5. Explain your answer.
6. How could you increase the voltage produced by a single fuel cell?
    a. Increase the amount of platinum catalyst
    b. Increase the surface area of the plates
    c. Increase the number of plates
    d. Use only one plate

7. Hydrogen fuel cells can be tiny. If the average size of the fuel cell in the FCV is 200 µM how many fuel cells do you need to have an output voltage of 200 V? The amount of electricity produced by a fuel cell is dependent on both the temperature and the pressure of the **system.** Above is a schematic of the FCV. Identify the components of the **system.** Select all that apply.



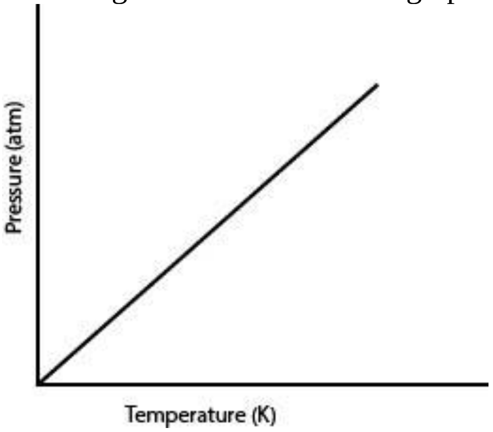
- a. Hydrogen tank
- b. Fuel cell stack
- c. Electric drive motor
- d. Wheels
- e. Car

8. Now focusing on the system, what chemical reaction is occurring in the fuel cell?
- a. Hydrogen → water vapor
- b. Hydrogen → liquid water
- c. Hydrogen + oxygen → water vapor
- d. Hydrogen + oxygen → liquid water

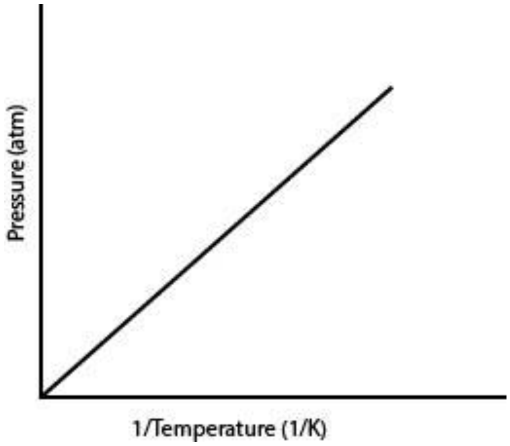
9. You notice that the fuel gauge on the FCV is showing low fuel. What does this mean?
- a. You are running low on hydrogen gas.
- b. You are running low on oxygen gas.
- c. You are running low on water vapor.

10. The fuel in your car is stored as a gas. How is the temperature of the surroundings related to the pressure of the gas? Select the correct graph that shows this relationship.

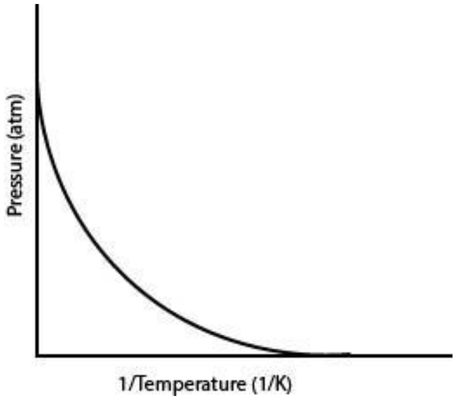
a.



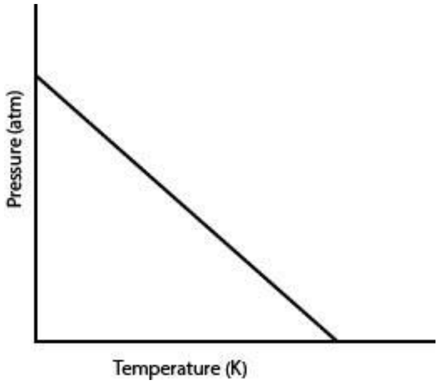
b.



c.



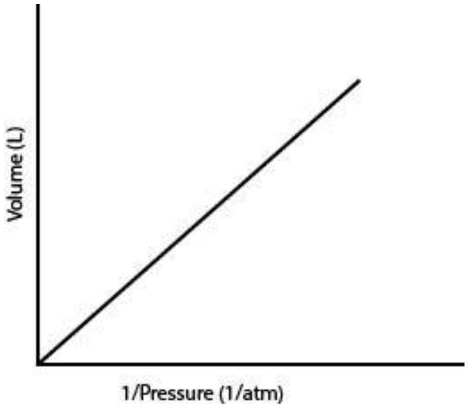
d.



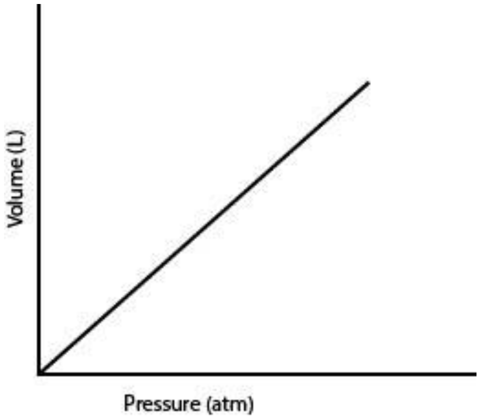
11. Today is a particularly cold day so would you expect the pressure to be higher, lower, or the same as a summer day?
    - a. Higher
    - b. Lower
    - c. Stay the same

12. Explain your answer.

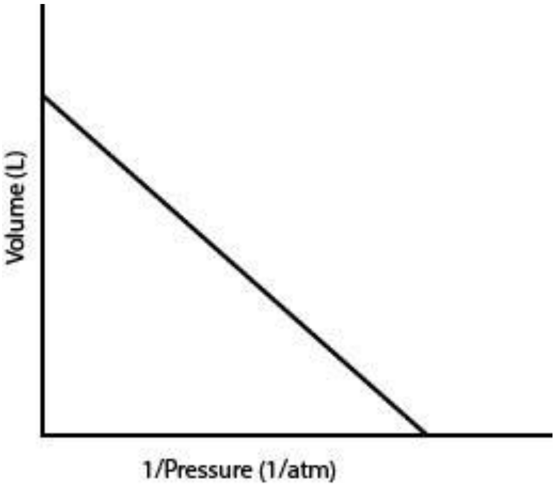
13. Fuel tanks on a FCV are flexible and adjust the volume to keep the pressure constant. What is the relationship between pressure and volume? Select the correct graph that shows this relationship.


Volume (L)
1/Pressure (1/atm)

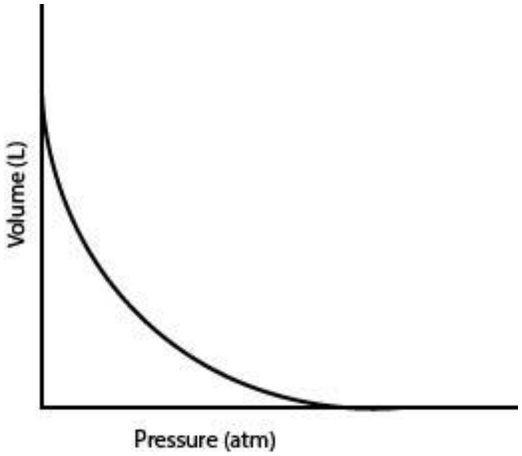
a.


Volume (L)
Pressure (atm)

b.


Volume (L)
1/Pressure (1/atm)

c.



d.

14. Today is a particularly cold day, what happens to the volume of the flexible tank?
    a. Contract
    b. Expand
    c. Stay the same
15. Explain your answer.

Scoring *(1 point each)*: 9-15 (>=60%) – move on
0-8 (<60%) – repeat with a note to make sure to check the hints.

Scenario 1 Questions (4):

1. Galvanic Cells (1)
2. Cell Potential (1)
3. System/Surroundings (1)
4. Macroscopic – Gases (1)

Scoring *(1 point each)*: 3-4 (>=75%) – Scenario 2 0-2 (<=50%) repeat

---

Scenario 2: Introduction

Your focus in this scenario will be on symbolic representations which will involve some calculations.

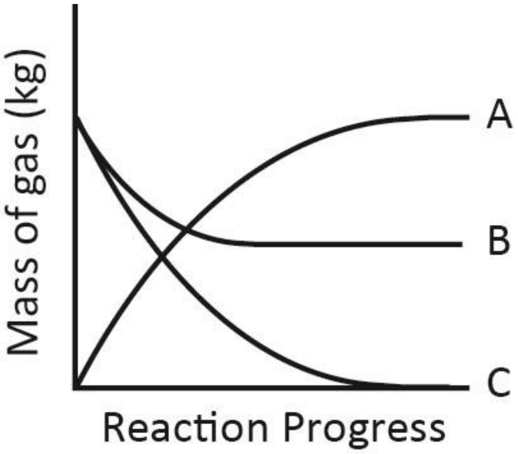
Please use the hints provided as they are designed to help you with answering the questions. Good luck!

You have been chosen to test drive a hydrogen fuel cell car (referred to as fuel cell vehicle or FCV). You may have heard that these cars are more efficient and better for the environment than a car that runs on gasoline as the fuel (referred to as a standard vehicle or SV). The hydrogen used in your car is stored in a flexible tank that keeps the pressure at 10,000 psi. The reactant gases undergo catalytic reactions that produce energy that powers your car. The energy output is less than 60% efficient and results in a fuel economy

of roughly 70 mpk (miles per kilogram of hydrogen). The car is rated for 300 miles per tank of gas with a maximum temperature rating of 85°C (185°F). Based on your experience in your chemistry class you are going to figure out how big the tank is.

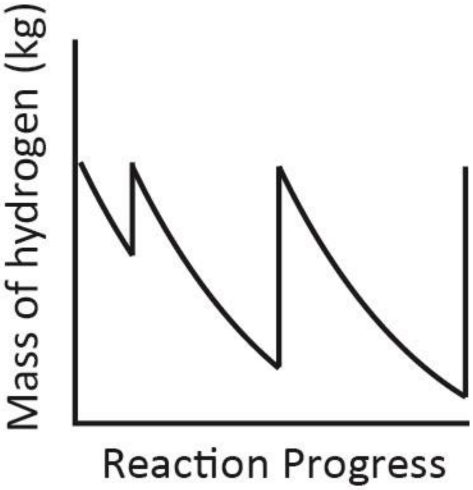
The hydrogen gas used for fuel is expensive. Thinking about the efficiency of the fuel cell, you contemplate the reaction between hydrogen and oxygen that allows your car to run.

