


2015

# The testing effect in general chemistry: effects of repeated testing on student performance across different test modes

Anna Agripina Prisacari  
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>

 Part of the [Cognitive Psychology Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Instructional Media Design Commons](#)

## Recommended Citation

Prisacari, Anna Agripina, "The testing effect in general chemistry: effects of repeated testing on student performance across different test modes" (2015). *Graduate Theses and Dissertations*. 14928.  
<https://lib.dr.iastate.edu/etd/14928>

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

**The testing effect in general chemistry: Effects of repeated testing on student performance across different test modes**

by

**Anna Agripina Prisacari**

A thesis submitted to the graduate faculty  
in partial fulfillment of the requirements for the degree of  
MASTER OF SCIENCE

Major: Human Computer Interaction

Program of Study Committee:  
Thomas Holme, Major Professor  
Joseph Burnett  
Stephen Gilbert

Iowa State University

Ames, Iowa

2015

Copyright © Anna Agripina Prisacari, 2015. All rights reserved.

## DEDICATION

This work is dedicated to all who helped me reach this point in my professional career: my beloved mother, Svetlana Prisacari, sister, Vera Prisacari, and my grandparents, Pavel and Vera Prisacari, who consistently loved and supported me in my life.

I also would like to dedicate this thesis to all my teachers; thank you for sharing your knowledge with me and inspiring me to learn more.

## TABLE OF CONTENTS

	Page
DEDICATION .....	ii
LIST OF FIGURES .....	v
LIST OF TABLES .....	vi
ACKNOWLEDGMENTS .....	viii
ABSTRACT .....	ix
CHAPTER 1. INTRODUCTION .....	1
Testing Effect.....	1
Computer-based Testing .....	3
Statement of the Research Problem .....	5
Overview of the Study .....	7
Significance of the Study .....	8
Overview of the Chapters .....	9
CHAPTER 2. LITERATURE REVIEW .....	10
Testing Effect.....	10
How and why the testing effect works.....	12
Testing effect in the classroom .....	14
Testing effect benefits.....	16
Test Mode .....	16
Test mode order effect .....	18
Types of questions .....	19
Student test mode preferences .....	21
Chemistry Question Types.....	22
Summary .....	24
CHAPTER 3. METHODS .....	26
Pilot Study.....	26
Participants.....	27
Material and design.....	27
Procedure .....	28
Results.....	29
Changes to the study .....	31

Main Study.....	35
Participants.....	35
Materials and design.....	39
Procedure.....	44
Scoring.....	48
CHAPTER 4. RESULTS.....	49
Preparing Data.....	49
Evaluating Equivalency.....	50
Descriptive Statistics.....	52
Student testing preferences.....	53
Research Question 1: Overall Test Mode.....	56
Research Question 2: Gain by Question Type.....	60
Algorithmic gain.....	61
Conceptual gain.....	62
Definition gain.....	64
Repeated Versus Non-repeated Gains.....	66
Summary.....	67
CHAPTER 5. CONCLUSIONS.....	70
Empirical Contributions.....	70
Practical Contributions.....	77
Limitations.....	79
Future Directions.....	80
Conclusion.....	81
REFERENCES.....	83
APPENDIX A. APPROVED IRB FORM.....	94
APPENDIX B. QUESTIONS FOR ALL TESTS.....	95
APPENDIX C. INSTRUCTIONS TO COMPUTER-BASED TESTS.....	124
APPENDIX D. INSTRUCTIONS TO PAPER-BASED TESTS.....	125

## LIST OF FIGURES

	Page
Figure 1 Smith, Nakhleh, and Bretz (2010) framework for general chemistry questions .....	23
Figure 2 Mean proportions of performance on Test 1 and Test 2 by conditions.....	30
Figure 3 Diagram of the main study procedure .....	47
Figure 4 The means of student responses to a question that asked the students whether they had taken online tests for a class before .....	53
Figure 5 The means of student responses to a question that asked the students about their test mode preferences.....	54
Figure 6 The means of student responses to a question that asked the students why they preferred paper to online modes .....	55
Figure 7 The means of student responses to a question that asked the students why they preferred online to paper test modes. ....	56
Figure 8 Mean normalized gains by condition for Group I and Group II.....	59
Figure 9 Mean normalized algorithmic gains by condition for Group I and Group II..	62
Figure 10 Mean normalized conceptual gains by condition for Group I and Group II..	63
Figure 11 Mean normalized definition gains by condition for Group I and Group II.....	65

## LIST OF TABLES

	Page
Table 1 Pilot study participant information by the conditions and time of the sessions.....	28
Table 2 Frequency and percentage of participants by gender, ethnicity, year in school, and instructor and average of age and percentage received on a final exam for each course.....	36
Table 3 Number and percentage of participants by conditions for each course and percentage of participants over all courses.....	37
Table 4 Student information by course .....	39
Table 5 Chapters and topics for each chapter from a first-semester general chemistry course.....	40
Table 6 Distribution of algorithmic, conceptual, and definition questions .....	41
Table 7 Example of counterbalancing question formats and test order for Chapter 1 ...	43
Table 8 Mean proportions of correct answers and standard deviations on Test 1 for the four courses .....	51
Table 9 Post-hoc comparisons using the Tukey HSD test for the Test 1 mean proportions for the four courses .....	51
Table 10 Frequency and percentage of participants by gender for Group I and Group II .....	52
Table 11 Student distribution among four conditions for Group I and Group II for overall normalized gain analysis .....	57
Table 12 Summary of Tukey's HSD results for Group I and Group II pairwise comparisons of normalized gain means .....	59
Table 13 Sample size for Group I and Group II used for different analyses.....	61
Table 14 Summary of Tukey's HSD results for Group I and Group II pairwise comparisons of normalized conceptual gain means. ....	64
Table 15 Summary of Tukey's HSD results for Group II pairwise comparisons of normalized definition gain means .....	65

Table 16 Means and standard deviations of mean gains by condition for Group I  
and Group II .....74



## ACKNOWLEDGMENTS

I would like to acknowledge and thank those who supported me during my graduate school career. First and foremost, I would like to thank my advisor, Dr. Thomas Holme, for introducing me to educational research. His expertise, guidance, and enthusiasm have been invaluable. I would also like to thank my committee members, Dr. Joseph Burnett and Dr. Stephen Gilbert for their guidance and support throughout my graduate school journey. I appreciate all their contributions of time and ideas to this project. I am especially grateful to Dr. Armstrong, Dr. Dark, and Dr. Carpenter for their scholarly advise on this work and Dr. Cotos, Chris Nelson, and Dr. Huffman for sharing their knowledge on academic writing with me.

I also would like to thank the members of the Holme's research group at Iowa State University for their help in designing questions for my tests, collecting data, and scoring. Their constructive criticism and support have contributed to my professional growth at Iowa State University. Finally, to my mother, Svetlana Prisacari, sister, Vera Prisacari, and friends: thank you for cheering me on!

## ABSTRACT

Research on the testing effect shows that practice tests are more effective than additional studying for enhancing learning. However, there has been little research directly addressing the role of additional testing when students take paper-based or computer-based tests in college courses. Accordingly, the purpose of this study was to investigate the role of the testing effect and test mode on student performance.

The participants were 664 general chemistry students from two large universities in the Midwest and the Pacific Northwest. After covering the test material in their course, students completed two proctored practice tests that included 17 algorithmic question pairs, 5 conceptual pairs, and 2 definition pairs. Each practice test was delivered on computer or paper according to one of four conditions that were defined by the mode of the initial test and the mode of the final test. These conditions were: Computer-Computer, Computer-Paper, Paper-Computer, and Paper-Paper. After completing the initial practice test, students repeated half of the items. Feedback was provided after each test and performance was measured with normalized gains. After completing all of the tests, students were asked to indicate and explain their test mode preferences for general chemistry tests.

Four major conclusions resulted from the study. First, the testing effect was found in all conditions, but varied in terms of the test mode. Paper-Computer showed the lowest gains and these gains were significantly lower than those for Paper-Paper. Gains from Computer-Paper and Paper-Paper were not significantly different from each other. Second, the test mode did not affect the students' performance on algorithmic questions, but affected their performance on conceptual and definition questions. Third, the gains

from repeated items were significantly higher than the gains from non-repeated items. However, this testing effect was not consistent across all areas of chemistry content. Fourth, a majority of the students indicated a strong preference to take their next general chemistry test on paper since this mode allowed them to write on the test and show their work.

Overall, this study demonstrates the promising effects of testing. The results contribute to the understanding of the testing effect in a college classroom and the role of test mode for enhancing learning in general chemistry courses.

## CHAPTER 1. INTRODUCTION

Testing is a powerful tool for student learning. It can be used not solely to assess student learning and assign grades, but also to evaluate the instruction of the classroom teacher, deliver feedback on student's progress, review programs and curricula, and guide policy decisions (Seymour, 2002). In effort to increase student engagement in courses, instructors begin to instill their lectures with a number of activities, such as clicker tests, thus increasing and diversifying the use of testing in school. Therefore, it comes as no surprise that for over 100 years, there has been a strong interest to investigate testing and the effects it has on student learning (Abbott, 1909; Gates, 1917; Glover, 1989; Little & McDaniel, 2015; H. A. Peterson, 1944; Pyc & Rawson, 2010; Rock, 1957; Spitzer, 1939; Stenlund, Sundström, & Jonsson, 2014; Tulving, 1967; Wheeler, Ewers, & Buonanno, 2003).

Because testing is the core in education, the demand to study the effect of testing continues to grow. However, knowing that testing is important is not sufficient; knowing how to use testing effectively under different classroom conditions is an essential addition to educational research. To move forward, it is important to recognize and integrate the newfound knowledge of diverse and interdisciplinary studies that empirically explored the effect of testing with student-level data.

### Testing Effect

Ample evidence in cognitive psychology (Butler & Roediger, 2007; Jason C. K. Chan, McDermott, & Roediger III, 2006; Hogan & Kintsch, 1971; Roediger & Butler, 2011; Roediger III, Agarwal, McDaniel, & McDermott, 2011) suggests that although

most students prefer re-reading to repeated testing (Clark & Svinicki, 2014), testing is a more effective tool in storing information over long periods of time, also known as long-term retention (Squire, 1992). This phenomenon is known as the testing effect (Roediger & Karpicke, 2006a, 2006b) and is currently the most evidence-based learning strategy (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Studies on the testing effect have demonstrated that when students take a test after studying the material it enhances their memory and learning more than restudying that material alone (Butler, 2010; C. I. Johnson & Mayer, 2009; Kang, McDermott, & Roediger, 2007; J. D. Karpicke & Roediger, 2008; Toppino & Cohen, 2009). These results are typically found using the following design: students in one condition are instructed to study the material twice, whereas students in the second condition are instructed to study the material only once and then take a test. Next, students in both conditions take the final test. The results of these studies have demonstrated that when performance on the final test is compared, students from the second condition significantly outperform students from the first condition. Why do testing and additional studying opportunities lead to different learning results? In his book, William James (1890) explained this difference as follows:

“A curious peculiarity of our memory is that things are impressed better by active than by passive repetition. I mean that in learning by heart (for example), when we almost know the piece, it pays better to wait and recollect by an effort from within, then to look at the book again. If we recover the words in the former way, we shall probably know them the next time; if in the later way, we shall very likely need the book once more” (pp. 445-446).

The difference in student test performance as the result of frequent testing or studying is explained by the idea that each condition involves different cognitive processes. To successfully learn material, the student's memory must complete three steps: (1) encode, or register the material, (2) store it, and (3) find and retrieve it (Baddeley, 2009). While testing requires student to engage in effortful retrieval (i.e., final step), studying or re-studying the material involves encoding (i.e., initial step) (J. D. Karpicke & Roediger, 2008). For example, if a student is studying some material and learning occurs due to reading, the material that is read is encoded and then stored in an individual's long-term memory. As the student rereads that material, he or she marginally enhances the memory storage. However, when the student is presented with a test after initial learning, he or she is required to search and retrieve the information from his or her long-term memory, thus engaging his or her memory in all three steps. Evidence from studies on the testing effect suggest that this effortful information retrieval tends to strengthen the capacity to retrieve this information in the future more than the rereading process. Consequently, practicing repeated testing as measured by testing produces more learning than re-studying the material and, therefore, the testing effect can be classified as a beneficial strategy for studying.

### Computer-based Testing

While traditional, paper-and-pencil testing continues to dominate as the main method of test delivery mode, the interest and need for computer-based testing (CBT) is growing. A report by Allen, Survey, and Seaman (2014) showed that based on data from 2,831 US colleges and universities, about 80% of these institutions offered some level of

online instruction. The same report indicated that the proportion of higher education students taking at least one course online considerably increased from about 10% in 2003 to 33.5% in 2012 and according to about two-thirds of academic leaders, this number will continue to grow (Allen et al., 2014). Another reason behind the drive to switch from paper-and-pencil testing to CBT is a change in student demographics. According to a National Center for Education Statistics (NCES) report, postsecondary student enrollment will become more diverse by 2022; more non-traditional, part-time, and working students will go to college in the future (Hussar & Bailey, 2014). This change in student demographics is critically important, because non-traditional students often manage multiple responsibilities such as family and work and thus are more likely to enroll in online classes than typical students. Therefore, for many students CBT may become a more common method of assessment.

The rise in CBT interest can also be explained by a set of unique advantages that are not available with paper tests. From a teacher's perspective, online assessments may decrease grading time and reduce human error associated with grading, which may be especially important for instructors who teach large-enrollment introductory classes. In addition, the CBT permits teachers to customize their tests for different students (Wang, 2014) and incorporate dynamic and interactive features such as embedded videos and online simulations (for a review of CBT in basic science and medicine see Kuo & Wu, 2013). Such features provide an additional opportunity to study the relationship between the student performance and different format of online testing (DeBoer et al., 2014). What about the CBT benefits for the test users? When students were asked why they prefer online to paper tests, students frequently mentioned features such as immediate

feedback, display of additional information such as remaining time and built-in calculator, and flexibility to take a test anytime and anywhere (Engelbrecht & Harding, 2004; Hochlehnert, Brass, Moeltner, & Juenger, 2011; Noyes & Garland, 2008; Steinberg et al., 2014).

Because of high demand in online courses, the topic of CBT in the educational environment has gained considerable relevance and attention from teachers (Dragow, 1999). In particular, this has led to the need to understand not only how CBT is used, but also how it compares to traditional modes of testing (Alexander, Bartlett, Truell, & Ouwenga, 2000; Brallier, Schwanz, Palm, & Irwin, 2015; Mead & Dragow, 1993; Neuman & Baydoun, 1998; Russell & Haney, 1997).

#### Statement of the Research Problem

Despite the large body of research on the testing effect, the majority of these studies have been conducted in laboratory settings, asking participants to recall material such as word lists (Wheeler et al., 2003; Zaromb & Roediger, 2010) or paired associates (Carpenter, 2011; Keresztes, Kaiser, Kovács, & Racsmány, 2014; Toppino & Cohen, 2009). However, there are several differences between the laboratory and the classroom setting. Participants in the laboratory studies often receive a free recall test on material studied during the experiment session, whereas the tests administered in class may use different types of questions and assess content studied days, weeks, or even several months ago. For example, the content of general chemistry first-semester university course is frequently assessed at the end of semester with a comprehensive exam that includes algorithmic, conceptual, and definition questions (Smith, Nakhleh, & Bretz,



2010). Kingston (2008) reviewed the impact of computer and paper administration mode on test scores of students in grades 1-12 and found that discipline appeared to affect differences in student performance between paper-based and computer-based tests. Due to the differences between laboratory settings and classroom settings and limited information on chemistry test comparability, no conclusions could be drawn with regard to whether chemistry students benefit from practice testing before taking a comprehensive exam with different types of questions such as algorithmic, conceptual, and definition items.

As technology continues to supplement and enhance teaching practices in post-secondary education (Kirkwood & Price, 2014), there is still much to be learned about its role in testing. Specifically, under which circumstances student performance may be maximized when practice tests for a course' exam are available in paper and online versions to students? The research on test mode comparability reports mixed findings. Whereas some studies posit that a paper-and-pencil test is the more appropriate mode for testing (Bennett et al., 2008), others show no significant differences between paper and online test performances (Alexander et al., 2000; Hochlehnert et al., 2011). However, these studies compared the test mode only between two groups; one group took a test on paper and one group took the same test online. The design of the previous studies does not allow investigating the effects of multiple tests such as the mode change or the learning gains because they used only one test. Therefore, we do not know yet whether a change in the test mode would lead to higher student learning gains than a one-mode condition only. Further research needs to explore the outcome of taking practice tests using one or two test modes before a comprehensive exam of a course. Knowing what

students gain from taking their first test on paper or online and how it impacts their test performance on a subsequent test may be beneficial to instructors who teach blended courses or have access to deliver their tests in both modes.

### Overview of the Study

The testing effect has been widely studied in the laboratory, but few studies have explored the benefits of testing in classroom environments for course-relevant material. Although the results of the testing effect studies show positive changes in student performance, additional research that focuses on the test delivery mode in educational settings is needed. The purpose of this study was to replicate the testing effect, but for content more consistent with a realistic classroom environment using general chemistry questions in paper and online test modes.

Based on identified gaps in the literature surrounding this research, the study focused on the following two research questions:

- (1) Is there a difference in the measured performance on general chemistry tests for students across different test mode groups?
- (2) Is there a difference in average student gains for algorithmic, conceptual, and definition questions based on the mode by which the test was delivered?

To answer these questions, 664 students from four general chemistry courses were assigned to take two practice tests in one of four conditions: Computer – Computer, Paper – Paper, Computer – Paper, and Paper – Computer. The four conditions were defined by the mode of the first practice test (Computer vs. Paper) and the mode of the second practice test (Computer vs. Paper). Each practice test included a set of

algorithmic, conceptual, and definition items and was based on material that students learned in their general chemistry course. All sessions were conducted near the end of the first semester after all tested topics were covered in the course. Additionally, the sessions were proctored and took place in the testing environment similar to course's final exam.

### Significance of the Study

There are many studies that investigated the testing effect and test mode. However, the current study is the first known study to investigate test delivery mode and mode change as a contributor to the effects of testing. By integrating several concepts, this study addresses a gap in both the testing effect and test modality literatures and, thus, extends the previous work in several ways. First, the testing effect is measured in the context of general chemistry. So extending the testing effect to general chemistry materials the findings of this study would have face validity and could prove useful to chemistry instructors in formatting their current or future tests. Because this study used students from general chemistry courses as participants and questioned their knowledge in that subject, it allowed the replication of testing effect studies but in a more applied setting, thus helping measure the possible role of the testing effect in student learning in a classroom setting.

Second, and more importantly, the study investigates the effects of test mode on students' performance using algorithmic, conceptual, and definition questions within the content area of chemistry. Computer use in educational assessment continues to rapidly increase. Exploring the effect of test mode on different question types common to the field of chemistry may yield new recommendations for general chemistry instructors

about the benefits of tests and how student learning may be enhanced as the result of testing. Measuring which test mode promotes student test performance can allow instructors to design and deliver their tests in general chemistry courses more effectively. This, in turn, will make course assessments more useful for student learning. Thus, by bridging the concepts of the testing effect, test modality, and question types, this work extends the understanding of the use and benefit of testing and provides important evidence-based recommendations for the use of testing to educators.

### Overview of the Chapters

This thesis presents the background for this study, the methods used to collect data, the results, an analysis of the results, and a discussion. The introduction chapter provides a brief background for the study, problem statement, overview of the study including its purpose and research questions, and study's significance. Chapter 2 reviews the current literature on the testing effect, test modality, and type of material used in general chemistry testing settings, all topics that are vital to the purpose and design of this study. After a literature review, Chapter 3 presents the methods used to collect data and information on participants. Chapter 4 reports the results. Finally, this thesis concludes with Chapter 5 that presents a summary of the results, followed by implications for chemistry instructors, a discussion on study's limitations, and opportunities for future directions.

## CHAPTER 2. LITERATURE REVIEW

This chapter summarizes the achievements and limitations of three topics that are most relevant to this study. These topics are the testing effect, test mode, and chemistry question types. Specifically, how the testing effect is typically studied in the laboratories and classroom settings and what benefits it offers to students. Also, it reviews studies that empirically investigated the comparability of paper and online test modes mode at the overall and item level test levels. In some studies students received only one test either on paper or online and then the performances of both tests were compared. In other studies, students received two tests in different modes, which allowed exploration of the interaction of test mode and test mode order effect. Finally, it discusses the types of questions commonly found in the field of general chemistry.

### Testing Effect

Learning usually occurs during the study phase and testing is one of the methods to assess it. In the surveys on study strategies, college students report using rereading as their primary study method (Carrier, 2003; J. Karpicke, Butler, & Roediger, 2009). The phenomenon known as the testing effect is generally studied by comparing the student performance that results from conditions which vary in amount of studying (i.e., rereading) and testing. The research on the testing effect has shown that repeated testing enhances long-term memory development better than only rereading (Roediger & Karpicke, 2006b). Karpicke and Roediger (2007) asked 60 undergraduate students to learn 40 unrelated words in one of three ways. In the standard learning condition, subjects studied (S) the word then recalled it by taking a free-recall test (T) (i.e., STST). In the

repeated-study condition, subjects studied the item three times and recalled it once (i.e., SSST). Finally, in the repeated-test condition, subjects studied the item once and recalled it three times consecutively (i.e., STTT). A final free-recall test was conducted one week later to examine the testing effect on long-term retention. The results showed that during the learning phase students under the standard condition (STST) outperformed students subjected to the other two conditions. However, when performance was tested one week later, standard and repeated test conditions outperformed the repeated study condition (i.e., 68%, 64%, and 57% of the words were recalled, respectively). These findings tend to suggest that repeated studying and repeated testing influence long-term learning in different ways. More importantly, the frequent testing slows down forgetting (Spitzer, 1939).

Multiple studies have replicated the findings of Karpicke and Roediger's (2007) study using a wide variety of learning materials, including word lists (Carpenter, 2009, 2011; Halamish & Bjork, 2011; Kornell, Bjork, & Garcia, 2011; D. J. Peterson & Mulligan, 2013; Zaromb & Roediger, 2010), foreign language vocabulary (Carpenter, Pashler, Wixted, & Vul, 2008; Finn & Roediger, 2011; Kang & Pashler, 2014; Pyc & Rawson, 2010; Vaughn, Rawson, & Pyc, 2013), text passages (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Butler, 2010; Clark & Svinicki, 2014; Einstein, Mullet, & Harrison, 2012; Little & McDaniel, 2015), narratives (Jason C K Chan, Thomas, & Bulevich, 2009), pictures (Pastötter, Weber, & Bäuml, 2013), symbol-word pairs (Coppens et al., 2011) and video-recorded lectures (Butler & Roediger, 2007; Szpunar, Khan, & Schacter, 2013). Therefore, the testing effect appears to be a very robust and replicable phenomenon.

Most of the studies on the testing effect focus on the benefits of repeated testing of the material that was learned shortly before the test. However, it is commonplace to test college students on the material that they have learned several weeks or months ago (e.g., cumulative test). Instead of asking participants to study the material during the session and before the first practice test, participants in this study were recruited from a general chemistry course where they studied the material for several weeks prior to their session. Thus, the study phase occurred during the regular course period and not during the experiment' session. This difference in the testing effect design allowed to study the benefits of testing under conditions characteristic of college settings.

### **How and why the testing effect works**

Given that frequent testing improves long-term retention, how can the benefits of the testing effect may be explained? Research points out that a transfer appropriate processing (TAP) concept may be a mechanism to account for the testing effect. TAP suggests that performance on the final test is enhanced when type of processing during encoding (i.e., learning) and retrieval (i.e., testing) stages match (Morris, Bransford, & Franks, 1977). For example, students can retrieve course information easier when teachers provide similar cues to students during homework, quizzes, and a final test. McDaniel (2007) illustrated this idea with a simple example: a teacher who actively engages students in discussion on concepts and their relationship to each other, but who tests students' knowledge with a set of specific definitions in multiple-choice format creates a mismatch between encoding and retrieval processing, thus leading to students' poor performance on the test.

Veltre, Cho, and Neely (2014) furthered this idea by testing test-based TAP and using the procedures common in the testing effect experiments. First, authors assigned 48 students to the Restudy Review group and 48 students to the Test Review group. Second, subjects learned 50 English words. After a 5-minute filler task, subjects in the Restudy Review group restudied all items one more time while students in the Test Review group recalled the material with half items cued semantically and half items cued orthographically. For example, if the target word were “ABOVE”, its semantic cue was “BEYOND” and its orthographic cue was “AB\_V\_”. Next, all students reviewed the items with different cues. For students in the Restudy Review group, items were recalled with half semantic and half orthographic cues. However, for students in the Test Review group, in addition to half semantic and half orthographic cues, cues were either the same or new. Finally, after two days, all students were invited to take the final test during which they were asked to recall all items as they did in the review session. These manipulations allowed researchers to not only study the testing effect (i.e., comparing the performance of two groups), but also measure the level of the testing effect relative to cue similarity. A comparison of the correctly recalled items on the final test showed that students from the Test Review group recalled more items than students from the Restudy Review group, indicating a positive testing effect. Evidence for test-based TAP was strong, also. Students recalled more items correctly as cues from the review and the final test became more similar. This finding suggests that students’ performance on the final test may be linked to presence and similarity of cues. In other words, student performance can improve when conditions during the final test resemble the conditions that were present when learning that material.



