Measuring the testing mode in general chemistry: The effect of computer versus paper mode on test performance, cognitive load, and scratch paper

Anna Agripina Prisacari

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Chemistry Commons, Educational Psychology Commons, and the Psychology Commons

Recommended Citation


http://lib.dr.iastate.edu/etd/15399

This Dissertation 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.
Measuring the testing mode in general chemistry: The effect of computer versus paper mode on test performance, cognitive load, and scratch paper

by

Anna Agripina Prisacari

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Jared A. Danielson, Major Professor
   Eunjin Bahng
Amanda J. Fales-Williams
   Jonathan Kelly
   Craig A. Ogilvie

Iowa State University
Ames, Iowa
2017

Copyright © Anna Agripina Prisacari, 2017. All rights reserved.
DEDICATION

I dedicate this work to my best friend and my mother, Svetlana Prisacari, who gave me the opportunity and support to achieve my American Dream.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1. Motivation</td>
<td>1</td>
</tr>
<tr>
<td>My previous research</td>
<td>1</td>
</tr>
<tr>
<td>My current research</td>
<td>4</td>
</tr>
<tr>
<td>2. Organization of the Dissertation</td>
<td>8</td>
</tr>
<tr>
<td>Article 1. Comparing student performance using computer and paper—based tests. Results from two studies in general chemistry</td>
<td>8</td>
</tr>
<tr>
<td>Article 2. Rethinking testing mode: Should I offer my next chemistry test on paper on computer?</td>
<td>9</td>
</tr>
<tr>
<td>Article 3. Computer-based versus paper-based testing: Investigating the testing mode with cognitive load and scratch paper use</td>
<td>10</td>
</tr>
<tr>
<td>References</td>
<td>12</td>
</tr>
<tr>
<td>CHAPTER 2: COMPARING STUDENT PERFORMANCE USING COMPUTER AND PAPER—BASED TESTS. RESULTS FROM TWO STUDIES IN GENERAL CHEMISTRY</td>
<td>16</td>
</tr>
<tr>
<td>1. Abstract</td>
<td>16</td>
</tr>
<tr>
<td>Graphical Abstract</td>
<td>17</td>
</tr>
<tr>
<td>Keywords</td>
<td>17</td>
</tr>
<tr>
<td>Introduction</td>
<td>17</td>
</tr>
<tr>
<td>Online mode in chemistry</td>
<td>18</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>19</td>
</tr>
<tr>
<td>Study 1</td>
<td>20</td>
</tr>
<tr>
<td>Methods</td>
<td>20</td>
</tr>
<tr>
<td>Results</td>
<td>27</td>
</tr>
<tr>
<td>Study 2</td>
<td>30</td>
</tr>
<tr>
<td>Methods</td>
<td>30</td>
</tr>
<tr>
<td>Results</td>
<td>32</td>
</tr>
<tr>
<td>Discussion</td>
<td>35</td>
</tr>
<tr>
<td>Limitations</td>
<td>37</td>
</tr>
<tr>
<td>Implications</td>
<td>39</td>
</tr>
<tr>
<td>Author information</td>
<td>40</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>40</td>
</tr>
<tr>
<td>References</td>
<td>40</td>
</tr>
<tr>
<td>Supplemental information</td>
<td>47</td>
</tr>
<tr>
<td>Item format effects</td>
<td>47</td>
</tr>
<tr>
<td>Example test (Areas for open response answers are removed for space reasons.)</td>
<td>49</td>
</tr>
</tbody>
</table>
CHAPTER 3. RETHINKING TESTING MODE: SHOULD I OFFER MY NEXT CHEMISTRY TEST ON PAPER OR COMPUTER?

Abstract 54
Introduction 55
   Paper-based versus computer-based testing 56
   The present study 58
Method 60
   Participants 60
   Design 61
   Material 63
   Procedure 70
   Data analysis 72
Results 75
   Overall gains 75
   Algorithmic gains 75
   Conceptual gains 76
   Definition gains 76
Discussion and Conclusions 78
Limitations and Future Research 82
References 83

CHAPTER 4: COMPUTER-BASED VERSUS PAPER-BASED TESTING: INVESTIGATING THE TESTING MODE WITH COGNITIVE LOAD AND SCRATCH PAPER USE 89

Abstract 89
Highlights 90
1. Introduction 90
2. Literature Review 91
   2.1. Cognitive load 91
   2.2. Use of scratch paper 96
3. Present Study 97
4. Methods 99
   4.1. Participants 99
   4.2. Design 100
   4.3. Material 101
   4.4. Scratch paper 104
   4.5. Procedure 104
5. Results 106
   5.1. Cognitive load 106
   5.2. Use of scratch paper 108
   5.3. Question type: Algorithmic, conceptual, and definition 108
6. Discussion and Conclusion 112
   6.1. Limitations and future research 116
Acknowledgments 117
References 118
ACKNOWLEDGEMENTS

Success is incremental; it is a product of hard work, dedication, and help from other people. There are many people who are part of my success story. They are professors, colleagues, family members, and friends who taught me how to think, confront challenges, and, most importantly, not give up.

• My advisor, Dr. Jared Danielson: thank you for sharing your knowledge with me, providing guidance, and giving valuable advice on professional and personal matters.
• My POS committee members, Dr. Eunjin Bahng, Dr. Amanda Fales-Williams, Dr. Jonathan Kelly, and Dr. Craig Ogilvie: thank you for giving me helpful feedback.
• Dr. Ana-Paula Correia: thank you for your constructive criticism, continuous support, and always being available when I needed help.
• Dr. Thomas Holme: thank you for igniting the interest to be a researcher in me.
• Mrs. Karen Bovenmyer: thank you for sharing your contagious energy with me.
• Dr. Jonathan Kelly, Dr. Stephen Gilbert, Dr. Eliot Winer, and Amy Carver: thank you for giving me the opportunities to teach and mentor students. No textbook can prepare you to be a great teacher and is better than the actual classroom with students.
• Chemistry instructors, teaching assistants, and students: thank you for helping me to conduct studies and collect valuable data for my research.
• Friends and colleagues from the Graduate and Professional Student Senate (GPSS): thank you for endlessly challenging me. Especially, I would like to thank Dr. Brian Tlach: thank you for being a true leader, a good listener, and a reliable friend.
• Friends and colleagues at the Center for Communication Excellence, especially Dr. Elena Cotos, Dr. Sarah Huffman, and Chris Nelson: you taught me a never expiring skill - how
to communicate research clearly. Also, thank you for offering me the opportunity to be one of the interdisciplinary writing consultants; helping other graduate students helped me to become a better writer and thinker.

• Members of the HCI Peer Review Group, Dan Bell, Bhaskar Bhattacharya, Holly Rae Bemis-Schultz, Chase Meusel, and Desarae Veit: thank you for your helpful comments on my articles and presentations.

• Students in my classes: thank you for asking me tough questions; they helped me to learn, grow, and be a better teacher.

• School teachers, especially my Moldavian middle school teacher Doamna Vera Mahu, US high school director Mr. Manny Fritz, and high school math teacher Mrs. Stark: you knew about my talents before I discovered them.

• Gauri Ramasubramanian: thank you for lifting my spirits on gloomy days. When I was entering data at 2AM, to cheer me up you danced outside my window.

• Friends, Betsa Mantilla and Darlene Sanchez: thank you for celebrating with me my birthdays, professional milestones, and personal achievements.

• My dearest and best friends, mom Svetlana Prisacari and sister Vera Prisacari: thank you for nonstop encouragement and unconditional support. Spending time together helped me to charge up and continue my work.

• My grandparents, Pavel and Vera Prisacari: although you are not with me anymore, it was you who raised a fighter in me. Thank you for teaching me to dream and work hard to turn dreams into reality.

All, thank you for helping me create this success story!
ABSTRACT

The purpose of this dissertation was to study the effect of testing mode (giving a test on computer versus paper) in authentic classroom environments. Specifically, this work sought to broaden the pedagogical implications of testing mode by examining variations between computer and paper mode in student test performance, cognitive load, and use of scratch paper. Data were collected from students who were enrolled in undergraduate general chemistry courses. Students took multiple proctored general chemistry tests that included a variety of chemistry topics and three question types: algorithmic, conceptual, and definition. After each test students were provided with correct-answer feedback. The tests were either graded and part of the class (i.e., class quizzes) or served as a practice test for the upcoming test in the course. To measure testing mode, tests were delivered on computer or paper, forming four conditions. The conditions, defined by the mode of the initial and final test, were: Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper conditions. A series of studies were conducted and their findings are presented in three separate articles. Article 1 discusses test performance of two groups of students ($N = 207$ and $N = 215$) who took two similar practice tests in one session. Article 2 discusses test performance of new students ($N = 221$) who took two quizzes in class on Day 1 and Day 8 and one practice test on Day 12. Article 3 examines testing mode with students ($N = 221$) who reported their cognitive load and provided scratch paper. The results revealed that overall there seems to be little difference between online and paper-based testing. No significant differences were found in student test performance nor in the cognitive load measures between the computer and paper modes. However, a significant difference between two modes was detected with use of scratch paper, though the effect was very small ($\eta^2 = .04$). Although students used scratch paper more when taking a chemistry test on paper than online, it is not
clear whether the difference of approximately one question represents any practical significance. The present dissertation supports the conclusion that online testing is a promising alternative to the traditional paper-and-pencil mode most often used in general chemistry courses and changing from paper to online mode would not impose an additional cognitive load on students.
CHAPTER 1. INTRODUCTION

The focus of this dissertation is testing mode and its effect on multiple testing outcomes in the context of general chemistry. The work described here is presented in three, related papers targeted at peer-reviewed journals in the fields of chemical education, computers in education, and human computer interaction. While these papers are written using third-person voice, this chapter uses a first-person narrative to illustrate how this work was created and developed over the years.

Motivation

My primary motivation for this research came from the results of my master’s thesis (Prisacari, 2015) in which I examined the role of testing mode (delivering a chemistry practice test on computer versus paper) on student performance. Although my previous research contributes to the broader literature on testing mode by examining it with different combinations of computer-based and paper-based modes, the current research extends this work to a more authentic classroom environment while exploring the testing mode with cognitive load and students’ use of scratch paper.

My previous research

From conducting the literature review for my master’s thesis I learned that testing is advantageous for long-term learning (Carpenter, 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger, Agarwal, Kang, & Marsh, 2010; Roediger, Putnam, & Smith, 2011; Roediger & Butler, 2011). Although the benefit of testing has been well documented, I found that testing mode was a relatively unexplored topic in the testing research.
The testing mode refers to the method of delivering a test to students (e.g., on computer or paper) and the testing mode effect refers to differences in student performance between a computer-based and paper-based test with intent to measure equivalent outcomes. Reviewing the literature, I found that the testing mode has been mostly explored in one of two ways. In some studies, all participants take one test, with one group of students taking a test on computer and another group of students taking the same test on paper (Alexander, Bartlett, Truell, & Ouwenga, 2000; Clariana & Wallace, 2002; Emerson & MacKay, 2011; Hochlehnert, Brass, Moeltner, & Juenger, 2011; Jackel, 2014; Tsai & Shin, 2012). In other studies, students take two tests: one in computer mode and one in paper mode. These tests can be either identical (Meyer, Innes, Stomski, & Armson, 2015) or different, but match in topic and difficulty (Johnson & Green, 2006). Not many examples of testing mode under more diverse conditions can be found in the literature. Escudier, Newton, Cox, Reynolds, and Odell (2011) have extended the testing mode research by exploring the order of multiple tests, but their study included only computer to paper and paper to computer conditions. Thus, the gap in the testing mode literature remains. We are still left with an important question: does student performance vary when students take two tests using different modes versus when students are taking both tests only via one mode? Given that students are often exposed to multiple tests in a class and instructors have access to build their tests in either mode, I found this question to be worthy of further investigation. I wanted to compare previously studied conditions (i.e., computer to paper and paper to computer) to situations when students take multiple tests only on computer or only on paper. Will students benefit if all of their tests are delivered online? Finally, I found no studies that aimed to investigate testing mode in the general chemistry setting. As more universities begin to offer opportunities to their students to learn chemistry online (Leontyev & Baranov, 2013; O’Malley,
Agger, & Anderson, 2015; Pienta, 2013), it became clear to me that more research in the testing mode area was needed.

To address the significant gaps in the literature, for my master’s thesis, I asked 664 students from two universities to take two general chemistry practice tests in subsequent order before their end-of-semester general chemistry exam. Each test was delivered either on computer or paper, resulting in four experimental conditions (Test 1 – Test 2): Computer – Computer, Computer – Paper, Paper – Computer, and Paper – Paper. This design allowed me to investigate the testing mode with all possible mode combinations while proctoring two different tests that were similar in content. Each test included 17 algorithmic questions, 5 conceptual questions, and 2 definition questions. All three question types are common to general chemistry course (Smith, Nakhleh, & Bretz, 2010). Several interesting results emerged from this study. First, I found that improvements in test performance from Test 1 to Test 2 varied by condition, suggesting that student performance might be affected by the mode of their initial and subsequent test. It was particularly disadvantageous to take the first practice test on paper and the second practice test on computer relative to taking both practice tests on paper. Second, I found that the testing mode was significant with conceptual and definition questions and not significant with algorithmic questions.

Although my master’s results contributed to the testing mode literature, I still remained puzzled by them. I wanted to know how testing mode behaves under more authentic classroom settings and why one testing mode might help students to obtain a higher test score than the other mode. These questions guided my current research.
My current research

My current research involved several studies. First, to understand the testing mode better and ensure reliability, I repeated my master’s study with new students (i.e., Chapter 2). Second, I designed and conducted a new experiment in which I addressed several limitations of my master’s study and measured testing mode with cognitive load and scratch paper. The results of this experiment are divided in two separate papers. In Chapter 3, I present the results of student test performance and in Chapter 4, I discuss results of cognitive load and scratch paper.

To design the new study, I began with limitations of my previous work. In my master’s work, I proctored two tests on the same day using only a 20-minute interval between two practice tests. Although proctoring a test immediately after the study is easy to do, this strategy shows little benefit for long-term learning. If a student tests his or her memory shortly after studying, he or she is most likely to recall the material studied last than material from the middle of the test. This effect is known as the recency effect (Baddeley & Hitch, 1993). Furthermore, taking one test after another stimulates participants to use and rely on their short-term memory. A considerable amount of empirical evidence has shown that learning is enhanced when it is spaced apart rather than massed together (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Karpicke & Roediger, 2007; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). One of the most extensive studies that empirically measured the effect of frequent testing with different time intervals is the study by Spitzer (1939). In this classic experiment, Spitzer asked the entire sixth-grade student population from 91 elementary schools in Iowa (i.e., 3605 students) to read two articles (Article A and Article B). After reading Article A, children took Test A and then read Article B. Next, based on their condition, students took Test B, each condition being defined by the number of repeating Test B and when it was taken. The results of this study,
shown in Figure 1, revealed that using a combination of frequent tests with some time intervals and proctoring the initial test shortly after learning seems to be the most advantageous method. For example, students who took Test B on Day 21 after practicing it twice already (i.e., Condition 1) recalled about twice as more as students who took Test B on Day 21 for the first time (i.e., Condition 6).

**Figure 1.** Proportion of correct answers on multiple-choice Test B taken at different days after initial learning as reported by Spitzer (1939).
The implementation of additional time between tests resembles a more realistic classroom environment where students take tests with some time intervals. Learning about the benefit of spacing tests, I designed my new study accordingly. Instead of two unique practice tests delivered on the same day, in my current work I used three different assessments; each being delivered on different days. Another limitation of my previous work was the unbalanced material. The material of my current work includes an equal number of algorithmic, conceptual, and definition questions. This modification to the material allowed me to investigate whether testing mode equally affects all chemistry question types or if it is predominant in one or more, and not others. I present these results in Chapter 3.