1. Ignoring the catalyst, what is the symbolic representation (balanced equation) for the reaction that occurs between hydrogen and oxygen in the fuel cell?
   a. $H(g) + O(g) \rightarrow H_2O(g)$
   b. $2H_2(g) + O_2(g) \rightarrow 2H_2O\ (g)$
   c. $H_2(g) + O_2(g) \rightarrow H_2O\ (g)$
   d. $2H(g) + O(g) \rightarrow 2H_2O\ (g)$
   e. $H(g) + O(g) \rightarrow H_2O\ (l)$
   f. $2H_2(g) + O_2(g) \rightarrow 2H_2O\ (l)$
   g. $H_2(g) + O_2(g) \rightarrow H_2O\ (l)$
   h. $2H(g) + O(g) \rightarrow 2H_2O\ (l)$
2. What type of reaction is your previous answer?
   a. Double displacement reaction
   b. Combustion reaction
   c. Decomposition reaction
   d. Oxidation-reduction reaction
3. Fill in the various elements and coefficients of the reduction reaction taking place. You **must** enter a numerical value for a coefficient (including if the coefficient is 1, but remember this can also be 0).
4. Fill in the various elements and coefficients of the oxidation reaction taking place. You **must** enter a numerical value for a coefficient (including if the coefficient is 1, but remember this can also be 0).
5. What substance corresponds to each letter?

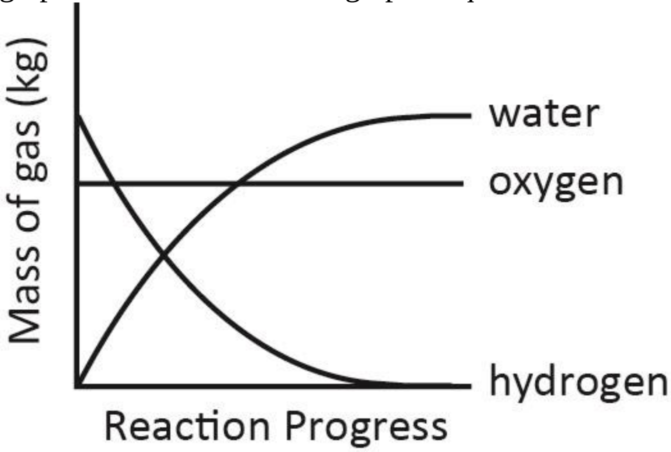


   a. Hydrogen
   b. Water vapor
   c. Oxygen

6. Looking at the image in the previous question, which letter corresponds to the limiting reactant?
    a. A
    b. B
    c. C
7. How many times did you refuel the FCV?



8. Describe how this graph is different from the graph in question 5.



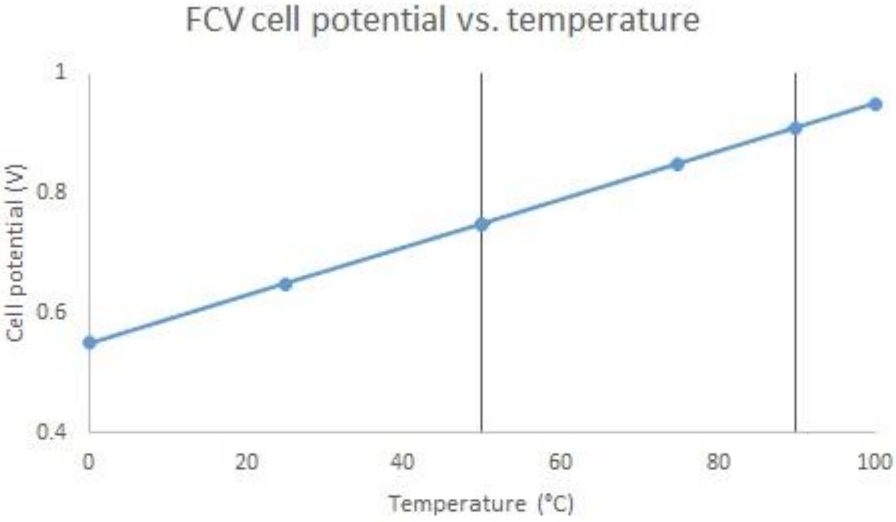
    a. Oxygen is the limiting reactant
    b. Oxygen is being constantly supplied
    c. Oxygen is not present
    d. Oxygen is now the product

In chemistry class you've been learning about Galvanic cells and remember that the definition of a galvanic cell is "an electrochemical cell that generates electricity by means of a spontaneous redox reaction" (p. 669).

9. Is the fuel cell in your car a galvanic cell?
    a. Yes
    b. No
10. Explain your answer.

11. Galvanic cells use spontaneous oxidation-reduction reactions to produce electrical energy. The amount of energy produced by the cell that "is available to do work" is called Gibbs free energy (p. 644). Given a constant energy output and using the sign convention for Gibbs free energy that you are familiar with what's the relationship between Gibbs free energy and efficiency?
    a. The more positive the Gibbs free energy the more efficient the reaction
    b. The more negative the Gibbs free energy the more efficient the reaction
    c. Gibbs free energy is not related to the reaction efficiency
12. Each car comes with an efficiency rating. You know that your car runs at about 80% efficiency meaning that 80% of the hydrogen fuel can successfully be converted to usable energy. What is one reason the efficiency is not 100%?
13. Calculate change in Gibbs free energy for one mole of the system of hydrogen and oxygen combining to form water vapor at room temperature (25°C) where the change in enthalpy is -241.8 kJ and the change in entropy is -147.3 J/K.
    a. 43680 kJ
    b. 3441 kJ
    c. -197.9 kJ
    d. -238.1 kJ
14. Is this reaction spontaneous based on the number you calculated in the previous problem?
    a. Yes
    b. No

15. As you know from class, Gibbs free energy is related to cell potential. Use the plot to describe what's happening in a normal FCV?



    a. At the normal FCV operating temperature, the cell potential is lower and the efficiency is lower.
    b. At the normal FCV operating temperature, the cell potential is lower and the efficiency is higher.
    c. At the normal FCV operating temperature, the cell potential is higher and the efficiency is lower.
    d. At the normal FCV operating temperature, the cell potential is higher and the efficiency is higher.

Gases behave ideally at sufficiently low pressure and high temperature.

16. Using the information above, assuming the tank in your car has a maximum temperature rating of 125°C, how big is the tank in your car if you have 3.0 kg of hydrogen in the tank?

   a. 0.14 L
   b. 9.8 L
   c. 45 L
   d. 72 L

17. Assuming just a volume of 1.00 L, how many hydrogen molecules are in this tank at STP? 1.8×10^{22} hydrogen molecules

   a. 5.9×10^{22} hydrogen molecules
   b. 6.2×10^{24} hydrogen molecules
   c. 2.0×10^{25} hydrogen molecules

18. Assuming just a volume of 1.00 L, how many hydrogen molecules are in this tank at 10,000 psi (680 atm)?

   a. 1.2×10^{25} hydrogen molecules
   b. 4.0×10^{25} hydrogen molecules
   c. 1.8×10^{26} hydrogen molecules
   d. 5.9×10^{26} hydrogen molecules

19. At STP hydrogen molecules are approximately 3800 pm far apart and at 680 atm they compress to approximately 440 pm far apart. How many times closer together are the molecules at high pressure than at low pressure?

20. If hydrogen is stored at 10,000 psi in your vehicle, is it realistic to consider hydrogen as an ideal gas at this pressure?

   a. Yes
   b. No

21. Explain your previous answer.

Scoring *(1 point each)*: 13-21 (>60%) – move on
0-12 (<60%) – repeat with a note to make sure to check the hints.

Scenario 2 Questions (4):

1. Symbolic - Reactions (1)
2. Nernst Equation (1)
3. Spontaneity and Temperature (1)
4. Symbolic - Gases (1)

Scoring *(1 point each)*: 3-4 (>=75%) – Scenario 3 0-2 (<=50%) repeat

---

Scenario 3: Introduction

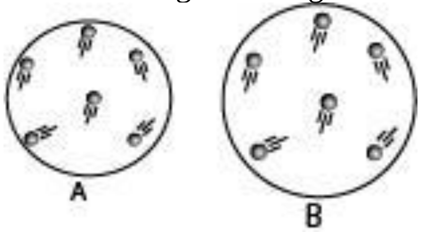
Your focus in this scenario will be on the particulate level.

Please use the hints provided as they are designed to help you with answering the questions. Good luck!

You have been chosen to test drive a hydrogen fuel cell car (referred to as fuel cell vehicle or FCV). You may have heard that these cars are more efficient and better for the environment than a car that runs on gasoline as the fuel (referred to as a standard vehicle or SV). Today you are car-pooling to chemistry class with a friend and discussing your upcoming chemistry exam on energy. Your friend says that because your car is using energy to drive, the reactions occurring inside the fuel cell must all be exothermic. He says lots of chemical energy gets released when bonds are broken due to the energy stored in the bonds that the car then coverts into electrical energy. You tell your friend that you remember hearing your chemistry professor say that even though a reaction may overall be exothermic, energy is still required to break the bonds of the reactants involved in the reaction before the atoms can rearrange and form new bonds. You aren't sure who is right, but start to discuss both the enthalpy and entropy involved in the reactions occurring in your FCV.

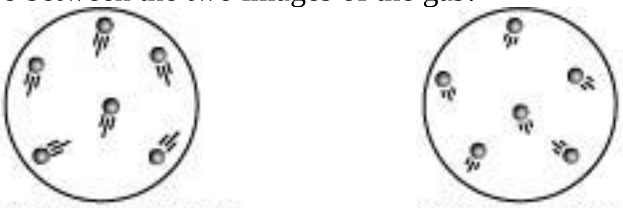
A reaction is the result of molecular collisions. Reactions cannot occur without sufficient kinetic energy and proper orientation of the molecules. As the temperature increases the gas particles gain more energy which causes a greater number of collisions. If we compare the reaction inside a fuel to a much simpler process, the combustion reaction of hydrogen, $H_2 + \frac{1}{2} O_2 \rightarrow H_2O$ then $\Delta G° = -228.6$ kJ/mol .

1. What is the difference between the two images of the gas?



a. At high pressures molecules are closer together and are less likely to collide.
b. At high pressures molecules are farther apart and are less likely to collide.
c. At high pressures molecules are closer together and are more likely to collide.
d. At high pressures molecules are farther apart and are more likely to collide.