### **Testing effect in the classroom**

Although some experiments have investigated the testing effect using educationally relevant materials (e.g., prose material), only a few studies have been carried out in the classroom settings using material from participants' classes. In laboratory experiments variables such as student' study time are usually well controlled whereas in an actual classroom these variables may greatly vary. Therefore, it is important to review empirical evidence from the testing effect studies that were conducted in the classroom settings to generalize the testing effect to the classroom.

To investigate the testing effect experimentally in school, the majority of the studies examined whether frequent testing (i.e., quizzing) enhanced the retention of the material students had to learn in their class relative to non-tested material. The measure was performance on a criterial test (e.g., unit test, cumulative test) administered sometime after the learning phase. In an attempt to replicate the testing effect results in an educational setting and using class-related material, Carpenter, Pashler, & Cepeda (2009) found that 36 weeks after the learning session, 8<sup>th</sup> grade students retained better US history facts when material was tested (i.e., reading the question and writing down the answer) than restudied (reading both question and answer). Roediger, Agarwal, McDaniel, and McDermott (2011) also observed significant benefits of testing on middle school students' retention of material from their social studies classes. Students who answered in-class questions via individualized response units, or "clickers" retained the material better on their course exams than students who read the questions with the answers provided. McDaniel, Wildman, and Anderson (2012) also found that completing online quizzes (compared to reading the questions and answers) enhanced exam scores in

an online college course on brain and behavior (see also McDaniel, Anderson, Derbish, & Morrisette, 2007).

One limitation these classroom studies have is that they used only factual content as material (e.g., “Who assassinated President Abraham Lincoln?” in Carpenter et al., 2009). Recently, Dirx, Kester, and Kirschner (2014) were able to address this limitation by measuring the testing effect with the factual content and application of principles and procedures. Authors had 38 high school students either repeatedly study four times (i.e., SSSS) text on probability calculations or study, take a test, study again, and take the test a second time (i.e., STST). The final test was administered one week later and included five factual and five procedural questions that were previously used in STST condition. The results of the final test showed that students in STST condition significantly outperformed students in SSSS condition on both factual and application questions.

The study by Dirx et al. (2014) is lacking in some important respects. First, the final performance was measured with questions to which students in STST have been exposed twice during the learning phase. This design does not mirror well real classroom setting where students can be exposed to material multiple times through class activities, quizzes, additional reading, homework, and tests. Second, authors did not include any information about whether students received any feedback on their tests. Previous studies report that providing the correct answers particularly after a test enhances the testing effect (Roediger & Butler, 2011). Butler and Roediger (2008) found that when using multiple-choice tests delayed feedback is more effective than immediate feedback (e.g., displaying the correct answer after each question) and no feedback. Sharing the correct

answers with students after a test increased the proportion of correct answers and decreased the proportion of incorrect answers on a later test.

Despite limitations, the implications of studies discussed in this section are worth noting: the testing effect promotes long-term learning in educational settings using diverse and authentic classroom material.

### **Testing effect benefits**

In addition to improved learning, previous research has shown that the testing effect has been found to produce direct and indirect benefits for students. When students are tested with only two semester tests and one final test they are less likely to study consistently between the tests, leading to massed studying right before the test (Mawhinney, Bostow, Laws, Blumenfeld, & Hopkins, 1971). Frequent testing encourages students to more actively engage in their learning (Szpunar et al., 2013) and space their studying more equally over time (Mawhinney et al., 1971). This practice of frequent testing reduces student test anxiety and helps students to practice the material and identify what they need to study before their next test. Lastly, after experiencing frequent testing students report greater learning (Leeming, 2002) and satisfaction with the course (Bangert-Drowns, Kulik, & Kulik, 1991; Leeming, 2002).

### Test Mode

The test mode refers to studies that aim to compare student performance on a test that was administered online or with paper and pencil. The experiments on the test mode comparability have produced mixed and complex results. Multiple studies have aimed to

answer which test mode (e.g., traditional paper-and-pencil or web-based) has a greater effect on student learning by comparing student performance of a test taken on paper to performance of a test taken online. In their review paper, Mazzeo and Harvey (1988) analyzed the test mode effect by examining the findings of 27 research articles that compared paper and online versions of different tests, including Slossen Intelligence Test, Minnesota Multiphasic Personality Inventory, California algebra test, and others. Based on their review, authors concluded that performance on paper-and-pencil tests was not analogous to online performance. For example, out of 27 cited studies, 11 reported no significant differences between paper-and-pencil and online scores, three studies showed higher scores for online testing, and 13 showed higher scores for traditional testing (for another review, also see Bunderson, Inouye, & Olsen, 1988).

Recently, the test mode has been investigated in classroom environments using different proctoring settings. For example, Spivey and McMillan (2014) compared the test performance and study efforts of 174 students who took an upper-level finance class taught by the same instructor. Student performance was measured by the grades on tests taken on paper or online and study efforts were measured by tracking the number of times students accessed the course material via Blackboard course management software. Despite the fact that students who received all tests online were not proctored and students who took closed-book tests on paper were proctored, results showed that test mode did not affect study effort or student performance. Students who took the tests online did not exhibit significantly more or less study effort or score differently than students who took the tests in paper mode.

Other studies have reported similar results of no significant difference between paper and online test mode versions in different disciplines. Alexander, Bartlett, Truell, and Ouwenga (2000) found no significant difference between proctored online tests and proctored paper tests in a computer technology class. Likewise, Tsai and Shin (2012) demonstrated the comparability of paper-based and computer-based versions of National Board Dental Hygiene Examination test. Yet, contradictory results have been reported by Brallier, Schwanz, Palm, and Irwin (2015). In their study, researchers found that students in upper psychology course who took non-proctored online tests significantly outperformed students who took the proctored tests on paper. Although proctoring could be a plausible answer to explain these mixed results, it is yet unclear whether student performance depends on the test mode for a proctored science test.

Another area of research that lacks empirical evidence is the test mode order effect. Since student learning is often measured with multiple tests, it would be appropriate to investigate how the test mode affects performance when students take several tests.

### **Test mode order effect**

Often, students are exposed to several tests in a class. Yet, relatively few studies have explored the effect of test mode order. In other words, is there an effect of different test modes on student learning when several tests are given to a student? The implication of the test mode order is that teacher might need to consider the mode of his or her first and last test in order to promote student learning.

Johnson and Green (2006) examined the role of test mode using 104 eleven-year old students as participants and two mathematical tests, Test A and Test B, as the material. One of the tests was delivered via computer whereas the other was taken on paper. The items on tests A and B were of equal difficulty and the order of the mode and the order of the tests were counterbalanced (Goodwin, 2010). This design resulted in four groups. Group 1 took Test A on paper first and then took Test B online; Group 2 took Test A first online and then received Test B on paper; Group 3 took Test B on paper first and then Test A online; and finally, Group 4 had Test B online followed by Test A on paper. Data analysis showed that the order of test mode, or whether students took a paper or online test first, did not affect their performance on the second test. Although this experiment sheds light on the test mode order effect, it has some limitations. Since its design compared only two groups (computer first and paper mode second and paper first and computer mode second) other groups in which modes of both tests remain unchanged should be added for a more complete test mode order effect analysis. Second, this study used only eight questions per test. These limitations reinforce the need to further investigate the test mode effect using a more comprehensive design and tests with number of questions most commonly used in a discipline.

### **Types of questions**

Although the test mode studies report mixed findings, they share an important limitation that focuses on how the equivalency is analyzed. Often authors use the overall test scores as their dependent variable and mode as the independent variable when calculating the statistical significance (Alexander et al., 2000; Escudier, Newton, Cox,

Reynolds, & Odell, 2011). Even though this is a simple procedure, it does not allow the identification of some patterns in item characteristics that may reveal new information about test mode effect. For example, when Johnson and Green (2006) compared the overall performance of paper and online mathematical tests, no statistically significant difference was found between the tests, yet some differences were detected for individual questions. After some in-depth item-by-item analysis, three out of 16 total items were found to be easier on paper than online and one item was found to be easier online than on paper.

Jackel (2014) furthered this idea by examining the scores of 1652 students who took a college entrance test either online or on paper. The test consisted of three types of questions: critical reasoning (making logical decisions based on scientific, technical, and business stimuli), verbal reasoning (interpreting context from arts, humanities, and social sciences), and quantitative reasoning (solving mathematical and scientific problems). Looking at student performance by question type, the quantitative reasoning items showed the highest difference in scores. On average, online questions were 5% more difficult than the same questions delivered on paper. In particular, the highest difference between online and paper questions was observed for questions for which students had to use diagrams.

Mixed test mode effects were also reported by Steinberg, Brenneman, Lin, Carlson, Bridgeman, and Golub-Smith (2014) who examined data of four years of paper-based and computer-based Praxis test, which was taken in a mode chosen by each test-taker. This test fulfills the initial teaching licensure and passing it is required in 28 states. Comparing the average scores of three sections (reading, writing, and mathematics) while

controlling for test takers' demographics, it was found that most scores of computer-based tests were higher than the scores of paper and pencil tests for reading items. Four out five online demographic groups outperformed paper-based groups on reading items whereas only two online demographic groups outperformed paper-based groups on writing and mathematics items. The remaining demographics groups showed no significant difference between two testing modes. Further analysis showed that the level of education associated with the test mode. That is, undergraduate test-takers are more likely to choose paper tests and test-takers with at least a bachelor's degree – online tests.

### **Student test mode preferences**

Several studies have investigated students' test preferences for computer vs. paper-based tests. In their study, using a questionnaire Engelbrecht and Harding (2004) asked 106 first-year calculus students what mode of testing they preferred. The results showed that students preferred the most online assessment, (56.6%), followed by paper assessment (21.7%) and no specific preference (21.75%). In addition, students were asked to give reasons why they preferred one mode of testing to another. Students who preferred online mode, most frequently mentioned reduction in stress, flexibility of time and location, and availability of immediate feedback. Students who preferred paper mode, most frequently mentioned availability of showing their work that could lead to receiving some partial credit and familiarity of taking the test on paper. In contrast, Hochlehnert, Brass, Moeltner, and Juenger (2011) reported that 63% of fifth-year medical students preferred to take their test on paper. However, student explanations for their test mode preferences were similar to Engelbrecht and Harding's (2004) findings.



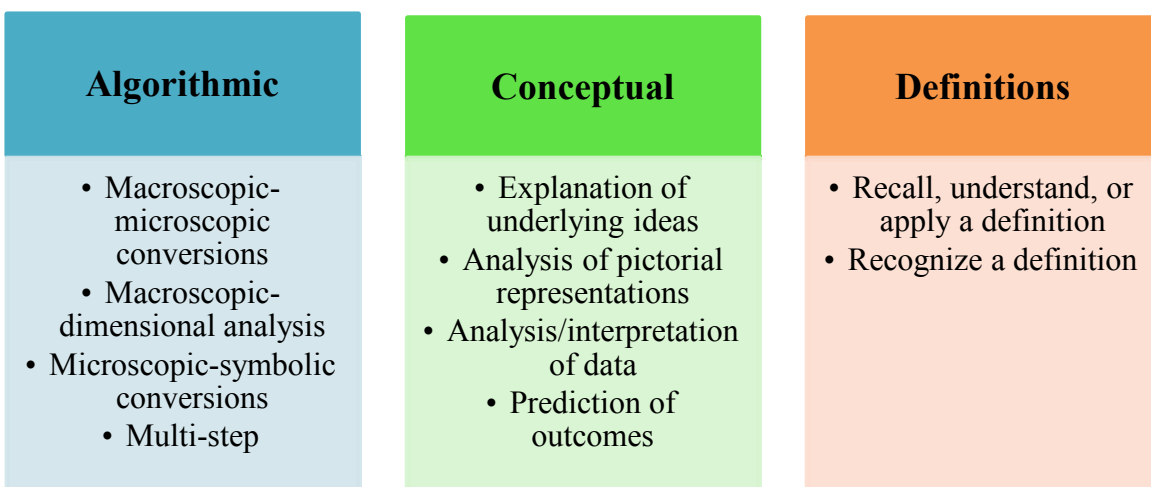
On the basis of the results of these studies, one might conclude that the test mode comparability in classroom setting has not been fully established and needs more empirical evidence. Although student test mode preferences have been studied, there is still relatively little known about students' current testing mode preferences and specifically in the field of chemistry where students may be exposed to both modes.

### Chemistry Question Types

Although previous studies provide insights into the mode effects on tests for different disciplines such as biology (D. Kim & Huynh, 2007), English (Emerson & MacKay, 2011; D. Kim & Huynh, 2008), mathematics (S. Wang, Jiao, Young, Brooks, & Olson, 2007a), and reading (S. Wang, Jiao, Young, Brooks, & Olson, 2007b), none of them specifically focused on chemistry. There is a major need to address this gap because large numbers of students take at least one chemistry course in college to fulfill requirements for their degree (e.g., biology, engineering, medical sciences, physics). Therefore chemistry test developers (e.g., American Chemical Society), test publishers, teachers, and students need to know the degree of test mode effect on student learning and test performance.

Using the previous studies (Nakhleh & Mitchell, 1993; Nakhleh, 1993; Nurrenbern & Pickering, 1987; Pickering, 1990; Sawrey, 1990), Smith, Nakhleh, and Bretz (2010) redefined the framework for general chemistry questions by analyzing the American Chemical Society (ACS) general chemistry tests. These tests were developed by a group of chemical education experts and included questions on a variety topics that are typically taught in general chemistry courses. Based on their in-depth analysis,

researchers identified three primary categories for chemistry questions: algorithmic, conceptual, and definitions. Algorithmic questions imply use of memorized process to obtain an answer, conceptual questions – non-algorithmic material, and definitions – recalling or recognizing a definition. The descriptions for each category are summarized in Figure 1.



**Figure 1.** Smith, Nakhleh, and Bretz (2010) framework for general chemistry questions

Holme and Murphy (2011) explored the differences in student performance on algorithmic and conceptual chemistry questions by reviewing 40 item pairs of two ACS Exams Institute's tests. The first test, the first-term general chemistry exam, included data from 3073 students and the second test, the second-term general chemistry exam, included data from 3557 students. An item pair consisted of two different questions that tested student knowledge on the same topic. One item tested the knowledge algorithmically and another one conceptually. The results revealed some interesting patterns; students do not answer algorithmic and conceptual questions in a similar manner even if the questions are on the same topic. While the performance of some pairs showed

better results for algorithmic questions than conceptual questions, other questions are better answered in a conceptual form. Even though this study did not comprise definition-type questions, its findings suggest that student performance on a general chemistry test items may vary by question type.

### Summary

Most evidence points to the conclusion that testing generates greater benefits for long-term learning than re-reading. However, the literature is not totally consistent on the test mode effect and still lacks empirical evidence on the test mode order effect. Previous studies lack empirical data on comparisons between groups in which testing modes changed and remained identical. With this information, a more complete picture of the test mode effects can be formed, showing when two modes and in what order can be used to produce greater learning.

Prior reviews of the literature suggest that the test scores (i.e., differences in student performance on paper-and-pencil vs. online tests) may be attributed to differences in the characteristics of the test items. Typically, studies that assess mode comparability of paper-based tests versus computer-based tests conduct their comparisons at the test level, that is, combining results of the entire paper test and compare them to the overall online test results. When comparing performances of paper and online tests using averages, valuable information could be lost. When only aggregated scores of entire tests are considered, statistically significant results of individual items could cancel out at the overall test score level. Therefore, future test mode comparability studies need to

examine the impact of test mode at the overall test score level as well as at the item level using relevant and useful parameters.

One way to investigate the testing mode at the item level is to consider a discipline's types of questions. Several studies have explored the test mode equivalency using question type as a variable of interest and found that student performance on a paper-based vs. online-based test was not always similar for each question type. In the field of general chemistry, students are frequently tested with algorithmic, conceptual, and definition questions. Little is known about the effects of chemistry question types, test mode, and test mode order on student test performance so it still remains a hypothesis for future investigation.

## CHAPTER 3. METHODS

While previous chapter highlighted limitations of studies on the testing effect and test mode, this chapter discusses how they were addressed with the methods of the present study. More specifically, in the studies reported here the testing effect was investigated using material from a general chemistry course. Within this content area, students' algorithmic and conceptual skills were tested in addition to their factual knowledge, delayed feedback was incorporated, all sessions were proctored, and tests were delivered using one or two test modes.

To pre-test the effects of test mode, test mode order, and performance of different general chemistry question types in a classroom setting, a pilot study and an experiment were conducted in undergraduate general chemistry courses. First, this chapter describes the pilot study and its impact on the design of the main experiment. Second, the chapter discusses methodology of the experiment regarding participants, materials, design, procedure, and test scoring. Participants from all studies gave their informed consent and all procedures were approved by the Institutional Review Board of the universities where studies took place (see Appendix A).

### Pilot Study

The pilot study was conducted with 102 General Chemistry (Chem 177) students at Iowa State University in fall 2012. The goal was to study how students' performance changes on items on a practice exam as a factor of test mode and test mode order, verify the experimental design, identify any constraints caused by the procedure (e.g., unclear

instructions, inappropriate use of time, etc.), and assess the validity and level of difficulty of items used on the practice exams.

### **Participants**

Students from a general chemistry (Chem 177) course were recruited for this pilot study. Out of 168 registered students, 106 participated in the study, in which they were provided with iterations of practice exams in advance of their final exam for the Chem 177 course. Due to age (i.e., being under 18 years old), four students were excluded from the data entry and analysis, thus resulting in the final sample of 102 students. The majority of participants were females (i.e., 67% - female, 33% - male) and the average age was 18.7. All sessions were conducted in small groups of one to eight students and were completed within 90 minutes. Upon session completion, each participant received a complementary access to an online American Chemical Society (ACS) general chemistry practice exam.

### **Material and design**

The database of general chemistry questions, including some from the out-of-print ACS Exams General Chemistry Test Bank and newly composed items, was used to compose practice tests. These tests consisted of 24 question pairs or 48 unique multiple-choice (MC) questions. Each pair consisted of two different questions that tested student knowledge on the same topic. To design open-ended (OE) questions, all 48 MC questions were transformed to OE format by keeping the stem and removing four alternative options, thus resulting in a total of 96 items (48 MC and 48 OE items). This step will be

discussed in more detail in the methods section of the main experiment. Five test versions were generated in which the order of items was randomized for each test and participant.

For the experimental section of the study, participants were randomly assigned to one of the nine conditions that indicated the mode of Test 1 and Test 2. Tests were given either on paper-and-pencil (abbreviated as PP), computer (abbreviated as Com), or iPads (designated as iPad) that were provided to students by the researcher. All sessions were conducted either in the morning, afternoon, or evening. Table 1 depicts the number of participants by condition and time of the day.

**Table 1.** Pilot study participant information by the conditions and time of the sessions ( $N = 102$ )

Conditions	Morning				Afternoon					Evening			N
	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	
Test1 - Test2													
Com-Com	3				6					6			15
Com-iPad	2	1	2		1		2			1			9
Com-PP	1				4					3			8
iPad-Com	1	3	2		2		1			1			10
iPad-iPad	4				3					1			2
iPad-PP	1						3	4				2	10
PP-Com	4				3					8			15
PP-iPad	3				2					4			9
PP-PP	5				2					6			3
N	6	13	6	7	3	6	12	11	8	8	8	14	102
	Total participants: 32				Total participants: 40					Total participants: 30			

## Procedure

All sessions were proctored by a graduate student and occurred 1 – 14 days prior to the course's final examination. Students assigned to condition in which at least one test mode was on the computer took place in a computer laboratory whereas sessions in the other conditions were proctored in a standard classroom. Therefore, five conditions

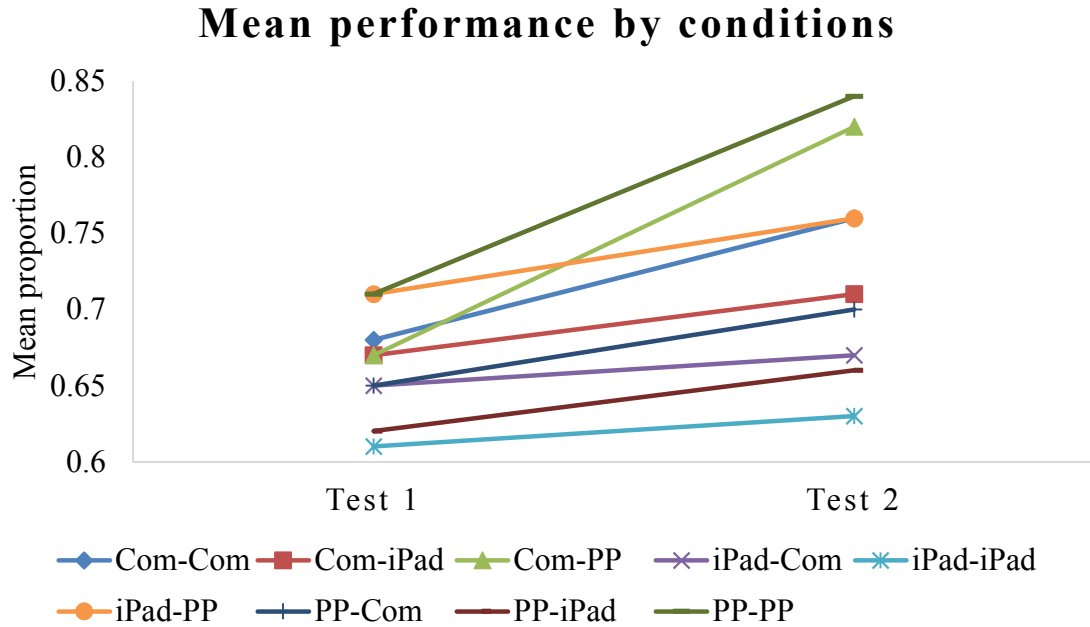
were held in the computer lab (Com-Com, Com-iPad, Com-PP, iPad-Com, and PP-Com) and four conditions were held in the classroom (iPad-iPad, iPad-PP, PP-iPad, and PP-PP). Online tests were delivered via Blackboard software and paper tests were pre-printed before each session. A 35-minute time limit was imposed for each practice test. During the test students were permitted to use only a basic calculator, scratch paper, and general chemistry data sheet provided by the proctor. After each test, students were given a few minutes to verify their own performance by comparing their answers to the list of correct answers. Before beginning the second test, students were asked to complete a non-test related activity for about five minutes. After the second test was complete, subjects received their code and directions for the online practice exam as compensation and were dismissed.

## Results

To analyze the results, the proportion of correct responses on Test 1 and Test 2 was calculated for each student. Overall, student performance on Test 2 was higher than performance on Test 1, yet these changes in performance were different for each condition. While the mean proportion of students in iPad-iPad condition improved by .03 on average, the mean proportion of students in Com-PP condition changed by .15 from Test 1 to Test 2 (Figure 2). These results tend to suggest that changes in student performance may vary by the mode in which tests are delivered to students.

An analysis of variance (ANOVA) showed that the mode of Test 1 (computer, paper, or iPad) had no statistically significant effect on Test 1 performance,  $F(2, 99) = .07, p = .9359$ . This finding is consistent with Tsai and Shin's (2012) results who





**Figure 2.** Mean proportions of performance on Test 1 and Test 2 by conditions ( $N = 102$ )

compared overall student performance of computer-based or paper-based comprehensive National Board Dental Hygiene Examination test and found no statistically significant differences between two test versions. There was also a non significant effect of Test 1 mode on Test 2 performance,  $F(2, 99) = 1.96, p = .1467$ . However, one-way ANOVA revealed a significant effect on Test 2 mode on performance of Test 2,  $F(2, 99) = 8.41, p = .0004$ , with a large effect size ( $\eta^2 = .15$ ). Post hoc analyses using Tukey's HSD test indicated that students who took Test 2 on paper ( $M = .81, SD = .12$ ) performed better on Test 2 than students who took Test 2 using iPad ( $M = .66, SD = .15$ ) and students who took Test 2 on computer ( $M = .71, SD = .16$ ). There was no significant difference on Test 2 performance between students who took Test 2 on computer ( $M = .71, SD = .16$ ) and iPad ( $M = .66, SD = .15$ ). These results indicated that student performance on Test 1 and

Test 2 did not depend on the mode of Test 1, but the performance on Test 2 depended on the mode of Test 2.

Some observations were noted during the testing sessions. First, not all students were familiar with the iPad. This caused some students to spend several minutes learning about iPad's features and, as the result, having less time to complete the test. Second, most students completed Test 2 within 20 minutes, so the time limit for Test 2 was adjusted in subsequent sessions of the pilot study. Students still had 35 minutes for Test 2, but if finished earlier, students were allowed to move to the next step.

### **Changes to the study**

Among many advantages, conducting a pilot study allows researchers to collect preliminary data and identify potential problems that might occur using the proposed design (van Teijlingen & Hundley, 2001). Data analysis and personal observations recorded during the pilot study indicated the need to make several improvements to the study material, design, and research procedure. Next, this section summarizes the applied adjustments and their impact on the overall methods of the main study.

#### Material changes

First, item difficulty was calculated for each item by dividing the number of correct responses by the number of total students who saw that item. Ding and Beichner (2009) suggest revising the items with item difficulty values lower than 0.3 or higher than 0.9. Therefore, pairs whose at least one question exhibited either floor effect (1% - 14% of students answered a particular question correctly) or ceiling effect (91% - 100% of

students answered a particular question correctly) were modified by either re-writing the stem (e.g., stem's 3D image was replaced with the Lewis structure), replacing at least one of the alternatives, or substituting the item with a new item (e.g., Redox Chemistry topic was replaced with Net Ionic Equation).