For my third article, I measured the cognitive load (Sweller, Ayres, & Kalyuga, 2011) and use of scratch paper while students took multiple tests in different modes to shed more light on differences between the two modes (i.e., Chapter 4). Currently, little is known about the role of cognitive load in online testing. It is still unclear whether the testing mode interferes with learning and imposes unnecessary cognitive load on students. To remedy this problem, I assessed cognitive load using subjective rating scales of student mental effort (Paas, 1992) and level of difficulty (Ayres, 2006) for each item. After students completed their third assessment, I collected, coded, and analyzed the scratch paper.

Therefore, the present body of research extended my previous work with the following four research goals:

1. To replicate testing mode findings of my master’s work
2. To explore testing mode using more authentic classroom settings
3. To study testing mode using balanced general chemistry material
4. To investigate testing mode differences using cognitive load and student use of scratch paper

The list of research questions that I derived from my previous work and addressed in the current work is highlighted in Figure 2. The results of this work are particularly relevant to instructors who currently teach or are considering teaching undergraduate general chemistry courses in blended or online formats.

Figure 2. List of research questions that emerged out of Prisacari (2015) work and their categorization by chapters that comprise the current work.

In summary, (1) I repeated my previous work on testing mode; (2) I studied testing mode under conditions similar to classroom environment (3) with balanced material, while (4) investigating the testing mode in conjunction with cognitive load and scratch paper. All studies were approved by the Iowa State University Institutional Review Board and informed consent was obtained from all participants (Appendix A).
Organization of the Dissertation

This dissertation consists of five chapters: this introductory chapter, three journal articles, and a concluding chapter. Next, I introduce each article, indicating which out of four research goals it accomplishes and my role in each project. Article 1 is presented in Chapter 2, Article 2 – Chapter 3, and Article 3 – Chapter 4 of this dissertation.

**Article 1. Comparing student performance using computer and paper—based tests. Results from two studies in general chemistry**

Chapter 2 is a paper that was submitted to the *Journal of Chemical Education*. The central goal of this article was to replicate my previous findings, that is, finding a significant difference between Paper – Computer and Paper – Paper conditions (Goal 1). Therefore, this paper presents two studies aimed at uncovering potential differences in online versus paper testing in the context of general chemistry. The first study (Study 1) represents my master’s work and the second study (Study 2) is the replication of Study 1. Data for Study 1 were collected in fall 2013 ($N = 207$) and data for Study 2 were collected in the same class in fall 2015 with new students ($N = 215$). All students took two practice tests in one of four conditions that were identified by the testing mode of Test 1 and Test 2: Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper. Each test included 17 algorithmic, 5 conceptual, and 2 definition questions. The list of all items is shown in Appendix B. The results of both studies showed that the effect of taking the first practice test on computer and the second practice test on paper (Computer – Paper) appears to be indistinguishable from the condition in which students took both practice tests on paper (Paper – Paper).
For this work I recruited participants, scheduled and proctored the testing sessions, coded and entered data, performed data analysis, and wrote this article. The test material was written and practice tests were graded with help from Dr. Baluyut, Dr. Brigham, Dr. DeSilva, Dr. Linenberger, Dr. Luxford, Dr. Raker, and Dr. Reed. Dr. Holme assisted me in gathering additional data from the American Chemical Society Examinations Institute (ACS-EI). Both co-authors, Dr. Holme and Dr. Danielson, provided comments on initial draft and helped me to address the reviewers’ comments.

**Article 2. Rethinking testing mode: Should I offer my next chemistry test on paper on computer?**

Chapter 3 is a paper that was published in the *Computers & Education* journal in 2017. In this article, I present a new study that addresses the limitations of my previous work. That is, I compare student performance on three assessments (i.e., two proctored quizzes and one practice test) taken in paper-based or computer-based testing mode using an equal number of algorithmic, conceptual, and definition questions. The design of this study helped me to examine the effect of testing mode using a more authentic classroom setting (Goal 2) and with balanced chemistry material (Goal 3). Each assessment was delivered on computer or paper and was proctored on a different day. All participants ($N = 221$) took Quiz 1 on Day 1, Quiz 2 on Day 8, and one practice test on Day 12. The list of all chemistry items is presented in Appendix C. Both quizzes represented required class assignments so they were proctored in the actual chemistry class (Appendix D and Appendix E), while the practice test was voluntary and took place outside the class. Since the content of Quiz 1 and Quiz 2 was different, but the practice test included both Quiz 1 and Quiz 2 content, I formed the testing mode conditions based on the mode combination of Quiz 1 – practice test and Quiz 2 – practice test. All students took one quiz on computer and
one quiz on paper; this design allowed me to assign two conditions to each student. For example, if a student took Quiz 1 on computer, Quiz 2 on paper, and the practice test on paper, then his or her test performance would fall into two conditions: Computer – Paper for Quiz 1 content and Paper – Paper for Quiz 2 content. The results showed that students in one testing mode condition did not perform significantly better than students in other conditions.

To make sure the material of this study aligns well with the material of the course which set the authentic classroom setting for this study, I closely worked with the course instructor and teaching assistants. I designed the study, wrote the material for the study with assistance from Dr. Appy, Dr. Baluyut, Dr. Reed, and Gauri Ramasubramanian. Both the class instructor and the lead teaching assistant reviewed all material before the study. Before the initial data collection, I attended the meeting with all teaching assistants and provided instructions on how to proctor and grade the quizzes. Logan Fischer helped me to proctor the practice test and I graded the practice test. I coded, formatted, and analyzed data and wrote this article. Dr. Danielson provided input regarding the study design, provided comments on the initial draft and helped me to address the reviewers’ comments.

**Article 3.** *Computer-based versus paper-based testing: Investigating the testing mode with cognitive load and scratch paper use*

Chapter 4 is a paper that was submitted to *Computers in Human Behavior* journal. This paper explored testing mode with cognitive load and scratch paper (Goal 4). Chemistry students \((N = 221)\) completed three chemistry assessments. Each assessment was delivered on computer or paper. To measure cognitive load, students reported their mental effort and level of difficulty after answering each question. To analyze the optional use of scratch paper, I collected student
scratch paper after the third assessment (as it was the longest test out of the three assessments) and analyzed it for completeness (Appendix F and Appendix G). The results on cognitive load showed no significant difference in the cognitive load measures between the computer and paper modes at the overall test level or by question type. For scratch paper, students used this resource more during a paper-based test than online test, especially when the questions were conceptual.

For this study, I designed the study, recruited the participants, performed all analyses, and wrote the paper. The first two assessments were proctored by the teaching assistants and the third assessment was proctored by Logan Fischer and me. The data were entered with the help from teaching assistants. I scored all scratch paper and Logan Fischer served as the second rater who independently scored only scratch paper used by students in the computer-based condition. Dr. Danielson provided input regarding the study design, and reviewed the initial draft before this article was submitted to the journal.

In summary, this dissertation presents findings from several studies on testing mode conducted in general chemistry settings with chemistry students and material. Results are presented in three journal articles: (1) testing mode studied with Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper conditions and two practice tests taken in one session, (2) testing mode studied with the same four conditions, new and balanced chemistry material, and three assessments, each given on a different day, and (3) testing mode explored with cognitive load and scratch paper.
References


CHAPTER 2: COMPARING STUDENT PERFORMANCE USING COMPUTER AND PAPER—BASED TESTS. RESULTS FROM TWO STUDIES IN GENERAL CHEMISTRY

A paper submitted to the Journal of Chemical Education

Anna A. Prisacari†, Thomas A. Holme*†‡, Jared Danielson†

†Human Computer Interaction, Iowa State University, Ames, Iowa 50011, United States
‡Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States
*Corresponding author

Abstract

Taking a test online rather than on paper is becoming increasingly common. However, there has been little research directly addressing the testing mode (taking a test on paper or online) in chemistry courses, particularly when students take multiple practice tests before an exam. Two studies were conducted to investigate student performance on two proctored general chemistry practice tests as a function of the testing mode. Data were collected in 2013 (i.e., Study 1) and again in 2015 (i.e., Study 2). The participants were 422 undergraduate students (Study 1 $N = 207$ and Study 2 $N = 215$) from a first-semester general chemistry at a Midwestern university. In each study students took two practice tests. Each test included 17 algorithmic, 5 conceptual, and 2 definition questions and was administered on computer or paper. The mode combination of Test 1 – Test 2 identified the four conditions: Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper. The results show minor differences between online and paper modes. In particular, no significant difference was found between Computer – Paper and Paper –
Paper conditions. This pattern suggests that online testing is a promising alternative to the traditional paper-and-pencil mode most often used in chemistry.

Graphical Abstract

Keywords
First-Year Undergraduate/General, Testing/Assessment, Chemistry Education Research

Introduction

Recently it has become more common to assess students’ learning in online settings. For the last several years, the American Chemical Society Examinations Institute (ACS-EI) has been working to design online modes of its tests. In addition to norm-referenced exams, ACS-EI offers several student practice tests online such as full-year general chemistry, full-year organic chemistry, analytical chemistry, and first-term general chemistry. With such an increase in computer-based assessment this study was undertaken to examine the equivalency of paper-based and computer-based general chemistry practice tests by measuring and comparing performance of students who take the same practice test in one of two modes. The results of this
study may provide specific insights into how to deliver general chemistry practice tests so student learning from them is maximized and be of interest to instructors who are considering or already using online platform to test their students.

**Online mode in chemistry**

As technology becomes more affordable and powerful, computer-based activities in chemistry classroom settings provide powerful means for fostering student learning\(^1,2\) and enhancing the quality of assessment.\(^3\) While teaching chemistry has traditionally been done in face-to-face classroom settings using hands-on lab experience and paper-and-pencil methods for assessing student learning, recent research shows alternative and reasonably effective strategies to teach chemistry and deepen student understanding by using technology. For example, computer-based animations and simulations can help students learn abstract concepts.\(^4,5\)

The opportunities to assess student learning have also changed with the development of new tools that offer unique advantages over traditional, paper-based tests\(^6\). Notably, the growing body of research regarding testing mode, which typically involves comparing student test performance on a paper-based test to test performance on the same test delivered online, has often shown mixed findings. In addition to the mixed results, these studies have three important limitations. First, the effect of the testing mode has not been explored with multiple tests. Is there a testing mode order effect? Even though students routinely take multiple tests to illustrate their learning in a class, little empirical data has been gathered regarding whether taking one test on paper and the next test online is as effective as taking both tests only via paper or online. Investigating the testing mode order would help to construct a more complete picture about technology infused student learning. Such information is essential to instructors who teach
blended classes or have opportunities to assess their students using either paper-based or computer-based testing modes. In addition, chemistry instructors would benefit from knowing in which mode to deliver practice tests and exams in order to maximize student performance.

Second, testing mode has not been explored in detail in the context of general chemistry. Efforts to evaluate the comparability of a paper-based and computer-based modes have been demonstrated with material from disciplines like anatomy,7,8 biology,9,10 English,11,12 mathematics,10,13 and reading,14 but these studies produced mixed results. While some studies report no significant differences between computer-based and paper-based test,7,8,15–18 other studies report that two modes do not produce equivalent test performance19 and differences generally favor the paper mode.11,13,20 Because many undergraduate students take general chemistry courses to fulfill their degree requirements, it would be worthwhile to examine whether the effect of the testing mode is present in a general chemistry setting. Lastly, Cumming21 suggests that “a single study is rarely, if ever, definitive; additional related evidence is required” to increase precision and robustness of the original work. To our knowledge, a direct replication of testing mode study has never been tested. Therefore, two years after the initial study, we replicated it with new students to ensure reliability of the results.

Purpose of the Study

The goals of this study were to examine whether (a) the testing mode and the testing mode order are related to student performance on assessments in a general chemistry practice exam setting and (b) the results of the initial study can be replicated with new students. To investigate the testing mode and its order, general chemistry students took two proctored practice tests in one of four conditions that were defined by the testing mode of the initial practice test
and the second practice test. The conditions were: Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper. Course-relevant material included items classified as algorithmic, conceptual, and definition questions, question types that are common in general chemistry. Different question types allowed us to examine the testing mode not only at the broad level, but also at the level of individual items, based on their type. Using the settings that are common to undergraduate general chemistry tests, all sessions were timed, proctored, and conducted in classrooms. To examine whether the testing mode results can be replicated, the experiment was repeated after a two year delay with new students using the same design, material, and classroom setting. Thus, the data were collected twice: in 2013 (i.e., Study 1) and 2015 (i.e., Study 2) fall semesters. The research questions of Study 1 and Study 2 were:

1. Is there a difference in test performance for students across the four testing mode conditions?
2. Is there a difference in test performance for algorithmic and conceptual questions based on the mode of tests?

Study 1

Methods

Participants

A total of 221 students were recruited from a first-semester general chemistry at a large Midwestern university. However, three students were excluded for the following reasons: one student left in the middle of the session and two students experienced some issues with their laptops that prevented their data from being fully recorded, resulting in 218 students as original sample size. Results from eleven more students were removed because they had answered some component set of the test items, such as algorithmic, or conceptual questions correctly in their
first practice. Such a score on the first test precludes a student demonstrating any gain from repeated practice for that type of question. Thus, the final sample size contained 207 students (male = 37.2%, female = 61.4%, no response = 1.4%; mean age = 19). The course from which students were recruited lasted 15 weeks and included four sessions a week: three 50-minute lectures and one 50-minute recitation that was led by a senior undergraduate or graduate student. Notably, in addition to receiving the opportunity to practice before final course examination, all students were given complementary access to the online ACS general chemistry practice test for participating in the project. The study was approved by the Institutional Review Board (IRB) and the consent form was obtained from all participants at the beginning of the session (see procedure section).

Materials and design

Because the final exam of this course was either the 2012 (in Study 1) or 2015 (in Study 2) ACS first-semester general chemistry exam, the practice test items were constructed using items that had items constructs and content coverage commensurate with an ACS Exam. The items were not taken from any released ACS Exam. To compose the test items, we used six areas that are typically taught at the first-semester general chemistry university level. For convenience, these areas will be referred to as “chapters” in the subsequent descriptions of the project. As depicted in Figure 1, each chapter contained four topics. For example, the four topics of chapter on matter were quantum numbers, periodicity, atomic structure, and isotopes. For each topic, two closely related multiple-choice questions were composed, resulting in four question pairs for each chapter or 24 question pairs in total (see Figure 1 where each cell represents a question pair). Since students took two practice tests, one question of each pair was displayed on Test 1 and another question on Test 2. Thus, each practice test contained 24 questions and each test displayed different questions on the same set of topics. An example of a complete 24-item test is
included with the Supplementary Material. We illustrate an example of a “related pair” of multiple choice items here, where the following pair of questions was composed to test knowledge of lone pairs electrons (i.e., Chapter 3, Topic 3):

Question A

When the correct Lewis structure is drawn for acetylide ion, \( \text{C}_2^2^- \), what is the total number of electron lone pairs present?

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

Question B

When the correct Lewis structure is drawn for acetylide ion, \( \text{C}_2^2^- \), what is the total number of bonding electron pairs present?

A. 0       B. 1       C. 2       D. 3
As chemistry students are often tested with multiple-choice and open-ended questions, half of the questions were displayed to students in multiple-choice format and half in open-ended format. Open-ended versions of all items (algorithmic, conceptual, and definition) were composed by removing the answer options from multiple-choice items. We counterbalanced the material so all topics appeared equally often in multiple-choice and open-ended format. Differences based on item construct (open-ended versus multiple-choice) were observed and were consistent with prior work that generally finds open-ended items are more difficult. A detailed description of how the material was counterbalanced, and discussion of the item construct findings are presented in other publication. Because this component of the study is largely a replication of

**Figure 1.** List of 17 algorithmic (white), 5 conceptual (green), and 2 definition (yellow) question pairs by 6 chapters. Each chapter contains 4 topics and each topic includes 1 question pair that is composed of 2 questions similar in content and level of difficulty.
earlier work, graphical summary of the item construct results are included only in Supplementary Material.