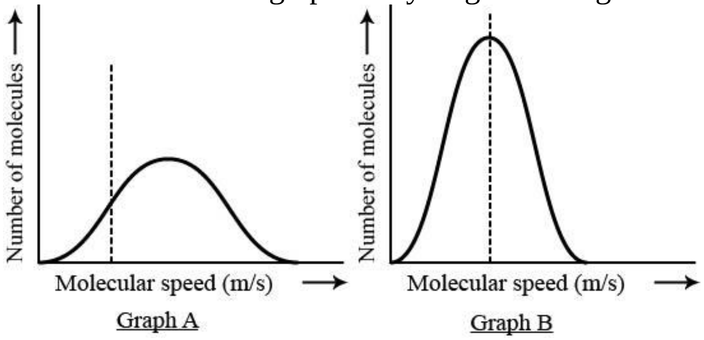
2. What is the difference between the two images of the gas?


High temperatures Low temperatures

a. At higher temperatures the particles, on average, are moving fast.

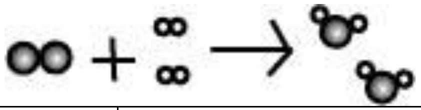
b. At lower temperatures the particles, on average, are moving fast.
c. At higher temperatures every particle is moving fast.
d. At lower temperatures every particle is moving fast.

3. What is the difference between the two graphs if hydrogen is the gas in both graphs?



a. Graph A has a greater fraction of gas particles moving at or above the marked speed.
b. Graph B has a greater fraction of gas particles moving at or above the marked speed.
c. Graph A and Graph B have the same fraction of gas particles moving at or above the marked speed.

4. Which particles must collide for the above reaction to start?
a. 1 molecule of H_2 and 1 molecule of O_2
b. 1 molecule of H_2 and 1 O atom
c. 1 H^+ ion and 1O^{2-} ion
d. 2 H^+ ions and 1 O^{2-} ion



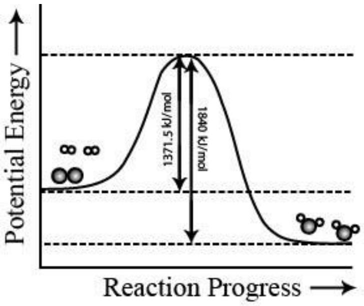
Type of bonds broken	Number of bonds broken	Bond enthalpy (BE) (kJ/mol)	Energy change (kJ/mol)
H-H	2	436.4	872.8
O=O	1	498.7	498.7
Type of bonds formed	Number of bonds formed	Bond enthalpy (BE) (kJ/mol)	Energy change (kJ/mol)
O-H	2	460	1840

$\Delta H° = \Sigma BE \text{ (reactants)} - \Sigma BE \text{ (products)}$

$\Delta H° = (872.8 \text{ kJ/mol} + 498.7 \text{ kJ/mol}) - 1840 \text{ kJ/mol} = -469 \text{ kJ/mol}$

5. Is this reaction endothermic or exothermic?
a. Exothermic because the enthalpy is negative.
b. Exothermic because the enthalpy is positive.
c. Endothermic because the enthalpy is negative.
d. Endothermic because the enthalpy is positive.

6. Another way to display the information in the table above is with an energy diagram. Looking at this energy diagram, how much energy is needed to reach the transition state?



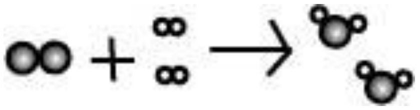
a. -469 kJ/mol
b. 1371.5 kJ/mol
c. 1840 kJ/mol
d. Not enough information

7. Where does this value come from?

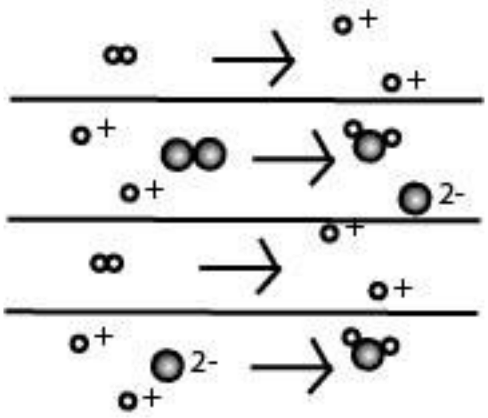
8. The Ea for this reaction is not 1371.5 kJ/mol. What does that mean in terms of the energy diagram below? Include in your answer an explanation of why the Ea and the ΣBE are not the same in terms of the intermediate(s) formed.

9. If the rate law for this reaction is rate = k $[H_2][O_2]$, select the most plausible mechanism for this reaction.

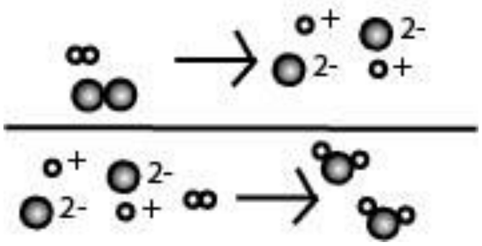
a.



b.



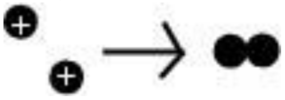
c.



d. None of these

10. Explain your answer.

11. Simplifying this process and just looking at forming one H-H bond, what is the sign of the entropy change for this reaction?



- a. Positive because the number of microstates is reduced.
- b. Positive because the number of microstates is increased.
- c. Negative because the number of microstates is reduced.
- d. Negative because the number of microstates is increased.

12. Is this reaction spontaneous?

- a. Yes because the reaction decreases the entropy of the universe.
- b. Yes because the reaction increases the entropy of the universe.
- c. No because the reaction decreases the entropy of the universe.
- d. No because the reaction increases the entropy of the universe.
- e. Cannot be determined from the information given.

13. What is the sign of the enthalpy change for this reaction?

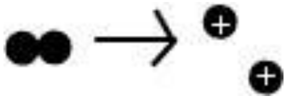
- a. Positive because heat is released from the system.
- b. Positive because heat is absorbed from the surroundings.
- c. Negative because heat is released from the system.
- d. Negative because heat is absorbed from the surroundings.

14. Based on your answers to 14 and 15 above, is this reaction exothermic or endothermic?

- a. Exothermic
- b. Endothermic

15. Explain you answer.

16. Reversing this process and thinking about breaking one H-H bond. Is this reaction exothermic or endothermic?



- a. Exothermic
- b. Endothermic

17. Explain your answer.

18. Based on your answers to 14-17, who was right? You or your friend?

- a. You
- b. Your friend

19. Explain your answer.

Scoring *(1 point each)*: 12-19 (>60%) – move on
0-11 (<60%) – repeat with a note to make sure to check the hints.

Scenario 3 Questions (4):

1. Particulate - Gases
2. Energy Diagrams
3. Reaction Mechanisms
4. Energy/Bonding

Scoring *(1 point each)*: 3-4 (>=75%) – Final questions 0-2 (<=50%) repeat

---

Final Questions (12):

1. Galvanic Cells
2. Cell Potential
3. System/Surroundings
4. Macroscopic – Gases
5. Symbolic – Gases
6. Particulate - Gases
7. Symbolic - Reactions
8. Nernst Equation
9. Spontaneity
10. Energy Diagrams
11. Reaction Mechanisms
12. Energy and Bonding

D.2 Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items

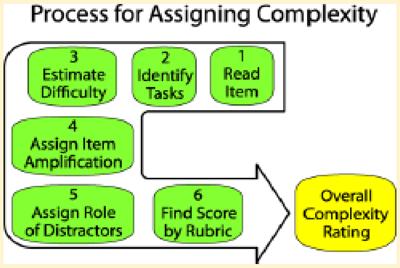
JOURNAL OF CHEMICAL EDUCATION
Article
pubs.acs.org/jchemeduc

Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items

Jeffrey R. Raker,[†] Jaclyn M. Trate,[‡] Thomas A. Holme,[†] and Kristen Murphy*,[‡]

[†]Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States
[‡]Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201, United States

ABSTRACT: Experts use their domain expertise and knowledge of examinees' ability levels as they write test items. The expert test writer can then estimate the difficulty of the test items subjectively. However, an objective method for assigning difficulty to a test item would capture the cognitive demands imposed on the examinee as well as be assignable by any domain expert familiar with the examinee group. One such instrument for assigning objective complexity of general chemistry exam items has already been reported. A revised instrument for assigning objective complexity of organic chemistry exam items is presented including the reliability and validity studies of the rubric.



KEYWORDS: *Testing/Assessment, Organic Chemistry, Chemical Education Research*
FEATURE: Chemical Education Research

INTRODUCTION

A fundamental premise in any form of content testing is that the test items constitute a set of cognitive tasks required of students. In 2011, Knaus et al. reported the development of an instrument for the expert-based, objective determination of cognitive complexity for general chemistry exam items.[1] The basis of the instrument was the identification of concepts or skills needed to answer an exam item, the relative difficulty (i.e., easy, medium, or hard) for each concept or skill, and the interactivity of the concepts and skills (i.e., nonsignificant, basic, or complex). Each concept or skill and the interactivity were assigned a numerical value that when summed represents a numerical measure of objective exam item complexity. Knaus et al. further report inter-reliability statistics for the complexity measures and associated their objective complexity with general chemistry exam item performance and student mental effort ratings (i.e., subjective complexity), thereby establishing validity and reliability for the complexity rubric for general chemistry exam items.

Cognitive complexity is a concept that was first proposed by Bieri[2] where "cognitive complexity—simplicity" designated "the degree of differentiation of the construct system" (ref 2, p 263), or a more cognitively complex system differentiates well where a cognitively simple system poorly differentiates. Cognitive complexity was further quantified by examining the number of independent constructs as proposed by Crockett.[3] Cognitive complexity has been examined with mathematics items,[4] on science[5] and chemistry[6] assessments, between multiple-choice and constructed-response formats,[7] and in relation to spatial tasks,[8] to name a few. Within cognitive complexity one can specifically focus on the complexity related to a task where task complexity is predicated on considering the cognitive demand a test item (the cognitive task) imposes on students. This idea expands on the foundational concepts of cognitive load theory wherein the cognitive demand on students manifests primarily in the working memory.[9−13] The fundamental concepts of cognitive load have factored into a number of theories of cognition and learning;[14] therefore, the combination of information processing and cognitive load can serve as a useful organizational theme for understanding task complexity of chemistry test items.[6,15,16]

The mental workload or cognitive demand on the working memory is separated into three components: the intrinsic, extraneous, and germane cognitive load. The extraneous cognitive load is the component that can be altered, depending on the learner and the environment presented to the learner. For example, presenting multiple representations to explain a concept that require a learner to integrate information can pose a higher extraneous cognitive load as opposed to presenting a single, integrated figure with explanation. Intrinsic cognitive load is the inherent difficulty of the material from the learner perspective and cannot be altered. The germane cognitive load is the portion of the load remaining for processing in the working memory. The "intrinsic cognitive load refers to the internal complexity of the task being attempted" and this should be able to be estimated by experts both in the domain and with the level of learner and their understanding.[17] The tasks must then be considered in terms of the elements of knowledge required to complete the task from the perspective of the learner. These elements have varying levels of difficulty and may be related to one another in the successful completion of the task. Therefore, both the estimation of the difficulty of