Second, data analysis suggested the need to reexamine the correct answers. Changing the question format from MC to OE originally led to all OE questions contain the same correct answer (e.g., "Question: What is the oxidation number of manganese in  $[\text{MnO}_4]^{2-}$ ?" Alternatives : "+4", "+6", "+8", "+10" Answer: "+6"); vs. Question: "What is the oxidation number of manganese in  $[\text{MnO}_4]^{2-}$ ?" Answer: "+6"). However, as the data analysis showed, the correct answers to OE periodicity questions were inaccurate. Hence, the answers to these questions were changed and differed from answers of MC questions (e.g., "Question: Which halogen atom has the largest radius?" Alternatives: "Br", "Cl", "F", "I" Answer: "I"); vs. Question: "Which halogen atom has the largest radius?" Original Answer: "I", Updated Answer: "At"). Even though "At" is larger than "I", it could not be the correct answer for the MC version because it was not given as a choice, which cases the subject to evaluate all options and pick the best answer among the presented alternatives. Such question designs (i.e., excluding the best answer from four alternatives) are common to general chemistry MC tests. As the result of these alterations, 33 changes were made (12 stems rewritten, 11 changes to item alternatives, 8 item replacements with new ones, and 2 answer changes to OE questions). The complete list of the final 96 questions is reported in Appendix B.

Finally, rather than re-generating several test trials, a technique called counterbalancing was used. When running an experiment using repeated measures

design, researchers frequently use the counterbalancing method (Butler & Roediger, 2008; McDaniel et al., 2012; Zaromb & Roediger, 2010). Counterbalancing helps researchers to control order and learning effects by spreading out the effects evenly over all conditions (Goodwin, 2010; Howell, 2002). All counterbalancing steps will be outlined in the upcoming materials section of the main study.

### Design changes

Since computer labs were not always available, the computer mode was changed from desktop computers to personal laptops so sessions could be conducted in a regular classroom. Due to no significant difference between iPad and computer modes and the observation that some students had limited familiarity with using iPads, the number of modes was reduced from three (e.g., paper, computer, and iPad) to two (paper and computer). This change also resulted in reducing the number of conditions from nine to only four (i.e., PP – PP, PP – Com, Com – PP, and Com – Com), thus, simplifying the design. In addition, these changes in design allowed recruiting more students and increasing the overall sample size. Due to a limited number of iPads, only five students could be scheduled for any condition that included iPad as one of its test modes.

Eliminating iPads as the mode and being able to conduct sessions in a regular classroom, allowed the scheduling of up to 80 students for any session.

Another change in design included adding an extra practice test between Test 1 and Test 2. This additional test contained half of items from Test 1. Observing improvement in student performance from Test 1 to Test 2 (Figure 1), additional practice test allowed to examine whether it would lead to additional improvements in student

performance on items that were practiced twice relative to the ones that were practiced once. More detailed information about intermediate test is presented in the design section under the main study.

### Procedure changes

The time for the non-test related activity was increased from five minutes to 20 minutes. This permitted students to take a longer break before completing the final test and helped to reduce potential item and answer memorization. Next, changes were made to time instructions for Test 2. Even though equal times (i.e., 35 minutes) were allocated for Test 1 and Test 2, students were permitted to move to the next phase of the study if Test 2 was completed earlier. Lastly, due to the limited ability to insert and view scientific symbols and images in Blackboard, all online tests of the main study were transferred and delivered using Qualtrics software (Qualtrics, 2015).

### Scoring changes

While the calculation and use of proportion of correct responses as a dependent measure is a frequent practice in testing effect studies where participants take several tests (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Roediger III, Agarwal, McDaniel, & McDermott, 2011), it requires additional steps to classify responses into various equivalent groups so analyses and recommendations could be made for each group. For example, high-performing students may respond to the test mode differently than low-achieving students. To address this concern, a normalized gain using Hake's (1998) formula was used as the dependent measure in place of the proportions.

Calculation of the normalized gain is useful because it considers student's initial (pre) and final (post) state of knowledge and then normalizes his or her pre/post gap, thus accounting for students' high scores (ceiling effect) and low scores (floor effect).

### Main Study

Following the pilot study, the main study was conducted with three undergraduate first-semester general chemistry courses at two universities.

### Participants

A total of 664 students were recruited from general chemistry courses: two taught at the Iowa State University and one taught at a university in the Pacific Northwest. Both institutions are four-year public universities with undergraduate enrollment above 20,000 students, offering undergraduate and graduate programs in chemistry, with traditional general chemistry courses taught by several faculty members. The similarity of three courses was established based on the material covered in each class. All courses included in the study covered nomenclature, states of matter, chemical reactions, atomic structure, chemical bonding, the properties of gases, and thermochemistry. General Chemistry (Chem 177) at Iowa State University and third quarter general chemistry (3Q GenChem) course at the university in the Pacific Northwest were designed for science students and General Chemistry for Engineering Students (Chem 167) at Iowa State University covered the same content, but with an emphasis on applications in engineering. Table 2 provides demographic information related to age, gender, ethnicity, year in school, and

**Table 2.** Frequency and percentage of participants by gender, ethnicity, year in school, and instructor and average of age and percentage received on a final exam for each course ( $N = 664$ )

Courses				
Course and semester	Chem 167 Fall 2013	Chem 177 Fall 2013	Chem 177 Spring 2014	3Q GenChem Spring 2014
<i>N</i>	222	218	73	151
Student Demographics				
Age (mean)	18.6	19.0	19.4	19.1
Gender				
Male	182 (82%)	82 (37.6%)	30 (41.1)	40 (26.5%)
Female	39 (17.6)	133 (61%)	43 (58.9)	109 (72.2)
No response	1 (.4)	3 (1.4%)	0	2 (1.3%)
Ethnicity				
American Indian	2 (.9%)	0	0	0
Asian or Pacific Islander	29 (13.1%)	21 (9.6%)	8 (11%)	33 (21.8%)
African American	5 (2.3%)	9 (4.1%)	3 (4.1%)	1 (.7%)
Caucasian	163 (73.4%)	163 (74.8%)	50 (68.5%)	93 (61.6%)
Hispanic/Latino	8 (3.6%)	7 (3.2%)	7 (9.6%)	7 (4.6%)
Biracial/Multicultural	4 (1.8%)	8 (3.7%)	4 (5.5%)	12 (8%)
Prefer to not disclose	10 (4.5%)	8 (3.7%)	1 (1.3%)	3 (2%)
No response	1 (.4%)	2 (.9%)	0	2 (1.3%)
Year in school				
Freshman	205 (92.3%)	163 (74.8%)	44 (60.3%)	102 (67.6%)
Sophomore	9 (4%)	40 (18.4%)	24 (32.9)	34 (22.5%)
Junior	5 (2.3%)	6 (2.7%)	5 (6.8%)	9 (6%)
Senior	0	5 (2.3%)	0	2 (1.3)
Graduate student	0	0	0	0
Other*	1 (.5%)	2 (.9%)	0	2 (1.3%)
No response	2 (.9%)	2 (.9%)	0	2 (1.3%)
Course Information				
% on the final exam (mean)	79.57%	66.5%	59%	72.3%
Instructor				
A	98 (44.1%)			
B	120 (54.1%)			
C		123 (56.4%)		
D		0	29 (39.7%)	
E		95 (43.6%)	43 (58.9%)	
F				38 (25.2%)
G and H**				112 (74.2%)
Don't know	4 (1.8%)	0	1 (1.4%)	1 (.6%)
No response	0	0	0	0

\* 1 returning alumni, 1 transfer student, 1 preference to not disclose, and 2 post-baccalaureate

\*\* Instructors G and H taught their sections together

course information, such as the mean percentage obtained by students in each course during a final examination given shortly after the study and the range of instructors in different sections for three courses from two universities.

Students from both universities were recruited through class visits, recommendations from faculty, and emails. At the time of recruitment, students were informed about the focus of the study, procedure for each session, and benefits each participant will receive. A few days before the sessions were conducted, all students received a link to a survey asking them to indicate their interest in participation by registering for a session. Based on this information, students were assigned to one of four conditions in a manner that would allow equal sample sizes among conditions to the best extent possible (see Table 3).

**Table 3.** Number and percentage of participants by conditions for each course and percentage of participants over all courses ( $N = 664$ )

Conditions (Test 1 – Test 2)	Chem 167 Fall 2013 $N = 222$	Chem 177 Fall 2013 $N = 218$	Chem 177 Spring 2014 $N = 73$	3Q GenChem Spring 2014 $N = 151$	Total $N = 664$
Com-Com	51 (23%)	48 (22%)	0	74 (49%)	173 (26%)
Com-PP	54 (24.3%)	45 (20.6%)	29 (39.7%)	37 (24.5%)	165 (25%)
PP-Com	48 (21.6%)	73 (33.5%)	44 (60.3%)	0	165 (25%)
PP-PP	69 (31.1%)	52 (23.9%)	0	40 (26.5%)	161 (24%)

The data were collected at three different times: at the end of the fall semester of 2013 the data were collected from Chem 167 and Chem 177 students at Iowa State University, at the end of spring quarter of 2014 the data were collected from the third quarter general chemistry students (3Q GenChem) at the university in the Pacific Northwest, and at the end of spring semester of 2014 from Chem 177 students at Iowa State University again.



### Iowa State University participants

A total of 523 Iowa State University students were recruited from General Chemistry for Engineering Students (Chem 167) in the fall semester of 2013 and students from first-semester General Chemistry (Chem 177) courses in the fall semester of 2013 and spring semester of 2014. Chem 167 was taught by two instructors (Instructors A and B), Chem 177 (fall 2013) by three (Instructor C, D, and E), and Chem 177 (spring 2014) by two (Instructor D and E) (see Table 2).

Data from ten students who participated during the fall semester of 2013 were excluded. Three students left in the middle of the session, two students were below 18 years of age, and five students experienced some internet issues that prevented their data from being fully recorded. Therefore, the total sample size from Iowa State University was 513 students (Chem 167, fall 2013 – 222 students; Chem 177, fall 2013 – 218; and Chem 177, spring 2014 - 73). Table 4 summarizes how many students were enrolled in each course, how many students signed up for the sessions, how many students participated, and how many students were excluded from the study due an incomplete test, young age, or online issues. In addition to the opportunity to practice the course content prior to the final examination, all students received a complimentary access to the online ACS general chemistry practice exam.

### University in the Pacific Northwest participants

An additional 151 participants were recruited from a general chemistry course taught at the university in the Pacific Northwest (see Table 4). This class was taught by three instructors (Instructor F, G, and H), however, two instructors (Instructor G and H)

co-taught their sections together (see Table 2). Names of student participants from the university in the Pacific Northwest were entered into a raffle for the chance to win a mini iPad.

**Table 4.** Student information by course ( $N = 664$ )

Course and semester	Iowa State University			University in the Pacific Northwest
	Chem 167 Fall 2013	Chem 177 Fall 2013	Chem 177 Spring 2014	3Q GenChem Spring 2014
<i>N</i> of students enrolled in class	1221	663*	395	789
<i>N</i> of students who signed up for study**	597		120	205
<i>N</i> of students who participated	450		73	151
<i>N</i> of students excluded	10		N/A	N/A
Total	222	218	73	151

\* Students were recruited from 32 out of 64 sections only.

\*\* Students were given a link to a survey in which they indicated their interest to participate in the study and preference for day and time of the session.

## Materials and design

### Test material

The content of the first-semester general chemistry taught at a university level is typically presented in six chapters each containing four topics, as illustrated in Table 5.

**Table 5.** Chapters and topics for each chapter from a first-semester general chemistry course

	Chapters					
	1. Matter	2. Moles & Reactions	3. Structure & Bonding	4. Stoichiometry	5. Gases	6. Thermochemistry
1.	Quantum (Numbers) Mechanics	Naming Inorganic Compounds	Bonding	Mole to Mole Stoichiometry	Kinetic Molecular Theory	Specific Heat
2.	Periodicity	Percent Composition	Bond Length	Limiting Reagent	Ideal Gas Law	Heat of Formation
3.	Atomic Structure	Acid/Base Chemistry	Lone Pairs Electrons	Molarity	Partial Pressure	Exothermic/ endothermic
4.	Isotopes	Net Ionic Equation	Shape	Balancing Equation	Boyle's Law	Heat of Reaction

A database of ACS questions was used as a guide to compose one pair of MC questions for each of the 24 topic/chapter combinations. For example, to test students' understanding of naming inorganic compounds (Chapter 2, Topic 1), the following pair of questions was composed:

**Question A:**

What is the formula of chromium(III) carbonate?

- A.  $\text{Cr}_3\text{CO}_3$       B.  $\text{Cr}(\text{CO}_3)_3$       C.  $\text{Cr}_2(\text{CO}_3)_3$       D.  $\text{Cr}_3(\text{CO}_3)_2$

**Question B:**

What is the formula of cobalt(II) phosphate?

- A.  $\text{Co}_2\text{PO}_4$       B.  $\text{Co}(\text{PO}_4)_2$       C.  $\text{Co}_2(\text{PO}_4)_3$       D.  $\text{Co}_3(\text{PO}_4)_2$

Twenty-four question pairs were constructed. Since students took two practice tests in one session, one question from each pair was assigned to Test 1 while the other was assigned to Test 2. For example, if students saw Question A on Test 1, they saw Question B on Test 2, whereas other students saw Question B on Test 1 and Question A on Test 2. It is important to note that this study aimed to measure the effects of testing on

knowledge within a domain, rather than on the ability to remember the answer to a specific question that was asked previously.

Since general chemistry students are frequently exposed to both MC and OE questions on their tests, all students received half of the questions in MC format, and half in OE format on both tests. The OE questions were written by removing the four answer options so that only the question stem remained (e.g., “What is the formula of cobalt(II) phosphate?”). Thus, 48 OE questions were created by transforming each MC question into its corresponding OE version, resulting in a pool of 96 unique questions (Appendix B). In total, each question pair included four questions: two MC and their two replicated OE questions. Chemistry experts were asked to classify all 96 questions into three categories: algorithmic (68 questions or 17 question pairs), conceptual (20 questions or 5 question pairs), and definitions (8 questions or 2 question pairs). Table 6 lists the distribution of algorithmic, conceptual, and definition questions among chapters and

**Table 6.** Distribution of algorithmic, conceptual, and definition questions  
17 algorithmic questions are highlighted in dark grey, 5 conceptual – medium grey, and 2 definitions (quantum mechanics and periodicity) – light grey.

	Chapters					
	1. Matter	2. Moles & Reactions	3. Structure & Bonding	4. Stoichiometry	5. Gases	6. Thermochemistry
1.	Quantum (Numbers) Mechanics	Naming Inorganic Compounds	Bonding	Mole to Mole Stoichiometry	Kinetic Molecular Theory	Specific Heat
2.	Periodicity	Percent Composition	Bond Length	Limiting Reagent	Ideal Gas Law	Heat of Formation
3.	Atomic Structure	Acid/Base Chemistry	Lone Pairs Electrons	Molarity	Partial Pressure	Exothermic/ endothermic
4.	Isotopes	Net Ionic Equation	Shape	Balancing Equation	Boyle's Law	Heat of Reaction

topics.

In order to reduce the effect of question format order (i.e., whether a student sees a question from a pair in MC or OE format first), the format of question pairs was counterbalanced. Counterbalancing was used so that 24 questions appeared equally often as MC and OE questions, and question forms A and B appeared equally often on Test 1 and Test 2. Consequently, eight different test versions were formed. Table 7 shows how question format and test order were counterbalanced for Chapter 1 questions. The same method was used for all chapters. At the time of a session, each student received a test in one of these eight test versions. Lastly, the order of all 24 questions for Test 1 and Test 2 was always randomized for each student.

One of the challenges of classroom-based research is that some of the students may not have a strong background on a topic, and may perform poorly on the tests. If so, tests may not be expected to benefit learning as much as they have been shown to do in laboratory experiments that rely less on any specific content knowledge. Anticipating this potential issue and to explore the differences in gains of repeated items and non-repeated items, an additional practice test including half of the questions from Test 1 was given to each student before they moved on to Test 2. All students thus completed Test 1 containing 24 questions, and then after receiving feedback, they completed Test 1A or Test 1B in which 12 of these questions appeared again, followed by feedback. The 12 questions that appeared on Test 1A consisted of those representing Chapters 1, 3, and 5 and Test 1B consisted of items from Chapters 2, 4, and 6. Thus, all students completed Test 1 (consisting of four questions each from six chapters), followed by Test 1A or Test 1B (consisting of four repeated questions each from three of those chapters—either 1-3-5

or 2-4-6), followed by Test 2 containing a new set of four questions from each of the six chapters. Students were randomly assigned to receive Test 1A or Test 1B.

The final manipulation regarded the test mode and test mode order. To investigate how test performance is affected by the mode in which a test is delivered, tests were

**Table 7.** Example of counterbalancing question formats and test order for Chapter 1. These eight versions illustrate how four questions from Chapter 1 were counterbalanced with question format (multiple-choice “MC” and open-ended “OE”) and test order (A vs. B). All “A” items appeared together on one test and all “B” items appeared on another test.

#### Version I

Chapter 1: Topics	Order of Tests	
	Test 1	Test 2
1. Q. Mechanics	A (MC)	B (MC)
2. Periodicity	A (MC)	B (OE)
3. Atomic Structure	A (OE)	B (MC)
4. Isotopes	A (OE)	B (OE)

#### Version V

Chapter 1: Topics	Order of Tests	
	Test 2	Test 1
1. Q. Mechanics	B (MC)	A (MC)
2. Periodicity	B (OE)	A (MC)
3. Atomic Structure	B (MC)	A (OE)
4. Isotopes	B (OE)	A (OE)

#### Version II

Chapter 1: Topics	Order of Tests	
	Test 1	Test 2
1. Q. Mechanics	A (MC)	B (OE)
2. Periodicity	A (OE)	B (MC)
3. Atomic Structure	A (OE)	B (OE)
4. Isotopes	A (MC)	B (MC)

#### Version VI

Chapter 1: Topics	Order of Tests	
	Test 2	Test 1
1. Q. Mechanics	B (OE)	A (MC)
2. Periodicity	B (MC)	A (OE)
3. Atomic Structure	B (OE)	A (OE)
4. Isotopes	B (MC)	A (MC)

#### Version III

Chapter 1: Topics	Order of Tests	
	Test 1	Test 2
1. Q. Mechanics	A (OE)	B (MC)
2. Periodicity	A (OE)	B (OE)
3. Atomic Structure	A (MC)	B (MC)
4. Isotopes	A (MC)	B (OE)

#### Version VII

Chapter 1: Topics	Order of Tests	
	Test 2	Test 1
1. Q. Mechanics	B (MC)	A (OE)
2. Periodicity	B (OE)	A (OE)
3. Atomic Structure	B (MC)	A (MC)
4. Isotopes	B (OE)	A (MC)

#### Version IV

Chapter 1: Topics	Order of Tests	
	Test 1	Test 2
1. Q. Mechanics	A (OE)	B (OE)
2. Periodicity	A (MC)	B (MC)
3. Atomic Structure	A (MC)	B (OE)
4. Isotopes	A (OE)	B (MC)

#### Version VIII

Chapter 1: Topics	Order of Tests	
	Test 2	Test 1
1. Q. Mechanics	B (OE)	A (OE)
2. Periodicity	B (MC)	A (MC)
3. Atomic Structure	B (OE)	A (MC)
4. Isotopes	B (MC)	A (OE)

given to each student either on paper or online. The study also explored the test mode order. Using two practice tests, allowed forming four experimental conditions that varied in test mode of Test 1 and Test 2. Two of these conditions, as in Johnson and Green (2006) were the groups in which the mode of Test 1 was different from mode of Test 2 (i.e., Test 1 delivered on paper and Test 2 delivered online using a computer). The other two conditions were the groups in which the mode of both tests remained the same throughout the session (e.g., both tests delivered on paper or online). Thus, students were assigned to one of four between-subject conditions manipulating test mode and mode order: Computer – Computer, Paper – Paper, Computer – Paper, and Paper – Computer (also see Table 3). The design of such conditions allowed the investigation of not only the comparability of two test modes, but the effect of mode change, too.

#### Test mode student preferences survey

In addition to the test material, all students completed a test mode preference survey. In the survey, students were asked to indicate whether they have taken a test online previously, in which mode they would prefer to take their next general chemistry test, and to explain their responses.

#### **Procedure**

After approximately two-thirds of the course material was covered in class, students were invited via class visits and emails to participate in this study. To sign up for one of the sessions, students were given a link to an online calendar that listed multiple options for testing days and times. The days and times for each condition were

determined before student recruitment. Therefore when students were signing up for the sessions they were not aware what condition they would be in. Students were notified of their actual testing schedule by email, which included an electronic consent form for them to fill out. Reminder emails were sent out to students one day prior to their scheduled testing session.

To mimic the conditions in which college students usually take general chemistry tests, all sessions were conducted on campus in chemistry classrooms and were proctored by a researcher. Before the session, students were each given a copy of the periodic table and a general chemistry data sheet that contained some general chemistry formulas. The same or a similar data sheet was allowed during the course's final examination. Students were also allowed to bring and use a basic scientific calculator and scratch paper. Students assigned to one of the conditions that involved performing the experiment on the computer were instructed to bring their fully charged laptops to the session.

Upon their arrival, students were seated at a desk and were given the instructions according to the test mode condition they were assigned. Students in the Computer – Computer condition were instructed to log in to Qualtrics where they completed the consent form, followed by Test 1, Test 1A or Test 1B, and Test 2, and then the demographics survey. Students in the Computer – Paper condition were instructed to log in to Qualtrics where they completed the consent form, followed by Tests 1 and Test 1A or Test 1B. Test 2 and the demographics survey were given on paper to them after they completed a distractor task. Students in the Paper – Computer condition received a copy of the printed consent form, Test 1 and Test 1A or Test 1B, and a distractor task. Then they were instructed to log in to Qualtrics where they completed Test 2 and the



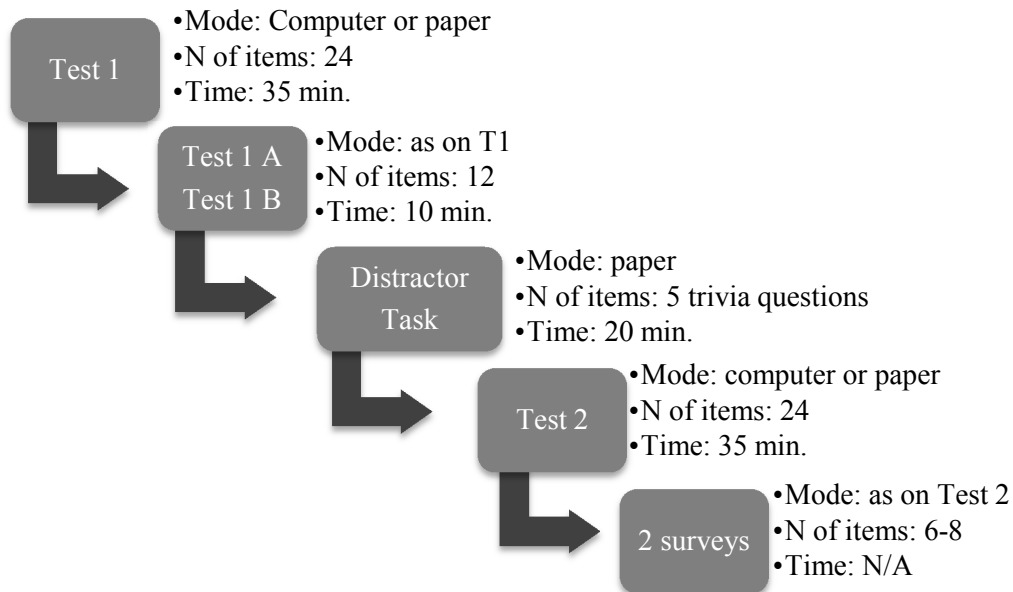
demographics survey. Finally, students in the Paper – Paper condition received a copy of printed material that included a copy of the consent form, Test 1, Test 1A or Test 1B, Test 2, and the demographics survey.

After students signed the consent form the proctor asked students to carefully read the instructions to Test 1. Appendix C illustrates the instructions given to all computer-based tests and Appendix D illustrates the instructions given to all paper-based tests. Students were then given 35 minutes to complete Test 1. After completing Test 1, students received immediate feedback on Test 1 in the form of the correct answers. Feedback was delivered in the same mode as the test had been administered. For example, if the student completed Test 1 online, the student received feedback online after he submitted his answers online along with a copy of his submitted answers. If the student took Test 1 via paper-and-pencil, he or she received feedback in the form of printed answers and could review the answers by flipping the pages of the exam back and forth.

After students were done viewing the feedback, the proctor collected students' scratch paper so students could not use their notes for the subsequent test. Students were then given ten minutes to complete Test 1A or Test 1B. As in Test 1, after completing Test 1A or Test 1B students received feedback on the correct answers, and the proctor collected students' scratch papers.

Following Test 1A or Test 1B, students completed a 20-minute distractor task in which they were asked to write down the answers to five trivia questions. Next, students were given up to 35 minutes to complete Test 2. Students were allowed to check their scores when they were done with the test and then were asked to complete two short

surveys: the demographics survey and test mode student preferences survey. When a student had finished with his or her test, the proctor collected the student's packet or checked the student's online submission upon completion, gave the student his or her complimentary access to the ACS practice test if the student was from Iowa State University, thanked, and dismissed the student. For students from the university in the Pacific Northwest, after each session the names of participants were recorded on a separate sheet for a drawing that was completed shortly after the last session.