Lastly, all questions were classified by five experts into three categories: algorithmic, conceptual, and definition. Figure 1 shows the distribution of the six chapters and their topics, indicating 17 algorithmic, 5 conceptual, and 2 definition question pairs. The material was presented on either paper or computer, and students were given a particular form depending on the counterbalanced condition to which they were assigned. All students were assigned to one of four conditions (see procedure section for more information) that were identified by the testing mode of Test 1 and Test 2: Computer - Computer ($N = 22\%$), Computer - Paper ($N = 21\%$), Paper - Computer ($N = 33\%$), and Paper - Paper ($N = 24\%$) (see Table 1). While some students experienced only one testing mode (i.e., Computer – Computer and Paper – Paper), other students took two tests in different modes (i.e., Computer – Paper and Paper – Computer). This design allowed for examining not only the differences in test performance by the testing mode, but the effect of testing mode order as well.

**Procedure**

Students in the general chemistry class were invited to participate in this study, after the class had covered most of the content. The value of participation was described only in terms of preparation for the final exam. No course points were offered for participation. Students who were interested in participating received a link that contained a scheduling interface that listed multiple options of when they could take the practice tests and served as a registration process. The days and times for each condition were determined prior to student recruitment but not disclosed to students. Thus, students signed up for conditions based on their time preference rather than the testing mode preference. After registration was closed, all students received an email that notified them of the location of the study, included a copy of the consent form, and
listed items that they needed to bring with them to the session. All students were asked to bring a basic calculator, scratch paper, and a pencil. In addition to these items, students in the conditions that included at least one online practice test, were instructed to bring a laptop. To replicate the setting where testing commonly occurs, all sessions were conducted in classrooms and were proctored by the main researcher.

All practice tests were designed to mimic timed tests that the students would take as their final exam in the course. Before the session started, the researcher distributed a copy of the periodic table and a general chemistry data sheet that contained some general chemistry formulas, because these materials were permitted during the class’s final examination. After signing the consent form, students received instructions according to their condition. For example, for the paper-based test, students were informed when to open the test and at which page to stop whereas for the online-based test, students received instructions regarding how to log on. All online tests were delivered via Qualtrics software and were protected by a password. The password was given to students before Test 1 and could be used only during the session time, which did not permit students to share password with other students nor retake the test after the session. The timeline of the student experience of each session is shown in Figure 2.

![Timeline of the student experience during the study.](image)

After receiving the printed package or activating Test 1, students were given 35 minutes to complete Test 1. The proctor announced when there were ten minutes remaining. After Test
students received feedback in the form of correct answers and feedback was given in the same testing mode as Test 1. If a student took Test 1 online, then the correct answers along with the student’s submitted answers appeared on the screen. If the student took Test 1 on paper, then he or she needed to flip to the feedback page to see the list of the correct answers and flip the pages of the test back and forth to compare the answers. Students were allowed to review their test performance for several minutes and were instructed to begin next test only after the proctor confirmed that all students in the session had finished their review. After Test 1, all students repeated half of Test 1 (i.e., 12 questions; time = 10 minutes) and again were provided with feedback. This step in the design was included to assess memory factors that were reported on elsewhere.

Next, all students participated in a 20-minute distractor task that involved answering five, non-chemistry related trivia questions. This task allowed students to take a short break and also served as a check on the impact of short-term recall of answers to Test 1. Following the distractor task, students took Test 2 and after completing it, were given the feedback as in Test 1. Finally, students completed a short demographics survey after which the researcher collected students’ scratch paper and paper-based test or verified the online submission, gave access to the ACS online practice test, and thanked and dismissed the students.

All sessions took place 1-15 days before the course’s comprehensive exam, were conducted in groups of 3-33 students, and lasted approximately 1.5-2 hours.

**Scoring**

Using a rubric, each question was scored as 1 (i.e., correct) or 0 (i.e., incorrect) by the principal investigator and several chemistry graduate students and post-doctoral associates. Next, using proportion of correctly answered questions of Test 1 and Test 2 individual normalized gains were calculated for overall performance on the practice exams. In addition individual normalized gains were determined for the subsets of items characterized as algorithmic and
conceptual. Definition questions were excluded from the subset calculation because with only two such questions many students answered both correctly on the first practice test and could not demonstrate gains in this category as a result.

Results

To test if the assumptions of ANOVA have been met, several tests were performed. The Levene's test for homogeneity of variance between conditions showed a non-significant result (i.e., \( p\text{-value} = .95 \)) and gains appeared to show a reasonably normal distribution in the histogram and Normal Q-Q Plot. Because this project was conceived to look at student performance gains due to practice tests, a few students who scored 100% either on Test 1 or a particular question type set from Test 1 (for question type analysis) were removed from the analysis where they could not show any gain. Thus, the final sample size contains students who have overall, algorithmic, and conceptual gains. Table 1 shows the original and final sample sizes by four conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample size N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
</tr>
<tr>
<td>Computer - Computer</td>
<td>48 (22%)</td>
</tr>
<tr>
<td>Computer - Paper</td>
<td>45 (21%)</td>
</tr>
<tr>
<td>Paper - Computer</td>
<td>73 (33%)</td>
</tr>
<tr>
<td>Paper - Paper</td>
<td>52 (24%)</td>
</tr>
<tr>
<td>Total</td>
<td>218 (100%)</td>
</tr>
</tbody>
</table>

The \( \alpha \)-level for all analyses was set at .05 and all results were depicted graphically with confidence intervals instead standard errors, as suggested by Cumming. First we examined the effect of testing mode on the overall gains. To answer the first research question, a one-way
analysis of variance (ANOVA) was performed with conditions as the independent variable and overall normalized gains as the dependent variable. The results, \( F(3, 203) = 7.47, p = .0001, \eta^2 = .1 \), indicated a significant difference among the conditions and a medium effect\(^{30}\). Post-hoc comparisons using the Tukey HSD test indicated that student gain in the Paper - Paper condition \((M = .32, SD = .30)\) was significantly higher than the gain in the Computer - Computer condition \((M = .12, SD = .29)\), the gain in the Computer - Paper condition \((M = .27, SD = .33)\) was significantly higher than the gain in the Paper - Computer condition \((M = .08, SD = .30)\), and the gain in the Paper – Paper condition \((M = .32, SD = .30)\) was significantly higher than the gain in Paper – Computer \((M = .08, SD = .30)\). The results are displayed in Figure 3.

![Figure 3. The boxplots for overall gains by conditions for Fall 2013.](image)

To answer the second research question, two ANOVA analyses were performed by grouping items as algorithmic and conceptual and considering gain score differences in the four
previously defined conditions for each group. For the algorithmic gain, results showed significance, $F(3, 203) = 3.72, p = .01$ and a small effect size ($\eta^2 = .05$). The post-hoc comparisons showed that Paper – Paper gain ($M = .31, SD = .42$) was significantly higher than Paper – Computer gain ($M = .06, SD = .42$). The mean algorithmic gains are presented in Figure 4.

Figure 4. The boxplots for algorithmic item gains by conditions for Fall 2013.

Similar to algorithmic gain, analysis of conceptual gain revealed a significant effect of the testing mode with a small effect, $F(3, 203) = 2.85, p = .04, \eta^2 = .04$, but the post-hoc Tukey test showed no significant differences among conditions ($p > 0.05$) (see Figure 5). While the F test is significant, the Tukey test may not always show significance, because it is a conservative test that attempts to control the overall alpha level. These results in our initial study suggest
that although the effect of testing mode was found to be significant at the overall level with gains for both algorithmic and conceptual items, the difference was small, effectively arguing that the testing mode did not produce meaningful changes in gains. These 2013 results are intriguing, so a replication study with a new cohort of general chemistry students was carried out in fall 2015.

Figure 5. The boxplots for conceptual item gains by conditions for Fall 2013.

Study 2

Methods

Prior to student recruitment, a total sample size was calculated to determine how many students were needed to detect any differences in gains among testing mode conditions. First, we estimated the means of overall gain for different conditions (Computer – Computer = .12; Computer – Paper = .27; Paper – Computer = .08; and Paper – Paper = .32) and error variance (i.e., .09) of Study 1. Using these estimates we computed the required sample assuming 5%
significance level and 80% power in STATA 13.1. The required minimum sample size was 108. Since dropout was anticipated (i.e., not all participating students may complete the study successfully), the required sample size was inflated from 108 to 122 or 31 participants per condition using the following formula: \( n_d = n/(1 - R_d)^2 \), where \( n_d \) is the inflated estimated sample size, \( n \) is the original estimated sample size, and \( R_d \) is the rate of dropout that was calculated using Study 1 data (out of 221 students who participated, only 207 were included in the analysis = 6% dropout rate).

A total of 250 new students were recruited from students taking the same general chemistry course 4 semesters later than those who participated in the initial study. However, 7 students were not included in the analyses (3 students were under the age of 18, 1 student forgot to charge their laptop and thus had to switch the condition in the middle of the session, 1 student did not follow the instructions correctly, 2 students left in the middle of the session). Thus, the original sample size was 243. Next, 28 students were removed as they answered either all test questions correctly, all algorithmic questions, or all conceptual questions correctly on Test 1. Therefore, the final sample size was 215 students (male = 35%, female = 59%, no response = 6%; mean age = 18.6). None of these students participated in Study 1 and the only benefit they received was the additional opportunity to practice before the final exam, in other words students in the replication study did not receive an access code to the online ACS practice exam in addition to the preparations arising from the participation in the project. Student sample size by condition is shown in Table 2. All sessions took place 3 - 7 days before the course’s comprehensive exam and were conducted in groups of 58 - 73 students. The material, design, procedure, and scoring were the same as in Study 1.
Results

Before conducting analyses, ANOVA assumptions have been verified. The Levene’s test for homogeneity of variance between conditions showed a non-significant result (i.e., $p$-value = .26) and the histogram and Normal Q-Q Plot graphs of gains appeared to show a normal distribution. By contrast to Study 1, the results of overall normalized gains showed no support for the testing mode ($F(3, 211) = 2.05, p = .1077$) as illustrated in Figure 6.

Table 2. Original and final sample size distribution among four conditions of Study 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sample size $N$ (%)</th>
<th>Original</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer - Computer</td>
<td>57 (23.5%)</td>
<td>52 (24%)</td>
<td></td>
</tr>
<tr>
<td>Computer - Paper</td>
<td>57 (23.5%)</td>
<td>55 (26%)</td>
<td></td>
</tr>
<tr>
<td>Paper - Computer</td>
<td>56 (23%)</td>
<td>48 (22%)</td>
<td></td>
</tr>
<tr>
<td>Paper - Paper</td>
<td>73 (30%)</td>
<td>60 (28%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>243 (100%)</td>
<td>215 (100%)</td>
<td></td>
</tr>
</tbody>
</table>
Next, analyses were performed to identify gains for only the algorithmic and conceptual item sets. For algorithmic gain, the results revealed no significant difference among conditions $F(3, 211) = 1.23, p = .3001$ (Figure 7), while significance was detected with gain of conceptual questions $F(3, 211) = 3.67, p = .01, \eta^2 = .05$ (Figure 8). The post-hoc comparisons showed that
Paper – Paper gain ($M = .48, SD = .60$) was significantly higher than Computer – Paper gain ($M = .07, SD = .76$).

**Figure 7** The boxplots for algorithmic item gains by conditions for Fall 2015.
Figure 8. The boxplots for conceptual gains by conditions for Fall 2015.

Discussion

Students face learning in computerized environments in many contexts including chemistry. The first chemistry massive open online course, or MOOC was offered in 2012 and since then many chemistry education researchers have explored the landscape of online environment and its effect on student learning. Even though technology offers unique opportunities to support assessment activities in science, little is known whether students taking online assessments in general chemistry can produce similar test performance as students who take the same test in traditional, paper-and-pencil mode.

An interest in the relative impact of computer-based assessment derives from well-known learning effects, such as the testing effect. In this case, the investigations were carried out in a
low-stakes environment using a practice test methodology. Because the students self-selected and were practicing for an upcoming class exam, however, they were likely to be sufficiently motivated to perform well on the practice tests. Not surprisingly, performance gains were always present when comparing Test 1 to Test 2. The results of Study 1 in 2013 showed the presence of testing mode effects, indicating the benefit of taking the second test on paper regardless of the testing mode of Test 1. The highest gains were observed in Computer - Paper and Paper - Paper conditions and no significant difference was found between these two conditions. This study suggested that, although a combination of online and paper testing can be attractive to both instructors and students, the testing mode order, or which test mode comes first, might be worth consideration in designing study activities. The results showed a significant difference between Paper - Computer and Computer - Paper conditions, suggesting that student test performance would be enhanced if students take the first practice test online and the next one on paper. Further analyses based on categorizing the question type revealed that the testing mode differences were potentially attributable to performance gains on conceptual questions. Finally, it would be ideal to include student performance on the final exam in this study, but all students in the class had a variety of preparation tools available in addition to these practice tests, so within the design parameters of the study, it is not possible to make meaningful comparisons and attribute them to a testing effect for participants.

Study 1 was replicated with 215 new general chemistry students two years later (e.g., Study 2). Importantly, while some differences were detected among student performances in Study 1 with a medium effect, no testing mode effect was found in Study 2. The fact that the second study, in 2015 showed no significant effects associated with testing mode, while the earlier study detected the testing mode effect is worth further investigation. Ultimately, there are
a number of factors that are likely to influence student performance differences such as students’ prior knowledge. It is possible that students in one study were better prepared for the tests than students in other study. To test this hypothesis, we compared the mean performance on the initial test and found that students in Study 2 ($M = .67$) significantly outperformed students on test 1 in Study 1 ($M = .59$). Thus, this group appears to have been better prepared. Because we used a convenience sample, it is difficult to attribute this difference to any other external factor such as instructors or content coverage of the two courses. Nonetheless, prior research suggests that the effect of testing may be moderated by students’ prior knowledge.\textsuperscript{38–40} Future research could investigate further whether testing mode differences are more predominant with students who are less prepared versus students who are better prepared for general chemistry tests.

Limitations

While this study is informative regarding the effect of testing mode on student performance in general chemistry, it has some limitations. First, the participants were self-selected students who might be different from those students who chose to not participate. Second, data for some students were excluded from the analyses as normalized gains could not be calculated due to their perfect scores on the initial test. This was particularly problematic when calculating performance gains for items categorized as definition items. Because there were only two definition questions, a high percentage of students (34\% - Study 1 and 65\% - Study 2) answered both questions correctly, so gain scores could not be defined for those students. If student performance on definition style items is a concern for chemistry instruction, a higher number of items would be needed to study this type of question in subsequent studies. Third, the material contained an unequal number of questions in each question type category.
Out of 24 questions, there were 17 algorithmic, 5 - conceptual, and 2 - definition questions. While this distribution of questions was a good representation of type of questions general chemistry students typically see on their final exam, it complicates investigations as to whether the testing mode equally affects all chemistry question types or it is predominant for one type of item more than others.

Finally, the time period between two practice tests was short. Students took both tests on the same day with a 20-minute time interval. Such design does not reflect well the authentic classroom setting as students do not typically take back to back tests that cover the same content. Inserting an additional time gap between learning activities (in this case a practice exam) represents a more common teaching strategy and, as shown by previous research, enhances long-term learning more than back-to-back or massed learning. Known as the spacing effect, it represents the desirable difficulty, or better condition of learning, for learners. The more difficult the retrieval task, the more it benefits learning. If the material is easily available and still accessible in working memory, as is the case in massed learning, the individual only needs to review this information. However, when some time interval is inserted between the initial learning and testing, some deactivation or forgetting occurs. Accordingly, the person must regenerate the full retrieval process and thus recall becomes more challenging. Because the spacing process requires time to process new material, during which the student rehearses and connects new material to knowledge already stored in his or her long-term memory, spaced practice works better than multiple learning sessions that take place one after another. The experiments by Karpicke and Roediger showed that when participants recalled the material two days after the study session, delaying the initial test increased the difficulty of retrieval and, consequently, boosted student performance on the final test. To study ways to maximize student
learning, future studies on the testing mode should extend the time interval between the tests. That is learning activities should be separated by a period of at least 12 hours that include a night sleep. Such studies are more challenging to recruit student participants, but would likely show effects unachievable with the shorter term experiments described here. The literature on distributed practice clearly suggests that separating learning activities by one day rather than conducting all learning on the same day aids students to remember content for an extended period of time.