Published: September 18, 2013

dx.doi.org/10.1021/ed400373c | J. Chem. Educ. 2013, 90, 1290−1295

Table 1. Chronology of the Rubric Development Process

Time	Location	Intent of Focus Group
Spring 2010	ACS National Meeting, San Francisco, CA	Discuss visual−spatial components of common organic chemistry items and experts ability to relate to what students interpret from these representations
Fall 2010	ACS National Meeting, Boston, MA	(1) Discuss the general chemistry rubric and how to adapt for organic chemistry (2) Trial the new ideas for assigning complexity with organic chemistry items
Winter 2010	Combined Southwestern and Southeastern Regional Meetings of the ACS, New Orleans, LA	Testing of the complexity rubric with two groups of participants (faculty and graduate students). Rated components of approximately 50 organic chemistry exam items
Spring 2011	ACS National Meeting, Anaheim, CA	Testing of the complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items
Fall 2011	ACS National Meeting, Denver, CO	Testing of the complexity rubric with faculty; Rated components of approximately 50 organic chemistry exam items.
Winter 2011	Southwestern Regional Meetings of the ACS, Austin, TX	Testing of the complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items
Spring 2012	ACS National Meeting, San Diego, CA	Testing of the numeric complexity rubric with faculty; Rated components of approximately 50 organic chemistry exam items
Summer 2012	Biennial Conference on Chemical Education, Pennsylvania State University	Testing of the numeric complexity rubric with faculty. Rated components of approximately 50 organic chemistry exam items

the elements of knowledge and their interactivity are key to a successful estimation of the cognitive complexity of an exam item. Fundamentally, this estimation is based on the expertise of the raters' knowledge of the knowledge of the learner.

Initial attempts to use an unmodified version of the general chemistry complexity rubric on multiple-choice organic chemistry exam items found that expert raters were unable to directly apply this rubric to exam items in organic chemistry. Through a series of focus groups with organic chemistry instructors, a portion of those who use and write organic chemistry examinations for the ACS Examinations Institute, a modified complexity rubric has been designed that offers a valid and reliable instrument for multiple-choice organic chemistry exam items. This paper will describe: (i) the process used to develop the modified instrument, (ii) a method for using the instrument to evaluate complexity of exam items, and (iii) studies to establish instrument validity and reliability.

■ INSTRUMENT DEVELOPMENT

Over the course of three years, eight focus groups were held at national and regional meetings of chemists and chemical educators. The location, time, and details about the various focus groups are listed in Table 1.

Two major themes related to ways to enhance the complexity rubric emerged during these activities. First, the "concept/skill interactivity" factor of the general chemistry complexity rubric was not supported by the organic chemistry workshop participants. The interactivity between the concepts and skills necessary to answer an exam item was associated with an idea participants defined as "amplification". The term amplification attempts to capture the sense that students confront the learning of the many chemical reactions they encounter in organic chemistry using strategies that cluster similar reactions that may be quickly recognized, or require more effort to recognize. As such the rankings were set as easy, medium, and hard levels rather than nonsignificant, basic, and complex, which were used in the general chemistry complexity rubric. Discussions during the development of the instrument rarely included methods of translating qualitative information into a numeric rating system. As such, when the discussion of "amplification" began at the first focus group, the association of this concept as what students must do with elements of knowledge to successfully solve exam items was the predominate feature of this component, rather than any mathematical operation or translation into a numeric rating. Participants, drawing on their instructional experiences, may have inferred a multiplicative factor to describe the interactivity and this differs from the additive factor as used in the general chemistry complexity rubric.

The second rubric enhancement was related to ways to characterize answer options for the multiple-choice items. This theme emerged spontaneously at the second focus group and was routinely reaffirmed by subsequent raters as an important inclusion in a rubric for assigning complexity to organic chemistry test items. For general chemistry, multiple-choice exam items consist of a task prompt and possibly a table, graph, or representation; an answer can often be derived independent of the exam item's multiple answer options. In other words, a student most often can develop an answer prior to looking for that answer in the item's answer options. This is not as often the case with multiple-choice organic chemistry exam items. Some exam items could be solved with the "selection" process used in general chemistry; however, some items required an "elimination" of answer options to determine the correct answer and some items required an "evaluation" of every answer option to determine the correct, best answer. Elimination and evaluation of options may represent multiple-choice test taking strategies for any content topic when a "selection" answer is not known, but elimination and evaluation are not optional test-taking strategies for the evaluated multiple-choice organic chemistry exam items; they were required.

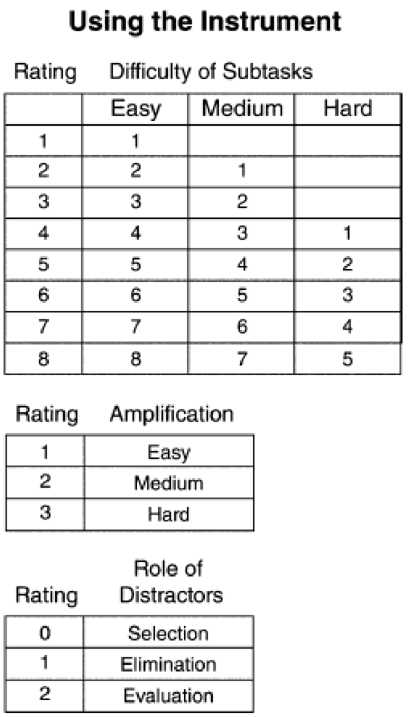
As an example of an elimination item on an organic chemistry test, consider a question that requires a student to determine which compound (from the multiple-choice responses) produces a given infrared spectrum. To answer this question, an examinee generally needs to determine the functional groups present (or not present) in the infrared spectrum and eliminate answer options based on this information. Answer options are eliminated based on the information inferred from the spectrum until only one answer option, the correct option, remains.

For an example of an evaluation item, consider a question that expects a student to determine which of a series of reactions leads to the formation of a desired product. To answer this question, an examinee must normally evaluate each of the series of reactions to determine which would lead to the formation of the desired product. In comparing evaluation to elimination, elimination is the use of information in the task prompt to eliminate answers, whereas evaluation is the use of information in the answer options to determine the correct answer.

To produce a more functional complexity rubric for applications to organic chemistry test items, each of these concerns must be considered. While the workshop participants tended to view an "amplification factor" as multiplying the complexity (e.g., "this makes the questions twice as hard") in terms of the concepts of working memory and cognitive load, this impression is not consistent with learning theories, so the amplification factor is considered as an additive component of complexity. This choice may be at odds with the linguistic prompts associated with amplification, but the label was chosen by a group of experts within the content domain. The idea of making an exam item more difficult by including more elements maps to their cognitive structure of the domain; therefore, maintaining a word that prompts a useful categorization is arguably more important than the specific mathematical construct this factor holds within the model for assigning the complexity. Next, the adjustment of the complexity rubric model to account for the differing role of distractors is also considered an additive component of complexity. Items that can be solved using only selection will have an additive value of "0", thus maintaining coherence with the general chemistry rubric. Items that use elimination have an additive value of "1", and items that required evaluation of responses have an additive value of "2". This ranking of the role of distractors assumes that the cognitive demand of eliminating answer options is less onerous than evaluating answer options. Each of these factors will be discussed in more detail in the next section on using the instrument to evaluate exam items.

Figure 1 presents the rubric by which the cognitive complexity of an organic chemistry exam item or other chemistry task can be determined. Analysis of an exam item is achieved through a seven-step process in which the rater must

1. Analyze the item.
2. Determine the subtasks (i.e., student processes) used in solving the problem as "what the student needs to do" and "what the student needs to recognize".
3. Estimate from the perspective of a student the relative difficulty (i.e., easy, medium, or hard) of each of the subtasks.
4. Estimate from the perspective of a student how difficult (i.e., easy, medium, or hard) the extent of the interactivity of the subtasks required to answer the problem and assigns an amplification score from the rubric.
5. Determine the role of the distractors (i.e., selection, elimination, or evaluation) and assign a score from the rubric.
6. Use the rubric to assign a score to each subtask process.
7. Determine the overall complexity rating by summing all of the values from steps 4–6.



Using the Instrument

Rating	Difficulty of Subtasks		
	Easy	Medium	Hard
1	1		
2	2	1	
3	3	2	
4	4	3	1
5	5	4	2
6	6	5	3
7	7	6	4
8	8	7	5

Rating	Amplification
1	Easy
2	Medium
3	Hard

Rating	Role of Distractors
0	Selection
1	Elimination
2	Evaluation

Figure 1. Cognitive complexity rating rubric for multiple-choice organic chemistry test items.

The design of the "difficulty of subtasks" component of this rubric follows that of the previously published rubric in general chemistry[1] where overlap exists between the easy, medium, and hard subtask ratings. This feature allows for two or more raters to come to similar complexity values through different parsing of the subtasks. For example, one rater may state that a given problem is composed of "three" easy tasks, whereas another rater may state that a given problem is composed of "two" medium tasks. In both these instances, the overall subtask complexity score would be 3. Both sets of subtasks and corresponding difficulty ratings lead to the same complexity score. This feature of the rubric allows for greater inter-rater reliability, as will be discussed in the next section.

As noted earlier, when compared to the general chemistry complexity rubric, this new organic chemistry complexity rubric renames the "interactivity" factor to the "amplification" factor. Because both factors are additive, the difference in labels can be viewed as semantic, as is justified by the difference in language usage of groups of chemists with different subdisciplinary specialization. In determining instrument reliability, several numeric models of the rubric were tested to determine whether the amplification factor should be considered a multiplicative factor; these models provided similar reliability statistics. Therefore, in an effort to keep the organic chemistry complexity rubric similar to the general chemistry complexity rubric, the amplification factor adds a value of 1, 2, or 3 to the overall problem complexity corresponding to easy, medium, or hard. Note that the organic chemistry complexity rubric does not include a 0 value corresponding to "nonsignificant" amplification as was the case with the general chemistry rubric. This difference suggests a difference in perspective among instructors in the two types of courses, where organic chemists perceive that test items in their subject inevitably lead to having students piecing together concepts.

Within this template, Figure 2 provides an example of the assignment of complexity for a multiple-choice organic chemistry item. For this item, which seeks to determine the ability of students to identify an aromatic molecule, the example shows the first analysis step as one way that the required cognitive tasks can be parsed. It is important to realize that, while this example is drawn from workshop participants, it is

Step 1. Analyze the item.

Which compound or ion is aromatic?

A 
B 
C 
D 

Steps 2 and 3. Determine the subtasks necessary for a student to answer the question (Step 2) and assign a corresponding relative difficulty value to each one (Step 3).