**Figure 3.** Diagram of the main study procedure

All sessions were conducted in groups of 3 - 40 students and lasted approximately 90 - 120 minutes. Figure 3 shows the timeline of the session, indicating the sequence, the mode of test, the number of items, and time allocated for each test.

## Scoring

The data for all participants were scored by the principal investigator and several chemistry graduate students and post-doctoral associates. The maximum total score for Test 1 and Test 2 was 24 and for Test 1A or Test 1B was 12. Each student's response was coded 0 for an incorrect answer and 1 for a correct answer. With the assistance of several chemistry experts, a grading rubric was created for the open-ended questions. According to the rubric, OE responses were coded as correct if (1) students produced the exact correct answer; (2) students wrote an acceptable answer, a short answer that closely resembled the exact correct answer, or (3) student's calculations were within the acceptable numeric range. Appendix B lists all questions along with the exact and acceptable answers.

After scoring all questions, a normalized difference score (Test 2 Score – Test 1 Score / 100% - Test 1 Score) as in (Hake, 1998) was calculated for each participant. For example, if a student answered 14 out of 24 questions correctly on Test 1 (i.e., .58) and 18 on Test 2 (i.e., .75), his or her normalized gain would be calculated as follows:  $(.75 - .58) / (1 - .58) = .4$ . This number indicates how much a student's performance increases out of his or her maximum possible increase. If a student's mean of Test 1 was 1 (i.e., student answered all questions correctly on Test 1), his or her response was removed from the analysis, as student's normalized gain could not be calculated. Lastly, normalized gains for algorithmic, conceptual, and definition questions were calculated for each participant. Therefore, each participant had four scores: overall normalized gain, algorithmic normalized gain, conceptual normalized gain, and definition normalized gain.

## CHAPTER 4. RESULTS

This chapter presents the results of the data analysis that answered the following two research questions:

- (1) Is there a difference in the measured performance on general chemistry tests for students across different test mode groups?
- (2) Is there a difference in average student gains for algorithmic, conceptual, and definition questions based on the mode by which the test was delivered?

Before answering each research question, data were prepared for analyses and group equivalency among four courses was established.

### Preparing Data

The data consisted of scores for Test 1, Test 1 A or Test 1B, Test 2, and students' self-reported demographics and test mode preferences. The data from the paper tests and surveys were first manually entered into an Excel spreadsheet. The data from the online tests and surveys were downloaded and then merged with data from the paper groups. All data were analyzed using STATA 13 statistical software (StataCorp, 2013).

A significance level of  $p \leq .05$  was used for all analyses (i.e.,  $\alpha = 0.05$ ). When significant effects were found, eta squared ( $\eta^2$ ) (Pearson, 1911) was used for one-way analysis of variance (ANOVA) tests. To interpret the strength or magnitude of shared variance between variables, guidelines by Cohen (1988) were used (also see Ferguson, 2009; Pallant, 2010).

### Evaluating Equivalency

In order to verify whether participants from different courses had similar prior knowledge, group equivalence by course was investigated. If differences were found, then participants would be separated into groups defined by equivalency. Testing group equivalence may indicate the presence of a confounding variable (Lewis & Lewis, 2005). For this study, checking equivalence is important to help establish that the independent variables in the experiment are the source of the observed effects. While prior chemistry knowledge of students is inevitably important, it was controlled by an appropriate choice of groups.

In order to investigate equivalency across the four courses, the proportions of correct answers from Test 1 were used as the dependent variable and the course served as the independent variable. The data were processed to examine whether the difference in mean proportions of correct answers on Test 1 existed among the four courses. Participants were divided by course into four groups: Chem 167 ( $N = 222$ ), Chem 177 fall ( $N = 218$ ), 3Q GenChem ( $N = 151$ ), and Chem 177 spring ( $N = 73$ ). The mean proportions and standard deviations for Test 1 mean scores are presented in Table 8. A one-way between-groups ANOVA showed a statistically significant effect for course,  $F(3, 660) = 13.66, p < .001$ . The effect size, ( $\eta^2$ ), was .06, indicating a medium effect size. Tukey honestly significant difference (HSD) post-hoc tests revealed that the mean proportions for Test 1 were significantly different for the three pair-wise comparisons. (1) The mean proportion of Chem 167, fall 2013 ( $M = .54, SD = .16$ ) was significantly different from the mean proportion of Chem 177, fall 2013 ( $M = .60, SD = .17$ ). (2) The mean proportion of Chem 167, fall 2013 ( $M = .54, SD = .16$ ) was significantly different

from the mean proportion of 3Q GenChem, spring 2014 ( $M = .65$ ,  $SD = .16$ ). Lastly, (3) the mean proportion of 3Q GenChem, spring 2014 ( $M = .65$ ,  $SD = .16$ ) was significantly different from the mean proportion of Chem 177, spring 2014 ( $M = .55$ ,  $SD = .18$ ). The  $p$ -values of all Tukey's HSD pair comparisons are presented in Table 9.

**Table 8.** Mean proportions of correct answers and standard deviations on Test 1 for the four courses ( $N = 664$ )

Course	$N$	Mean proportions of correct answers on Test 1	Standard deviation
Chem 167, Fall 2013	222	.54	.16
Chem 177, Fall 2013	218	.60	.17
Chem 177, Spring 2014	73	.55	.18
3Q GenChem, Spring 2014	151	.65	.16

**Table 9.** Post-hoc comparisons using the Tukey HSD test for the Test 1 mean proportions for the four courses. The values in the table indicate the  $p$ -values for each comparison.

Courses	Chem 167 Fall 2013 $M = .54$	Chem 177 Fall 2013 $M = .60$	Chem 177 Spring 2014 $M = .55$	3Q GenChem Spring 2014 $M = .65$
Chem 167 Fall 2013 $M = .54$		.001*	.980	<.001*
Chem 177 Fall 2013 $M = .60$	.001*		.107	.051
Chem 177 Spring 2014 $M = .55$	.980	.107		<.001*
3Q GenChem Spring 2014 $M = .65$	.000*	.051	<.001*	

\* significant at .05 level

Since students demonstrated differences in performance on Test 1 and these differences appear to define two similar groups, the data were divided into two groups according to the results of the Tukey's HSD test for all subsequent analyses. Group I included Chem 167 and Chem 177 spring semester data and Group II included Chem 177 fall semester and 3Q GenChem data. Consequently, all subsequent statistical analyses were based on these groups, defined by their similarity in performance on the first component of the experiment in order to control for prior knowledge.

### Descriptive Statistics

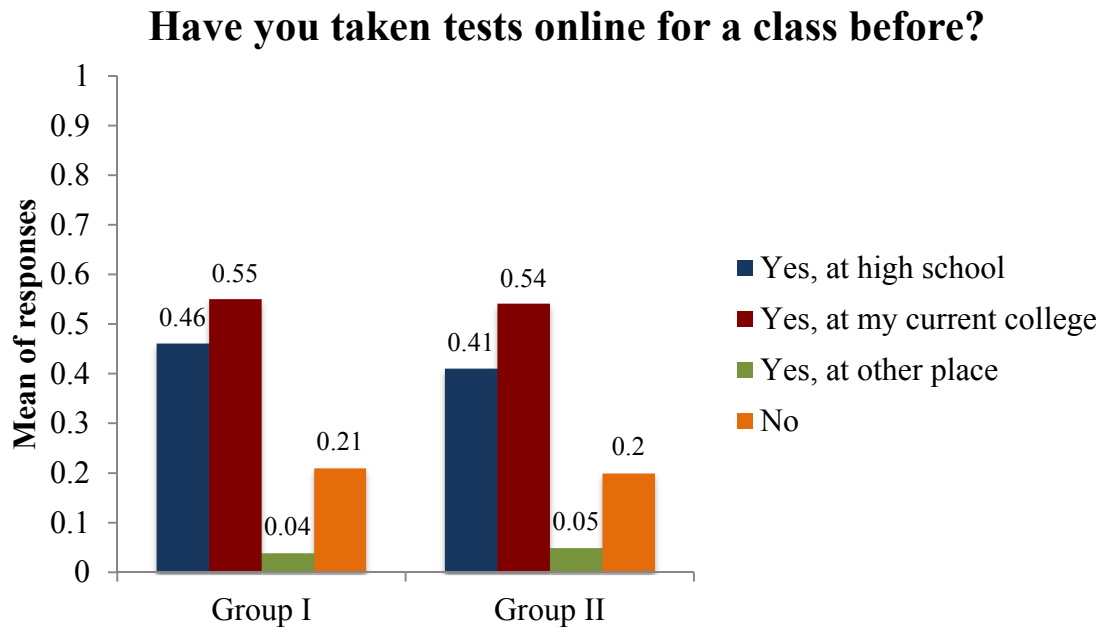
Overall, the average age, ethnicity, year in school, and percentage on the final exam were similar between Group I and Group II. Based on a visual inspection of the frequency distribution, the only factor that showed differences in demographics between the two groups was gender. While Group I was mostly characterized by male students, Group II had more female students (Table 10).

**Table 10.** Frequency and percentage of participants by gender for Group I and Group II ( $N = 664$ )

<b>Gender</b>	<b>Group I <math>N = 295</math></b>	<b>Group II <math>N = 369</math></b>
Male	212 (71.86%)	122 (33.06%)
Female	82 (27.80%)	242 (65.58%)
No response	1 (.34%)	5 (1.36%)

### Student testing preferences

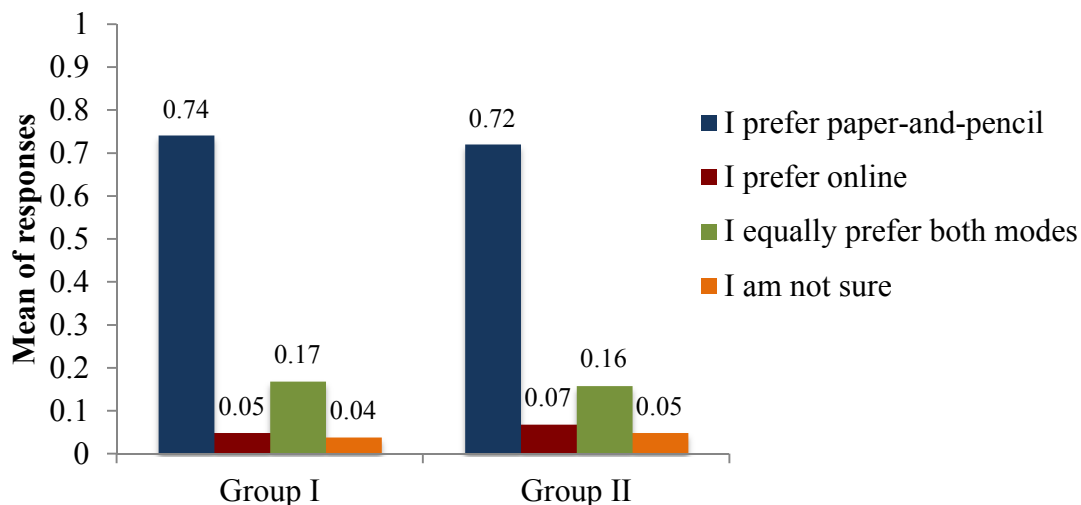
At the end of the session, all of the students were asked to indicate their testing preferences by answering several multiple-choice questions. Some of these questions were single answer questions and some of these questions were multiple answer questions. First, students were asked about their previous experiences using online tests



**Figure 4.** The means of student responses to a question that asked the students whether they had taken online tests for a class before. Students were allowed to select more than one answer. Among other places, students mentioned community college, work, previous college, placement exams, or did not provide an answer.

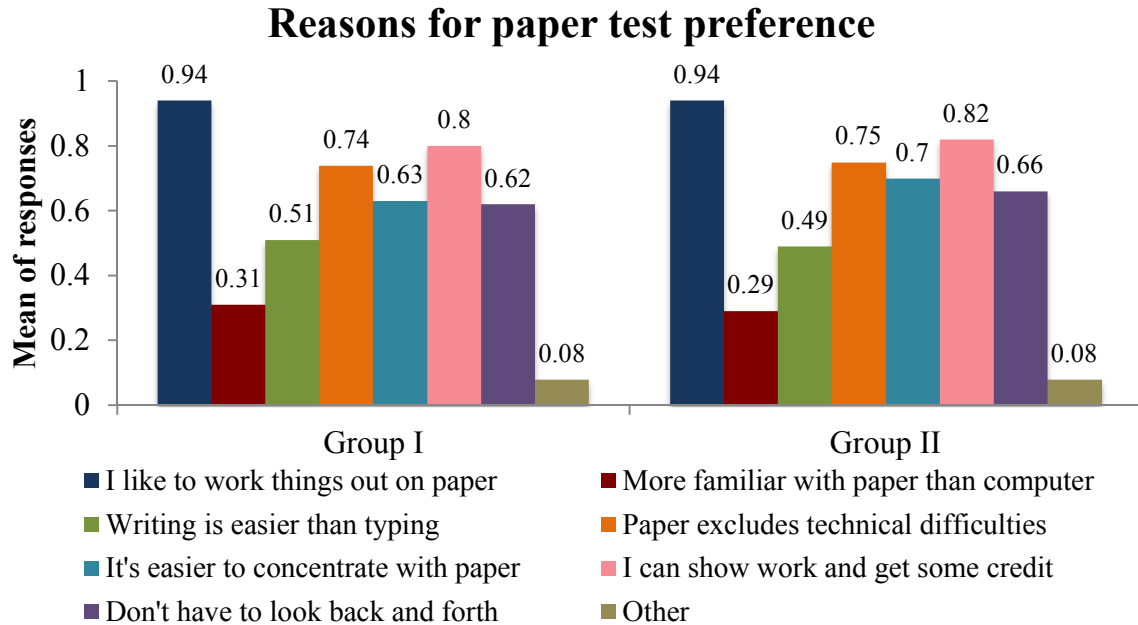


### If your next chemistry test were offered in both modes, which method would you prefer?



**Figure 5.** The means of student responses to a question that asked the students about their test mode preferences. Students could select only one answer.

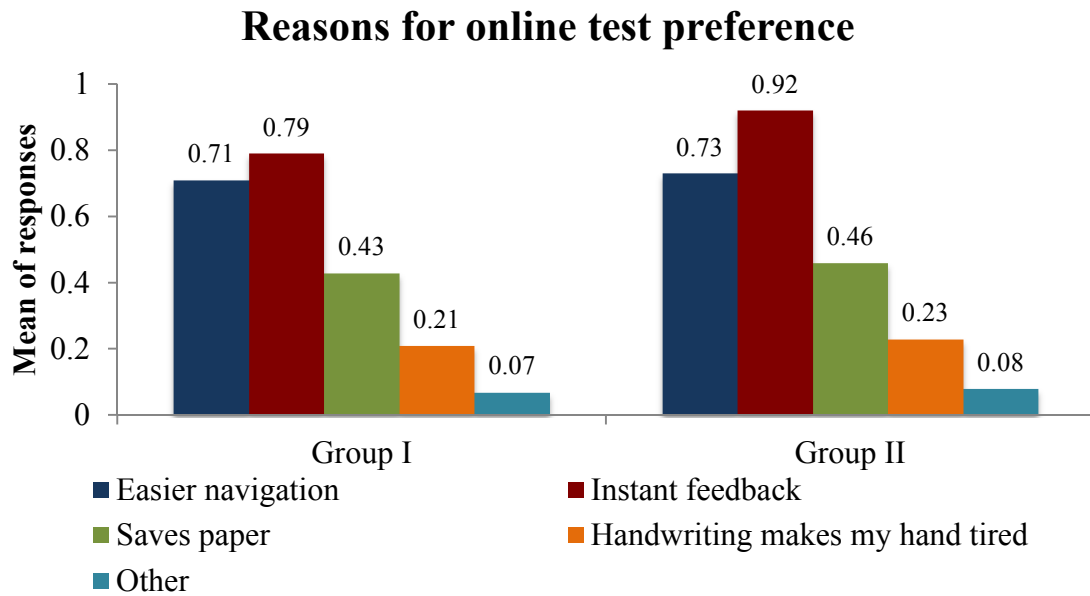
(see Figure 4). Next, students were asked to indicate their test mode preference (see Figure 5) and explain their choice (see Figures 6 – 7). If a student indicated a preference for paper or online modes, he or she then explained his or her choice by selecting one or more options from a list or provide his or her own answer. Figure 6 summarizes the choices of the students who preferred paper tests and Figure 7 summarizes the choices of the students who preferred online tests.



**Figure 6.** The means of student responses to a question that asked the students why they preferred paper to online modes. Students were allowed to select more than one answer. Among other reasons, students mentioned preference to write on the questions or cross them out, difficulty to write formulas online, inability to score online open-ended questions, easier navigation using paper, and eye pain from reading online material for long durations.

Four main themes have emerged from these data. First, most students were familiar with online testing. When asked whether the student had taken tests online for a class before, only 21% from Group I and 20% from Group II indicated that they had never taken a test online before (see Figure 4). Second, students in both groups expressed a very strong preference for paper mode (see Figure 5). Third, approximately all students who preferred the paper mode based their decision on the ability to write on the test (94% for Group I and 94% Group II) (see Figure 6). Fourth, students who indicated a preference for online tests often justified their answer by stating that online testing offers easier navigation (71% for Group I and 73% for Group II) and instant feedback (79% for

Group I and 92% for Group II) (see Figure 7). This pattern of testing preferences and student justifications of paper or computer tests was consistent with previous findings, which explored student test mode preferences in reading, writing, and mathematics tests (Steinberg et al., 2014). The consistency between two groups and consistency with prior published research helps provide validity for the data collected.



**Figure 7.** The means of student responses to a question that asked the students why they preferred online to paper test modes. Students were allowed to select more than one answer. Among other reasons, students mentioned that typing was faster than handwriting.

#### Research Question 1: Overall Test Mode

The first research question of this study was to explore the effect of Test 1 and Test 2 test modes on student performance, as measured by normalized gains. Normalized gains were calculated using the means of correctly answered 24 questions on Test 1 and Test 2. To answer this question, a one-way between-groups ANOVA was performed for Group I and Group II participants to compare the means of the normalized gains across

the four conditions. In other words, is there a difference in student gains for students who took general chemistry tests using different test modes or one mode? Because one student in Group II received 100% on Test 1, a normalized gain could not be calculated for this student. As a result, the data from this student was excluded from this analysis, thus reducing the sample size for Group II from 369 to 368. Participants were classified according to the mode of Test 1 and Test 2, which represented four conditions. Table 11 shows the frequency of participants for whom normalized gains were computed across the four conditions for Group I and Group II.

**Table 11.** Student distribution among four conditions for Group I and Group II for overall normalized gain analysis ( $N = 663$ )

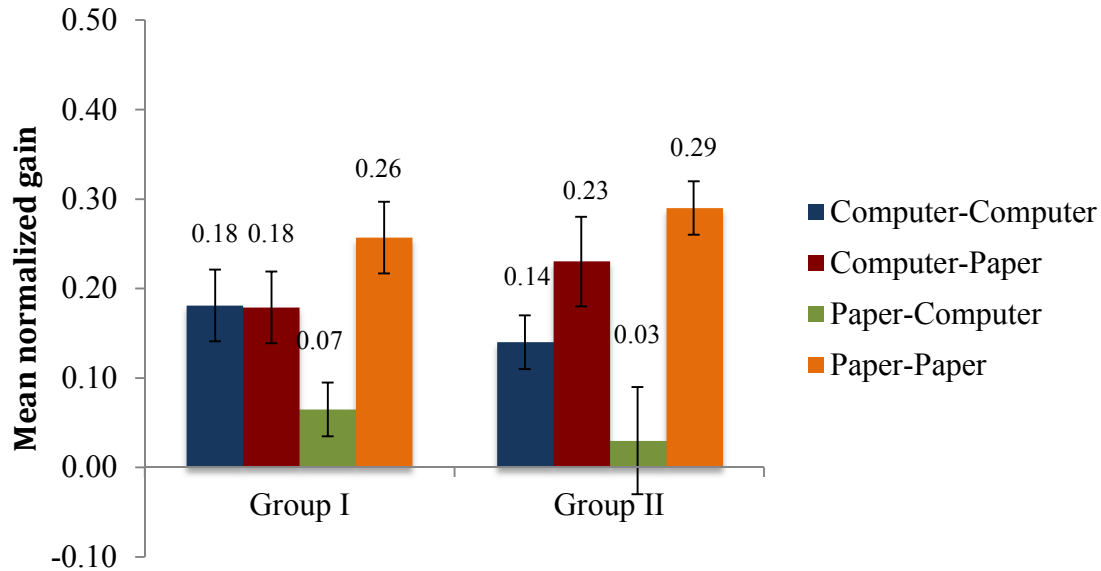
Test mode of Test 1 – Test 2	Group I $N = 295$	Group II $N = 368$
Com-Com	51	122
Com-PP	83	82
PP-Com	92	72
PP-PP	69	92

There was a statistically significant difference in both groups. For Group I, the results were  $F(3, 291) = 4.32, p = .0053, \eta^2 = .04$  and for Group II, the results were  $F(3, 364) = 6.49, p = .0003, \eta^2 = .05$ . Despite finding statistical significance, the actual difference in mean gains, calculated using  $\eta^2$ , among conditions was small (.04 for Group I and .05 for Group II). Post-hoc comparisons using the Tukey HSD test revealed one significant comparison for Group I and three significant comparisons for Group II. For Group I, the mean gain for the Paper-Paper condition ( $M = .26, SD = .31$ ) was significantly different from the Paper-Computer condition ( $M = .07, SD = .34$ ). For Group II, the same pairwise comparison showed significance between the Paper-Paper

condition ( $M = .29, SD = .29$ ) and the Paper-Computer condition ( $M = .03, SD = .47$ ). In addition, two more comparisons were found significant for Group II: (1) The Paper-Paper condition ( $M = .29, SD = .29$ ) was significantly different from the Computer-Computer condition ( $M = .14, SD = .37$ ) and (2) the Paper-Computer condition ( $M = .03, SD = .47$ ) was significantly different from the Computer-Paper condition ( $M = .23, SD = .46$ ).

Although students from all conditions showed overall positive gains, these results provide evidence that gains may vary due to the test mode. Furthermore, even though Paper-Paper showed the highest gains among all conditions in both groups, these gains were not statistically different from the Computer-Paper gains. This pattern of findings suggests that the highest gains were observed among the students who took Test 2 on paper regardless of what mode they took Test 1. Means of normalized gains for each condition are shown in Figure 8 and all pairwise comparisons for Group I and Group II are summarized in Table 12. As illustrated in Table 12, if a confidence interval did not include zero, the means of normalized gains were significantly different.

### Overall gains by test mode



**Figure 8.** Mean normalized gains by condition for Group I and Group II. Error bars represent standard error of the gain means.

**Table 12.** Summary of Tukey's HSD results for Group I and Group II pairwise comparisons of normalized gain means. The table reports the difference in means, standard error,  $p$ -value, and confidence intervals for each pair comparison. The  $p$ -value column indicates which pairs of means are significantly different at the 5% significance level.

Group I (N = 295)					
Pair comparisons A vs. B	Mean diff. A - B	Standard error	$p$ - value	[95% Conf. Interval]	
Com-PP vs. Com-Com	-.002	.06	1.000	-.160	.156
PP-Com vs. Com-Com	-.116	.06	.217	-.272	.039
PP-PP vs. Com-Com	.076	.06	.629	-.088	.241
PP-Com vs. Com-PP	-.114	.05	.129	-.249	.021
PP-PP vs. Com-PP	.079	.06	.504	-.067	.223
PP-PP vs. PP-Com	.192	.05	.003*	.051	.334
Group II (N = 368)					
Com-PP vs. Com-Com	.088	.06	.410	-.059	.234
PP-Com vs. Com-Com	-.106	.06	.276	-.259	.046
PP-PP vs. Com-Com	.152	.05	.030*	.010	.294
PP-Com vs. Com-PP	-.194	.06	.014*	-.360	-.028
PP-PP vs. Com-PP	.064	.06	.713	-.092	.220
PP-PP vs. PP-Com	.258	.06	.000*	.097	.420

\* significant at 0.05

### Research Question 2: Gain by Question Type

Although the results reported in Figure 8 suggest that student gains are statistically different among the four conditions, it still remains unknown whether these differences vary by question type. Accordingly, additional analyses were conducted to explore whether student gains on different question types varied across conditions. Normalized gains for each question type were calculated using the means of correctly answered 17 algorithmic questions, 5 conceptual questions, and 2 definition questions on Test 1 and Test 2. To answer the second research question, a one-way ANOVA was conducted on the algorithmic, conceptual, and definition gains with condition (Computer-Computer, Computer-Paper, Paper-Computer, Paper-Paper) as the independent variable. Since several students answered all algorithmic, conceptual, or definition questions correctly on Test 1, the normalized algorithmic, conceptual, and definition gains could not be calculated for these students. It is important to note that because gains are calculated based on means of two tests and it is easier to obtain a perfect score with fewer questions, it is possible to detect smaller sample sizes with fewer number of items on a test. Since the overall gain was calculated using 24 questions, algorithmic gain – 17 questions, conceptual gain – 5 questions, and definition gain – 2 questions, the sample size for each gain varied. While the overall gain was missing data only from one student in Group II, the sample size of the definition gain, which was based on only two questions, contained the most missing data. This change in sample is a potential source of bias, especially when the number of missing values is large. The importance of missing data will be discussed in more detail in the next chapter. Table 13 provides a summary of

the original sample sizes and the sample sizes used for algorithmic, conceptual, and definition gain analyses for each group.