Implications

The results of our studies have interesting educational implications for chemistry instructors. Frequent and distributed tests represent an effective instructional method and the work presented here suggests that testing methods can be used with either online or paper testing modes. Although university instructors tend to adopt technology at a slow rate, this study indicates that online practice tests, which may be easier to administer and grade, can work as an instructional method. Giving a practice test online and delivering the final exam on paper, represents a good alternative to assessing students’ knowledge with only the traditional paper-and-pencil mode. Using the mix of online and paper modes would allow instructors and students to derive benefits of both modes without compromising students’ test performance. For instance, the online mode would permit instructors to effortlessly document and store student test scores for future content evaluation and comparison purposes, whereas it would provide instant feedback to students, a feature that chemistry students have mentioned as their top reason to prefer online to paper mode. Giving the last test on paper would permit chemistry students to
take the test in what is commonly their preferred testing mode and work things out on paper,
another common feature that was previously reported by chemistry students.\textsuperscript{28}

Author information

Corresponding Author
*E-mail: taholme@iastate.edu

Acknowledgments

Funding for this project was provided by the ACS Exams Institute as part of its research into the
development of online testing methods. We would like to thank chemistry instructors who allowed recruiting their general chemistry students and students who were part of this study. Also, we would like to thank John Baluyut, Jennifer Brigham, Chamila DeSilva, Kimberly Linenberger, Cynthia Luxford, Jeffrey Raker, and Jessica Reed for their help with material writing and test grading.

References


(31) Simon, S. When the F test is significant, but Tukey is not


Supplemental information

Comparing Student Performance Using Computer and Paper—based Tests. Results from Two Studies in General Chemistry

By

Anna A. Prisacari†, Thomas A. Holme*‡, Jared Danielson‡
†Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States
‡Human Computer Interaction, Iowa State University, Ames, Iowa 50011, United States

Item format effects

The impact of item format, multiple-choice versus open ended items, was also tracked in this study. Each student had six items where the item format was multiple-choice on both test 1 and test 2; six items with multiple-choice on test 1 and open-ended on test 2; six items with open-ended on test 1 and multiple-choice on test 2; and six items with open-ended for both items. The specific items in each case were randomized and counter-balanced. The results of item format are summarized in the following four figures.

Figure S1 – Computer-Computer Condition Results: Items that have open-ended responses in test 2 show less gain, or even a decrease in student performance, while items that have multiple-choice responses in test 2 show stronger performance increases.
**Figure S2 – Computer-Paper Condition Results:** All item formats show increased performances on test 2 when it is taken on paper. Performance increases are lower, however, when the item format on test 2 is open-ended.

**Figure S3 – Paper-Computer Condition Results:** Items that switch from multiple-choice to open-ended responses in test 2 show a decrease in student performance, while items that have multiple-choice responses in test 2 show stronger performance increases.
Figure S4 – Paper-Paper Condition Results: Performance increases are seen for all item format conditions, with open-ended on test 1 and multiple choice on test 2 showing the largest performance gain.

Example test (Areas for open response answers are removed for space reasons.)

1. Which quantum number primarily dictates the energy of the atomic orbital?
   A. \( n \)  
   B. \( \ell \)  
   C. \( m_\ell \)  
   D. \( m_s \)

2. Which noble gas atom has the largest radius?

3. An atom of strontium–90 (\(^{90}\)Sr) contains how many electrons, protons, and neutrons?

4. Indium has two naturally occurring isotopes, \(^{113}\)In and \(^{115}\)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. \(^{113}\)In and \(^{115}\)In are present in nearly equal amounts
   B. \(^{113}\)In is present in much greater amounts than \(^{115}\)In
   C. \(^{113}\)In is present in much smaller amounts than \(^{115}\)In
   D. Cannot be determined without more information about the mass of each isotope
5. What is the formula of chromium(III) carbonate?
   A. \( \text{Cr}_3\text{CO}_3 \)      B. \( \text{Cr(CO}_3)_3 \)      C. \( \text{Cr}_2(\text{CO}_3)_3 \)      D. \( \text{Cr}_3(\text{CO}_3)_2 \)

6. 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?

7. 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

8. What is the net ionic equation for the reaction?
   \( \text{NaCl} + \text{AgNO}_3 \rightarrow \text{NaNO}_3 + \text{AgCl} \)

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

10. 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

11. When the correct Lewis structure is drawn for acetylide ion, \( \text{C}_2^{2-} \), what is the total number of electron lone pairs present?
    A. 0      B. 1      C. 2      D. 3
12. What is the molecular shape of dichloromethane?

13. How many moles of Fe are needed to produce 10.0 mol of H₂?

\[ 4 \text{H}_2\text{O}(g) + 3 \text{Fe}(s) \rightarrow \text{Fe}_3\text{O}_4(s) + 4 \text{H}_2(g) \]

| A. 7.50 mol | B. 13.3 mol | C. 15.0 mol | D. 30.0 mol |

14. How many moles of Al₂O₃ can be produced from 27 g of aluminum and 16 g of oxygen?

<table>
<thead>
<tr>
<th>Atomic molar masses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
</tr>
<tr>
<td>O</td>
</tr>
</tbody>
</table>

15. What mass of MgCl₂ is required to prepare 2.00 L of 0.550 M solution?

<table>
<thead>
<tr>
<th>Molar mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgCl₂</td>
</tr>
<tr>
<td>95.2 g/mol</td>
</tr>
</tbody>
</table>

| A. 1.10 g | B. 26.2 g | C. 28.9 g | D. 105 g |

16. Balance this equation and then answer the following question:

\[ ? \text{S} + ? \text{HNO}_3 \rightarrow ? \text{H}_2\text{SO}_4 + ? \text{NO}_2 + ? \text{H}_2\text{O} \]

How many moles of H₂O are formed per mole of sulfur consumed?

17. 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?

18. 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?
19. The partial pressures of a gaseous mixture are given in the table. What is the mole fraction of hydrogen in the mixture?

<table>
<thead>
<tr>
<th>Partial Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
</tr>
<tr>
<td>CO₂</td>
</tr>
<tr>
<td>CH₄</td>
</tr>
<tr>
<td>C₂H₄</td>
</tr>
</tbody>
</table>

A. 0.135  B. 0.194  C. 0.258  D. 0.413

20. A student collected 40.0 mL of N₂ gas when the temperature was 20.0°C and the pressure was 720. torr. The next day the temperature was still 20.0°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

21. What is the final temperature when a 21.5 g sample of gold (specific heat = 0.129 J·g⁻¹·°C⁻¹) emits 233 J of heat when it cools from 125°C?

22. Calculate ΔH° 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

\[ \text{H}_2(g) + \text{Cl}_2(g) \rightarrow 2\text{HCl}(g) \quad \Delta H° = -185 \text{ kJ} \]

\[ \text{H}_2(g) + \frac{1}{2} \text{O}_2(g) \rightarrow \text{H}_2\text{O}(g) \quad \Delta H° = -242 \text{ kJ} \]

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

23. 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
24. Given the bond energies in the table, what is the $\Delta H^\circ$ for the chemical reaction, $\text{Cl}_2(g) + \text{F}_2(g) \rightarrow 2\text{ClF}(g)$?

<table>
<thead>
<tr>
<th>Bond Energies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl–Cl</td>
</tr>
<tr>
<td>F–F</td>
</tr>
<tr>
<td>Cl–F</td>
</tr>
</tbody>
</table>
CHAPTER 3. RETHINKING TESTING MODE: SHOULD I OFFER MY NEXT CHEMISTRY TEST ON PAPER OR COMPUTER?

A paper published and reproduced with permission from *Computers & Education*, 2017, 106, pg 1 – 12
DOI: 10.1016/j.compedu.2016.11.008

Anna Agripina Prisacari*, Jared Danielson

*Corresponding author
Email addresses: annacari@iastate.edu (A. A. Prisacari), jadaniel@iastate.edu (J. Danielson)

Abstract

The purpose of this study was to compare student performance on two proctored quizzes and one practice test taken in paper-based or computer-based testing mode using an equal number of algorithmic, conceptual, and definition questions to examine the effect of testing mode when students take multiple tests in class. Data were collected from 221 students enrolled in a general chemistry course at a Midwestern university. After learning the material in lecture, students completed two quizzes: Quiz 1 and Quiz 2. One quiz was delivered on paper and another quiz was taken on computer. Each quiz tested student understanding of different chemistry concepts and the quizzes were proctored on different days. Several days after Quiz 2, students took a
practice test that tested the knowledge of Quiz 1 and Quiz 2 with different yet similar questions, and again, the testing mode of the practice test was either paper-and-pencil or computer. After each quiz and the practice test, students received feedback on their performance. Differences in performance between the quizzes and the practice test were measured with normalized gains and the differences between the normalized gains for each condition were analyzed using an Analysis of Variance. The results showed no significant testing mode effects among the four conditions overall, or for algorithmic, conceptual, or definition type of questions. Altogether, the results of the present study do not provide evidence to suggest that instructors need to be concerned about testing mode (paper versus computer) when designing and administering chemistry tests.

Keywords: improving classroom teaching; postsecondary education; teaching/learning strategies

Introduction

New technologies are quickly gaining popularity in education. Today’s incoming college freshmen use technology on a larger scale than ever before. In 2014, 33 US states offered full-time virtual schools, enrolling close to 262,000 students (Miron & Gulosino, 2016) and several US states required their high school students to engage in online learning (e.g., United States. State of Michigan., 2006; United States. State of Florida., 2015). As students become more accustomed to using computers for educational purposes, many instructors begin to teach their courses in online or blended formats that may result in transferring all or some of the class instruction and testing activities from face-to-face to online platforms. For example, according to the director of the Testing Center at Iowa State University, the number of exams proctored online in the Testing Center each semester has increased from 2,176 in Fall 2003 to 83,963 in Fall 2015

For the purposes of the present study, computer-based testing refers to any test taken via any kind of computer (e.g., desktop computer, laptop, tablet). Computer-based tests are frequently delivered via a network (online testing) though testing is possible via computers that are not connected to any network. While computer-based testing is gaining popularity in schools, some educators question whether changing from traditional paper-and-pencil testing to computer-based testing is effective and offers benefits for student learning. Computer-based testing offers several unique benefits to instructors (e.g., reducing human error associated with grading and embedding interactive media) and students (e.g., receiving feedback quickly), yet some empirical studies show some convincing support for paper-based methods. In the next section, we review literature that directly addresses differences in paper-based versus computer-based testing.

**Paper-based versus computer-based testing**

The testing mode refers to the method of delivering a test to students (e.g., on paper or computer) and the testing mode effect refers to differences in student performance between tests given in different modes. One strategy for examining the testing mode effect is to compare students’ test performance on a test taken on paper to performance on the same test taken on a computer. Studies making this comparison have produced mixed results, with some studies detecting no significant difference between computer-based and paper-based test results, (Hochlehnert, Brass, Moeltner, & Juenger, 2011; Horkay, Bennett, Allen, Kaplan, & Yan, 2006; Meyer, Innes, Stomski, & Armson, 2015; Tsai & Shin, 2012), and other studies identifying
significant differences (Bennett et al., 2008; Clariana & Wallace, 2002; Goldberg & Pedulla, 2002; Keng, McClarty, & Davis, 2008).

Some researchers have explored the testing mode effect by comparing computer-based versus paper-based test performance on two tests that were administered at different times. For example, in their testing mode comparability study, Chua and Don (2013) found that when students took the test only once, there was no significant difference between computer-based and paper-based performance. However, when students took the same test twice, the mean performance on the final test was significantly higher for the computer group ($M = 73.06$) than for the paper group ($M = 63.44$). These findings suggest that the testing effect (i.e., learning as a result of taking a test) (Dirkx, Kester, & Kirschner, 2014; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Karpicke, 2006) may be a factor that contributes to differences found between test performance of a computer and paper test. However, Chua and Don investigated the testing mode by repeating one test twice, whereas in actual classroom students are typically assessed with different tests.

Recently, Prisacari (2015) investigated the testing mode effect using two different chemistry practice tests. Each test was administered on paper or computer and both tests were proctored in one session with a 20-minute gap between tests. The practice tests were given in one of four conditions defined by the mode of the first test and the mode of the second test. Four conditions (First Test – Second Test) were: only computer (Computer – Computer), only paper (Paper – Paper), and a combination of computer and paper modes (Computer – Paper and Paper – Computer). The results of this study showed that when students took two back-to-back chemistry practice tests, increase in student performance from the initial test to the second test was the highest when students took the final test on paper regardless of the testing mode of the
first test. However, this study had two limitations that are important to consider. One limitation is that both tests were proctored on the same day. Since students are often exposed to multiple tests taken over days or weeks in a class, the question of whether the testing mode effect of multiple tests taken on different days contributes anything to learning is worthy of further investigation.

The implementation of additional time between tests resembles a more realistic classroom environment where students take tests with time intervals such as several days or weeks. This practice of spreading out learning activities such as taking tests with some time intervals is known as the spacing effect, which suggests that learning is enhanced when recall is spaced apart rather than massed together (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Karpicke & Roediger, 2007; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). Similar to the testing effect, the spacing effect represents a desirable difficulty, because massed practice leads to better performance on a test that immediately follows learning, but spaced practice leads to better long-term retention.

The second limitation of Prisacari's (2015) study is that each test contained three types of questions in unequal numbers: 17 questions were algorithmic, 5 questions were conceptual, and 2 questions were definitions. The results of this study have suggested that testing mode effect may be more prevalent in tests that include more algorithmic than conceptual or definition questions. Therefore, the question remains whether testing mode is present in tests that contain equal numbers of algorithmic, conceptual, and definition questions.

The present study

The main purpose of the present study was to explore the testing mode in more authentic settings (e.g., when a period of time elapsed between tests) and with a variety of question types.
First, the study explored whether student performance varies between multiple chemistry paper-based and computer-based tests taken on different days as part of a general chemistry class. Instead of giving practice tests on the same day, this study used three different assessments delivered on different days. All participants took two quizzes (i.e., Quiz 1 on Day 1 and Quiz 2 on Day 8) and one practice test on Day 12 (see Figure 1). This design allowed students to have several days between any two assessments, which could be used to review the content and feedback before the next test. Given the power of the spacing effect, students’ test performance may show pronounced testing mode effects when students take tests on different days instead of taking all tests on the same day.

Second, the testing mode (computer-based versus paper-based testing) was investigated using tests with equal numbers of three question types that are common to chemistry: algorithmic, conceptual, and definition (Smith, Nakhleh, & Bretz, 2010). In this study, we used material that included six algorithmic question pairs, six conceptual question pairs, and six definition question pairs (see section 2.3). A combination of all three question types was presented to students on each assessment. On Quiz 1, students saw 2 algorithmic, 4 conceptual, and 4 definition questions; on Quiz 2 students saw 4 algorithmic, 2 conceptual, and 2 definition questions; and on the practice tests students saw 6 algorithmic, 6 conceptual, and 6 definition questions. Each assessment was administered using a paper mode or computer mode. Controlling the number of questions allowed us to investigate whether testing mode is present in one or more, and not other question types. The present study, therefore, sought to answer two questions. The first question was which testing mode (only on paper, only on computer, or using a combination of two testing modes) would benefit learning in undergraduate general chemistry students. The second question was whether the testing mode would affect student test
performance when questions were only algorithmic, conceptual, or definition. In other words, is there a benefit to using a particular testing mode for a test that contains more algorithmic, conceptual, or definition questions?