Subtasks	Level of Difficulty
Recall the definition of an aromatic compound	Easy
Determine if each structure has a cyclic array of p-orbitals	Medium
Determine if each structure has requisite "$4n + 2$" electrons in the p-orbitals	Medium

Step 4. Next, consider the subtasks identified in Step 2 and determine their level of amplification (easy, medium, or hard).

Description	Amplification (Rating)
These subtasks together are the basic algorithm for determining aromaticity of a compound	Easy

Step 5. Next, consider the distractors in the problem and determine the role of the distractors in answering the problem (selection, elimination, or evaluation)

Description	Role of Distractors (Rating)
Distractors are eliminated based on the definition of an aromatic compound	Elimination

Steps 6 and 7. Use the rubric to determine the rating of the subtasks (Step 6) and the overall objective complexity rating for this item (Step 7).

Components	Count	Component Set Rating
Easy Subtasks	1	1
Medium Subtasks	2	3
Hard Subtasks	0	0
Amplification	NA	1
Role of the Distractors	NA	1
Overall cognitive complexity rating (sum of ratings) for this item	**NA**	**6**

Figure 2. Example of a rater assignment of cognitive complexity of an organic chemistry item.

not unique and the rubric does not require uniqueness. Once the student steps are identified and classified in terms of their challenge level, the amplification and distractor role aspects are considered. When all steps are taken, a numerical value can be determined.

■ INSTRUMENT RELIABILITY

This instrument is designed to be an expert-based rating system whereby multiple organic chemistry instructors rate a given test item. For any such instrument, measures of inter-rater agreement must be determined. Faculty who write and use ACS Organic Chemistry Examinations, who were not part of the rubric design team, were recruited to serve as expert raters. These faculty members were asked to rate the complexity of all exam items from a 50-question online Organic Chemistry Practice Exam offered by the ACS Examinations Institute; of these 50 items, 42 items were rated by all raters (and thus included in the analysis). This practice exam was independently administered to a trial group of students in Spring 2012. The faculty received a 25-min training session on using the instrument; the training included the rating of two exam items as a workshop group. Independent ratings of exam item complexity were then made. Inter-rater reliability for the cognitive complexity rating rubric was established using ratings from 42-items of the practice exam. Two-way, mixed intraclass correlations were calculated using Stata 12 statistical analysis software.[18] Additionally, Cronbach α values were calculated. In exploratory research, ICC values should be at least 0.70 or higher to retain an "adequate" scale; many researchers require a cutoff of 0.80 for a "good" scale. Table 2 provides the data of the calculated intraclass correlation coefficients for ratings collected from the practice exam. These data establish an inter-rater agreement of approximately 83%. This value is sufficient, per accepted cutoffs, for establishing internal reliability.

Table 2. Inter-Rater Reliability Statistics for Revised Cognitive Complexity Ratings

Measure	ICC Values[a]	Lower Bound	Upper Bound	F Values	Significance
Single Rater	0.3801	0.2630	0.5246	5.91	<0.001
Average of Raters	0.8307	0.7406	0.8982	5.91	<0.001

[a] Two-way, mixed intraclass correlation coefficient values for 8 raters (42 items rated); Cronbach α = 0.8349.

■ INSTRUMENT VALIDITY

To determine the instrument validity, correlation studies were performed between the complexity ratings derived from the rubric and both student performance on exam items and average student mental effort on the exam items. Item complexity, as determined by this rubric, is a proxy for objective complexity because ratings are derived from experts who parse the cognitive steps a student must likely take independent of the action of any particular student. Because a more complex item has a greater cognitive demand, a negative correlation between objective complexity, thus measured, and student performance is hypothesized. In a similar way, unless most students use an unexpected pathway that subverts the estimated objective complexity, a positive correlation between objective complexity and average student mental effort (i.e., subjective complexity) can be predicted. Similar correlations were observed in a study on general chemistry practice exams.[19] These hypotheses were tested by recording the performance (correctness of item responses) and load on working memory (measured by student reports of mental effort) from 80 students who participated in an online trial offering of an ACS Examinations Institute Organic Chemistry Practice Exam during the Spring 2012 semester. During Spring 2012, students who purchased "ACS Organic Chemistry Exams—The Official Guide" received an enclosed card that offered the use of the online practice exam. Additionally, one instructor of organic chemistry was provided with codes for the students in the organic chemistry course to participate in the online practice exam in preparation for their final exam. Therefore, the student data collected were presumably from students preparing to take a final examination or similar in organic chemistry.

To process this information into validity evidence, student data were included in this study only if they responded to all 50 exam questions and provided mental effort ratings for each of those items. Performance was calculated for each exam item as the fraction of students answering the question correctly, which defines the item difficulty in classical test theory.[20] Students were also asked to rate their mental effort on each exam item by responding to the prompt "how much mental effort did you expend on this question?" that immediately followed each item. Students could respond with "very little", "little", "moderate amounts", "large amounts", or "very large amounts". To put these responses onto a numerical scale for comparison basis, they were scored as 1, 2, 3, 4, and 5, respectively. Average student mental effort ratings were calculated for each exam item as the sum of mental effort values divided by the total number of students.

Once the data were collected, several steps were taken to process that data. First, student performance versus cognitive (objective) complexity was plotted, as shown in Figure 3. As hypothesized, a negative relationship exists between these two variables; lower cognitive complexity values, in general, correspond to better student performance (i.e., values closer to 1). A Pearson correlation between these two measures was determined to be −0.4414 ($p = 0.0013$); this is considered to be a moderate and significant correlation. Visual inspection of the scatter plot suggests a handful of items for which student performance is lower than might be predicted based on task complexity alone. Items that include common misconceptions as distractors, for example, can fall into this category. The complexity of the item is not increased by the availability of these incorrect answers, but they nonetheless collect a relatively high fraction of students, thereby lowering student performance on such items. Faculty members who regularly teach organic chemistry and serve on exam writing committees have developed expertise in identifying the common distractors that students may select based on incorrect processes. These distractors can include misconceptions, although whether a distractor specifically addresses a misconception is not requested of the committee when the exam items are constructed. Additionally, through the editing process, exam items are routinely altered as the committee contributes their collective expertise, thus making it difficult to pinpoint one expert's knowledge of students' misconceptions and how this is translated into particular distractors.[21] It is also important to note that it is not inherent in the nature of the distractor that represents a misconception whether or not the student "finds" that distractor with more, less, or about the same amount of mental effort, that is, experiences similar or different cognitive complexity. Further investigation of outliers, particularly with qualitative methods, to elucidate this could provide more insight into what a student is experiencing while working through these items; however, this was beyond the scope of this project.

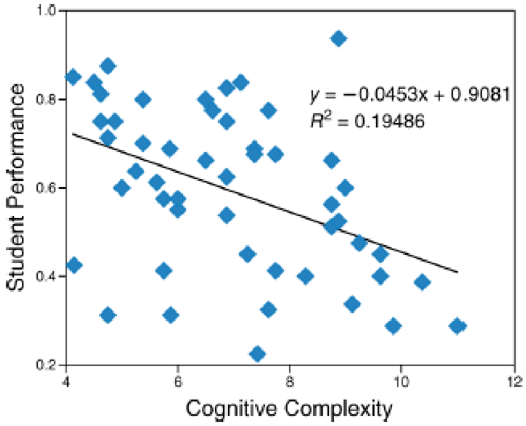


Figure 3. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = -0.4414$; $p = 0.0013$).

Second, student performance versus average mental effort was plotted, as shown in Figure 4. Once again, as predicted, a negative relationship exists between these two variables; lower average mental effort ratings, in general, correspond to better student performance (i.e., values closer to 1). A Pearson correlation between these two measures was determined to be −0.7308 ($p < 0.0001$); this is considered to be a high and significant correlation.

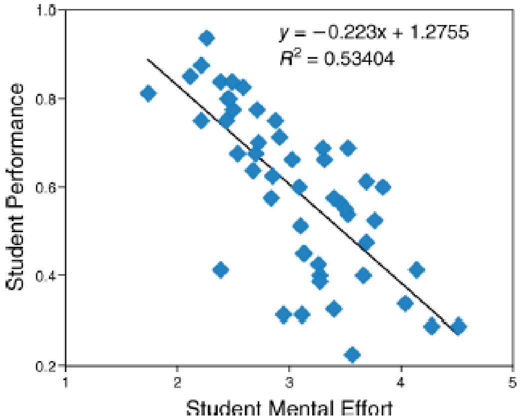


Figure 4. Correlation of student performance (using statistical item difficulty) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = -0.7308$; $p < 0.0001$).

Lastly, cognitive complexity versus average mental effort was plotted as shown in Figure 5. In this case, the prediction is that the two forms of complexity, objective and subjective, that are postulated to be measured should mirror each other. As predicted, a positive relationship exists between these two variables; high cognitive complexity ratings tend to correspond

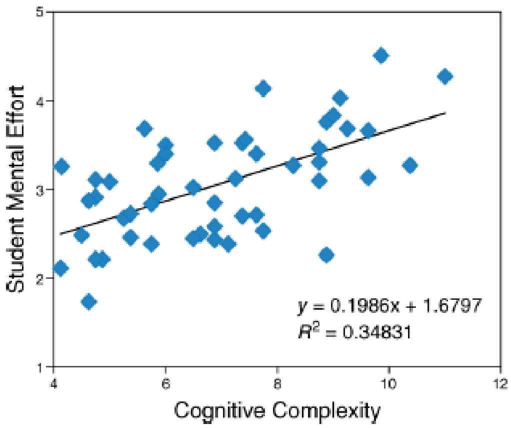


Figure 5. Correlation of student mental effort (using average student ratings) with cognitive complexity (using average item ratings) for 80 students on the ACS organic chemistry online practice exam (Pearson correlation $r = 0.5902$; $p < 0.0001$).

to high average mental effort ratings. A Pearson correlation between these two measures was determined to be 0.5902 ($p <$ 0.0001); this is considered to be a moderate and significant correlation.

These moderate to high correlations between student performance, cognitive complexity, and student mental effort suggest that the developed rubric is valid. The correlation values obtained with this study are higher than those found in the previously described general chemistry complexity rubric.[19] This observation suggests that the revised rubric is an improved tool for measuring the cognitive complexity and is more inclusive of a broader set of examination items.

■ CONCLUSION AND IMPLICATIONS FOR INSTRUCTION

This study has reported a revised instrument for the assignment of cognitive complexity of organic chemistry exam items. A rubric that helps content experts parse cognitive demands on students of multiple-choice items is presented. High inter-rater reliability measures were obtained for this revised rubric instrument. The cognitive complexity assigned to a set of organic chemistry items using this tool correlates at moderate to high levels with similar constructs (i.e., student performance on exam items and average mental effort ratings). As with the initial publication of the general chemistry complexity rubric, this revised rubric will provide an enhanced window into "exploring the relationship between the complexity of content taught and student cognition and learning".[1] In addition, studies of complexity can contribute to the development and design of chemistry assessment materials.