**Table 13.** Sample size for Group I and Group II used for different analyses

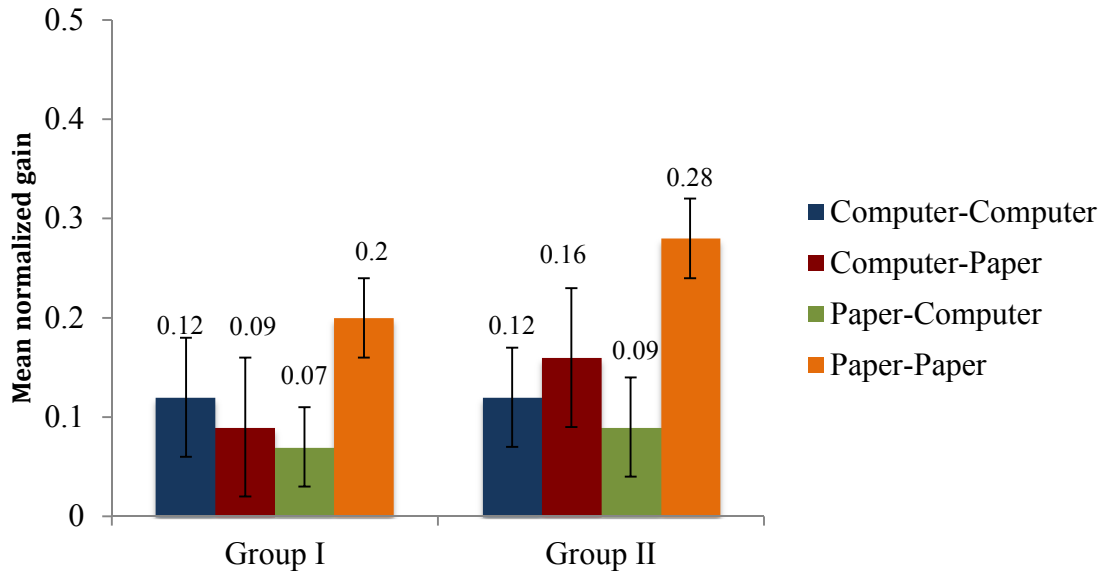
Group I Sample Size					
Test mode of Test 1 – Test 2	Original sample <i>N</i> = 295	Overall gain <i>N</i> = 295	Algorithmic gain <i>N</i> = 294	Conceptual gain <i>N</i> = 288	Definition gain <i>N</i> = 204
Number of questions		24	17	5	2
Com-Com	51	51	51	49	38
Com-PP	83	83	83	81	57
PP-Com	92	92	91	91	60
PP-PP	69	69	69	67	49
Group II Sample Size					
Test mode of Test 1 – Test 2	Original sample <i>N</i> = 369	Overall gain <i>N</i> = 368	Algorithmic gain <i>N</i> = 367	Conceptual gain <i>N</i> = 342	Definition gain <i>N</i> = 229
Com-Com	122	122	122	113	78
Com-PP	82	82	82	73	50
PP-Com	73	72	71	69	42
PP-PP	92	92	92	87	59

### Algorithmic gain

For Group I there was no statistically significant difference among gains for the algorithmic questions,  $F(3, 290) = 1.27, p = .287$ . The results of a one-way ANOVA performed on Group II data yielded a similar pattern,  $F(3, 363) = 2.66, p = .05$ . These results suggest that in the case of introductory university-level general chemistry, algorithmic questions are fairly unaffected by the test mode. The mean normalized algorithmic gains are presented in Figure 9.



### Algorithmic gain by test mode

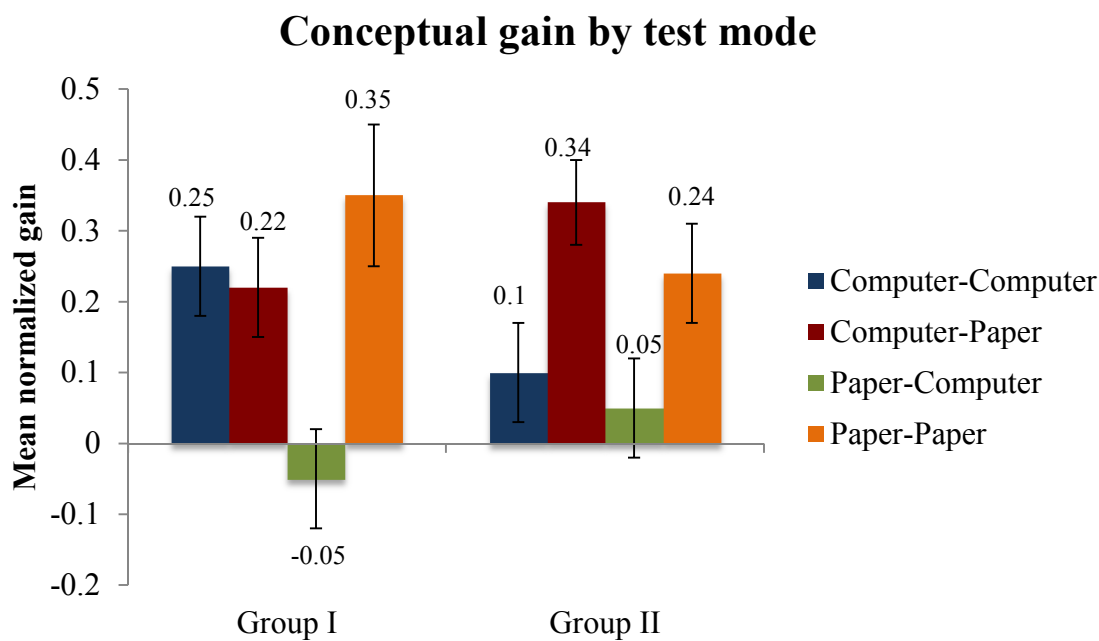


**Figure 9.** Mean normalized algorithmic gains by condition for Group I and Group II. Error bars represent standard error of the gain means.

### Conceptual gain

A one-way ANOVA analysis revealed a significant difference among the conceptual gains in Group I,  $F(3, 284) = 5.07, p = .0019, \eta^2 = .05$  and in Group II,  $F(3, 338) = 3.15, p = .025, \eta^2 = .03$ . The effect size for these tests was small ( $\eta^2 = .05$  for Group I and  $\eta^2 = .03$  for Group II). For Group I, Tukey HSD test showed that the mean normalized conceptual gains for Paper-Computer ( $M = -.05, SD = .71$ ) was significantly lower than the mean normalized conceptual gains for Paper-Paper ( $M = .35, SD = .81$ ). The results for Group II indicated that the mean normalized conceptual gains for Paper-Computer ( $M = .05, SD = .62$ ) was significantly lower than the mean normalized conceptual gains for Computer-Paper ( $M = .34, SD = .53$ ). The analyses of both groups showed that student gains for conceptual questions were the lowest in the Paper-Computer condition. In Group I, the gains were even negative, suggesting that students,

on average, answered fewer conceptual questions correctly on Test 2 than on Test 1. For this category of item, the routinely observed testing effect is somehow mitigated by the change in the mode from paper to computer. It is not clear what makes a conceptual item more difficult in a computer mode, but this is an area worth of future exploration. The difference among mean normalized conceptual gains for Computer-Computer, Computer-Paper, and Paper-Paper were not statistically significant, suggesting that conceptual questions in general chemistry could be delivered in any of these formats. Mean normalized conceptual gains as a function of condition and group are displayed in Figure 10 and all Tukey's HSD pairwise comparisons are shown in Table 14.



**Figure 10.** Mean normalized conceptual gains by condition for Group I and Group II. Error bars represent standard error of the gain means.

**Table 14.** Summary of Tukey's HSD results for Group I and Group II pairwise comparisons of normalized conceptual gain means. The table reports the difference in means, standard error,  $p$ -value, and confidence intervals for each pair comparison. The  $p$  - value column indicates which pairs of means are significantly different at the 5% significance level.

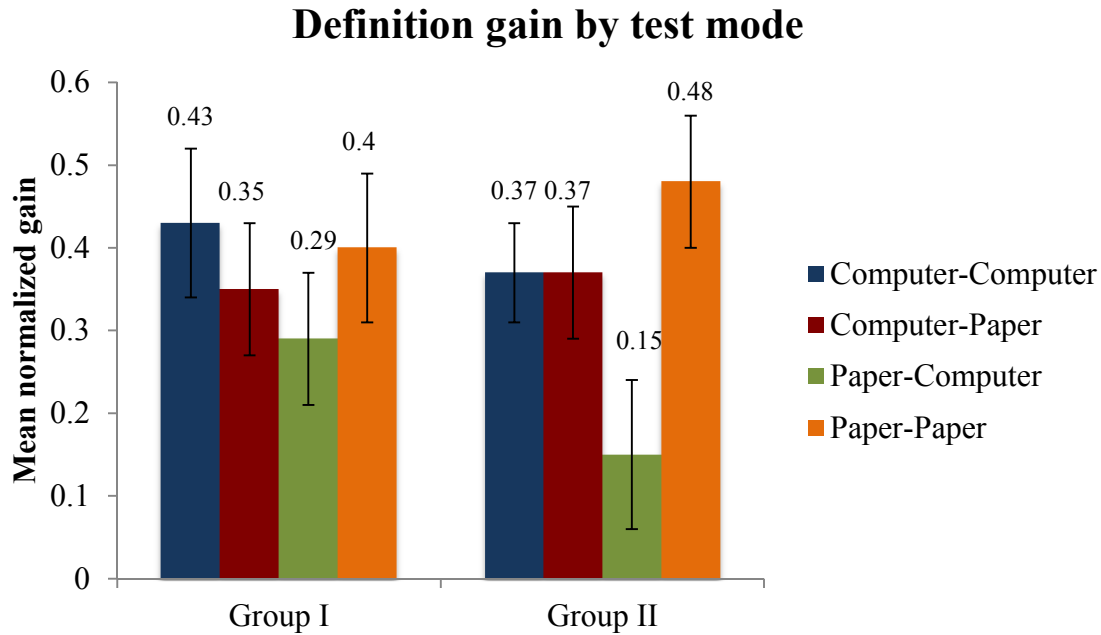
<b>Group I (N = 288)</b>					
Pair comparisons A vs. B	Mean diff. A - B	Standard error	$p$ - value	[95% Conf. Interval]	
Com-PP vs. Com-Com	-.035	.125	.992	-.357	.286
PP-Com vs. Com-Com	-.302	.122	.065	-.618	.013
PP-PP vs. Com-Com	.103	.129	.858	-.232	.437
PP-Com vs. Com-PP	-.267	.105	.056	-.539	.005
PP-PP vs. Com-PP	.138	.114	.618	-.156	.432
PP-PP vs. PP-Com	.405	.111	.002*	.119	.691
<b>Group II (N = 342)</b>					
Com-PP vs. Com-Com	.232	.096	.076	-.016	.480
PP-Com vs. Com-Com	-.050	.098	.96	-.302	.202
PP-PP vs. Com-Com	.139	.091	.424	-.096	.374
PP-Com vs. Com-PP	-.282	.107	.045*	-.559	-.004
PP-PP vs. Com-PP	-.093	.101	.797	-.355	.169
PP-PP vs. PP-Com	.189	.103	.260	-.077	.455

\* significant at 0.05

### Definition gain

A one-way ANOVA revealed mixed results about items that related to definitions. For condition, there were no significant effect found in Group I, but there was a significant effect shown in Group II,  $F(3, 225) = 2.67$ ,  $p = .0486$ , with a small effect size ( $\eta^2 = .03$ ). Post-hoc comparisons using the Tukey HSD test indicated that the mean normalized definition gain for the Paper-Paper condition ( $M = .48$ ,  $SD = .61$ ) was statistically higher than the mean normalized definition gain for the Paper-Computer condition ( $M = .15$ ,  $SD = .59$ ) (see Table 15). The comparisons of mean normalized definition gain data indicates that Paper-Computer leads to positive yet, less superior gains than the Computer-Computer, Computer-Paper, and Paper-Paper conditions when

compared to other test mode conditions (see Figure 11). Although this pattern was found in both groups, Paper-Computer was significantly lower than Paper-Paper only in Group II.



**Figure 11.** Mean normalized definition gains by condition for Group I and Group II. Error bars represent standard error of the gain means.

**Table 15.** Summary of Tukey's HSD results for Group II pairwise comparisons of normalized definition gain means. The table reports the difference in means, standard error, *p*-value, and confidence intervals for each pair comparison. The *p* – value column indicates which pairs of means are significantly different at the 5% significance level.

<b>Group II (N = 229)</b>					
Pair comparisons A vs. B	Mean diff. A - B	Standard error	<i>p</i> - value	[95% Conf. Interval]	
Com-PP vs. Com-Com	-.002	.105	1.000	-.273	.270
PP-Com vs. Com-Com	-.217	.111	.207	-.504	.070
PP-PP vs. Com-Com	.111	.100	.681	-.147	.370
PP-Com vs. Com-PP	-.215	.121	.287	-.529	.098
PP-PP vs. Com-PP	.113	.111	.740	-.175	.401
PP-PP vs. PP-Com	.328	.117	.027*	.026	.631

\* significant at 0.05

### Repeated Versus Non-repeated Gains

To explore the benefits of testing, all students repeated half of Test 1. While some students repeated items from chapters 1, 3, and 5, other students repeated items from chapter 2, 4, and 6. This design allowed us to compute and compare normalized gains for repeated chapters and non-repeated chapters for each student. More specifically, the purpose was to examine whether the additional practice would result in higher gains for repeated than non-repeated chapters. Although the mean normalized gain of repeated items was higher than the mean normalized gain of non-repeated items, a paired t-test revealed that the means of both groups were comparable. For Group I there was no statistically significant difference between repeated chapters ( $M = .13$ ,  $SD = .73$ ) and non-repeated chapters ( $M = .11$ ,  $SD = .47$ ),  $t(290) = .36$ ,  $p = .72$ . Also, there was no statistically significant difference between Group II repeated chapters ( $M = .16$ ,  $SD = .54$ ) and non-repeated chapters ( $M = .10$ ,  $SD = .57$ ),  $t(354) = 1.64$ ,  $p = .10$ .

Additional paired t-tests were conducted to determine whether there was a benefit of additional test when repeating items from chapters 1, 3, and 5 or from chapters 2, 4, and 6. A paired t-test was conducted to compare the mean normalized gain for repeated items from chapters 1, 3, and 5 and non-repeated items from chapters 2, 4, and 6. Consistent with the testing effect literature, the results for Group I showed there was a statistically significant difference between the mean normalized gains for repeated items from chapters 1, 3, and 5 ( $M = .34$ ,  $SD = .40$ ) and the mean normalized gains for non-repeated items from chapters 2, 4, and 6 ( $M = .002$ ,  $SD = .45$ ),  $t(133) = 6.75$ ,  $p < .001$ , with a large effect size ( $\eta^2 = .26$ ). Similar results were found for Group II where the mean normalized gain for repeated items from chapters 1, 3, and 5 ( $M = .32$ ,  $SD = .47$ ) was

higher than the mean normalized gain for non-repeated items from chapters 2, 4, and 6 ( $M = .01$ ,  $SD = .67$ ),  $t(175) = 5.26$ ,  $p < .001$ , with a large effect size ( $\eta^2 = .14$ ).

Lastly, a paired t-test was conducted to compare the mean normalized gain for repeated items from chapters 2, 4, and 6 and non-repeated items from chapters 1, 3, and 5. A paired t-test for Group I revealed an effect of additional testing between the repeated items from chapters 2, 4, and 6 ( $M = -.05$ ,  $SD = .89$ ) and non-repeated items from chapters 1, 3, and 5 ( $M = .20$ ,  $SD = .47$ ),  $t(156) = -3.57$ ,  $p = .0005$ ,  $\eta^2 = .08$ . Group II data also showed a statistical significance between the items from chapters 2, 4, and 6 ( $M = .009$ ,  $SD = .56$ ) and non-repeating items from chapters 1, 3, and 5 ( $M = .18$ ,  $SD = .44$ ),  $t(178) = -3.49$ ,  $p = .0006$ ,  $\eta^2 = .06$ . The effect size of these tests was medium ( $\eta^2 = .08$  for Group I and  $\eta^2 = .06$  for Group II). These results failed to support the proposition that additional testing inherently enhances the testing effect. It seems that the content itself plays an important role in determining whether or not repeated testing enhances the testing effect.

### Summary

This chapter presented the results of several analyses in order to determine the effect of test mode and test mode order on student performance using general chemistry practice tests. First, group equivalency was established by analyzing the proportions of correct answers from Test 1. According to the results, all data were divided into two groups. Group I included students from Chem 167 and Chem 177 spring semester and Group II included students from Chem 177 fall semester and 3Q GenChem.

Student test mode preferences showed the four main themes. Although most students were familiar with online testing, over 70% of students in both groups preferred to take their next general chemistry test on paper. This choice was largely supported by the ability to write on the test. On the other hand, students who preferred online to paper test, justified their mode choice with easier navigation and instant feedback.

To answer the first research question, a one-way ANOVA was performed for each group. Both groups revealed a statistically significant difference. Across both groups, the Paper-Paper condition led to a higher mean normalized gain than the Paper-Computer condition. In addition, the findings from Group II showed that the mean normalized gain from the Paper-Paper condition was higher than the mean normalized gain from the Computer-Computer condition and the mean normalized gain from the Computer-Paper condition was higher than the mean normalized gain from the Paper-Computer condition. These results show that although all conditions led to positive gains, these gains were not the same. The highest gains were observed in the Paper-Paper condition and the lowest gains were observed in the Paper-Computer condition.

For the second research question, additional one-way ANOVA analyses were performed to explore whether student gains of algorithmic, conceptual, and definition questions varied across conditions. In both groups, no statistically significant difference was found for the gain of algorithmic items. That is, student improvement from Test 1 to Test 2 on algorithmic items was not affected by the test mode and the test mode order. Analysis of the normalized conceptual gains revealed that the gains for the Paper-Computer were once again the lowest among conditions. In Group 1, the gains were even negative, which indicated that students performed better on conceptual questions on Test

1 than they did on Test 2. Significantly lower means of Paper-Computer condition were observed between Paper-Computer and Paper-Paper condition in Group I and between Paper-Computer and Computer-Paper condition in Group II. In both groups, no significant differences were found among mean normalized conceptual gains for Computer-Computer, Computer-Paper, and Paper-Paper conditions, suggesting that conceptual general chemistry questions may be delivered to students using one of these formats. Mixed results were found for the definition gains. While no significant effect for condition was found in Group I, the mean normalized definition gain of the Paper-Computer condition was significantly lower than the mean normalized definition gain of the Paper-Paper condition in Group II.

Lastly the testing effect was explored by comparing the normalized gains of repeated and non-repeated chapters. Although student mean normalized gain of repeated chapters was higher than gains of non-repeated chapters, no statistically significant difference was found between two groups. However, the evidence for the testing effect was found when students repeated items from chapters 1, 3, and 5. In both groups, the normalized gains of repeated chapters 1, 3, and 5 were higher than the normalized gains of non-repeated chapters 2, 4, and 6. Once again, statistical significance between repeated and non-repeated chapters was detected when students repeated items from chapters 2, 4, and 6. The results of both groups indicated that student normalized gains were higher for the non-repeated items from chapters 1, 3, and 5 than gains of the repeated items from chapters 2, 4, and 6. Possible inconsistencies between these findings and results of previous studies along with implications and limitations of this study's results will be discussed in Chapter 5.



## CHAPTER 5. CONCLUSIONS

The purpose of this study was to investigate the effect of the test mode and test mode order in a formal chemistry setting. More specifically, this study examined whether the paper-and-pencil and computer modes equally affected student performance on general chemistry practice tests. For this reason, students from general chemistry courses took two practice tests in one of the four conditions that varied in test mode and the order of mode. The gains on the tests were normalized and used as measures to compare student performance among four conditions.

Additionally, the testing effect was examined with the material that students learned in their general chemistry course before participating in this study. This investigation was performed to determine if additional testing promoted better retention of general chemistry information. For this analysis, the gains on questions from chapters that were repeated once and twice were compared. This chapter summarizes this study's empirical results, discusses its practical contributions and limitations, and provides directions for future research in this field.

### Empirical Contributions

The first goal of this study was to determine whether the test mode and the test mode order influenced student performance. Based on the performance of the initial test, students were divided into two groups: Group I included students from courses required for non-chemistry majors and Group II included students from courses required for chemistry majors. Overall, students from Group II outperformed students from Group I. Furthermore, the results of this study demonstrated that students in all conditions

generally experienced positive, yet different gains. A negative gain was found only in the Paper-Computer condition, that is, when students first took the practice test on paper and then switched to a computer, for the conceptual questions in Group I (see Table 16).

While the detected differences among gains were often found to be significant, the sizes of significant differences were small, indicating that 3% - 5% of variance in gains was accounted for by the conditions. Finding small effect sizes was not surprising; the literature on the testing effect suggests that in order for testing benefits to occur, more time must pass between the tests. The research in the domain of learning suggests that spacing the tests, known as the spacing effect, helps students in retaining information better in the long term than information that was studied repeatedly with no time intervals between sessions (Carpenter & DeLosh, 2005; Kornell & Bjork, 2008). Given the power of spacing, larger effects could have been detected if students took Test 1 and Test 2 on different days.

The present study revealed higher gains for conditions in which students took the last test on paper (i.e., Computer-Paper and Paper-Paper). Such a pattern of results is not surprising, because a vast majority of the students indicated a strong preference to take their next chemistry test using paper mode. Studying students' annotations of 150 university textbooks that are required for different disciplines, Marshall (1997) found that organic chemistry and calculus textbooks had more personalized annotations than many other textbooks. Marshall suggests that when students approach problems in context, they tend to solve them near the space that contains the question stem. It seems possible that a student's mode preference may influence their test performance. Student gains may be larger when the last test is delivered in the mode that matches student's preference. As

chemistry students' test mode preferences may be different from preferences of students from other disciplines such as dentistry (Escudier et al., 2011) and calculus (Engelbrecht & Harding, 2004), future studies should investigate this relationship in more detail within specific fields.

In all analyses, the lowest gains came from the Paper-Computer condition. In other words, it appears that Paper-Computer format presented the smallest testing benefits to students in general chemistry. These results also extend on Escudier's et al. (2011) work by adding conditions in which students take both tests on paper or computer. Several factors may explain this finding. First, studies that compared the effect of two test modes on the reading speed report that students generally read 15-30% faster on paper than on a computer (H. J. Kim & Kim, 2013). A potential alternative explanation may be in line with differences in interaction between two modes. As Jackel (2014) notes, the computer mode does not change the process of solving a problem, but it may change the manner in which student solves it. In a focus group by Escudier, Newton, Cox, Reynolds, and Odell (2011), students revealed that for each mode they adopted different approaches. In the paper-based testing, students preferred to choose what questions they wanted to answer first, whereas in the computer-based testing, students answered questions more or less in the order they were given. Therefore, experiencing a possible decrease in the reading speed and shift in test behavior when answering chemistry questions, may have produced the lowest gains.

Even though both Group I and Group II showed positive gains across conditions, some observations about the gain distribution between two groups were noted. First, the gains of Group II were more often higher than the gains of Group I. In particular, of the

four conditions, algorithmic gains of Group II were higher than the gains of Group I for three conditions and identical for Computer-Computer condition. Second, Group I had higher gains than Group II for the Computer-Computer condition and Group II had higher gains than Group I for the Computer-Paper and Paper-Paper conditions. As shown in Table 16, these patterns suggest that differences in gains may relate to course the student was in. It is possible that students in each course have experienced paper-based and computer-based testing to a different extent and so students' prior experience impacts their performance. For example, chemistry students from Group II often outperformed students from Group I when the final test was delivered on paper. This could indicate that Group II students might have more experience taking general chemistry tests on paper than online so when they were given the second test in a more familiar test mode, their gains exceeded the gains of Group I students, who might have experienced similar exposure to both modes. In other words, student performance on a final general chemistry test may vary based on the level of exposure to that test mode in his or her chemistry course. Since the use and frequency of online and paper testing in each course was not documented, it may be useful to consider this relationship in the future.

When significant differences were identified, the gains of the Paper-Computer were often found significantly lower than the gains of the Paper-Paper condition. For example, the results of the overall gains showed that when students took both tests on paper their gains were higher by .19 for Group I and .26 for Group II than the gains of students who switched from paper to computer. This finding suggests that if students take their first practice test on paper, it benefits their performance to take the next practice test on paper-and-pencil, too. Another interesting finding was found between Paper-Paper and

Computer-Paper conditions. Even though the Paper-Paper condition often showed superior gains, those gains were not statistically different from the gains of the Computer-Paper condition. This is an interesting finding; students may derive similar gains in performance regardless of mode of the first test as long as the last test is administered in the traditional mode.