Method

Participants

The participants in this study were students enrolled in an undergraduate General Chemistry course at a large Midwestern university. Out of 452 students enrolled in the course, 286 students registered to participate, 237 participated in the study, and 221 students were included in the final data analysis. Out of the 16 removed from the analysis, 2 students experienced some problems with their laptops during the session and completed the test partially on computer and on paper, 1 student forgot to bring a laptop and took all tests on paper, 9 students did not complete all tests, a test of one student was not graded correctly, and data from 3 students were removed as outliers. When the means with extreme cases and without them are very different, one recommendation is to remove outliers (Pallant, 2010). Due to the difference in means (with outliers: Quiz 1 gain = -0.13 and Quiz 2 gain = -0.18; without outliers Quiz 1 gain = -0.12 and Quiz 2 gain = -0.12), it was decided to exclude the outliers. Outliers were identified with studentized residual, Cook’s distance, and leverage measures (Field, 2009) using STATA 13 statistical software (StataCorp, 2013). Students were recruited via class announcements by the principal researcher and course instructor, emails, and reminders were posted on Blackboard. All participants received extra credit for participating in this study and gained an opportunity to practice for an upcoming test.
**Design**

In undergraduate general chemistry at this university, each week students attend three, 50-minute long lectures taught by the instructor and one, 50-minute long recitation facilitated by a student teaching assistant. While students attend lectures together, they are divided in 22 smaller groups (12 - 23 students) for recitations. During the recitation, the teaching assistant reviews the content of that week, answers students’ questions, and proctors a short quiz. Overall, in this class students are tested with 10-15 minute weekly quizzes in recitation, four one-hour tests, and one comprehensive final exam at the end of the semester. Each quiz tests student knowledge on the content that was covered within a week in class and each test focuses on the material covered in class from the last test, but may include some material covered for a previous test. For example, the second test could include content that appeared on the first test and material that was covered after the first test. This study was designed to prepare students for their third test and was conducted in the middle of the semester. Data were collected in two recitations that occurred after the second test and before the third test. Figure 1 illustrates the order of all assessments (two quizzes and one practice test) that were part of this study, the time period between any two assessments, and whether any assessment was a required part of class or voluntary.

**Days on which data were collected for this study**

| Day 1 | • Quiz 1  
Required |
| Day 8 | • Quiz 2  
Required |
| Day 12 | • Practice Test  
Optional |
| Day 15 | • Class Test  
Required |

**Fig. 1.** The timeline of two quizzes, practice test, and class test. Data from class test were not collected; the time of class test was added to the timeline only to illustrate the overall time period.
All students completed Quiz 1 and Quiz 2 during class recitation and the practice test was scheduled outside the class. Each assessment was delivered on paper or computer. The mode of Quiz 1 and Quiz 2 was randomly assigned before the study and the mode of the practice test was dictated by students’ time preference. About half of students (11 recitations; 227 students) were randomly assigned to take Quiz 1 on paper and Quiz 2 on computer while the other half of students (11 recitations; 225 students) took Quiz 1 on computer and Quiz 2 on paper. Since the practice test was voluntary, all interested students signed up to take the test either at Time 1 or Time 2, without knowing whether it was on paper or computer. This way, students signed up to take the practice test at the time that best fit their schedule and without selecting a particular mode. The practice test was composed of Quiz 1 and Quiz 2 content. This design allowed for creating four conditions defined by the mode (computer or paper) of each quiz and the practice test. The four conditions were: (1) Computer – Computer (Quiz 1 or 2 on computer and the practice test on computer), (2) Computer – Paper (Quiz 1 or 2 on computer and the practice test on paper), (3) Paper – Computer (Quiz 1 or 2 on paper and the practice test on computer); and (4) Paper – Paper (Quiz 1 or 2 on paper and the practice test on paper). If a student took Quiz 1 on computer, Quiz 2 on paper, and the practice test on paper, then his or her test performance would fall into two conditions: Computer – Paper for Quiz 1 content and Paper – Paper for Quiz 2 content (see Fig. 2). Thus, the testing mode of a quiz and the practice test was manipulated between subjects.
Fig. 2. Visual representation of four conditions defined by the computer or paper testing mode of Quiz 1/Quiz 2 and the practice test (A). Four conditions are: Computer – Computer, Computer – Paper, Paper – Computer, and Paper – Paper. If a student received Quiz 1 on paper, then he or she took Quiz 2 on computer. Each student participated in two conditions. For example, if a student took Quiz 1 on computer, Quiz 2 – paper, and the practice test on paper, then the student’s two conditions were Computer – Paper for Quiz 1 and Paper – Paper for Quiz 2 (B).

Material

The material was based on content covered in class and recitations that students needed to know for their next class exam. Overall, 18 question pairs covered nine topics. Since each topic is commonly tested using a specific question type, the topics determined the ratio among algorithmic, conceptual, and definition questions. Therefore, the number of algorithmic, conceptual, and definition questions varied by quiz. For example, Quiz 1 contained more conceptual questions, while Quiz 2 had more algorithmic questions. However, the practice test contained an equal number of questions of each type. Table 1 illustrates 18 question pairs and their distribution between Quiz 1 and Quiz 2 by topics and question types.
Table 1. Study material by quiz, topics, and question types.

<table>
<thead>
<tr>
<th>Quiz</th>
<th>Topics</th>
<th>Question Types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Definition</td>
</tr>
<tr>
<td>Quiz 1</td>
<td>Matter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intermolecular forces</td>
<td>Pair 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 7</td>
</tr>
<tr>
<td></td>
<td>Nonmolecular substances</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Balancing chemical equations</td>
<td>Pair 8</td>
</tr>
<tr>
<td></td>
<td>Types of reactions</td>
<td></td>
</tr>
<tr>
<td>Quiz 2</td>
<td>Solubility and precipitation reactions</td>
<td>Pair 12</td>
</tr>
<tr>
<td></td>
<td>Net ionic equation</td>
<td>Pair 13</td>
</tr>
<tr>
<td></td>
<td>Acid-base reactions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stoichiometry</td>
<td>Pair 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pair 17</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6 pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 pairs</td>
</tr>
</tbody>
</table>

Material covered in class that students were expected to know for their third class test was used to write questions. A total of 18 multiple-choice question pairs were composed by the course instructor and principal researcher. A question pair (Question A and Question B) consisted of two different questions that were based on the same topic and similar level of difficulty (see Table 2). For example, Pair 1 for Quiz 1 (i.e., Question A1) and practice test (i.e., Question B1) was:
Question A1

Which are condensed phases of matter?

i) solid    ii) liquid    iii) gas

a) i only
b) i and ii
c) ii and iii
d) All are condensed phases
e) None are condensed phases

Question B1

In which phase(s) are particles able to move around (i.e., not locked in place)?

a) Gas only
b) Liquid only
c) Solid only
d) Gas and liquid
e) Liquid and solid

The first question of each pair was administered during a quiz and the second question was administered during the practice test. All questions A appeared on the quizzes (i.e., questions A1 – A10 appeared on Quiz 1 and questions A11 – A18 appeared on Quiz 2) and all questions B (i.e., B1 – B18) appeared on the practice test. Questions B1 - B10 corresponded to
Quiz 1 content and questions B11 - B18 corresponded to Quiz 2 content. The total number of questions in this study was 36: 10 questions in Quiz 1, 8 – Quiz 2, and 18 – practice test (see Table 2).

Questions for each quiz and the practice test were composed by the course instructor and principal researcher and pre-tested by the course instructor and lead teaching assistant to make sure students would be able to answer all questions within 15 minutes for quizzes and 60 minutes for the practice test. Since students are given 60 minutes to complete a class test that may include both multiple-choice and open-ended questions, additional changes were made to the content of the practice test. The original practice test contained 18 questions, which could be answered in under 60 minutes. To replicate the test environment, three additional open-ended questions were added to the practice test. However, student performance on additional questions was excluded from the data analysis, because the gain of these questions could not be calculated.

When creating a new test, is important to evaluate its validity and reliability. Before the study began, the face validity, content validity, and level of difficulty of all questions were established by four chemistry experts (two senior chemistry doctoral students, one chemistry professor, and one chemistry post-doctoral researcher). Using experts for validity check is a simple and commonly used process in the chemistry education (Cooper, Underwood, & Hilley, 2012; McClary & Bretz, 2012). We elected not to employ a test-retest approach to estimating the reliability of the tests, because student learning is likely to change from one test to the next. Cronbach’s $\alpha$ was calculated to estimate internal consistency of each quiz, producing the following results: Quiz 1 ($\alpha = 0.46$), Quiz 2 ($\alpha = 0.47$), and Practice Test ($\alpha = 0.43$). While low internal consistency can indicate low reliability, it can also result from other factors.
Table 2. List of 18 question pairs. Questions A1 - A10 represented the material for Quiz 1; A11 - A18 for Quiz 2; and B1 - B18 for the practice test. Questions within a pair were different questions yet on the same topic.

<table>
<thead>
<tr>
<th>Content</th>
<th>Question Pairs</th>
<th>Quiz Questions</th>
<th>Practice Test Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quiz 1</strong></td>
<td>Pair 1</td>
<td>A1</td>
<td>B1</td>
</tr>
<tr>
<td></td>
<td>Pair 2</td>
<td>A2</td>
<td>B2</td>
</tr>
<tr>
<td></td>
<td>Pair 3</td>
<td>A3</td>
<td>B3</td>
</tr>
<tr>
<td></td>
<td>Pair 4</td>
<td>A4</td>
<td>B4</td>
</tr>
<tr>
<td></td>
<td>Pair 5</td>
<td>A5</td>
<td>B5</td>
</tr>
<tr>
<td></td>
<td>Pair 6</td>
<td>A6</td>
<td>B6</td>
</tr>
<tr>
<td></td>
<td>Pair 7</td>
<td>A7</td>
<td>B7</td>
</tr>
<tr>
<td></td>
<td>Pair 8</td>
<td>A8</td>
<td>B8</td>
</tr>
<tr>
<td></td>
<td>Pair 9</td>
<td>A9</td>
<td>B9</td>
</tr>
<tr>
<td></td>
<td>Pair 10</td>
<td>A10</td>
<td>B10</td>
</tr>
<tr>
<td><strong>Quiz 2</strong></td>
<td>Pair 11</td>
<td>A11</td>
<td>B11</td>
</tr>
<tr>
<td></td>
<td>Pair 12</td>
<td>A12</td>
<td>B12</td>
</tr>
<tr>
<td></td>
<td>Pair 13</td>
<td>A13</td>
<td>B13</td>
</tr>
<tr>
<td></td>
<td>Pair 14</td>
<td>A14</td>
<td>B14</td>
</tr>
<tr>
<td></td>
<td>Pair 15</td>
<td>A15</td>
<td>B15</td>
</tr>
<tr>
<td></td>
<td>Pair 16</td>
<td>A16</td>
<td>B16</td>
</tr>
<tr>
<td></td>
<td>Pair 17</td>
<td>A17</td>
<td>B17</td>
</tr>
<tr>
<td></td>
<td>Pair 18</td>
<td>A18</td>
<td>B18</td>
</tr>
</tbody>
</table>
According to Streiner (2003) $\alpha$ can be low, even with reliable tests, when one or more of the following conditions are present: (1) tests measure number of items completed in a specific period of time; (2) the items are presented in order of difficulty; (3) the answer to one item depends on the answer to a different item; or (4) the test is multi-dimensional. In our case, each test was multi-dimensional, measuring a variety of concepts from chemistry across conceptual, algorithmic, and definitional questions. Therefore, we elected to compare scores from two similar groups as described by Cooper et al. (2012) and Bretz & McClary (2015). We performed a Kruskal-Wallis test to determine if responses from multiple recitations were different. Students from all recitations had the same instructor and were exposed to the same instruction. However, as each assessment was administered with different test modes, students were divided into four groups accordingly (as described in the design section). Table 3 summarizes the $p$-values for all groups, indicating no significant difference in scores across different recitations that took all assessments in the same mode ($p$-values > 0.05). The expert review, coupled with the comparison between similar groups makes a reasonable case for the validity and reliability of Quiz1, Quiz 2, and the practice test.
Table 3. List of p-values for four groups from Kruskal-Wallis test to determine reliability. Groups were identified by the testing mode of each quiz and practice test. For example, Group I included 58 students from ten recitations who received Quiz 1 on computer, Quiz 2 on paper, and practice test on computer.

<table>
<thead>
<tr>
<th>Group</th>
<th>N and # of sections</th>
<th>Quiz 1</th>
<th>Quiz 2</th>
<th>Practice Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Computer</td>
<td>Paper</td>
<td>Computer</td>
</tr>
<tr>
<td>1</td>
<td>58 (10)</td>
<td>0.46</td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td>2</td>
<td>50 (10)</td>
<td>0.78</td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>3</td>
<td>54 (11)</td>
<td>0.11</td>
<td>0.46</td>
<td>0.16</td>
</tr>
<tr>
<td>4</td>
<td>59 (11)</td>
<td>0.47</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

All computer assessments were delivered via Qualtrics and were protected by one-time only passwords that could activate the quiz or test only during a specific time and for a particular recitation. Since 11 recitations took Quiz 1 on computer on Day 1 and 11 recitations took Quiz 2 on computer on Day 8, unique passwords were generated for each recitation and day. This protocol restricted students from opening and beginning the assessment before the assigned time, accessing one assessment multiple times, opening any of the assessments after their assigned time, or using the password of one session to open the same quiz for a different session. If students did not come to their recitation, they were not able to open any online assessment and, consequently, received no points for the quiz that was proctored that day.

To match the visual appearance between paper and computer assessment, online questions were separated by page breaks and for the paper quizzes/practice test each question was printed on different page. This way each page contained one question and student had to either click a button or flip the page to move to the next question.
Procedure

Since teaching assistants were in charge of teaching recitations and conducting weekly quizzes, five days before Quiz 1, all 13 teaching assistants received quiz instructions from the principal researcher and were informed which of their upcoming two quizzes would be conducted on paper and which quiz would be on computer. Also, five days before Quiz 1, the principal investigator attended all classes and announced the study to students. Three days before Quiz 1, all students were notified via email about the mode of their quizzes so students could know on which day they needed to bring a laptop to their recitation. A reminder email was sent to students one day before each quiz.

Quiz 1

A few hours before Quiz 1, students who were assigned to take it on computer received individual links to their quiz. This email was sent before class so the proctor could verify at the beginning of the recitation whether students have received the link. If a student did not receive the link, the researcher had time to resend the link to the student before the quiz time. Since the quiz required a password, students were not able to begin the quiz before their assigned recitation time. This process ensured that all students had access to the quiz when it was time to take it. At the beginning of each recitation, the researcher delivered an envelope to the teaching assistant that contained the password for his/her recitation and extra copies of paper quizzes in case some students forgot to bring their laptop or experienced some technical problems. To verify their Quiz 1 performance, three days after taking Quiz 1 students received an email that contained the list of Quiz 1 questions along with their submitted answers. The answers to all questions were posted by the instructor on Blackboard a few days after the quiz, a method that instructor used for all quizzes. For recitations that took Quiz 1 on paper, the researcher delivered the envelope
that contained only paper quizzes. All paper quizzes were graded and returned to students during
the next recitation (i.e., Day 8). Thus, all students had a copy (emailed report or printed test) of
their submitted answers and could review their performance.

### Quiz 2
A week after Quiz 1, students took Quiz 2. Students who took Quiz 1 on paper received
Quiz 2 on computer and students who took Quiz 1 on computer took Quiz 2 on paper. Students
and teaching assistants received the Quiz 2 instructions in a similar manner as Quiz 1. The links
were emailed to students a few hours before Quiz 2 and passwords or paper quizzes were
delivered by the researcher at the beginning of each recitation. Student performance for the
online administration of Quiz 2 was emailed to students after three days and paper copies of Quiz
2 were returned to students during the next recitation. The answers to Quiz 2 questions were
posted on Blackboard after all students completed the quiz.