Beyond the research and development of an instrument to assign cognitive complexity, this instrument can also be useful for individual instructors in developing assessments. An experience that is quite likely for many instructors is the construction of a test thought to be easy and yet the students perform poorly on the items. One possible consideration for the "easy" items would be to conduct a complexity analysis of the items. Through this analysis, the instructor can gain a clearer picture of what the students must successfully recognize and do in order to perform well on the test. This analysis can be remarkably revealing in how difficult items may actually be for students. Beyond this, instructors can evaluate possible test items prior to giving an assessment for a better distribution of easy, medium, and hard test items. Although this process will always involve the human element and is not envisioned to be automated, fluidity in regular use of the rubric can underlie the process of test development. Indeed the rubric is used in similar efforts of the ACS-EI to assign difficulty of items by groups of raters as easy, medium, or hard. Ultimately, a better sensitivity of the difficulty of test items from the test-taker perspective can assist instructors in building and using better assessment tools.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kmurphy@uwm.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Early workshops in the development of this revised rubric benefited from the participation of A. Zenisky. We would like to thank the many organic chemistry faculty and ACS-EI exam writers who assisted in the development and revision of the complexity rubric, as well as the students who participated in the online practice exam. Funding for this project was derived from the National Science Foundation (NSF) under grants DUE-0618600, 0717769. Any opinions, findings, and conclusions communicated in this material are those of the authors and do not necessarily reflect the views of NSF.

■ REFERENCES

(1) Knaus, K. J.; Murphy, K. L.; Blecking, A.; Holme, T. A. *J. Chem. Educ.* **2011**, *88*, 554.
(2) Bieri, J. *J. Abnorm. Soc. Psychol.* **1955**, *51*, 263.
(3) Crockett, W. H. *Prog. Exp. Personality Res.* **1965**, *2*, 47.
(4) Daniel, R. C.; Embretson, S. E. *Appl. Psychol. Meas.* **2010**, *34*, 348.
(5) Baxter, G. P.; Glaser, R. *Educ. Meas: Issues Pract.* **1998**, *17*, 37.
(6) Johnstone, A. H. *Chem. Educ. Res. Pract.* **2006**, *7*, 49.
(7) Hancock, G. R. *J. Exp. Educ.* **1994**, *62*, 143.
(8) Thompson, E. G.; Mann, I. T.; Harris, L. *J. Br. J. Psychol.* **1981**, *72*, 249.
(9) Kalyuga, S.; Sweller, J. *J. Educ. Psychol.* **2004**, *96*, 558.
(10) Nadolski, R. J.; Kirshner, P. A.; van Merrienboer, J. J. G.; Woretschofer, J. *Educ. Res. Eval.* **2005**, *11*, 1.
(11) Paas, F.; Tuovinen, J. E.; Tabbers, H.; VanGerven, P. W. M. *Educ. Psychol.* **2003**, *38*, 63.
(12) Pelligrino, J. W.; Chudowsky, N.; Glaser, R. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academy Press: Washington, DC, 2001.
(13) van Merrienboer, J. J. G.; Sweller, J. *J. Educ. Psychcol.* **2005**, *97*, 147.
(14) Mayer, R. E. *Multimedia Learning*, 2nd ed.; Cambridge Univ. Press: New York, 2009.
(15) Benbasat, I.; Todd, P. *Decision Support Systems* **1996**, *17*, 241.
(16) Campbell, D. J. *The Academy of Management Review* **1988**, *13*, 40.
(17) Pawley, D.; Ayres, P.; Cooper, M.; Sweller, J. *J. Educ. Psychol. U.K.* **2005**, *25*, 75.
(18) StataCorp. *Stata Statistical Software: Release 12*; StataCorp LP: College Station, TX, 2011.
(19) Knaus, K J.; Murphy, K L.; Holme, T. A. *J. Chem. Educ.* **2009**, *86*, 827.
(20) Crocker, L.; Algira, J. *Introduction to Classical and Modern Test Theory*; Holt, Reinhart and Wilson: New York, 1986.
(21) Holme, T. A. *J. Chem. Educ.* **2003**, *80*, 594−598.

D.3 Assessment of NMR teaching and learning strategies in organic undergraduate labs

Introduction

Through an NSF funded research project (DUE – 1245666), a 300 MHz Bruker NMR with a 16-channel auto-sampler was purchased for use in undergraduate organic laboratories at Metropolitan State University in Denver, CO. Of interest to the members of the research team at MSU was the effective utilization of this instrument in laboratory instruction to help students bridge the gap that often exists between "authentic" NMR spectra and those shown in prepared instructional materials (i.e. textbooks and laboratory manuals). In order to determine if use of the new NMR instrument had a positive impact on student understanding of NMR, a scenario based interview activity (**Figure 1**) was developed. This activity was designed purposefully to include authentic spectra taken on the NMR instrument and to allow for the identification of instructional targets and misconceptions related to NMR instruction. Following completion of the activity, participants were asked a series of follow-up questions (**Table 1**) related to the capacity in which they had used the new instrument, as well as their perceived value of the new instrument

Table D3.1: Interview Participants

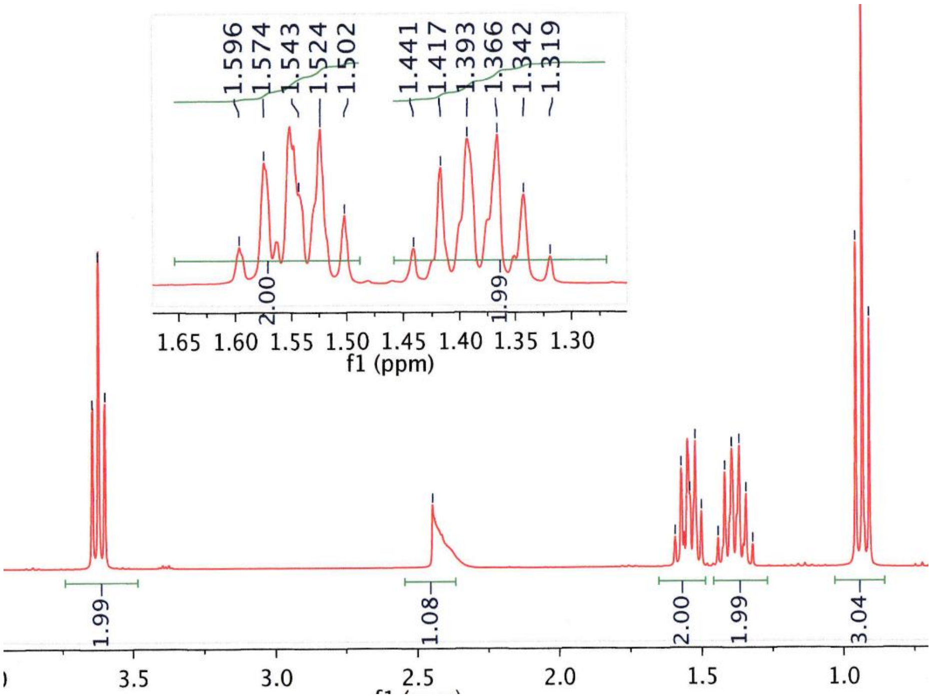
	Male	Female
Enrolled in organic chemistry I		3
Enrolled in organic chemistry II	5	4

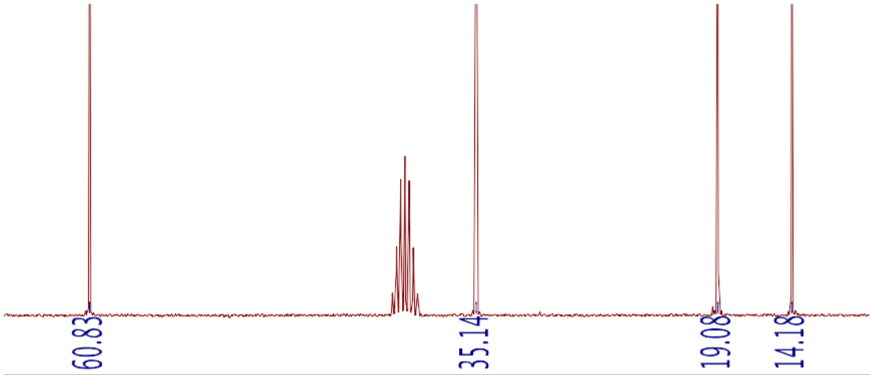
Figure D3.1: Scenario Based Interview Activity

Scenario 1: Al and Mo arrive at their organic chemistry lab and are told that their project for the day will be to identify an unknown and use it to perform a reaction. They are given very limited information about the reagent they are supposed to use and the product the reaction will form. They are told they will have access to both a 60 MHz NMR and a 300 MHz NMR for analysis.

The first reagent bottle Al and Mo pick up has a label that says “$C_4H_{10}O$ Molar Mass = 74 grams/mole”