**Table 16.** Means and standard deviations of mean gains by condition for Group I and Group II

		<b>Group 1 (N = 295)</b>		<b>Group 2 (N = 368)</b>	
<b>Test/Condition</b>	<b>M</b>	<b>SD</b>	<b>M</b>	<b>SD</b>	<b>SD</b>
<b>Overall gains (24 questions)</b>					
Com-Com	.18	.28	.14	.37	
Com-PP	.18	.41	.23	.46	
PP-Com	.07	.34	.03	.47	
PP-PP	.26	.31	.29	.29	
		<b>Group 1 (N = 294)</b>		<b>Group 2 (N = 367)</b>	
<b>Algorithmic gains (17 questions)</b>					
Com-Com	.12	.42	.12	.52	
Com-PP	.09	.63	.16	.65	
PP-Com	.07	.36	.09	.43	
PP-PP	.20	.35	.28	.39	
		<b>Group 1 (N = 288)</b>		<b>Group 2 (N = 342)</b>	
<b>Conceptual gains (5 questions)</b>					
Com-Com	.25	.46	.10	.72	
Com-PP	.22	.67	.34	.53	
PP-Com	-.05	.71	.05	.62	
PP-PP	.35	.81	.24	.63	
		<b>Group 1 (N = 204)</b>		<b>Group 2 (N = 229)</b>	
<b>Definition gains (2 questions)</b>					
Com-Com	.43	.56	.37	.54	
Com-PP	.35	.57	.37	.59	
PP-Com	.29	.63	.15	.59	
PP-PP	.40	.63	.48	.61	

The results of the ANOVA analyses for each question type showed mixed results.

While there was no effect from the test mode and test mode order for algorithmic

questions, these effects were detected in the Paper-Computer condition for conceptual and definition questions. The gains on conceptual questions were significantly lower in the Paper-Computer condition than in the Paper-Paper condition (Group I) and Computer-Paper condition (Group II). These results suggest that students, on average, may experience very low or even negative gains if they take general chemistry conceptual questions in paper mode first and then on computer as compared to Paper-Paper or Computer-Paper formats.

For the definition questions, the significance was found only in Group II. Once again, the gains of the Paper-Computer were statistically lower than the gains of the Paper-Paper condition. Students improved their performance from Test 1 to Test 2 on the definition questions across all conditions, but taking the first test on paper and the second test on computer led to the lowest gains. No significance was found from the gains on definition questions in Group I. This finding might be due to a small number of definition items (i.e., 2 questions) relative to 5 conceptual items and 17 algorithmic items.

The findings that the lowest gains are obtained in the Paper-Computer condition raise an important question. What causes the gains of this condition to be the lowest? Change in mode is one possible feature, however, in this study the Computer-Paper condition was not found to be significantly different from two conditions in which modes remained constant. Similar to the testing effect, it might be that a longer delay between repeated tests is needed to maximize the effect of matched versus non-matched mode. For example, Veltre, Cho, and Neely (2014) purposely used a two-day interval in their design to obtain evidence for the testing effect and test-based transfer appropriate processing

according to which retrieval of information is enhanced when conditions of encoding and retrieval stages match.

In sum, the results of the present study further support earlier studies reporting that test mode and changes in mode unequally affect student performance. While algorithmic questions are fairly unaffected by the test mode, conceptual and definition questions show significant differences in gains. These results both replicate and extend findings that chemistry students do not answer algorithmic and conceptual questions in the same manner. It is possible that differences in performance on algorithmic and conceptual items may have occurred as the result of previous exposure to both types of questions in class (Nakhleh, 1993).

The second goal of this study was to investigate whether the testing effect would extend beyond tests where participants learned the material shortly before the test. The practice tests were based on the material studied in a typical general chemistry semester-long course. Thus, tests used in this study required students to apply the knowledge they had accumulated for several weeks. Furthermore, rather than taking the same questions twice, items on Test 2 were different from items on Test 1. These factors in design allowed the investigation of the testing effect using settings more common in educational environment. The findings of this study illustrate that additional testing can benefit student performance; however, such benefits occurred for participants who were given additional testing opportunities on items from chapters 1, 3, and 5. One possible explanation why the testing effect was not present in all repeated chapters is the unequal distribution of general chemistry question type between chapters 1, 3, and 5 and 2, 4, and 6. Chapters 1, 3, and 5 included two definition questions, three conceptual questions, and

seven algorithmic questions, whereas chapters 2, 4, and 6 included two conceptual questions and ten algorithmic questions. To summarize, repeated items had a greater testing effect on student gains than non-repeated items, but only when students repeated questions from chapters 1, 3, and 5. This observation could be a consequence of the possibility that definition questions were present only in chapters 1, 3, and 5. Therefore, the testing effect could be more pronounced for the definition items.

### Practical Contributions

When used correctly, testing at the university level can be a highly effective tool for student learning. What are some useful teaching interventions when assessing students in general chemistry courses? The results of this study show promising effects of testing and, therefore, may encourage some chemistry instructors to use frequent testing before their course's final exam. The present study suggests that when students take two practice tests, performance improves. In the Glover's (1989) study (Experiment 3), for example, students recalled more information on the final test when they took at least one practice test before the final test than the group of students who had no practice test. However, on the final test students who received two practice tests on different days outperformed students who received two practice tests on the same day. Despite the fact that few students report using self-testing as a learning strategy (Bjork, Dunlosky, & Kornell, 2011; J. Karpicke et al., 2009) and students prefer re-studying material repeatedly (i.e., massed studying) over re-studying it after some intermission (i.e., spaced studying) (Bjork et al., 2011), instructors should promote self-testing and use repeated yet spaced tests regularly in their courses. This way, students can reap the benefits of the



testing effect technique such as improve scores on class exams (McDaniel et al., 2007), practice previously learned material so it could be more successfully retrieved at a later time relative to additional studying (Butler & Roediger, 2007), and more effectively learn new material whether it is from the same knowledge domain (Butler, 2010, Experiments 1b and 2) or different knowledge domain (Butler, 2010, Experiment 3). In addition, frequent testing forces students to rehearse material more often rather than engage in a less effective, massed studying before a cumulative exam and correct errors in information with feedback (Butler & Roediger, 2008). The need for feedback is especially important for tests in which students are exposed to incorrect information, as occurs inherently in multiple-choice questions where distractors must be incorrect. Since students enhance their learning through tests, taking a test like a multiple-choice test may lead to incorrect learning if correct answers are not provided after the test (Roediger & Butler, 2011).

Since the Computer-Paper condition was not found to be different from the Paper-Paper condition, the format of online mode for practice tests or quizzes and paper mode for the final test shows potential to be an effective strategy for general chemistry tests. Importantly, this mode combination would allow instructors and students to use the benefits of fast feedback, provide students the opportunity to write on tests, and deliver the final test in the mode that is preferred by most students without having a negative impact on student performance. In addition, feedback from initial online tests could help instructors in adjusting their teaching material and assignments so students could enhance their learning more before their next test.

### Limitations

While this study provides support for the benefit of testing in learning general chemistry, it is important to address its limitations. The first limitation is participant self-selection. Even though participants received compensation and were encouraged by their professors to take practice tests before the final examination, only a subset of class engaged in this study. This is problematic: the self-selected students may be different from those students who chose to not participate. As this study involved two practice tests, it is possible that students who were more motivated to receive a higher grade on the final exam chose to participate.

The second limitation is absence of a control group in which students would only read the test items and their answers. In testing effect studies the control group is integrated in order to establish the baseline for comparisons. Studies that compared the testing groups to a control group that spent time only re-studying the material, demonstrate the benefit of frequent testing (Butler, 2010; J. Karpicke & Roediger, 2007; McDaniel et al., 2007; Roediger et al., 2011). However, since the testing effect has been established and replicated by many researchers and to motivate more students to participate in the study by taking practice tests, the control group was omitted in the design of this study.

The third limitation is missing data that derived from using normalized gains and a small number of items. Even though a total of 664 students participated in this study, it was not possible to calculate the gains for all students, particularly when the analysis focused on subsets of items. According to the Hake's formula, students who answered all questions correctly on the initial test have reached the ceiling effect and, therefore, could

not improve their score. Another factor that led to missing data was the number of questions for each question type. Out of 24 questions, 17 questions were algorithmic, 5 were conceptual, and only 2 questions were definitions. As the number of items gets smaller, the higher is likelihood for student to answer all questions correctly within that category. Hence, data from 7 and 27 students from Group I and Group II, respectively, were omitted when analyzing conceptual gains and data from 91 and 140 students from Group I and Group II, respectively, were omitted when analyzing definition gains. Decrease in the sample size may result in lower statistical power, and thus calls for a careful interpretation of the results, particularly for the gains observed on definition items.

#### Future Directions

Building on the previous testing effect literature and the results of this study, future research can consider lengthening the time period between the practice tests. Such a design can shed more light into how the testing effect affects student performance when tests are delivered using one or two modes and taken on different days. Giving students a longer break between tests would allow them to use feedback from the initial test as a guide to what material they need to re-study before taking the next test.

Furthermore, while all chapters were repeated, the number of algorithmic, conceptual, and definition questions were not the same between chapters 1, 3, and 5 and chapters 2, 4, and 6. Since the testing effect was detected only with chapters 1, 3, and 5 future studies should equally distribute all questions types between the groups.

Most of the students expressed a preference for the paper mode. In this study, all scratch paper was collected after each test, yet the length and content of individual scratch notes were not analyzed. Although students who received tests on paper were allowed to use scratch paper, they still chose to show their work on the paper test. Future studies should examine a possible relationship between scratch paper and test mode by comparing the quality and quantity of student annotations who take a chemistry test on a paper or computer.

Last, whether these findings would hold for tests and exams of other chemistry courses or even disciplines is an interesting avenue for future research. The relationship between different test modes and other question types (e.g., essays) is an area that could offer more insights about the benefits of online and paper test mode, and when it may be appropriate to use them when assessing students' knowledge.

### Conclusion

Research in cognitive psychology has shown that testing is an effective tool to enhance student learning. Improvements in educational technologies permit educators to design and administer their tests online. Although computer-based testing offers a set of unique benefits, a major concern of many instructors when considering adapting online delivery mode is its effects on student learning. The small effect size and sample size of this study make it difficult to make any conclusive statements. Nonetheless, the results of this study add to the testing effect and test mode literature and can be used to make several recommendations about how to be aware of the role of test mode in enhancing learning in general chemistry courses. First, as the testing effect is apparent, chemistry

instructors should promote the use of testing in their classroom by testing students frequently and repeatedly. Since all conditions have generally illustrated the positive gains, even a small increase in performance might have a valuable effect in the long-term learning of general chemistry in college. Second, testing students first on paper and then on computer may have a small positive effect on student test performance. Given that each test mode offers its unique advantages to instructors and students, this combination permits to derive different benefits and address student test mode preferences. Finally, if paper-and-pencil testing is used throughout the course, the final test should be delivered on paper, as well. While testing is a powerful tool to improve retention, it is not yet clear whether these results have implications for instructors of other disciplines. Replicating this study using different material and with more questions within each question category is the next step needed.

## REFERENCES

- Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs: General and Applied*, 11(1), 159–177. Retrieved from <http://www.psycontent.com/content/e586546762237146/>
- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22(7), 861–876. <http://doi.org/10.1002/acp.1391>
- Alexander, M. W., Bartlett, J. E., Truell, A. D., & Ouwenga, K. (2000). Testing in a Computer Technology Course: An Investigation of Equivalency in Performance between Online and Paper and Pencil Methods. *Journal of Career and Technical Education*, 18(1), 69–80. Retrieved from <http://eric.ed.gov/?id=EJ636661>
- Allen, I. E., Survey, B., & Seaman, J. (2014). Grade change: Tracking online education in the United States, 45. Retrieved from <http://www.onlinelearningsurvey.com/reports/gradechange.pdf>
- Baddeley, A. (2009). What is memory? In A. Baddeley, W. Eysenck, Michael, & C. Anderson, Michael (Eds.), *Memory* (pp. 1–17). New York: Psychology Press. Retrieved from <https://owl.english.purdue.edu/owl/resource/747/06/>
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of Frequent Classroom Testing. *The Journal of Educational Research*, 85(2), 89–99. <http://doi.org/10.1080/00220671.1991.10702818>
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my writing test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning, and Assessment*, 6(9).
- Bjork, R. a., Dunlosky, J., & Kornell, N. (2011). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, 64(1). <http://doi.org/10.1146/annurev-psych-113011-143823>
- Brallier, S. A., Schwanz, K. A., Palm, L. J., & Irwin, L. N. (2015). Online Testing: Comparison of Online and Classroom Exams in an Upper-Level Psychology Course. *American Journal of Educational Research*, 3(2), 255–258. <http://doi.org/10.12691/education-3-2-20>
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1988). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 367–407). New York: Macmillan. <http://doi.org/10.1002/j.2330-8516.1988.tb00291.x>

- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. Retrieved from <http://psycnet.apa.orgjournals/xlm/36/5/1118>
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514–527. <http://doi.org/10.1080/09541440701326097>
- Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. <http://doi.org/10.3758/MC.36.3.604>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563–1569. Retrieved from <http://psycnet.apa.orgjournals/xlm/35/6/1563>
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547–1552. Retrieved from <http://psycnet.apa.orgjournals/xlm/37/6/1547>
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*(5), 619–636. <http://doi.org/10.1002/acp.1101>
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*(6), 760–771. <http://doi.org/10.1002/acp.1507>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*(2), 438–448. <http://doi.org/10.3758/MC.36.2.438>
- Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills*, *96*, 54–56. <http://doi.org/10.2466/pms.2003.96.1.54>
- Chan, J. C. K., McDermott, K. B., & Roediger III, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553–571. Retrieved from <http://psycnet.apa.orgjournals/xge/135/4/553>
- Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a Witnessed Event Increases Eyewitness Suggestibility. *Psychological Science*, *20*(1), 66–73.

- Clark, D. A., & Svinicki, M. (2014). The effect of retrieval on post-task enjoyment of studying. *Educational Psychology Review*, 27(1), 1–17. <http://doi.org/10.1007/s10648-014-9272-4>
- Cohen, J. (1988). The analysis of variance. In *Statistical power analysis for the behavioral sciences* (2nd editio, pp. 273–406). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Coppens, L. C., Verhoeijen, P. P. J. L., & Rikers, R. M. J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology*, 23(3), 351–357. <http://doi.org/10.1080/20445911.2011.507188>
- DeBoer, G. E., Quellmalz, E. S., Davenport, J. L., Timms, M. J., Herrmann-Abell, C. F., Buckley, B. C., Jordan, K. A., Huang, C., & Flanagan, J. C. (2014). Comparing three online testing modalities: Using static, active, and interactive online testing modalities to assess middle school students' understanding of fundamental ideas and use of inquiry skills related to ecosystems. *Journal of Research in Science Teaching*, 51(4), 523–554. <http://doi.org/10.1002/tea.21145>
- Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, 5(2), 1–17. <http://doi.org/10.1103/PhysRevSTPER.5.020103>
- Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The Testing Effect for Learning Principles and Procedures from Texts. *The Journal of Educational Research*, 107(5), 357–364. <http://doi.org/10.1080/00220671.2013.823370>
- Drasgow, F. & O.-B. (Ed.). (1999). *Innovations in Computerized Assessment*. Psychology Press. Retrieved from <https://books.google.com/books?hl=en&lr=&id=K415AgAAQBAJ&pgis=1>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <http://doi.org/10.1177/1529100612453266>
- Einstein, G. O., Mullet, H. G., & Harrison, T. L. (2012). The Testing Effect: Illustrating a Fundamental Concept and Changing Study Strategies. *Teaching of Psychology*, 39(3), 190–193. <http://doi.org/10.1177/0098628312450432>
- Emerson, L., & MacKay, B. (2011). A comparison between paper-based and online learning in higher education. *British Journal of Educational Technology*, 42(5), 727–735. <http://doi.org/10.1111/j.1467-8535.2010.01081.x>



- Engelbrecht, J., & Harding, A. (2004). Combing online and paper assessment in a web-based course in undergraduate mathematics. *Journal of Computers in Mathematics and ...*, 23, 217–231. Retrieved from <http://www.editlib.org/p/11487>
- Escudier, M. P., Newton, T. J., Cox, M. J., Reynolds, P. A., & Odell, E. W. (2011). University students' attainment and perceptions of computer delivered assessment; a comparison between computer-based and traditional tests in a "high-stakes" examination. *Journal of Computer Assisted Learning*, 27(5), 440–447. <http://doi.org/10.1111/j.1365-2729.2011.00409.x>
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538. <http://doi.org/10.1037/a0015808>
- Finn, B., & Roediger, H. L. (2011). Enhancing retention through reconsolidation: negative emotional arousal following retrieval enhances later recall. *Psychological Science*, 22(6), 781–6. <http://doi.org/10.1177/0956797611407932>
- Gates, A. I. (1917). *Recitation as a Factor in Memorizing*. Science Press. Retrieved from <https://books.google.com/books?hl=en&lr=&id=YinQhb3ZS8YC&pgis=1>
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81(3), 392–399. Retrieved from <http://psycnet.apa.orgjournals/edu/81/3/392>
- Goodwin, C. J. (2010). *Research in Psychology. Methods and design* (6th ed.). Hoboken, NJ, USA: Wiley & Sons, Inc.
- Hake, R. (1998). Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74.
- Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 801–812. Retrieved from <http://psycnet.apa.orgjournals/xlm/37/4/801>
- Hochlehnert, A., Brass, K., Moeltner, A., & Juenger, J. (2011). Does Medical Students' Preference of Test Format (Computer-based vs. Paper-based) have an Influence on Performance? *BMC Medical Education*, 11(1), 89. <http://doi.org/10.1186/1472-6920-11-89>
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 562–567. [http://doi.org/10.1016/S0022-5371\(71\)80029-4](http://doi.org/10.1016/S0022-5371(71)80029-4)

- Howell, D. (2002). Repeated-measures designs. In *Statistical Methods for Psychology* (5th edition, p. 792). Thomson Learning. Retrieved from <https://books.google.com/books?hl=en&lr=&id=CRYKAAAAQBAJ&pgis=1>
- Hussar, W. J., & Bailey, T. M. (2014). *Projections of Education Statistics to 2022*. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014051>
- Jackel, B. (2014). Item Differential in Computer Based and Paper Based Versions of a High Stakes Tertiary Entrance Test : Diagrams and the Problem of Annotation. *Diagrammatic Representation and Inference*, 8578, 71–77.
- James, W. (1890). *The principles of psychology*. Retrieved from [http://pages.uoregon.edu/koopman/courses\\_readings/phil123-net/identity/james\\_principles\\_2vols.pdf](http://pages.uoregon.edu/koopman/courses_readings/phil123-net/identity/james_principles_2vols.pdf)
- Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101(3), 621–629. Retrieved from <http://psycnet.apa.org/journals/edu/101/3/621>
- Johnson, M., & Green, S. (2006). On-Line Mathematics Assessment: The Impact of Mode on Performance and Question Answering Strategies. *Journal of Technology, Learning, and Assessment*, 4(5). Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=EJ843854> <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ843854>
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4-5), 528–558. <http://doi.org/10.1080/09541440601056620>
- Kang, S. H. K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition*, 3(3), 183–188. <http://doi.org/10.1016/j.jarmac.2014.05.006>
- Karpicke, J., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory (Hove, England)*, 17(4), 471–479. <http://doi.org/10.1080/09658210802647009>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science (New York, N.Y.)*, 319(5865), 966–8. <http://doi.org/10.1126/science.1152408>
- Karpicke, J., & Roediger, H. (2007). Repeated retrieval during learning is the key to long-term retention☆. *Journal of Memory and Language*, 57(2), 151–162. <http://doi.org/10.1016/j.jml.2006.09.004>

- Keresztes, A., Kaiser, D., Kovács, G., & Racsmány, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cerebral Cortex (New York, N.Y. : 1991)*, 24(11), 3025–35. <http://doi.org/10.1093/cercor/bht158>
- Kim, D., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *The Journal of Technology, Learning, and Assessment*, 6(4), 1–31.
- Kim, D., & Huynh, H. (2008). Computer-Based and Paper-and-Pencil Administration Mode Effects on a Statewide End-of-Course English Test. *Educational and Psychological Measurement*, 68(4), 554–570. <http://doi.org/10.1177/0013164407310132>
- Kim, H. J., & Kim, J. (2013). Reading from an LCD monitor versus paper: Teenagers' reading performance. *International Journal of Research Studies in Educational Technology*, 2(1), 15–24. <http://doi.org/10.5861/ijrset.2012.170>
- Kingston, N. M. (2008). Comparability of Computer- and Paper-Administered Multiple-Choice Tests for K–12 Populations: A Synthesis. *Applied Measurement in Education*, 22(1), 22–37. <http://doi.org/10.1080/08957340802558326>
- Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: what is “enhanced” and how do we know? A critical literature review. *Learning, Media and Technology*, 39(1), 6–36. <http://doi.org/10.1080/17439884.2013.770404>
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19(6), 585–592. <http://doi.org/10.1111/j.1467-9280.2008.02127.x>
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85–97. <http://doi.org/10.1016/j.jml.2011.04.002>
- Kuo, C.-Y., & Wu, H.-K. (2013). Toward an integrated model for designing assessment systems: An analysis of the current status of computer-based assessments in science. *Computers & Education*, 68, 388–403. <http://doi.org/10.1016/j.compedu.2013.06.002>
- Leeming, F. C. (2002). The Exam-A-Day Procedure Improves Performance in Psychology Classes. *Teaching of Psychology*, 29(3), 210–212. [http://doi.org/10.1207/S15328023TOP2903\\_06](http://doi.org/10.1207/S15328023TOP2903_06)

- Lewis, S. E., & Lewis, J. E. (2005). The Same or Not the Same: Equivalence as an Issue in Educational Research. *Journal of Chemical Education*, 82(9), 1408.  
<http://doi.org/10.1021/ed082p1408>
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition*, 43(1), 85–98.  
<http://doi.org/10.3758/s13421-014-0453-7>
- Marshall, C. C. (1997). Annotation: From paper books to the digital library. In *Proceedings of the second ACM international conference on Digital libraries - DL '97* (pp. 131–140). <http://doi.org/10.1145/263690.263806>
- Mawhinney, V. T., Bostow, D. E., Laws, D. R., Blumenfeld, G. J., & Hopkins, B. L. (1971). A comparison of students studying-behavior produced by daily, weekly, and three-week testing schedules. *Journal of Applied Behavior Analysis*, 4(4), 257–64.  
 Retrieved from  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1310703&tool=pmcentrez&rendertype=abstract>
- Mazzeo, J., & Harvey, A. L. (1988). *The Equivalence of Scores from Automated and Conventional Educational and Psychological Tests: A Review of the Literature*. New York. Retrieved from  
<http://research.collegeboard.org/publications/content/2012/05/equivalence-scores-automated-and-conventional-educational-and>  
<http://research.collegeboard.org/sites/default/files/publications/2012/7/research-report-1988-8-equivalence-scores-automated-conven>
- McDaniel, M. A. (2007). Transfer: Rediscovering a central concept. In H. L. Roediger, Y. Dubai, & M. Fitzpatrick (Eds.), *Science of memory: Concepts*. New York: Oxford University Press. Retrieved from  
[http://www.psychology.sunysb.edu/srajaram-/PDFs for Publication/concepts.pdf](http://www.psychology.sunysb.edu/srajaram-/PDFs%20for%20Publication/concepts.pdf)
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4-5), 494–513. <http://doi.org/10.1080/09541440701326154>
- McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, 1(1), 18–26.  
<http://doi.org/10.1016/j.jarmac.2011.10.001>
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449–458.  
<http://doi.org/10.1037/0033-2909.114.3.449>

- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533. [http://doi.org/10.1016/S0022-5371\(77\)80016-9](http://doi.org/10.1016/S0022-5371(77)80016-9)
- Nakhleh, M. B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry. *Journal of Chemical Education*, 70(1), 52–55. <http://doi.org/10.1021/ed070p52>
- Nakhleh, M. B., & Mitchell, R. (1993). Concept learning versus problem solving: There is a difference. *Journal of Chemical Education*, 70(3), 190–192.
- Neuman, G., & Baydoun, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement*, 22(1), 71–83.
- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352–1375. <http://doi.org/10.1080/00140130802170387>
- Nurrenbern, S., & Pickering, M. (1987). Concept learning versus problem solving: Is there a difference? *Journal of Chemical Education*, 508–510. <http://doi.org/10.1021/ed064p508>
- Pallant, J. (2010). *SPSS: Survival manual* (4th edition). New York: McGraw Hill. Retrieved from <https://statistics.laerd.com/stata-tutorials/one-way-anova-using-stata.php>
- Pastötter, B., Weber, J., & Bäuml, K.-H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, 27(2), 280–285. <http://doi.org/10.1037/a0031797>
- Pearson, K. (1911). Biometrika Trust. *Biometrika*, 8(1/2), 254–256.
- Peterson, D. J., & Mulligan, N. W. (2013). The negative testing effect and multifactor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(4), 1287–1293. Retrieved from <http://psycnet.apa.orgjournals/xlm/39/4/1287>
- Peterson, H. A. (1944). Recitation or recall as a factor in the learning of long prose selections. *Journal of Education*, 35(4), 220–228. Retrieved from <http://psycnet.apa.orgjournals/edu/35/4/220>
- Pickering, M. (1990). Further studies on concept learning versus problem solving. *Journal of Chemical Education*, 254–255. <http://doi.org/10.1021/ed067p254>
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science (New York, N.Y.)*, 330(6002), 335. <http://doi.org/10.1126/science.1191465>