### Practice Test
Six days before the practice test, students received an email with the link to register for
the practice test. Since the practice test was composed in the style of the class exam, students
may have been motivated to participate and take the practice test seriously as preparation for the
class exam. The practice test was administered on Day 12 at two different times. One session
was allocated for the paper mode and one session for the computer mode. During the registration
students were asked to sign up only for one session without knowing which session was paper or
computer. After registering for the practice test, students received a confirmation email that
indicated the mode of the practice test and included a copy of the consent form. At the beginning
of the practice test, all students received a copy of the periodic table and solubility table and were
asked to sign the consent form. Students who forgot to bring their own scratch paper also 
received blank paper as scratch paper. Next, the proctor announced that students had 60 minutes 
to complete the test, students would be notified when there were ten minutes remaining, students 
were encouraged to use the entire time and verify their answers if they were to finish early, and 
students would receive an answer key after the test. No students completed the test very quickly 
and instead took the time to answer the test questions thoroughly as seen by student notes that 
were collected with the test; this may indicate that students approached this test seriously and 
made an effort to perform well. After students completed the test, the proctor provided a printed 
answer key to each student. Students who took the practice test on paper checked their 
performance by reviewing their paper practice test and students who took the practice test on 
computer received a copy of their answers via email immediately after submitting the test. 
Lastly, when students were done verifying their test performance, they were asked to return the 
periodic table, answer key, scratch paper, and the practice test (only paper session) to the proctor. 
The practice test took place in a large chemistry classroom and was proctored by the principal 
researcher and one graduate teaching assistant from the class.

Data analysis

Paper quizzes were scored as correct (1) or incorrect (0) by class teaching assistants and 
online quizzes and all practice tests were scored by the principal investigator. Since each student 
was tested twice on the same material, we used Hake's formula (1998) of normalized gain 
instead of mean scores. The original formula was modified; instead of calculating the gain from 
the averages of the tests, we calculated normalized gain for each student (Bao, 2006). This 
modification allowed us to analyze average gains at the overall test level and by each question
type and examine the pattern of individual gains (see Fig. 7). Normalized gains were calculated as follows. Four proportions correct were calculated: two for content of Quiz 1 and two for Quiz 2 content for each student (see Table 2). Quiz 1 content included questions from Quiz 1 (questions A1 - A10) and questions from the practice test (questions B1 - B10). Quiz 2 content included questions from Quiz 2 (questions A11 - A18) and the second part of the practice test (questions B11 - B18). Using these proportions, overall normalized gains were calculated for each student twice: for Quiz 1 content and Quiz 2 content. Using this method, we calculated each student’s normalized gain score between pretest and posttest in each of the four conditions (Computer - Computer, Computer - Paper, Paper - Computer, and Paper - Paper) and for each question type (algorithmic, conceptual, and definition). The modified formula for calculating normalized gain of Quiz 1 and Quiz 2 content was:

\[
\frac{\text{Practice Test questions } B_1 - B_{10} - \text{ Quiz 1}}{1 - \text{ Quiz 1}} \quad \frac{\text{Practice Test questions } B_{11} - B_{18} - \text{ Quiz 2}}{1 - \text{ Quiz 2}}
\]

If a student scored 0.4 on a quiz and 0.9 on the practice test, then his or her normalized gain would be \((0.9 - 0.4) / (1 - 0.4) = 0.83\), suggesting that the student improved his or her performance by 83% out of all possible to-be learned content. Normalized gain, or \(g\), can be positive or negative and be classified as low \((g < 0.3)\), medium \((0.3 < g < 0.7)\), and high \((g > 0.7)\). One limitation using \(g\) is a potential risk of reducing the original sample size. If students
receive a perfect score on the initial test, their gain cannot be calculated and therefore, the gains of these students cannot be used for statistical analyses. Measuring student learning using normalized gains instead of means has been used in a variety of educational studies including biochemistry (Ojennus, 2015), chemistry (Bretz & McClary, 2015; Michinov, Morice, & Ferrières, 2015), engineering (Abdul, Adesope, Thiessen, & van Wie, 2016), geoscience (Anderson & Libarkin, 2016), and physics (Marshman & Singh, 2016). The advantage of using normalized gain is that it takes into account initial test performance and normalizes pre- to post-test gain.

After calculating normalized gain scores for each condition and for each question type, we classified each student into one out of four conditions based on his or her combination of the testing mode of two quizzes and the practice test. Since students took one quiz on paper and another on computer, each student was assigned to two conditions: a condition by the testing mode combination of Quiz 1 and Practice Test and another condition by the testing mode combination of Quiz 2 and Practice Test. For example, a student could be placed in the Computer – Paper condition for Quiz 1 – Practice Test and Paper – Paper condition for Quiz 2 - Practice Test (see Fig. 2B). The between-subject independent variable were the testing mode conditions and the dependent variable was the change in performance from a quiz to the practice test (i.e., normalized gain). Next, we checked that our data satisfied all statistical assumptions for parametric techniques (Pallant, 2010) and then we compared gains among conditions and question types using a series of analyses of variance (ANOVA). All ANOVA tests were performed twice: once for the content of Quiz 1 and once for the content of Quiz 2 with STATA 13.1.
Results

First we report the overall gains and then detail the results of gains for each question type.

Overall gains

Since 14 students received 100% on Quiz 1 and 7 students on Quiz 2, their normalized gain scores could not be calculated and data of these students were excluded from analysis. Thus, the sample size for Quiz 1 normalized gain was 207 and for Quiz 2 it was 214 students.

Overall, students produced a negative gain: the proportion of correct responses on the quizzes were higher than on the practice test. Figure 3 shows the average overall gains for Quiz 1 and Quiz 2 as a function of testing mode condition. The results of a one-way ANOVA revealed no statistically significant difference among conditions for Quiz 1, $F(3, 203) = 0.36, p = .78$ nor for Quiz 2 content $F(3, 210) = 0.95, p = .42$.

Algorithmic gains

After omitting students who answered all algorithmic questions correctly on Quiz 1 or Quiz 2, sample size was adjusted to 112 students and 199 for Quiz 1 and Quiz 2 respectively. While students improved their performance on Quiz 1 material, three out of four conditions had a negative gain on Quiz 2 material (see Fig. 4). Again, there was no significant effect of testing mode on gains, $F(3, 108) = 0.96, p = .41$ for Quiz 1 content and $F(3, 195) = 2.49, p = .06$ for Quiz 2 content, indicating that changes in test performance did not significantly differ among the four conditions.
Conceptual gains

One hundred and eighty nine students were included in this analysis for Quiz 1 and 183 students represented the sample size for Quiz 2. Although students in the Computer - Paper condition were the only group that exhibited a negative gain for conceptual questions, the ANOVA showed no significant effect of testing mode on performance of conceptual questions for Quiz 1 $F(3, 185) = 1.85, p = .14$ or for Quiz 2 $F(3, 179) = 0.47, p = .71$ (see Fig. 5).

Definition gains

Based on the initial performance on definition questions, 149 students and 139 students were included for Quiz 1 and Quiz 2 analysis respectively. As Figure 6 shows, all gains were positive. However, the ANOVA of definition gains indicated an $F(3, 145) = 1.41, p = .24$ for Quiz 1 and $F(3, 135) = 1.03, p = .38$ for Quiz 2. As in the previous analyses, the testing mode for chemistry definition questions was not significant.

Fig. 3. Mean normalized overall gains by conditions for Quiz 1 (N = 207) and Quiz 2 (N = 214). Error bars represent 95% confidence intervals.
Fig. 4. Mean normalized algorithmic gains by conditions for Quiz 1 (N = 112) and Quiz 2 (N = 199). Error bars represent 95% confidence intervals.

Fig. 5. Mean normalized conceptual gains by conditions for Quiz 1 (N = 189) and Quiz 2 (N = 183). Error bars represent 95% confidence intervals.
The aim of this study was to compare student performance on two proctored quizzes and one practice test taken on paper or computer for a general chemistry course using algorithmic, conceptual, and definition questions to find further evidence for the advantage of using paper or computer testing modes. This study yielded several interesting results.

One of the benefits of calculating individual gains is the ability to plot individual data and review its pattern. Figure 7 illustrates that more students received a negative gain on at least one quiz (i.e., Q2 - 4) than students who received no or positive gain on Quiz 1 and Quiz 2 material (i.e., Q1). These results suggest that although a subset of students improved their performance on both quizzes, most students showed a decline in performance and this decline was on a wider range (i.e., from -3 to 0) than range of improvement (i.e., from 0 to 1). This observation explains why the mean of Quiz 1 and Quiz 2 gains was also negative despite that more students showed positive gains than negative gains on each quiz. For example, 135 students had a gain equal or higher than 0 (i.e., Q1 and Q2) whereas only 68 students had a gain below 0 on Quiz 1 (i.e., Q3
and Q4). The mean of Quiz 1 and Quiz 2 gains is depicted as a triangle in Q3, the least preferred quadrant.

Results of the overall normalized gains indicate that no significant testing mode effects were found among the four conditions. In other words, computer-based tests may be used to assess general chemistry knowledge without any significant effect on student performance. Using a computer-based test instead of traditional paper-and-pencil test could permit instructors to provide feedback to students on their performance faster, a feature that students often cite as a reason to prefer computer testing mode (Engelbrecht & Harding, 2004; Prisacari, 2015). One possible explanation as to why no difference among testing mode conditions was found is that student familiarity with computers is increasing. As more schools begin to adopt and use technology (e.g., one-to-one laptop initiative), students are becoming more accustomed to using computers for learning purposes and, as a result, taking a test on computer might be becoming more familiar and comfortable for students. It is also possible that other factors influenced this finding. For instance, in the case of the present study the practice test was a low-stakes examination. Perhaps the findings would have been different had all of the tests contributed to the grade. Nonetheless, it is likely that students were motivated to do well on the practice test, as discussed earlier.

Contrary to our expectations, overall gains were negative. A study by Roediger, Agarwal, McDaniel, and McDermott (2011) indicated that students tend to score similarly on a subsequent test or improve their test performance when tests are taken with a time delay of several days or weeks and feedback is provided after the test because multiple effortful tests allow students to identify gaps in knowledge and strengthen learning during future study sessions (Roediger, Putnam, & Smith, 2011). While more students had a positive normalized gain than a
negative gain (Quiz 1 content: \(N = 91\) versus \(N = 69\) and Quiz 2 content: \(N = 93\) versus \(N = 78\)), the range of normalized gains was from -3 to +1. The average normalized gain for the Quiz 1 content was -0.12 and the average normalized gain for Quiz 2 content was -0.11, suggesting that some students answered correctly far more questions on a quiz than on the practice test, which led to average negative gains. For example, for Quiz 1 content five students answered 9 out of 10 questions correctly (i.e., Quiz 1 performance = 0.9) and only 6 out of 10 (i.e., Practice Test questions B1- B10 performance = 0.6), resulting in -3 normalized gain. A possible explanation as to why some students performed better on the quiz than on the practice test could be the time of quiz administration. In this study, students received quizzes at the end of their recitation after reviewing the content with their student teaching assistant. It is possible that some quiz content was reviewed in class so when students were taking the quiz, they were taking advantage of their short-term memory rather than testing their actual understanding of chemistry concepts.

However, the practice test was not preceded by the review session and, thus, could reflect student understanding more accurately. It is also possible that students put more effort into answering the quiz questions than the practice test questions, because the quizzes were graded and the practice test was not.
Fig. 7. Scatterplot of Quiz 1 and Quiz 2 individual gains. Each dot illustrates a combination of gains a student received. Students who had 0 or missing value for at least one gain were excluded (N = 203). Quadrant 1 (Q1) (N = 89) includes students who received Quiz 1 gain $\geq$ 0 and Quiz 2 gain $\geq$ 0; quadrant 2 (Q2) (N = 46) includes students who received Quiz 1 gain $\geq$ 0 and Quiz 2 gain < 0; quadrant 3 (Q3) (N = 29) includes students who received Quiz 1 gain < 0 and Quiz 2 gain < 0; and quadrant 4 (Q4) (N = 39) includes students who received Quiz 1 gain < 0 and Quiz 2 gain $\geq$ 0. A triangle depicts the sample mean of both gains.

The results on algorithmic, conceptual, and definition normalized gains showed that students in one testing mode condition did not perform significantly better than students in other conditions. These findings suggest that no particular combination of paper and computer testing mode leads to higher test performance. The algorithmic normalized gain finding was consistent with Prisacari's (2015) study that also indicated no educational benefit of testing mode for chemistry algorithmic questions. Therefore, the testing mode does not seem to affect the students’ performance on algorithmic, conceptual, and definition chemistry questions.
Limitations and Future Research

One limitation of this study was the self-selected sample. Since only 52% of students chose to take the practice test (i.e., the final data represented 49% as some students were removed), our data may not accurately reflect the entire student sample of this class. Another limitation pertains to the spacing effect. The benefit of spacing retrieval in education has been known for some time (Dempster, 1991). Although the spacing effect was accounted for in the design on this study (i.e., spacing apart quizzes from the practice test), learning was not spaced from testing during recitations. Work by Karpicke & Roediger (2007) suggests that delaying the first test aids long-term learning more than proctoring the first test immediately after learning, because delayed test represents higher difficulty. Scheduling the learning and testing activities so there is a time interval between them is beneficial to long-term memory, especially when the time gap includes sleep (Mazza et al., 2016). Future work should investigate the testing mode under more rigorous conditions, by inserting a period of time between learning (e.g., recitation) and initial testing. That is, instead of administering Quiz 1 at the end of recitation, this quiz should be scheduled for a different day (e.g., conducted at the beginning of next recitation) to maximize learning. This modification to design may result in obtaining positive normalized gains between quizzes and the subsequent test, and aid students to better retain the chemistry content in the long-term.

Another important question is whether the present findings would replicate with different students and different content from chemistry or other discipline. This study focused on content from one portion (approximatively 20%) of one course and it yet remains unknown if similar findings would be reported with other content of this class. Knowing if the testing mode is affected by content can lead to important implications for better understanding the effects of
testing in the educational settings. Therefore, additional research with new and diverse participants and material is desired to support the results of this study.

References


StataCorp. (2013). *STATA*. College Station: TX: StataCorp LP.


CHAPTER 4: COMPUTER-BASED VERSUS PAPER-BASED TESTING: INVESTIGATING THE TESTING MODE WITH COGNITIVE LOAD AND SCRATCH PAPER USE

A paper submitted to Computers in Human Behavior journal

Anna Agripina Prisacari\textsuperscript{a}, Jared Danielson\textsuperscript{b}

\textsuperscript{a}Iowa State University, Human Computer Interaction, 1620 Howe Hall, Ames, IA 50011, USA
\textsuperscript{b}Iowa State University College of Veterinary Medicine, 2256 Vet Med, Ames, IA 50011, USA

Abstract

The aim of the present study was to explore the relationship between testing mode (taking a test on computer versus paper) and two other factors: (1) cognitive load and (2) scratch paper use in an undergraduate general chemistry class setting. Cognitive load was measured with two self-report questions (perceived mental effort and level of difficulty) and scratch paper use was analyzed manually. All 221 students completed three assessments administered either on computer or paper. The assessments included a variety of chemistry topics with an equal number of three question types (algorithmic, conceptual, and definition). There was no significant difference in the cognitive load imposed by computer or paper-based tests at the overall test level or by question type. Students utilized scratch paper more on paper-based than online tests, especially when the questions were conceptual, and they used scratch paper the most for algorithmic questions. Altogether, these results provide further support that online testing can be implemented in educational settings without imposing additional cognitive load on students.
Highlights

- No significant differences were found in cognitive load between two modes.
- Students used scratch paper more on paper than online tests (conceptual questions).
- Students used scratch paper most for algorithmic questions.