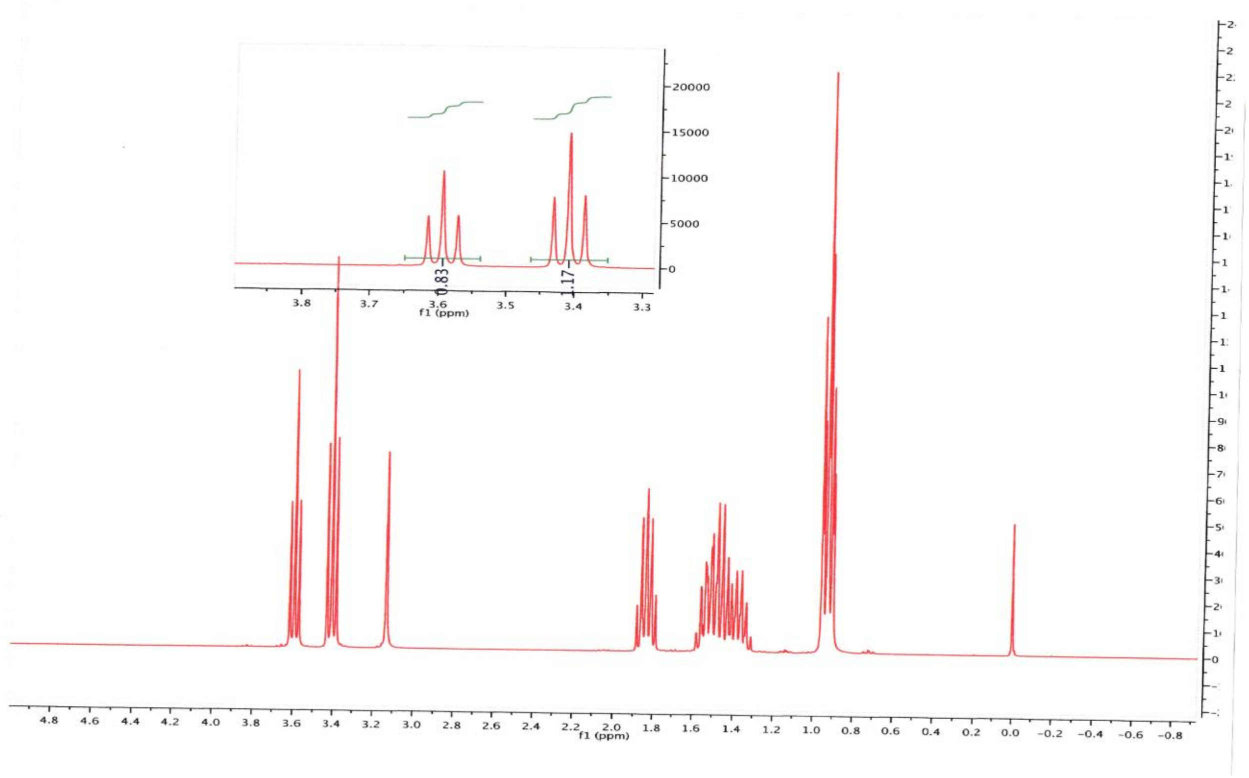
Al takes an ^{1}H NMR taken on a 300 MHz instrument and a ^{13}C NMR taken on a 75.5 MHz NMR and the collected spectra are shown below. What should Al and Mo conclude is the reagent in the bottle?



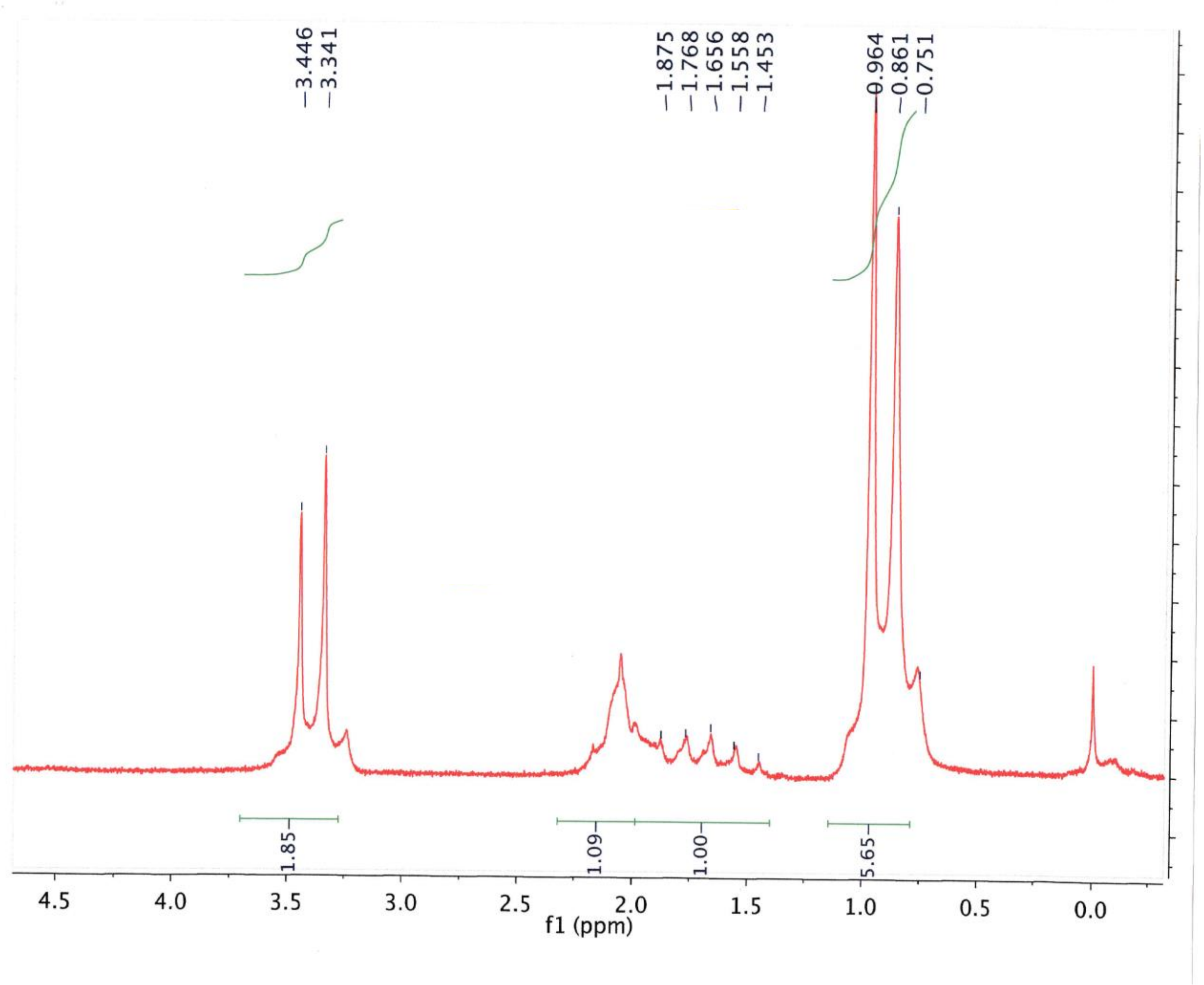


Scenario 2: Now that Al and Mo have figured out the identity of their reagent they are going to react it with NaBr salt and H_2SO_4 by following an experimental procedure they got from their TA. Upon

completion of the reaction and after work-up Al and Mo take a ^{1}H NMR on a 300 MHz instrument of the product they have isolated. What should Al and Mo conclude about their sample?



Scenario 3: Uh Oh! Al and Mo didn't know they were supposed to hand their product in to the TA and it got mixed up with another unlabeled bottle. There is a long line of students using the 300 MHz NMR so Al and Mo decide to use the 60 MHz NMR. Below is the ^{1}H NMR spectra they obtained. What should Al and Mo conclude about this sample?



Scenario 3: Al and Mo are pretty certain they know the identity of the sample they took on the 60 MHz NMR but want to take an NMR on the 300 MHz instrument to be sure. They collect the following ^{1}H NMR spectrum. What is the identity of this sample?

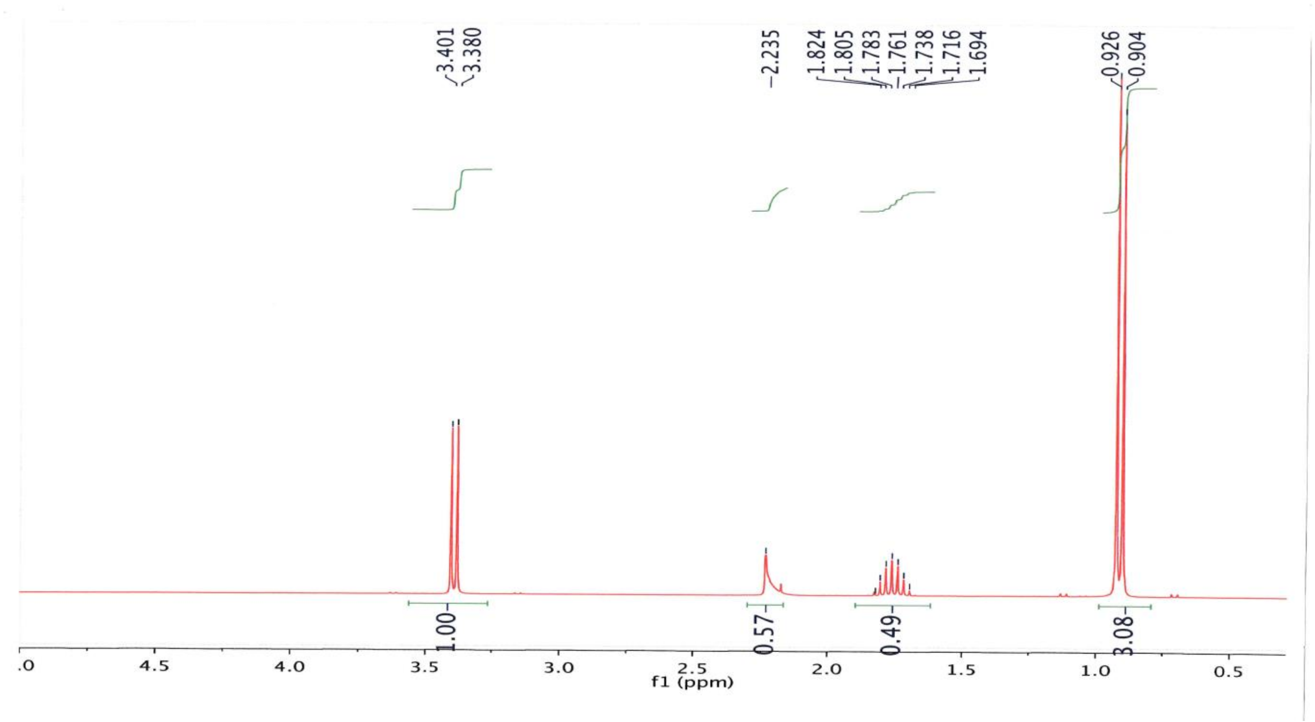


Table D3.2: Open-ended questions

- What value is there to being able to get an actual spectrum of a sample you submit as opposed to being given example data?
- How different would your lab experiences have been without having access to the data collected from the 300 MHz NMR?
- How did the data you collected from the NMR in lab differ from what you'd been taught in lecture?
- What value is there to have a 300 MHz NMR?
- Did utilizing the 300 MHz NMR change your opinion on the value of NMR as a characterization tool?
- Why do chemists use instrumentation like this?
- If you were doing an experiment and got back data that doesn't make sense, what would you do?
- Are you doing undergraduate research currently? Briefly describe your project.

Interview Assessment

Interview participants were graded real-time using a system based on the amount of support the student needs to use a feature of an NMR spectra to determine its structure (see Table D3.3).

Table D3.3 Interview grading rubric

None	Student identified a feature and used it correctly
I	Student needed generic support related to understanding given task
II	Student needed generic support related to understanding NMR (i.e. student was given generic information such as how to predict a splitting patteren)
III	Student needed task specific support but did not need help applying the knowledge (i.e. student looked information up in the provided text book)
IV	Student needed task specific support and needed help applying the knowledge (i.e. student was taught)

Results of the grading are summarized in Tables X-X and show that students demonstrated difficulty when completing these types of NMR related tasks. Most notably, only one student even attempted scenario 3 of the activity and no student completed all four scenarios of the activity. This observation was surprising given the fact that the spectra used in this activity directly replicated spectra given to students in organic chemistry I laboratory as part of an introduction to NMR activity. One student who chose to look back at her notes during completion of the activity was observed to have the NMR handout given during this laboratory meeting in her notebook but did not choose to reference it.