- Qualtrics. (2015). Qualtrics. Provo, Utah: Qualtrics. Retrieved from [www.qualtrics.com](http://www.qualtrics.com)
- Rock, I. (1957). The Role of Repetition in Associative Learning on JSTOR. *The American Journal of Psychology*, 70(2), 186–193. Retrieved from [http://www.jstor.org/stable/1419320?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/1419320?seq=1#page_scan_tab_contents)
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–7. <http://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–55. <http://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <http://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17(4), 382–395. Retrieved from <http://psycnet.apa.orgjournals/xap/17/4/382>
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, 5(3), 1–18.
- Sawrey, B. a. (1990). Concept learning versus problem solving: Revisited. *Journal of Chemical Education*, 67(3), 253. <http://doi.org/10.1021/ed067p253>
- Seymour, E. (2002). Tracking the Processes of Change in US Undergraduate Education in Science, Mathematics, Engineering, and Technology. *Science Education*, 86(1), 79–105. <http://doi.org/10.1002/sce.1044>
- Smith, K. C., Nakhleh, M. B., & Bretz, S. L. (2010). An expanded framework for analyzing general chemistry exams. *Chem. Educ. Res. Pract.*, 11(3), 147–153. <http://doi.org/10.1039/C005463C>
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641–656. Retrieved from <http://psycnet.apa.orgjournals/edu/30/9/641>
- Spivey, M. F., & McMillan, J. J. (2014). Classroom versus online assessment. *Journal of Education for Business*, 89(8), 450–456. <http://doi.org/10.1080/08832323.2014.937676>

- Squire, L. R. (1992). Nondeclarative Memory: Multiple Brain Systems Supporting Learning. *Journal of Cognitive Neuroscience*, 4(3), 232–243. <http://doi.org/10.1162/jocn.1992.4.3.232>
- StataCorp. (2013). STATA. College Station: TX: StataCorp LP.
- Steinberg, J., Brennehan, M., Lin, P., Carlson, J., Bridgeman, B., & Golub-Smith, M. (2014). A Comparison of Achievement Gaps and Test-Taker Characteristics on Praxis I® Tests. *ETS Research Report Series*, 2014(2), 1–20.
- Stenlund, T., Sundström, A., & Jonsson, B. (2014). Effects of repeated testing on short- and long-term memory performance across different test formats. *Educational Psychology*, 1–18. <http://doi.org/10.1080/01443410.2014.953037>
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 110, pp. 6313–7). <http://doi.org/10.1073/pnas.1221764110>
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, 56(4), 252–257. Retrieved from <http://psycnet.apa.org/journals/zea/56/4/252>
- Tsai, T.-H., & Shin, C. D. (2012). A Score Comparability Study for the NBDHE: Paper-Pencil Versus Computer Versions. *Evaluation & the Health Professions*, 36(2), 228–239. <http://doi.org/10.1177/0163278712445203>
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6(2), 175–184. [http://doi.org/10.1016/S0022-5371\(67\)80092-6](http://doi.org/10.1016/S0022-5371(67)80092-6)
- Van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, 35, 1–4. <http://doi.org/10.7748/ns2002.06.16.40.33.c3214>
- Vaughn, K. E., Rawson, K. A., & Pyc, M. A. (2013). Repeated retrieval practice and item difficulty: does criterion learning eliminate item difficulty effects? *Psychonomic Bulletin & Review*, 20(6), 1239–45. <http://doi.org/10.3758/s13423-013-0434-z>
- Veltre, M. T., Cho, K. W., & Neely, J. H. (2014). Transfer-appropriate processing in the testing effect. *Memory*, 1–9. <http://doi.org/10.1080/09658211.2014.970196>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007a). A Meta-Analysis of Testing Mode Effects in Grade K-12 Mathematics Tests. *Educational and Psychological Measurement*, 67(2), 219–238. <http://doi.org/10.1177/0013164406288166>

- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007b). Comparability of Computer-Based and Paper-and-Pencil Testing in K 12 Reading Assessments: A Meta-Analysis of Testing Mode Effects. *Educational and Psychological Measurement*, 68(1), 5–24. <http://doi.org/10.1177/0013164407305592>
- Wang, T. H. (2014). Developing an assessment-centered e-Learning system for improving student learning effectiveness. *Computers and Education*, 73, 189–203. <http://doi.org/10.1016/j.compedu.2013.12.002>
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11(6), 571–580. <http://doi.org/10.1080/09658210244000414>
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008. <http://doi.org/10.3758/MC.38.8.995>



## APPENDIX A. APPROVED IRB FORM

IRB ID: 12-520	
<b>INSTITUTIONAL REVIEW BOARD (IRB)</b>	
<b>Exempt Study Review Form</b>	
RECEIVED OCT 16 2012	
Title of Project: Test Effect as a Function of Test Format and Delivery	
By IRB	
Principal Investigator (PI): Anna Prisacari	Degrees: BA, MBA
University ID: [REDACTED]	Phone: [REDACTED]
Email Address: annacari@iastate.edu	
Correspondence Address: 1109 Gilman Hall, Ames, IA, 50011	
Department: Chemistry	College/Center/Institute: Chemistry
PI Level: <input type="checkbox"/> Tenured, Tenure-Eligible, & NTER Faculty <input type="checkbox"/> Adjunct/Affiliate Faculty <input type="checkbox"/> Collaborator Faculty <input type="checkbox"/> Emeritus Faculty <input type="checkbox"/> Visiting Faculty/Scientist <input type="checkbox"/> Senior Lecturer/Clinician <input type="checkbox"/> Lecturer/Clinician, w/Ph.D. or DVM <input type="checkbox"/> P&S Employee, P37 & above <input type="checkbox"/> Extension to Families/Youth Specialist <input type="checkbox"/> Field Specialist III <input type="checkbox"/> Postdoctoral Associate <input checked="" type="checkbox"/> Graduate/Undergrad Student <input type="checkbox"/> Other (specify: )	
FOR STUDENT PROJECTS (Required when the principal investigator is a student)	
Name of Major Professor/Supervising Faculty: Thomas Holme	
University ID: [REDACTED]	Phone: (515) 296-0570
Email Address: taholme@iastate.edu	
Campus Address: 0207 Gilman Hall	Department: Chemistry
Type of Project: (check all that apply)	<input checked="" type="checkbox"/> Thesis/Dissertation <input type="checkbox"/> Class Project <input type="checkbox"/> Other (specify: )
Alternate Contact Person: Shana Carpenter	Email Address: shacarp@iastate.edu
Correspondence Address: W112 Lagomarcino Hall	Phone: (515) 294-6385

**ASSURANCE**

- I certify that the information provided in this application is complete and accurate and consistent with any proposal(s) submitted to external funding agencies. Misrepresentation of the research described in this or any other IRB application may constitute non-compliance with federal regulations and/or academic misconduct.
- I agree to provide proper surveillance of this project to ensure that the rights and welfare of the human subjects are protected. I will report any problems to the IRB. See [Reporting Adverse Events and Unanticipated Problems](#) for details.
- I agree that modifications to the approved project will not take place without prior review and approval by the IRB.
- I agree that the research will not take place without the receipt of permission from any cooperating institutions, when applicable.
- I agree to obtain approval from other appropriate committees as needed for this project, such as the IACUC (if the research includes animals), the IBC (if the research involves biohazards), the Radiation Safety Committee (if the research involves x-rays or other radiation producing devices or procedures), etc.
- I understand that approval of this project does not grant access to any facilities, materials or data on which this research may depend. Such access must be granted by the unit with the relevant custodial authority.
- I agree that all activities will be performed in accordance with all applicable federal, state, local, and Iowa State University policies.

[REDACTED]  
Signature of Principal Investigator Date

[REDACTED]  
Signature of Major Professor/Supervising Faculty Date  
(Required when the principal investigator is a student)

- I have reviewed this application and determined that departmental requirements are met, the investigator(s) has/have adequate resources to conduct the research, and the research design is scientifically sound and has scientific merit.

[REDACTED]  
Signature of Department Chair Date

IRB Reviewer's Signature [REDACTED]

<input type="checkbox"/> No Human Participants	Review Date: October 18, 2012
EXEMPT Per 45 CFR 46.101(b):	1, 2, 4

Office for Responsible Research  
Revised: 07/05/12

1

## APPENDIX B. QUESTIONS FOR ALL TESTS

Chapter 1: Matter  
Topic 1: Quantum (Numbers) Mechanics  
Question Type: Definition

Q 1: Chapter 1, Question 1, MC

Answer: B

Which quantum number gives the shape of the atomic orbital?

A.  $n$

B.  $l$

C.  $m_l$

D.  $m_s$

Q 2: Chapter 1, Question 1, OE

Exact Answer:  $l$

Acceptable Answer: angular momentum/quantum #; L

Which quantum number gives the shape of the atomic orbital?

Q 3: Chapter 1, Question 1a, MC

Answer: A

Which quantum number primarily dictates the energy of the atomic orbital?

A.  $n$

B.  $l$

C.  $m_l$

D.  $m_s$

Q 4: Chapter 1, Question 1a, OE

Exact Answer:  $n$

Acceptable Answer: principal quantum/primary quantum #

Which quantum number primarily dictates the energy of the atomic orbital?

Chapter 1: Matter  
Topic 2: Periodicity  
Question Type: Definition

Q 5: Chapter 1, Question 2, MC

Answer: D

Which halogen atom has the largest radius?

**A.** Br

**B.** Cl

**C.** F

**D.** I

Q 6: Chapter 1, Question 2, OE

Exact Answer: At (Astatine)

Which halogen atom has the largest radius?

Q 7: Chapter 1, Question 2a, MC

Answer: C

Which noble gas atom has the largest radius?

**A.** Ar

**B.** He

**C.** Kr

**D.** Ne

Q 8: Chapter 1, Question 2a, OE

Exact Answer: Rn (Radon)

Which noble gas atom has the largest radius?

Chapter 1: Matter  
Topic 3: Atomic Structure  
Question Type: Algorithmic

Q 9: Chapter 1, Question 3, MC

Answer: A

An atom of strontium-90 ( ${}^{90}_{38}\text{Sr}$ ) contains how many electrons, protons, and neutrons?

- A.** 38 electrons, 38 protons, 52 neutrons    **B.** 38 electrons, 38 protons, 90 neutrons  
**C.** 52 electrons, 52 protons, 38 neutrons    **D.** 52 electrons, 38 protons, 38 neutrons

Q 10: Chapter 1, Question 3, OE

Exact Answer: 38 electrons, 38 protons, 52 neutrons

An atom of strontium-90 ( ${}^{90}_{38}\text{Sr}$ ) contains how many electrons, protons, and neutrons?

Q 11: Chapter 1, Question 3a, MC

Answer: A

An atom of niobium-93 ( ${}^{93}_{41}\text{Nb}$ ) contains how many electrons, protons, and neutrons?

- A.** 41 electrons, 41 protons, 52 neutrons    **B.** 41 electrons, 41 protons, 93 neutrons  
**C.** 52 electrons, 52 protons, 41 neutrons    **D.** 52 electrons, 41 protons, 41 neutrons

Q 12: Chapter 1, Question 3a, OE

Exact Answer: 41 electrons, 41, protons, 52 neutrons

An atom of niobium-93 ( ${}^{93}_{41}\text{Nb}$ ) contains how many electrons, protons, and neutrons?

Chapter 1: Matter  
 Topic 4: Isotopes  
 Question Type: Conceptual

Q 13: Chapter 1, Question 4, MC

Answer: C

Indium has two naturally occurring isotopes,  $^{113}\text{In}$  and  $^{115}\text{In}$ . The atomic mass of indium is 114.8 u. If one isotope of In is present in greater amounts, which is it, or are the two isotopes present in roughly the same amounts?

- A.  $^{113}\text{In}$  and  $^{115}\text{In}$  are present in nearly equal amounts
- B.  $^{113}\text{In}$  is present in much greater amounts than  $^{115}\text{In}$
- C.  $^{113}\text{In}$  is present in much smaller amounts than  $^{115}\text{In}$
- D. Cannot be determined without more information about the mass of each isotope

Q 14: Chapter 1, Question 4, OE

Exact Answer:  $^{113}\text{In}$  is present in much smaller amounts than  $^{115}\text{In}$  OR

$^{115}\text{In}$  is present in greater amounts

Indium has two naturally occurring isotopes,  $^{113}\text{In}$  and  $^{115}\text{In}$ . The atomic mass of indium is 114.8 u. If one isotope of In is present in greater amounts, which is it, or are the two isotopes present in roughly the same amounts?

Q 15: Chapter 1, Question 4a, MC

Answer: A

Europium has two naturally occurring isotopes,  $^{151}\text{Eu}$  and  $^{153}\text{Eu}$ . The atomic mass of europium is 151.9 u. If one isotope of Eu is present in greater amounts, which is it, or are the two isotopes present in roughly the same amounts?

- A.  $^{151}\text{Eu}$  and  $^{153}\text{Eu}$  are present in nearly equal amounts
- B.  $^{151}\text{Eu}$  is present in much greater amounts than  $^{153}\text{Eu}$
- C.  $^{151}\text{Eu}$  is present in much smaller amounts than  $^{153}\text{Eu}$
- D. Cannot be determined without more information about the mass of each isotope

Q 16: Chapter 1, Question 4a, OE

Exact Answer:  $^{151}\text{Eu}$  and  $^{153}\text{Eu}$  are present in nearly equal amounts

Acceptable Answer:  $^{151}\text{Eu}$  is present in slightly greater amount; or  $^{151}\text{Eu}$

Europium has two naturally occurring isotopes,  $^{151}\text{Eu}$  and  $^{153}\text{Eu}$ . The atomic mass of europium is 151.9 u. If one isotope of Eu is present in greater amounts, which is it, or are the two isotopes present in roughly the same amounts?

Chapter 2: Reactions  
Topic 1: Naming Inorganic Compounds  
Question Type: Algorithmic

Q 17: Chapter 2, Question 1, MC

Answer: C

What is the formula of chromium(III) carbonate?

- A.**  $\text{Cr}_3\text{CO}_3$       **B.**  $\text{Cr}(\text{CO}_3)_3$       **C.**  $\text{Cr}_2(\text{CO}_3)_3$       **D.**  $\text{Cr}_3(\text{CO}_3)_2$

Q 18: Chapter 2, Question 1, OE

Exact Answer:  $\text{Cr}_2(\text{CO}_3)_3$

What is the formula of chromium(III) carbonate?

Q 19: Chapter 2, Question 1a, MC

Answer: D

What is the formula of cobalt(II) phosphate?

- A.**  $\text{Co}_2\text{PO}_4$       **B.**  $\text{Co}(\text{PO}_4)_2$       **C.**  $\text{Co}_2(\text{PO}_4)_3$       **D.**  $\text{Co}_3(\text{PO}_4)_2$

Q 20: Chapter 2, Question 1a, OE

Exact Answer:  $\text{Co}_3(\text{PO}_4)_2$

What is the formula of cobalt(II) phosphate?

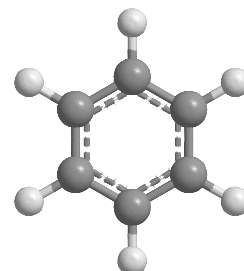
Chapter 2: Reactions  
Topic 2: Percent Composition  
Question Type: Algorithmic

Q 21: Chapter 2, Question 2, MC

Answer: D

Benzene is a hydrocarbon and is shown here. A hydrocarbon contains only carbon and hydrogen. What is the percent composition by mass of carbon in benzene?

- A. 8%
- B. 50%
- C. 80%
- D. 92%

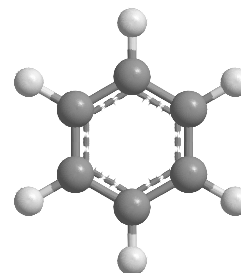


Q 22: Chapter 2, Question 2, OE

Exact Answer: 92%

Acceptable Answer: 91-93%; 93.01% - 93.99%

Benzene is a hydrocarbon and is shown to here. A hydrocarbon contains only carbon and hydrogen. What is the percent composition by mass of carbon in benzene?



Q 23: Chapter 2, Question 2a, MC

Answer: D

Propane is a hydrocarbon and is shown here. A hydrocarbon contains only carbon and hydrogen. What is the percent composition by mass of carbon in propane?

- A. 19%
- B. 27%
- C. 75%
- D. 81%



Q 24: Chapter 2, Question 2a, OE

Exact Answer: 81%

Acceptable Answer: 80-82%; 82.01% - 82.99%

Propane is a hydrocarbon and is shown here. A hydrocarbon contains only carbon and hydrogen. What is the percent composition by mass of carbon in propane?





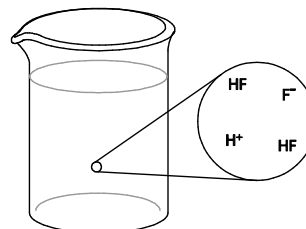
Chapter 2: Reactions  
 Topic 3: Acid/Base Chemistry  
 Question Type: Conceptual

Q 25: Chapter 2, Question 3, MC

Answer: C

In terms of acid/base chemistry, what is the classification of the substance shown? Water molecules are not shown for clarity.

- A.** strong acid      **B.** strong base  
**C.** weak acid      **D.** weak base

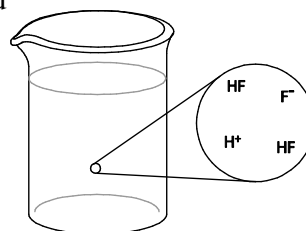


Q 26: Chapter 2, Question 3, OE

Exact Answer: Weak acid

Acceptable Answer: acid; acidic solution; or hydrofluoric acid

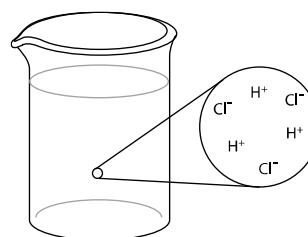
In terms of acid/base chemistry, what is the best classification of the substance shown? Water molecules are not shown for clarity.



Q 27: Chapter 2, Question 3a, MC

Answer: A

In terms of acid/base chemistry, what is the classification of the substance shown? Water molecules are not shown for clarity.



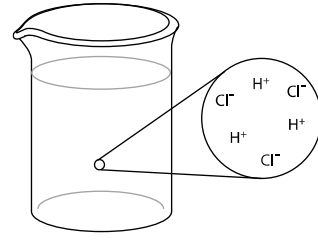
- A.** strong acid      **B.** strong base  
**C.** weak acid      **D.** weak base

Q 28: Chapter 2, Question 3a, OE

Answer: Strong acid

Acceptable Answer: acid, acidic solution, hydrochloric acid

In terms of acid/base chemistry, what is the most complete classification of the substance shown?  
Water molecules are not shown for clarity.

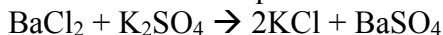


Chapter 2: Reactions  
 Topic 4: Net Ionic Equation  
 Question Type: Algorithmic

Q 29: Chapter 2, Question 4, MC

Answer: B

What is the net ionic equation for the reaction?



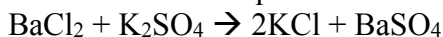
- (A)  $\text{K}^+(\text{aq}) + \text{Cl}^-(\text{aq}) \rightarrow \text{KCl}(\text{s})$   
 (B)  $\text{Ba}^{2+}(\text{aq}) + \text{SO}_4^{2-}(\text{aq}) \rightarrow \text{BaSO}_4(\text{s})$   
 (C)  $\text{Ba}^{2+}(\text{aq}) + 2\text{Cl}^-(\text{aq}) + 2\text{K}^+(\text{aq}) + \text{SO}_4^{2-}(\text{aq}) \rightarrow 2\text{KCl}(\text{s}) + \text{Ba}^{2+}(\text{aq}) + \text{SO}_4^{2-}(\text{aq})$   
 (D)  $\text{Ba}^{2+}(\text{aq}) + 2\text{Cl}^-(\text{aq}) + 2\text{K}^+(\text{aq}) + \text{SO}_4^{2-}(\text{aq}) \rightarrow 2\text{K}^+(\text{aq}) + 2\text{Cl}^-(\text{aq}) + \text{BaSO}_4(\text{s})$

Q 30: Chapter 2, Question 4, OE

Exact Answer:  $\text{Ba}^{2+}(\text{aq}) + \text{SO}_4^{2-}(\text{aq}) \rightarrow \text{BaSO}_4(\text{s})$

Acceptable Answer:  $\text{Ba}^{2+} + \text{SO}_4^{2-} \rightarrow \text{BaSO}_4$

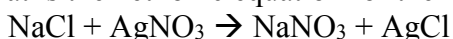
What is the net ionic equation for the reaction?



Q 31: Chapter 2, Question 4a, MC

Answer: D

What is the net ionic equation for the reaction?



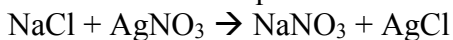
- (A)  $\text{Na}^+(\text{aq}) + \text{Cl}^-(\text{aq}) + \text{Ag}^+(\text{aq}) + \text{NO}_3^-(\text{aq}) \rightarrow \text{Na}^+(\text{aq}) + \text{NO}_3^-(\text{aq}) + \text{AgCl}(\text{s})$   
 (B)  $\text{Na}^+(\text{aq}) + \text{Cl}^-(\text{aq}) + \text{Ag}^+(\text{aq}) + \text{NO}_3^-(\text{aq}) \rightarrow \text{NaNO}_3(\text{s}) + \text{Ag}^+(\text{aq}) + \text{Cl}^-(\text{aq})$   
 (C)  $\text{Na}^+(\text{aq}) + \text{NO}_3^-(\text{aq}) \rightarrow \text{NaNO}_3(\text{s})$   
 (D)  $\text{Ag}^+(\text{aq}) + \text{Cl}^-(\text{aq}) \rightarrow \text{AgCl}(\text{s})$

Q 32: Chapter 2, Question 4a, OE

Exact Answer:  $\text{Ag}^+(\text{aq}) + \text{Cl}^-(\text{aq}) \rightarrow \text{AgCl}(\text{s})$

Acceptable Answer:  $\text{Ag}^+ + \text{Cl}^- \rightarrow \text{AgCl}$

What is the net ionic equation for the reaction?

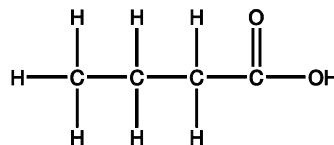


Chapter 3: Structure & Bonding  
 Topic 1: Bonding  
 Question Type: Algorithmic

Q 33: Chapter 3, Question 1, MC

Answer: C

How many sigma ( $\sigma$ ) and pi ( $\pi$ ) bonds are in one molecule of butyric acid (shown in the figure)?

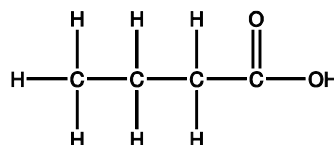


- (A) 12  $\sigma$  bonds and 1  $\pi$  bonds
- (B) 12  $\sigma$  bonds and 2  $\pi$  bonds
- (C) 13  $\sigma$  bonds and 1  $\pi$  bonds
- (D) 13  $\sigma$  bonds and 2  $\pi$  bonds

Q 34: Chapter 3, Question 1, OE

Exact Answer: 13  $\sigma$  bonds and 1  $\pi$  bonds

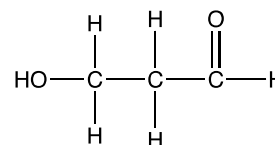
How many sigma ( $\sigma$ ) and pi ( $\pi$ ) bonds are in one molecule of butyric acid (shown in the figure)?



Q 35: Chapter 3, Question 1a, MC

Answer: C

How many sigma ( $\sigma$ ) and pi ( $\pi$ ) bonds are in one molecule of 3-hydroxy butanal (shown in the figure)?

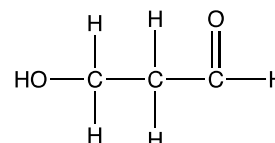


- (A) 9  $\sigma$  bonds and 1  $\pi$  bonds
- (B) 9  $\sigma$  bonds and 2  $\pi$  bonds
- (C) 10  $\sigma$  bonds and 1  $\pi$  bonds
- (D) 10  $\sigma$  bonds and 2  $\pi$  bonds

Q 36: Chapter 3, Question 1a, OE

Exact Answer: 10  $\sigma$  bonds and 1  $\pi$  bonds

How many sigma ( $\sigma$ ) and pi ( $\pi$ ) bonds are in one molecule of 3-hydroxy butanal (shown in the figure)?



## Chapter 3: Structure &amp; Bonding

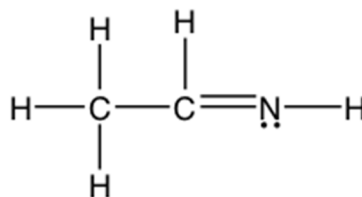
## Topic 2: Bond Length

## Question Type: Conceptual

Q 37: Chapter 3, Question 2, MC

Answer: A

Nitrogen may make single, double or triple bonds to carbon atoms. Which type of carbon-nitrogen bond will be **longer** than the one shown in this molecule?



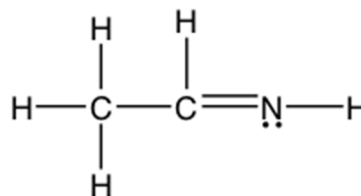
- A. single                      B. double                      C. triple  
D. None, all carbon-nitrogen bond lengths are equal

Q 38: Chapter 3, Question 2, OE

Exact Answer: single

Acceptable Answer: C-N

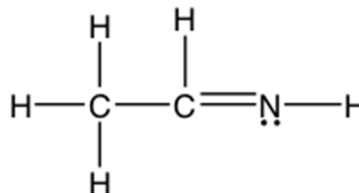
Nitrogen may make single, double or triple bonds to carbon atoms. Which type of carbon-nitrogen bond will be **longer** than the one shown in this molecule?