1. Introduction

Integration of computers in teaching and learning (Blin & Munro, 2008; Selwyn, 2007) and computer use by students are well discussed and researched topics (Castillo-Manzano, Castro-Nu, Opez-Valpuesta, & Teresa Sanz-Díaz, 2017; Kennedy, Rhoads, & Leu, 2016; Suhr, Hernandez, Grimes, & Warschauer, 2010; Zheng, Warschauer, & Farkas, 2013), particularly in high stakes testing contexts (Meyer, Innes, Stomski, & Armson, 2015; Prisacari & Danielson, 2017). Although testing remains a very common tool for assessing student learning and assigning grades, its features continue to change. Today, students may take a test via content management or other testing systems (e.g., Blackboard, Moodle, WebCT) that permit students to take a test anywhere and anytime and receive a test score almost immediately after submitting the test. However, as testing continues to evolve, new questions emerge. Are online tests as effective as paper-based tests for promoting learning? Are there any differences in test performance between computer-based and paper-and-pencil testing? If yes, what might cause these differences?

Some studies show that students perform differently on paper-based tests than computer-based tests (Bennett et al., 2008; Clariana & Wallace, 2002; Emerson & MacKay, 2011; Kim & Kim, 2013). Yet limited attempts have been made to explain any differences in the testing mode (i.e., administering a test on computer or paper). In this paper, we examine testing mode by proctoring three tests on computer or paper and measuring the cognitive load and use of scratch
paper associated with each testing mode in the context of an undergraduate general chemistry course. If factors such as cognitive load and/or student behavior (such as use of scratch paper) vary between testing modes, instructors might have reason to choose one mode over another. If the two modes produce equivalent cognitive loads and similar testing behaviors, instructors can have more confidence in choosing whichever mode is most convenient in their contexts. Before we describe the study, we present a literature review on each topic.

2. Literature Review

2.1. Cognitive load

One possible explanation for differences in scores between testing modes is that one mode imposes a greater extraneous cognitive load on the test taker than the other testing mode. If one testing mode produces high cognitive load, then working memory may be overloaded, which, in turn, may lead to poorer performance on that test. Originated in the 1980s, cognitive load theory (CLT) is an instructional theory that suggests that the success of learning depends on the availability of sufficient cognitive resources to meet the demands of a particular learning activity, learning being most successful when required and available cognitive resources match (Paas, Renkl, & Sweller, 2003, 2004; Sweller, Merrienboer, & Paas, 1998). In short, when a person experiences a high cognitive load, his or her attention is disrupted and performance on the task that demands attention suffers (Murphy, Groeger, & Greene, 2016). Overall, the total cognitive load consists of intrinsic and extraneous cognitive load (see Figure 1) (Kalyuga, 2011; Sweller, Ayres, & Kalyuga, 2011). Intrinsic cognitive load refers to the fundamental nature or difficulty of the learning task and extraneous cognitive load refers to load that is unnecessary or interferes with learning goals. Since the capacity of human working memory (Baddeley, 2002;
Barrett, Tugade, & Engle, 2004; Engle, 2002) is limited, CLT suggests that learning is obstructed when the resources of working memory are exceeded. Figure 2 provides a graphical representation of two additive categories (intrinsic and extraneous) of cognitive load. The figure illustrates four scenarios using an intrinsic cognitive load that remains unchanged while extraneous load changes in each scenario. When the instructional design is poor and requires a large amount of resources, there may be insufficient working memory resources for intrinsic cognitive load (Figure 2A). Learning may not even begin if the extraneous cognitive load consumes all available resources of the working memory. As extraneous cognitive load is reduced, more resources of working memory become available for learning (Figure 2B). Since intrinsic cognitive load is more relevant to learning and cannot be changed by instructional methods, literature on CLT suggests that extraneous cognitive load should be reduced as far as possible so more resources of working memory can be devoted to learning purposes (Figures 2C and 2D) (Sweller & Chandler, 1994). When total cognitive load exceeds working memory resources, learning may diminish (Figure 2B) or completely be omitted (Figure 2A). If intrinsic cognitive load cannot be modified and its level is high, reducing extraneous load becomes critical to avoid excessive total load. Working memory resources that no longer need to be directed towards extraneous load instead can be utilized by intrinsic load, which optimizes learning.
To study the effect of intrinsic or extraneous cognitive load on learning it is important to keep one category constant while altering the other (Sweller, 2010). For example, investigating the effect of extraneous load, Borges et al., (2016) tested the usability of iGeom, an interactive geometry software, by asking 69 undergraduate students to complete a series of geometry questions (i.e., keeping the intrinsic load constant) using a complete or reduced interface (i.e., varying extraneous load). The complete interface was defined by the presence of as many as possible features whereas the reduced interface included only a subset of features. Since intrinsic load remains unchanged, any differences in cognitive load measures must be due to extraneous load.

Procedures to measure cognitive load have considerably changed over time (Sweller, 1988; Sweller et al., 2011) and have been applied to a variety of environments (Hart, Sandra, 2006). All available independent variables to measure cognitive load can be classified into five categories: indirect measures, subjective measures, secondary task measures, physiological measures, and mixed methods. For a detailed description of each method we recommend Sweller.
et al., 2011. In this section we describe only subjective measures, as those methods were used in the current study, are used commonly, and considered to be reliable (Ayres, 2006; Paas, van Merrienboer, & Adam, 1994; Sweller et al., 2011).

Previous research suggests that tests delivered online may be associated with a higher cognitive load than paper-based tests. For example, in a study by Hochlehnert, Brass, Moeltner, and Juenger (2011) on a scale 1 to 5 students reported a higher additional mental effort when choosing computer-based mode (2.3) than paper mode (1.1; p < .0001). Similar results were reported by Noyes, Garland, and Robbins (2004) who found a significant difference in the perceived effort between two testing modes. Familiarity with paper-based tests and higher

**Figure 2.** Visual representation of why lowering extraneous cognitive load benefits learners. Dotted line represents total working memory resources available to process necessary information. (A) represents a task that imposes high extraneous load and impedes learning to occur; (B) represents a task that imposes less extraneous load, thus allowing some learning to occur; (C) represents a task that imposes just enough resources required for learning, and (D) represents a task that imposes very low extraneous load that leaves more than enough resources required for intrinsic load.
frequency of using paper for testing are some possible explanations of the differences in cognitive load measures. However, contrary to the previously mentioned results, Emerson and MacKay (2011) reported no difference in mean workload stress between the two modes.

Although cognitive load has been studied with testing mode, it is often measured after all tasks are completed in a form of a post-test questionnaire. This approach may not accurately estimate cognitive load while testing, because prior studies have shown differences between one-time (delayed) and immediate cognitive load ratings. Specifically, delayed mental effort (Schmeck, Opfermann, van Gog, Paas, & Leutner, 2015; van Gog, Kirschner, Kester, & Paas, 2012) and delayed difficulty ratings are higher than the average of ratings that followed immediately after each task (Schmeck et al., 2015). This suggests that students’ perception might contribute to differences in scores between immediate and delayed ratings. When providing delayed ratings, students may remember the more difficult problems rather than recalling all problems they received when estimating their cognitive load. Other possible question characteristics that students may take into account when reporting their cognitive load are the total number of problems or the total time taken to complete the test. Finally, any difference in cognitive load between paper and computer-based modes may vary by the type of question being asked. Because a test might include questions that vary in content, format, and type like in our study (see Section 4.3.1.), measuring cognitive load after each question could help shed light on how the testing mode may affect students’ cognitive load as questions change.

The present study contributes to the current literature by measuring the difference in cognitive load after each question. This approach addresses the potential problems of delayed ratings and differences among problem types, addressing that deficiency in the literature.
2.2. Use of scratch paper

During a test, students often use a variety of tools that may help them with answering test questions. For instance, when tests include questions that require some mathematical calculations, students may use a calculator. Another very popular tool that students tend to use during a test is scratch paper. In some instances, instructors may ask students to supplement their answers with hand-written notes students wrote during the test to assign partial points. Although scratch paper has been in use for many years, research on this topic is very limited. When discussing differences in student test performance by the testing mode, a few researchers have argued for the importance of studying the use of scratch paper. For example, when analyzing the difference in performance of a mathematics test taken on paper or online, Bennett et al. (2008) hypothesize that “students may have worked problems more frequently or more thoroughly in the paper booklets than students made use of scratch paper in the online condition.” Similarly, Keng, McClarty, and Davis (2008) pointed out that mathematical questions that involved graphing and geometric manipulation tended to favor the paper mode, which might suggest that students used more scratch paper while answering such questions. Nonetheless, we found no research directly addressing the use of scratch paper.

Theoretically, scratch paper use could have an important impact on the testing process and outcomes, which could vary by testing mode. When taking paper-based tests, students can easily add their notes in the test whereas students who take the test online may need to either reproduce the graph of the question on their scratch paper or use the drawing tools online. It is possible that both of these options add time to complete the test and introduce the possibility of errors (e.g., if students transcribe figures incorrectly). Furthermore, if a student takes a computer-based test, he or she might need to copy problem instruction or draw an image that is presented
on a test; all of which requires students to invest more time and effort. Hence, an understanding of student use of scratch paper is important from a practical perspective; knowing how and to what extent students rely on scratch paper as they take a test on paper versus online may aid instructors to select one mode over the other, especially if the test includes questions that encourage students to use scratch paper (e.g., calculation questions) or if the instructor asks students to submit their scratch paper with the test. Given these practical considerations, whether students taking an exam on paper are more or less likely to use the provided scratch paper than students taking the same test online is a focus of this study.

3. Present Study

This study was part of a wider research program exploring testing mode. While the focus of our previous work (Prisacari & Danielson, 2017) was student test performance among different testing mode conditions, the present study had three unique goals. The first goal was to extend the literature on testing mode by investigating whether testing mode (computer versus paper) produces differences in cognitive load. These two modes were selected because students taking undergraduate chemistry classes have previously reported being familiar with them (Prisacari, 2015). To investigate the testing mode effect, extraneous cognitive load (such as might be imposed by factors specific to testing mode) was measured while intrinsic cognitive load (i.e., the difficulty of questions) remained constant. All students were exposed to the same chemistry questions while taking the test on paper or on laptop. Cognitive load was measured at each testing session with two subjective (self-report) methods: mental effort (Paas, 1992) and perceived difficulty (Ayres, 2006). Such self-report measures are the most frequently used method to measure cognitive load (Naismith & Cavalcanti, 2015). As the material for this
experiment was composed of different chemistry topics and three question types (i.e., algorithmic, conceptual, and definition), we found it appropriate to measure cognitive load after each question instead of at the end of each assessment.

Another goal of this study was to measure the use of scratch paper. Taking a test on computer or paper prompts different student behavior. For example, taking a test online requires students to look up at the computer monitor and down at the scratch paper while taking a test on paper does not involve the process of looking back and forth. Hence, we anticipated each testing mode to produce different cognitive load and reflect different use of scratch paper.

The third goal was to determine whether the three question types differed on measures of cognitive load and use of scratch paper. Hence, we used equal number of three questions types: algorithmic, conceptual, and definition. These question types are very common to general chemistry and are frequently used for measuring student learning (Holme & Murphy, 2011; Prisacari & Danielson, 2017; Smith, Nakhleh, & Bretz, 2010). We predicted that algorithmic questions would reflect a higher use of scratch paper than conceptual and definition questions. We also predicted that we would find some differences in cognitive load by question type, because students may utilize different cognitive resources when answering each question type. For example, for an algorithmic problem students may use a heuristic, a shortcut that they had previously developed, whereas for a conceptual problem the students are likely to use analytical reasoning and thus engage more resources of working memory (Holme & Murphy, 2011). Therefore, this study addressed the following research questions:

- Do students report different levels of cognitive load during test taking when answering questions in one of two modes (computer or paper) and with different question types (algorithmic, conceptual, and definition)?
Do students use scratch paper differently when answering test questions in one of two modes (computer or paper) and with different question types (algorithmic, conceptual, and definition)?

Given the limited availability of research, and lacking a strong theoretical case for either, we initiated this research asserting the null hypothesis - that testing mode and question type would not produce different levels of cognitive load or scratch paper use.

4. Methods

4.1. Participants

The present study was conducted in a general chemistry class (N = 452) taught at a Midwestern university by a single instructor. Data were collected in three sessions: Quiz 1, Quiz 2, and the practice test. It is important to note that Quiz 1 and Quiz 2 data were collected in class and taking these assessments was part of students’ grades. However, data reported here is only from students who signed the consent form and participated in all three sessions including the voluntary practice test.

The participants in this study were 221 students (mean age = 19 years old; 57% - female; 85% - white, 4% - international, 2% - Black or African American, 2% - Hispanic or Latino, 1% - Asian, 1% - two or more races, and 5% - did not report their race). They were recruited by the first author and class instructor. As incentive, all participants received extra credit and were offered an opportunity to practice and receive feedback about content relevant to their course and an upcoming course test. Participants were randomly assigned to take one quiz on computer and one quiz on paper followed by the practice test that again was given on computer or paper. The number of students taking Quiz 1, Quiz 2, and the practice test on paper or via personal laptop
(i.e., computer) is shown in Figure 3. This study was reviewed and approved by the Institutional Review Board (IRB) at the institution where it was conducted and all participants signed a consent form before the study began.

4.2. Design

Students took required Quiz 1 and Quiz 2 on Day 1 and Day 8, and the optional practice test on Day 12. Quiz 1 and Quiz 2 data were collected during two consecutive recitations and therefore were spaced apart by one week. The testing mode of each assessment (Quiz 1, Quiz 2, and the practice test) was determined before the study. Half of class (i.e., 11 recitations) was randomly assigned to take Quiz 1 on computer and Quiz 2 on paper while the second half of the class took Quiz 1 on paper and Quiz 2 on computer. Before the practice test, students registered to take the test at Time 1 or Time 2, not knowing at which time the test would be administered on paper or computer. Only after the registration process was complete, all students were notified whether they would be taking the practice test online or on paper. Such precautions prevented student to select and take the practice test in the preferred testing mode.
4.3. Material

The material of this study consisted of chemistry test questions, two cognitive load questions, and scratch paper. To reduce the testing mode effect, chemistry and cognitive load questions were presented in each mode in a manner that permitted students in both modes to have an equitable experience. For example, only one question followed by two subjective measures was presented per page and students would need to either flip the page (paper mode) or click on a button (computer mode) to move to the next or previous question. This design allowed students to control access of questions.

4.3.1. Chemistry test questions

A total of 18 question pairs were composed for this study. A pair represents two questions that are similar in level of difficulty and test similar knowledge. The first question of
each pair appeared on a quiz and the second question of each pair appeared on the practice test. Thus, after seeing 10 questions on Quiz 1 and 8 questions on Quiz 2, students received 18 questions on the practice test.

Questions were written so there would be an equal number of algorithmic, conceptual, and definition questions on the practice test. Table 1 illustrates material for Quiz 1 and Quiz 2 and shows how the practice test contained an equal number of algorithmic, conceptual, and definition chemistry questions.

Table 1. Material by question type of the study. A pair represents two questions that are similar in topic and level of difficulty. One of these questions appeared on a quiz and another question appeared on the practice test. Quiz 1 included 10 questions, Quiz 2 included 8 questions, and the practice test included 18 questions (10 questions from Quiz 1 and 8 questions from Quiz 2). Although the number of questions and questions per question type varied from Quiz 1 to Quiz 2, the practice test contained an equal number of each question type (6 algorithmic, 6, conceptual, and 6 definition).