Table D3.4: Scenario 1 - Students use of features to identify unknown using ^{1}H NMR

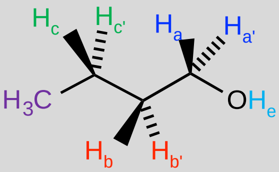
	Correct	Level I	Level II	Level III	Level IV	Not used
Identify 5 H environments	5				5	2
Calculate ° of unsaturation	5			1		6
Use splitting	7			1	4	
Use Integration	6				4	2
Identify alcohol functionality	9				3	
Identify CH_3	8			1	3	
Identify H_a	8				4	
Identify H_e	4				8	
Identify H_c and H_b	6		1		5	
Identify reagent as 1-butanol	8				4	

Table D3.5: Scenario 1 - Students use of features to identify unknown using ^{13}C NMR

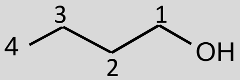
	Correct	Level I	Level II	Level III	Level IV	Not used
Identify 4 C environments	5				3	4
Identify septet as solvent	4				3	5
Identify C_1	2					10
Identify C_2	2					10
Identify C_3	2					10
Identify C_4	2					10

Table D3.6: Scenario 2 - Students use of features to identify product using ^{1}H NMR

	Correct	Level IV	Not used
Predict product	9	3	
ID spectra contains mixture	4	8	
Identify triplet of starting material	1	11	
Identify triplet of product	1	11	
Identify overlapping signals	5	7	
Identify hydrogen from -OH still present	2	10	
Recognize triplet ~1.0 from both species	2	9	1
Estimate relative amounts of SM and Product	3	7	2

Commonly employed student strategies

Students completing the interviews most commonly utilized a strategy of “guess and check” in which they would draw a structure and attempt to fit a spectrum to their drawing, as demonstrated by this student as they attempted to determine the unknown compound described in scenario 1:

Student: “I’m trying to organize it in my head cuz what I’m imagining is something like this and you have this OH like right here [draws 2-butanol]. That works for everything except for this.” [referring to signal for hydrogen of alcohol].

Another commonly observed strategy of students was to focus on regions of the spectra that didn’t convey meaningful information. This was most notable in the spectrum given in scenario 2 in which the product mixture gave rise to an uninterpretable region of the ^{1}H NMR spectrum. One student who exemplified this behavior explained the convoluted region as appearing when “**like a doublet is on top of a triplet or something. Okay these two peaks here are really tall and everything else this is for another signal and there’s another signal somewhere I can’t see because of the doublet so there’s probably 2 triplets and a doublet or hmm maybe more**”. Rather than identifying what usable information was conveyed in the spectrum, students found themselves speculating as to what would give rise to an observed “messy” signal. Other given explanations for the observation of this signal were contamination of the sample or that the NMR spectrum was not of the correct sample.

Conclusions

During the course of these interviews, organic chemistry I and II students demonstrated only a low level of proficiency in assigning typical NMR spectra obtained during laboratory experiments. While the scenarios presented in the activity do not replicate the experience a student gains by being in the laboratory and conducting NMR analysis, the activities do closely replicate the authentic spectra collected by students in the completion of laboratory experiments. Based on the results of these interviews, a more effective use of the new NMR instrument could be utilized in laboratory through the incorporation of explicit instruction using authentic spectra and experiments that more closely replicate authentic research. As students frequently commented that chemists utilize instruments such as NMR for the purpose of identifying unknowns, it is likely that students would benefit from experience in using NMR to confirm the identify of expected products or to determine the extent by which a reaction has occurred by learning to identify starting materials, solvent, product, or produced by-products.

CURRICULUM VITAE

Jaclyn Trate

Place of birth: Milwaukee, WI

EDUCATION

Doctor of Philosophy in Chemistry - Chemical Education Research 2017
University of Wisconsin-Milwaukee, Milwaukee, WI
"Integrating Scale-Themed Instruction across the General Chemistry Curriculum."
Advisor: Kristen Murphy
Minor focus: Organic Chemistry

Masters of Science in Chemistry – Organic Chemistry 2010
University of Mississippi, Oxford, MS
"Extension of the Metallation/Transmetallation Methodology to the Preparation of Pincer N-Heterocyclic Carbene Pd Complexes: Synthesis, Characterization, and Coordination Sphere Variations."
University of Mississippi, Oxford, MS
Advisor: T. Keith Hollis

Bachelor of Science, Chemistry 2006
St. Norbert College, De Pere, WI

WORK EXPERIENCE

Graduate Research Assistant

Department of Chemistry and Biochemistry, UW – Milwaukee 2012-2017

Developed and studied the efficacy of new instructional materials for the teaching of scale concepts in both semesters of a two semester general chemistry sequence.

Independent Contractor

American Chemical Society March 2016 – August 2016

Wrote answers to in-chapter learning activities for the 9th edition of the Chemistry in Context textbook.

Graduate Teaching Assistant

General Chemistry for Engineers, UW – Milwaukee Fall 2011, Spring 2012

Taught laboratory and discussion sections of a one-semester accelerated general chemistry course for engineers. Responsibilities included preparing lectures to reinforce lecture concepts, teaching laboratory techniques, writing and grading quizzes, grading lab reports, and proctoring and grading regular course exams.

Undergraduate research mentor Spring 2013 – Present

Benjamin Wartgow	“Targeting one component of science literacy: Integrating scale as a theme into laboratory experiments and assessments”
Brian Mohs	“Investigating students’ conceptual boundaries of scale”
Brooke Pirkov	“Investigating the connection between scale and intermolecular forces in the general chemistry laboratory”
Ann Hackl	“Investigation of absolute and relative scaling conceptions in students in an introductory college chemistry course”
Carly Otto	“Using pre-laboratory quizzes to investigate students’ understanding of scale”

PUBLICATIONS

4. Gerlach K.; Trate, J.; Blecking, A.; Geissinger, P.; and Murphy, K. (2014) “Investigation of Absolute and Relative Scaling Conceptions of Students in Introductory College Chemistry Courses”, Journal of Chemical Education, 91, 1526-1537.

3. Gerlach K.; Trate, J.; Blecking, A.; Geissinger, P.; and Murphy, K. (2014) “Valid and Reliable Assessments to Measure Scale Literacy of Students in Introductory College Chemistry Courses”, Journal of Chemical Education, 91, 1538-1545.

2. Raker, J.; Trate, J.; Holme, T.; and Murphy, K. (2013) “Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Items”, Journal of Chemical Education, 90, 1290-1295.

1. Cho, J.; Hollis, T. K.; Valente, E.; and Trate, J. (2011) “CCC-N-Heterocyclic Carbene Pincer Complexes: Synthesis, Characterization and Hydroamination Actitvity of a Hafnium Complex”, Journal of Organometallic Chemistry, 696, 373-377.

PRESENTATIONS – PRESENTING AUTHOR

13. “Integrating scale-themed instruction across the undergraduate general chemistry curriculum” (poster) Gordon Research Conference on Chemistry Education Research and Practice, Bates College, ME, June 2017.

12. “Response process validity study of a cross disciplinary assessment” 251st ACS National meeting, San Francisco, CA, March 2017

11. “Scale-themed instruction for the chemistry classroom” 2016 Biennial Conference on Chemical Education, University of Northern Colorado, Greeley, CO, August 2016

10. “Integrating scale-themed instruction across the general chemistry curriculum” 2016 Biennial Conference on Chemical Education, University of Northern Colorado, Greeley, CO, August 2016

9. “Assessment of NMR teaching and learning strategies in organic undergraduate labs” 251st ACS National meeting, San Diego, CA, March 2016

8. “Response process validity study of scale-themed assessments” 251st ACS National meeting, San Diego, CA, March 2016

7. “Integrating scale-themed instruction across the undergraduate general chemistry curriculum” 250th ACS National meeting, Boston, MA, August 2015.

6. “Integrating scale-themed instruction into the undergraduate general chemistry curriculum using active learning methodologies” 249th ACS National meeting, Denver, CO, March 2015.

5. “Investigating students’ conceptual boundaries of scale” 2014 Biennial Conference on Chemical Education, Grand Valley State University, August 2014.

4. “Integrating scale-themed lecture instruction into the undergraduate general chemistry curriculum” 247th ACS National meeting, Dallas, TX, March 2013.

3. “Integrating scale-themed (laboratory) instruction into the undergraduate general chemistry curriculum” 246th ACS National meeting, Indianapolis, IN, September 2013.

2. “Integrating scale-themed (laboratory) instruction into the undergraduate general chemistry curriculum” Great Lake Regional meeting of the ACS, La Crosse, WI, June 2013.

1. “Rating the Complexity of Organic Chemistry Items” 2012 Biennial Conference on Chemical Education, Pennsylvania State University, July 2012.

WORKSHOPS

4. “Scale interventions for lecture and laboratory” 2016 Biennial Conference on Chemical Education, University of Northern Colorado, August 2016

3. “Scale interventions for lecture and laboratory” 249th ACS National meeting, Denver, CO, March 2015.

2. “Scale interventions for lecture and laboratory” 2014 Biennial Conference on Chemical Education, Grand Valley State University, August 2014.

1. “Spatial scale: how understanding relative size is important in developing an understanding in science” Marie Curie High School Chicago, IL, August 2013.

AUTHORED EDUCATIONAL MATERIALS

General chemistry II: lecture activity workbook (with explicit scale themes)
General chemistry II: lecture activity workbook (w/out explicit scale themes)
General chemistry II: laboratory instructor manual
General chemistry II: laboratory manual
General chemistry I: lecture activity workbook (with explicit scale themes)
General chemistry I: lecture activity workbook (w/out explicit scale themes)
General chemistry I: laboratory instructor manual
General chemistry I: laboratory manual

PROFESSIONAL AFFILIATIONS

American Chemical Society (2006-present)
American Chemical Society Division of Chemical Education (2011-present)
National Association for Research in Science Teaching (NARST) (2014-present)

AWARDS AND HONORS

2014 Keulks award for graduate research

UNIVERSITY SERVICE

Event Coordinator, 2012-2013 Department of Chemistry and Biochemistry Graduate Student Council
Event Co-Organizer, 2012 and 2013 Younger Chemists Committee
Volunteer, 2013 Wisconsin science Olympiad
Presenter, 2015 Women in Science showcase, Kenosha Public Museum, Kenosha, WI.