Q 39: Chapter 3, Question 2a, MC

Answer: C

Nitrogen may make single, double or triple bonds to carbon atoms. Which type of carbon-nitrogen bond will be **shorter** than the one shown in this molecule?

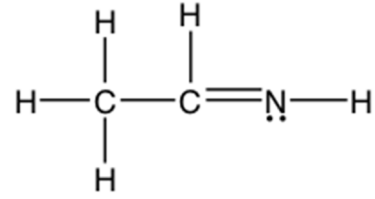


- A. single                      B. double                      C. triple  
D. None, all carbon-nitrogen bond lengths are equal

Q 40: Chapter 3, Question 2a, OE

Exact Answer: Triple

Nitrogen may make single, double or triple bonds to carbon atoms. Which type of carbon-nitrogen bond will be **shorter** than the one shown in this molecule?



Chapter 3: Structure & Bonding  
Topic 3: Lone Pairs Electrons  
Question Type: Algorithmic

Q 41: Chapter 3, Question 3, MC

Answer: C

When the correct Lewis structure is drawn for acetylide ion,  $C_2^{2-}$ , what is the total number of electron lone pairs present?

- A. 0**                      **B. 1**                      **C. 2**                      **D. 3**

Q 42: Chapter 3, Question 3, OE

Exact Answer: 2

Acceptable Answer:  $2e^-$

When the correct Lewis structure is drawn for acetylide ion,  $C_2^{2-}$ , what is the total number of electron lone pairs present?

Q 43: Chapter 3, Question 3a, MC

Answer: D

When the correct Lewis structure is drawn for acetylide ion,  $C_2^{2-}$ , what is the total number of bonding electron pairs present?

- A. 0**                      **B. 1**                      **C. 2**                      **D. 3**

Q 44: Chapter 3, Question 3a, OE

Exact Answer: 3

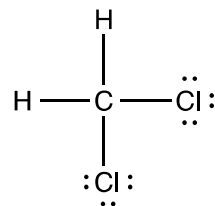
When the correct Lewis structure is drawn for acetylide ion,  $C_2^{2-}$ , what is the total number of bonding electron pairs present?

Chapter 3: Structure & Bonding  
 Topic 4: Shape  
 Question Type: Algorithmic

Q 45: Chapter 3, Question 4, MC

Answer: D

What is the molecular shape of dichloromethane?



A. seesaw

B. square planar

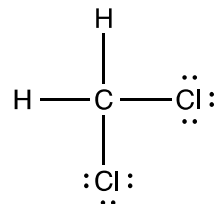
C. square pyramidal

D. tetrahedral

Q 46: Chapter 3, Question 4, OE

Exact Answer: tetrahedral

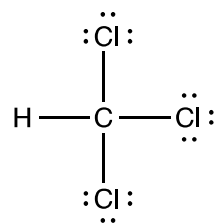
What is the molecular shape of dichloromethane?



Q 47: Chapter 3, Question 4a, MC

Answer: B

What is the molecular shape of trichloromethane?



A. seesaw

B. tetrahedral

C. square planar

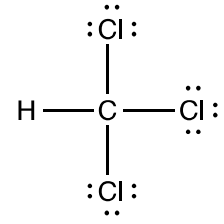
D. square pyramidal

Q 48: Chapter 3, Question 4a, OE

Exact Answer: tetrahedral



What is the molecular shape of trichloromethane?

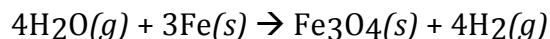


Chapter 4: Stoichiometry  
 Topic 1: Mole to Mole Stoichiometry  
 Question Type: Algorithmic

Q 49: Chapter 4, Question 1, MC

Answer: A

How many moles of Fe are needed to produce 10.0 mol of H<sub>2</sub>?



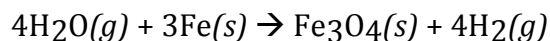
- A.** 7.50 mol                      **B.** 13.3 mol                      **C.** 15.0 mol                      **D.** 30.0 mol

Q 50: Chapter 4, Question 1, OE

Exact Answer: 7.50 mol

Acceptable Answer: 7.5 mol

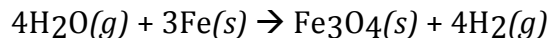
How many moles of Fe are needed to produce 10.0 mol of H<sub>2</sub>?



Q 51: Chapter 4, Question 1a, MC

Answer: B

How many moles of H<sub>2</sub> are produced by 10.0 mol of Fe?



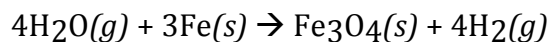
- A.** 7.50 mol                      **B.** 13.3 mol                      **C.** 15.0 mol                      **D.** 30.0 mol

Q 52: Chapter 4, Question 1a, OE

Exact Answer: 13.3 mol

Acceptable Answer: 13 mol – 13.34 mol

How many moles of H<sub>2</sub> are produced by 10.0 mol of Fe?



Chapter 4: Stoichiometry  
 Topic 2: Limiting Reagent  
 Question Type: Algorithmic

Q 53: Chapter 4, Question 2, MC

Answer: B

How many moles of  $\text{Al}_2\text{O}_3$  can be produced from 27 g of aluminum and 16 g of oxygen?

Atomic molar masses	
Al	27.0 g/mol
O	16.0 g/mol

- A.** 0.17 mol      **B.** 0.33 mol      **C.** 0.50 mol      **D.** 0.67 mol

Q 54: Chapter 4, Question 2, OE

Exact Answer: 0.33 mol

Acceptable Answer: 0.3 mol; 0.32 mol

How many moles of  $\text{Al}_2\text{O}_3$  can be produced from 27 g of aluminum and 16 g of oxygen?

Atomic molar masses	
Al	27.0 g/mol
O	16.0 g/mol

Q 55: Chapter 4, Question 2a, MC

Answer: C

How many moles of  $\text{Fe}_2\text{O}_3$  can be produced from 56 g of iron and 32 g of oxygen?

Atomic molar masses	
Fe	55.9 g/mol
O	16.0 g/mol

- A.** 0.17 mol      **B.** 0.33 mol      **C.** 0.50 mol      **D.** 0.67 mol

Q 56: Chapter 4, Question 2a, OE

Exact Answer: 0.50 mol

Acceptable Answer: 0.5 mol; 0.49mol

How many moles of  $\text{Fe}_2\text{O}_3$  can be produced from 56 g of iron and 32 g of oxygen?

Atomic molar masses	
Fe	55.9 g/mol
O	16.0 g/mol

Chapter 4: Stoichiometry  
 Topic 3: Molarity  
 Question Type: Algorithmic

Q 57: Chapter 4, Question 3, MC

Answer: D

What mass of  $\text{MgCl}_2$  is required to prepare 2.00 L of 0.550 M solution?

- A.** 1.10 g                      **B.** 26.2 g                      **C.** 28.9 g                      **D.** 105 g

Molar mass	
$\text{MgCl}_2$	95.2 g/mol

Q 58: Chapter 4, Question 3, OE

Exact Answer: 105 g

Acceptable Answer: 104.7 – 105 g

What mass of  $\text{MgCl}_2$  is required to prepare 2.00 L of 0.550 M solution?

Molar mass	
$\text{MgCl}_2$	95.2 g/mol

Q 59: Chapter 4, Question 3a, MC

Answer: D

What mass of anhydrous  $\text{Na}_2\text{S}$  is required to prepare 4.00 L of 0.270 M solution?

- A.** 1.08 g                      **B.** 5.27 g                      **C.** 72.2 g                      **D.** 84.2 g

Molar mass	
$\text{Na}_2\text{S}$	78.0 g/mol

Q 60: Chapter 4, Question 3a, OE

Exact Answer: 84.2 g

Acceptable Answer: 84.24 g; 84 g

What mass of anhydrous  $\text{Na}_2\text{S}$  is required to prepare 4.00 L of 0.270 M solution?

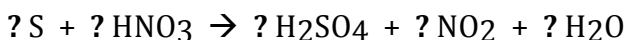
Molar mass	
$\text{Na}_2\text{S}$	78.0 g/mol

Chapter 4: Stoichiometry  
 Topic 4: Balancing Equation  
 Question Type: Algorithmic

Q 61: Chapter 4, Question 4, MC

Answer: B

Balance this equation and then answer the following question:



How many moles of H<sub>2</sub>O are formed per mole of sulfur consumed?

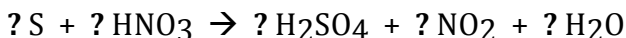
- A. 1**                      **B. 2**                      **C. 4**                      **D. 6**

Q 62: Chapter 4, Question 4, OE

Exact Answer: 2

Acceptable Answer: 1:2

Balance this equation and then answer the following question:

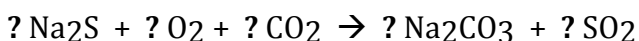


How many moles of H<sub>2</sub>O are formed per mole of sulfur consumed?

Q 63: Chapter 4, Question 4a, MC

Answer: A

Balance this equation and then answer the following question:



How many moles of SO<sub>2</sub> are formed per mole of sodium sulfide, Na<sub>2</sub>S, consumed?

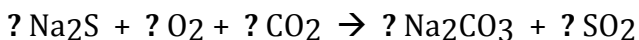
- A. 1**                      **B. 2**                      **C. 4**                      **D. 6**

Q 64: Chapter 4, Question 4a, OE

Exact Answer: 1

Acceptable Answer: 1:1; 2:2

Balance this equation and then answer the following question:



How many moles of SO<sub>2</sub> are formed per mole of sodium sulfide, Na<sub>2</sub>S, consumed?

Chapter 5: Gases  
Topic 1: Kinetic Molecular Theory  
Question Type: Conceptual

Q 65: Chapter 5, Question 1, MC

Answer: D

At room temperature, the group 18 elements, He, Ne, Ar, Kr, Xe are all gases. Which gas has the greatest average kinetic energy at room temperature?

- A.** Helium                      **B.** Argon                      **C.** Xenon  
**D.** None, the average kinetic energy is the same for each gas

Q 66: Chapter 5, Question 1, OE

Exact Answer: None, the average kinetic energy is the same for each gas

At room temperature, the group 18 elements, He, Ne, Ar, Kr, Xe are all gases. Which gas has the greatest average kinetic energy at room temperature?

Q 67: Chapter 5, Question 1a, MC

Answer: D

At 1500 °C, the group 1 elements, Li, Na, K, Rb, Cs are all monatomic gases. Which gas has the smallest average kinetic energy at 1500 °C?

- A.** Lithium                      **B.** Potassium                      **C.** Cesium  
**D.** None, the average kinetic energy is the same for each gas

Q 68: Chapter 5, Question 1a, OE

Exact Answer: None, the average kinetic energy is the same for each gas

Acceptable Answer: all have the same KE b/c they're all in the same column (correct answer with incorrect rationale)

At 1500 °C, the group 1 elements, Li, Na, K, Rb, Cs are all monatomic gases. Which gas has the smallest average kinetic energy at 1500 °C?

Chapter 5: Gases  
Topic 2: Ideal Gas Law  
Question Type: Algorithmic

Q 69: Chapter 5, Question 2, MC

Answer: A

An ideal gas sample occupies a volume of 16.4 L at 27 °C and 0.300 atm. How many moles of gas are present?

- A.** 0.200 mol      **B.** 0.450 mol      **C.** 3.50 mol      **D.** 10.0 mol

Q 70: Chapter 5, Question 2, OE

Exact Answer: 0.200 mol

Acceptable Answer: 0.199 – 0.204 mol

An ideal gas sample occupies a volume of 16.4 L at 27 °C and 0.300 atm. How many moles of gas are present?

Q 71: Chapter 5, Question 2a, MC

Answer: B

An ideal gas sample occupies a volume of 31.6 L at 27 °C and 0.350 atm. How many moles of gas are present?

- A.** 0.200 mol      **B.** 0.450 mol      **C.** 2.22 mol      **D.** 5.00 mol

Q 72: Chapter 5, Question 2a, OE

Exact Answer: 0.450 mol

Acceptable Answer: 0.447 – 0.460 mol

An ideal gas sample occupies a volume of 31.6 L at 27 °C and 0.350 atm. How many moles of gas are present?

Chapter 5: Gases  
 Topic 3: Partial Pressure  
 Question Type: Algorithmic

Q 73: Chapter 5, Question 3, MC

Answer: C

The partial pressures of a gaseous mixture are given in the table. What is the mole fraction of hydrogen in the mixture?

Partial Pressures	
H <sub>2</sub>	200. torr
CO <sub>2</sub>	150. torr
CH <sub>4</sub>	320. torr
C <sub>2</sub> H <sub>4</sub>	105. torr

A. 0.135

B. 0.194

C. 0.258

D. 0.413

Q 74: Chapter 5, Question 3, OE

Exact Answer: 0.258

Acceptable Answer: 25.8% or 0.260 or 26%

The partial pressures of a gaseous mixture are given in the table. What is the mole fraction of hydrogen in the mixture?

Partial Pressures	
H <sub>2</sub>	200. torr
CO <sub>2</sub>	150. torr
CH <sub>4</sub>	320. torr
C <sub>2</sub> H <sub>4</sub>	105. torr

Q 75: Chapter 5, Question 3a, MC

Answer: B

The partial pressures of a gaseous mixture are given in the table. What is the mole fraction of carbon dioxide in the mixture?

Partial Pressures	
H <sub>2</sub>	200. torr
CO <sub>2</sub>	150. torr
CH <sub>4</sub>	320. torr
C <sub>2</sub> H <sub>4</sub>	105. torr

A. 0.135

B. 0.194

C. 0.258

D. 0.413

Q 76: Chapter 5, Question 3a, OE

Exact Answer: 0.194

Acceptable Answer: 0.190 – 0.200 or 19% - 20%

The partial pressures of a gaseous mixture are given in the table. What is the mole fraction of carbon dioxide in the mixture?

Partial Pressures	
H <sub>2</sub>	200. torr
CO <sub>2</sub>	150. torr
CH <sub>4</sub>	320. torr
C <sub>2</sub> H <sub>4</sub>	105. torr



Chapter 5: Gases  
Topic 4: Boyle's Law  
Question Type: Algorithmic

Q 77: Chapter 5, Question 4, MC

Answer: C

A student collected 40.0 mL of  $N_2$  gas when the temperature was  $20.0^\circ C$  and the pressure was 720. torr. The next day the temperature was still  $20.0^\circ C$ , but there was only 38.4 mL of gas present. What was the pressure on this second day?

- A.** 691 torr      **B.** 700. torr      **C.** 750. torr      **D.** 760. torr

Q 78: Chapter 5, Question 4, OE

Exact Answer: 750 torr

Acceptable Answer: 0.98 – 0.99 atm

A student collected 40.0 mL of  $N_2$  gas when the temperature was  $20.0^\circ C$  and the pressure was 720. torr. The next day the temperature was still  $20.0^\circ C$ , but there was only 38.4 mL of gas present. What was the pressure on this second day?

Q 79: Chapter 5, Question 4a, MC

Answer: A

A student collected 115 mL of  $O_2$  gas when the temperature was  $25.0^\circ C$  and the pressure was 750. torr. The next day the temperature was still  $25.0^\circ C$ , but now there was 119 mL of gas present. What was the pressure on this second day?

- A.** 725 torr      **B.** 735 torr      **C.** 770. torr      **D.** 776 torr

Q 80: Chapter 5, Question 4a, OE

Exact Answer: 725 torr

Acceptable Answer: 724 – 725 torr; 0.950 – 0.955 atm

A student collected 115 mL of  $O_2$  gas when the temperature was  $25.0^\circ C$  and the pressure was 750. torr. The next day the temperature was still  $25.0^\circ C$ , but now there was 119 mL of gas present. What was the pressure on this second day?

Chapter 6: Thermochemistry  
Topic 1: Specific Heat  
Question Type: Algorithmic

Q 81: Chapter 6, Question 1, MC

Answer: B

What is the final temperature when a 21.5 g sample of gold (specific heat =  $0.129 \text{ J}\cdot\text{g}^{-1}\cdot\text{°C}^{-1}$ ) emits 233 J of heat when it cools from  $125\text{°C}$ ?

- A.  $-84\text{°C}$                       B.  $41\text{°C}$                       C.  $84\text{°C}$                       D.  $209\text{°C}$

Q 82: Chapter 6, Question 1, OE

Exact Answer:  $41\text{°C}$  or  $314.15 \text{ Kelvin}$

Acceptable Answer:  $314 \text{ Kelvin}$ ;  $40.6 - 41.4 \text{ C}$

What is the final temperature when a 21.5 g sample of gold (specific heat =  $0.129 \text{ J}\cdot\text{g}^{-1}\cdot\text{°C}^{-1}$ ) emits 233 J of heat when it cools from  $125\text{°C}$ ?

Q 83: Chapter 6, Question 1a, MC

Answer: C

What is the final temperature when a 13.7 g sample of aluminum (specific heat =  $0.900 \text{ J}\cdot\text{g}^{-1}\cdot\text{°C}^{-1}$ ) absorbs 421 J of heat when it is heated from  $22\text{°C}$ ?

- A.  $34\text{°C}$                       B.  $50\text{°C}$                       C.  $56\text{°C}$                       D.  $608\text{°C}$

Q 84: Chapter 6, Question 1a, OE

Exact Answer:  $56\text{°C}$  or  $329.15 \text{ Kelvin}$

Acceptable Answer:  $329 - 330 \text{ Kelvin}$

Accept:  $56.0 - 56.4 \text{ C}$

What is the final temperature when a 13.7 g sample of aluminum (specific heat =  $0.900 \text{ J}\cdot\text{g}^{-1}\cdot\text{°C}^{-1}$ ) absorbs 421 J of heat when it is heated from  $22\text{°C}$ ?

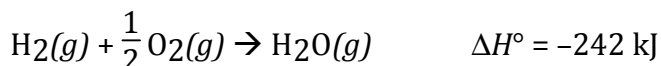
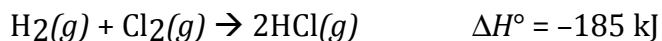
Chapter 6: Thermochemistry  
 Topic 2: Heat of Formation  
 Question Type: Algorithmic

Q 85: Chapter 6, Question 2, MC

Answer: A

Calculate  $\Delta H^\circ$  at 25 °C for  $2\text{HCl}(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{Cl}_2(g) + \text{H}_2\text{O}(g)$

given



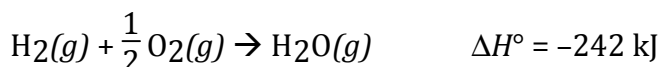
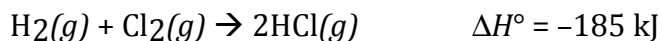
- A.** -57 kJ                      **B.** -185 kJ                      **C.** -306 kJ                      **D.** -427 kJ

Q 86: Chapter 6, Question 2, OE

Exact Answer: -57 kJ

Calculate  $\Delta H^\circ$  at 25 °C for  $2\text{HCl}(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{Cl}_2(g) + \text{H}_2\text{O}(g)$

given

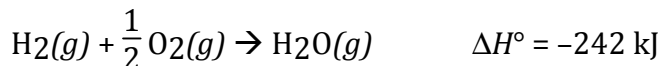
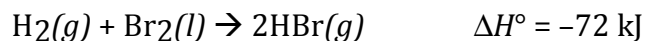


Q 87: Chapter 6, Question 2a, MC

Answer: B

Calculate  $\Delta H^\circ$  at 25 °C for  $2\text{HBr}(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{Br}_2(l) + \text{H}_2\text{O}(g)$

given

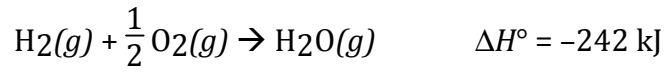
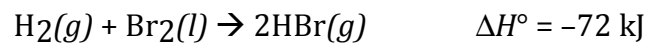


- A.** -98 kJ                      **B.** -170 kJ                      **C.** -278 kJ                      **D.** -314 kJ

Q 88: Chapter 6, Question 2a, OE

Exact Answer: -170 kJ

Calculate  $\Delta H^\circ$  at 25 °C for  $2\text{HBr}(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{Br}_2(l) + \text{H}_2\text{O}(g)$   
given



Chapter 6: Thermochemistry  
Topic 3: Exothermic/Endothermic  
Question Type: Conceptual

Q 89: Chapter 6, Question 3, MC

Answer: C

A material in the liquid state is first vaporized and then condensed. What are the respective steps in the process (in terms of endothermic/exothermic)?

- A. Endothermic then endothermic
- B. Exothermic then exothermic
- C. Endothermic then exothermic
- D. Exothermic then endothermic

Q 90: Chapter 6, Question 3, OE

Exact Answer: Endothermic and exothermic

A material in the liquid state is first vaporized and then condensed. What are the respective steps in the process (in terms of endothermic/exothermic)?

Q 91: Chapter 6, Question 3a, MC

Answer: A

A material in the solid state first melts and then evaporates. What are the respective steps in the process (in terms of endothermic/exothermic)?

- A. Endothermic then endothermic
- B. Exothermic then exothermic
- C. Endothermic then exothermic
- D. Exothermic then endothermic

Q 92: Chapter 6, Question 3a, OE

Exact Answer: Endothermic and endothermic

Acceptable Answer: endothermic process/reaction or endothermic

A material in the solid state first melts and then evaporates. What are the respective steps in the process (in terms of endothermic/exothermic)?

Chapter 6: Thermochemistry  
 Topic 4: Heat of Reaction  
 Question Type: Algorithmic

Q 93: Chapter 6, Question 4, MC

Answer: B

Given the bond energies in the table, what is the  $\Delta H^\circ$  for the chemical reaction,  $\text{Cl}_2(\text{g}) + \text{F}_2(\text{g}) \rightarrow 2\text{ClF}(\text{g})$ ?

Bond Energies	
Cl-Cl	243 kJ/mol
F-F	159 kJ/mol
Cl-F	255 kJ/mol

- A. -147 kJ      B. -108 kJ      C. +108 kJ      D. +147 kJ

Q 94: Chapter 6, Question 4, OE

Exact Answer: -108 kJ

Given the bond energies in the table, what is the  $\Delta H^\circ$  for the chemical reaction,  $\text{Cl}_2(\text{g}) + \text{F}_2(\text{g}) \rightarrow 2\text{ClF}(\text{g})$ ?

Bond Energies	
Cl-Cl	243 kJ/mol
F-F	159 kJ/mol
Cl-F	255 kJ/mol

Q 95: Chapter 6, Question 4a, MC

Answer: C

Given the bond energies in the table, what is the  $\Delta H^\circ$  for the chemical reaction,  $2\text{ClF}(\text{g}) \rightarrow \text{Cl}_2(\text{g}) + \text{F}_2(\text{g})$ ?

Bond Energies	
Cl-Cl	243 kJ/mol
F-F	159 kJ/mol
Cl-F	255 kJ/mol

- A. -147 kJ      B. -108 kJ      C. +108 kJ      D. +147 kJ

Q 96: Chapter 6, Question 4a, OE

Exact Answer: +108 kJ

Given the bond energies in the table, what is the  $\Delta H^\circ$  for the chemical reaction,  $2\text{ClF}(\text{g}) \rightarrow \text{Cl}_2(\text{g}) + \text{F}_2(\text{g})$ ?

Bond Energies	
Cl-Cl	243 kJ/mol
F-F	159 kJ/mol
Cl-F	255 kJ/mol

## APPENDIX C. INSTRUCTIONS TO COMPUTER-BASED TESTS

## Time:

You will have 35 minutes to complete this practice test. Proctor will notify you when there are ten minutes remaining. Submit your answers only when you are instructed to do so. If you finish earlier, take the time to verify your answers and check the spelling of your typed answers.

## Typing:

- When asked to type your answer, please spell out your answer (sigma, pi, Celsius degree).
- If your answer contains superscript please use ^ sign. For example,  $x^2 = x^2$ .
- If your answer contains subscript, please write your answer as it is. For example,  $H_2O = H2O$ .
- For calculation questions, please submit the final answer. For example, to 1/100 write 0.1 or 10%.

## Scoring:

- Each question is worth 1 point.
- Multiple-choice questions will be scored by the computer and your open-ended responses will be scored manually.
- Unanswered questions will be scored as incorrect.

## Extra:

During the test, you may use periodic table, general chemistry formulas, scratch paper, and basic calculator.

\*\*\*\*\*

## APPENDIX D. INSTRUCTIONS TO PAPER-BASED TESTS

## Time:

You will have 35 minutes to complete this practice test. Proctor will notify you when there are ten minutes remaining. When you see the message "PLEASE STOP HERE", please stop at that page. The proctor will notify you when you can move on. If you finish earlier, take the time to verify your answers and check the spelling of your typed answers.

## Writing:

- When asked to write your answer, please type it so the researcher can read it.
- For calculation questions, please submit the final answer. For example, to 1/100 write 0.1 or 10%.

## Scoring:

- Each question is worth 1 point.
- Unanswered questions will be scored as incorrect.

## Extra:

During the test, you may use periodic table, general chemistry formulas, scratch paper, and basic calculator.

\*\*\*\*\*