<table>
<thead>
<tr>
<th>Content</th>
<th>Question pairs</th>
<th>Question type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quiz 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Algorithmic</td>
<td>Pair 1</td>
<td>Definition</td>
</tr>
<tr>
<td>4 Conceptual</td>
<td>Pair 2</td>
<td>Definition</td>
</tr>
<tr>
<td>4 Definition</td>
<td>Pair 3</td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td>Pair 4</td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td>Pair 5</td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td>Pair 6</td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td>Pair 7</td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td>Pair 8</td>
<td>Definition</td>
</tr>
<tr>
<td></td>
<td>Pair 9</td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td>Pair 10</td>
<td>Definition</td>
</tr>
<tr>
<td><strong>Quiz 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Algorithmic</td>
<td>Pair 11</td>
<td>Conceptual</td>
</tr>
<tr>
<td>2 Conceptual</td>
<td>Pair 12</td>
<td>Algorithmic</td>
</tr>
<tr>
<td>2 Definition</td>
<td>Pair 13</td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td>Pair 14</td>
<td>Definition</td>
</tr>
<tr>
<td></td>
<td>Pair 15</td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td>Pair 16</td>
<td>Algorithmic</td>
</tr>
<tr>
<td></td>
<td>Pair 17</td>
<td>Conceptual</td>
</tr>
<tr>
<td></td>
<td>Pair 18</td>
<td>Definition</td>
</tr>
</tbody>
</table>

4.3.2. Cognitive load questions

Cognitive load was measured with two questions: perceived mental effort (Paas, 1992) and perceived level of difficulty (Bratfisch, Borg, & Dornic, 1972). This economical short measure allowed for collecting cognitive load data from multiple students in one setting without
adding too much time to the session. The time factor was particularly important to this study; two out of three sessions took place in the actual classroom setting and were part of the class so the length of these sessions was limited. Because self-reported measurements do not require much additional time and are considered to be the most common (Naismith & Cavalcanti, 2015), we found them to be the appropriate measurements in this study.

For both measures we used the 7-point subjective rating scale. Although the original scale of mental effort is a 9-point scale, previous studies have shown that a 7-point scale is also effective and permits measuring both the mental effort and difficulty with one, shorter scale, making the results more comparable (Schmeck et al., 2015). Similar to aforementioned studies (Bratfisch et al., 1972; Paas, 1992; Schmeck et al., 2015), cognitive load was assessed after students completed each quiz/practice test chemistry question as questions varied in topics and level of difficulty. Students were asked to report how much mental effort they invested in answering each question, ranging from (1) very low to (7) very high and how difficult they perceived each question, ranging from (1) very easy to (7) very difficult. Figure 4 depicts both cognitive load questions that students answered after each chemistry question.

<table>
<thead>
<tr>
<th>Rate how much mental effort you invested in answering Question 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Very low</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7 Very high</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rate how difficult you perceived Question 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Very easy</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7 Very difficult</td>
</tr>
</tbody>
</table>

Figure 4. Two cognitive questions that students answered after each chemistry question.
4.4. Scratch paper

To examine the use of scratch paper as natural student behavior students were encouraged to bring and use their own scratch paper to sessions, were provided with additional paper, if needed, and received no instructions on how to use the scratch paper. Due to time restrictions of class, scratch paper was collected only at the end of the practice test (see Procedure section for more information) and scored for completeness (i.e., how often student used scratch paper during the test). Each question was scored as 1 (student used scratch paper) or 0 (student did not use scratch paper). Text that was underlined, crossed, strikethrough, re-written from the instructions, circled, check marked, highlighted with symbols (e.g., arrows), and erased and could not be read was not counted. For the paper-based test, the scratch paper was scored by one rater as students used the empty space around questions as scratch paper and it was relatively easy to score whether students used scratch paper to answer each question. For the computer-based test, students used a white, blank paper as scratch paper and could write anywhere. Therefore, two raters independently reviewed all computer-based scratch paper to associate each note with a specific question. After the initial review there was a 94% agreement, but after the second review, the reviewers achieved consensus, ultimately resulting in 100% agreement.

4.5. Procedure

Each assessment was restricted by time, but all students managed to complete each assessment in time.

4.5.1. Quiz 1 and Quiz 2

To know when to bring a laptop, about a week before Quiz 1, all students were informed about the testing mode of their next two quizzes. The day before any online quiz, all students
received an email reminder to bring a laptop to their recitation. After the main portion of the recitation was covered and it was time for the quiz, students either received instructions how to log in to the online test or received the paper-based test. Students who forgot to bring their laptop were given a laptop or received the test on paper. All quizzes were password-protected and each recitation had its own password so students could not open the quiz before or after the class.

Three days after each online quiz, students received an email that contained the quiz questions, their submitted answers, and the list of correct answers. For paper-based quizzes, students received their graded quizzes in the next recitation.

4.5.2. Practice test

The practice test took place in a large chemistry classroom on a Sunday afternoon, after Quiz 2 and before the next main test in class. Participation in the practice test was not part of the class, but students were encouraged by the instructor to come and take the test since the practice test was written by the class instructor and in a similar manner to the class test. Additionally, students received extra credit for participating in the study. About a week before the practice test, students received an email with the link to register for one of two potential testing times for the practice test. One time was dedicated to the computer mode and another time to the paper mode. However, which time was allocated to computer or paper mode was not disclosed, because a lot of students would likely select the paper mode (Prisacari, 2015). Thus, students registered for the practice test without knowing in advance the testing mode of the test. After the registration was closed, students were informed of the testing mode and reminded to bring a basic calculator and scratch paper, and laptop (for the computer group only). Upon arrival, students were asked to sign the consent form and were provided with scratch paper, and a copy of the periodic table and solubility table. The test lasted 60 minutes, but students were allowed to leave earlier if they
finished the test. After students finished the test, the proctor provided a printed copy of the answer key to each student (paper-based condition) or the student received an email with his or her submitted answers and the correct answers (computer condition). Lastly, the proctor collected the scratch paper and the tables from all students, as well as the paper-based tests and answer key from the paper condition and thanked the participants.

5. Results

An alpha level of .05 was used for all tests of statistical significance unless otherwise noted. The statistical analyses were performed with STATA 13 and general assumptions that apply to the parametric techniques were verified before conducting any tests of significance.

5.1. Cognitive load

Multiple independent t-tests were conducted to compare mean mental effort and level of difficulty between computer-based and paper-based tests. A Bonferroni correction for Type I error was used to evaluate the results. After Bonferroni correction, no statistical significance was found in the cognitive load measures between the two testing modes. Table 2 reports mean mental effort and level of difficulty on Quiz 1, Quiz 2, and the practice test and the main analyses results.
Table 2. Statistics for cognitive load measures.

<table>
<thead>
<tr>
<th></th>
<th>$\mu_{\text{Computer}}$ [SD]</th>
<th>$\mu_{\text{Paper}}$ [SD]</th>
<th>p-value</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz 1</td>
<td>3.86 [1.17]</td>
<td>3.72 [1.12]</td>
<td>.3413</td>
<td>.9538</td>
</tr>
<tr>
<td>Quiz 2</td>
<td>4.17 [1.10]</td>
<td>4.38 [1.08]</td>
<td>.1550</td>
<td>-1.4269</td>
</tr>
<tr>
<td>Practice test</td>
<td>3.81 [.97]</td>
<td>3.79 [1.03]</td>
<td>.8792</td>
<td>.1522</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\mu_{\text{Computer}}$ [SD]</th>
<th>$\mu_{\text{Paper}}$ [SD]</th>
<th>p-value</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz 1</td>
<td>3.98 [1.07]</td>
<td>3.66 [1.05]</td>
<td>.0283</td>
<td>2.2080</td>
</tr>
<tr>
<td>Quiz 2</td>
<td>4.27 [1.10]</td>
<td>4.55 [1.00]</td>
<td>.0443</td>
<td>-2.0234</td>
</tr>
<tr>
<td>Practice test</td>
<td>3.90 [.98]</td>
<td>3.91 [1.01]</td>
<td>.9490</td>
<td>-.0641</td>
</tr>
</tbody>
</table>

Note: SD = standard deviation; t = t-statistic
Overall alpha = 0.05; Bonferroni-adjusted alpha = .017

It is worth noting that although subjective techniques for measuring cognitive load are common, as the name suggests, they yield subjective data. To address this limitation, we calculated the level of difficulty for each question using the item difficulty formula. The item difficulty represents the proportion of students who answered a question correctly (Ding & Beichner, 2009; Holme & Murphy, 2011). The higher the difficulty score, the easier the item. Item difficulty values between .3 to .9 are considered acceptable. Figure 5 illustrates item difficulty by question and testing mode for each assessment and shows that most questions were within the permitted level of difficulty. Once again, no significant difference was found in mean difficulty level between computer and paper mode for any of the three assessments. The results for Quiz 1 were computer mode ($M = 68.43$, $SD = 17.34$) and paper mode ($M = 66.11$, $SD = 17.34$).
19.78; \( t(18) = 0.28, p = .78 \). The Quiz 2 results were computer mode (\( M = 53.87, SD = 20.64 \)) and paper mode (\( M = 52.08, SD = 23.44; t(14) = 0.16, p = .87 \)). Finally, the practice test results were computer mode (\( M = 61.31, SD = 17.60 \)) and paper mode (\( M = 61.72, SD = 17.98; t(34) = -0.07, p = .94 \)).

### 5.2. Use of scratch paper

To examine student use of scratch paper by testing mode, an independent t-test was conducted with the number of questions for which students used scratch paper as the dependent variable and the testing mode (computer versus paper) as the independent variable. This test revealed a significantly fewer number of questions employing scratch paper in computer mode (\( M = 4.58, SD = 2.22 \)) than in paper mode (\( M = 5.52, SD = 2.33; t(219) = -3.08, p < .01 \)). The magnitude of the differences in the means (mean difference = -0.94, 95% CI: -1.55 to -0.34) was very small (\( \eta^2 = .04 \)). These results suggest that when students take a chemistry test on paper, they use scratch paper more than students who take the same test online, though it is not clear that the difference (of approximately one question on a given test) is an important one.

### 5.3. Question type: Algorithmic, conceptual, and definition

To investigate whether there is a difference in cognitive load and use of scratch paper by question type, we examined the practice test, which contained an equal number of algorithmic, conceptual, and definition items. A series of independent t-test analyses revealed that difference in perceived mental effort (Figure 6) and level of difficulty (Figure 7) for each question type was not significant between computer-based and paper-based mode. For scratch paper, there was a significant difference between computer and paper mode, but only for the conceptual questions.
and the effect size was small ($\eta^2 = .04$) (Figure 8). When answering conceptual questions on paper, students used scratch paper significantly more than students who answered the same conceptual questions online. As expected, students used scratch paper the most for algorithmic questions, yet the difference between computer and paper testing mode was not significant for algorithmic questions. Statistical summary of all analyses is presented in Table 3.
Figure 5. Mean difficulty level index for each Quiz 1, Quiz 2, and the practice test question by the testing mode.
Figure 6. Mean mental effort for algorithmic, conceptual, and definition question types on the practice test by the testing mode. Error bars represent 95% confidence intervals. Bonferroni-adjusted alpha = .017.

Figure 7. Mean level of difficulty for algorithmic, conceptual, and definition question types on the practice test by the testing mode. Error bars represent 95% confidence intervals. Bonferroni-adjusted alpha = .017.

Figure 8. Mean proportion of the practice test questions for which students used the scratch paper by the question type and testing mode. Error bars represent 95% confidence intervals. The asterisk indicates a statistically significant difference after Bonferroni adjustment ($p < .017$).
Table 3. Statistics for cognitive load measures and proportion of questions for which students used scratch paper by question type.

<table>
<thead>
<tr>
<th>Mental effort</th>
<th>( \mu_{\text{Computer}} ) [SD]</th>
<th>( \mu_{\text{Paper}} ) [SD]</th>
<th>( p )-value</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic</td>
<td>4.24 [1.01]</td>
<td>4.27 [1.05]</td>
<td>.81</td>
<td>-.2354</td>
</tr>
<tr>
<td>Conceptual</td>
<td>3.99 [1.05]</td>
<td>3.90 [1.09]</td>
<td>.57</td>
<td>.5760</td>
</tr>
<tr>
<td>Definition</td>
<td>3.21 [1.05]</td>
<td>3.20 [1.14]</td>
<td>.94</td>
<td>.0754</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
<th>( \mu_{\text{Computer}} ) [SD]</th>
<th>( \mu_{\text{Paper}} ) [SD]</th>
<th>( p )-value</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic</td>
<td>4.14 [1.06]</td>
<td>4.23 [1.08]</td>
<td>.53</td>
<td>-.6224</td>
</tr>
<tr>
<td>Conceptual</td>
<td>4.23 [1.10]</td>
<td>4.16 [1.10]</td>
<td>.65</td>
<td>.4538</td>
</tr>
<tr>
<td>Definition</td>
<td>3.34 [1.00]</td>
<td>3.35 [1.07]</td>
<td>.98</td>
<td>-.0240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of scratch paper</th>
<th>( \mu_{\text{Computer}} ) [SD]</th>
<th>( \mu_{\text{Paper}} ) [SD]</th>
<th>( p )-value</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic</td>
<td>.64 [.24]</td>
<td>.71 [23]</td>
<td>.02</td>
<td>-2.2888</td>
</tr>
<tr>
<td>Conceptual</td>
<td>.11 [16]</td>
<td>.19 [.21]</td>
<td>&lt;.01</td>
<td>-2.8507</td>
</tr>
<tr>
<td>Definition</td>
<td>.01 [07]</td>
<td>.03 [.06]</td>
<td>.17</td>
<td>-1.3741</td>
</tr>
</tbody>
</table>

Note: SD = standard deviation; \( t \) = t-statistic
Overall alpha = 0.05; Bonferroni-adjusted alpha = .017

6. Discussion and Conclusion

This study explored potential differences between testing modes in a chemistry classroom setting by comparing extraneous cognitive load and use of scratch paper with paper-based and computer-based tests. The differences in performance between computer and paper testing modes are reported in Prisacari and Danielson (2017) who found that on average students did not perform significantly better when taking a test on computer versus on paper. The focus of this
article was to measure and report the mental effort, level of difficulty, and extent to which students used scratch paper during a computer-based or paper-based test. A key component of this study was use of equal numbers of three questions types: algorithmic, conceptual, and definition. This material design permitted us to examine the cognitive load and use of scratch paper at the question type level.

Overall, the present study extends previous results that support no difference in cognitive load between the two modes (Emerson & MacKay, 2011). In other words, one testing mode does not place a higher or lower cognitive load on student chemistry learning than the other mode. This was supported by finding no significant differences in mental effort or level of difficulty between computer-based and paper-based tests at the overall test level and for each question type. Difficulty level was measured subjectively (asking students to report the level of difficulty after answering each question) and objectively (calculating the proportion of students who answered each question correctly). Table 4 shows the correlations between subjective and objective item difficulty. Overall, we found a strong, negative correlation; that is high objective scores (i.e., more students answer a question correctly) are associated with lower subjective difficulty scores. This shows that questions which students found to be more difficult (i.e., reporting higher cognitive load) were in fact more difficult (i.e., fewer students correctly answered difficult questions), which also provides support that subjective cognitive load is a valid measure.
Table 4. Pearson correlations between subjective and objective level of difficulty

<table>
<thead>
<tr>
<th></th>
<th>Computer</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz 1</td>
<td>-.53</td>
<td>-.62</td>
</tr>
<tr>
<td>Quiz 2</td>
<td>-.86*</td>
<td>-.86*</td>
</tr>
<tr>
<td>Practice test</td>
<td>-.73*</td>
<td>-.75*</td>
</tr>
</tbody>
</table>

* p < .05

When measuring the use of scratch paper, we found that scratch paper was used more during paper-based than online testing. This finding is not surprising; writing on a paper test represents a common student behavior, especially when the test contains questions that require calculation. For example, after studying annotations made by college students, Marshall (1997) found that organic chemistry and calculus textbooks contained more penciled-in annotations than any other textbooks. This result nicely complements our findings; algorithmic questions reflect a high use of scratch paper. Table 5 shows to what extent students utilized scratch paper for each question type while taking the practice test on paper or online. Since the practice test contained six algorithmic, six conceptual, and six definition questions, students could use scratch paper up to six questions for each question type. Relative to conceptual and definition questions, students used scratch paper more for algorithmic questions. For example, when taking the practice test online, 81.26% of students used the scratch paper to answer three or more algorithmic questions and hardly used it at all for definition questions (i.e., 94.64% of students did not use scratch paper). Interestingly, the difference in scratch paper use by testing mode was significant only for the conceptual questions. However, as Kruskal-Wallis test showed, there was no a statistically significant difference in mean difficulty scores among three question types (Computer: $\chi^2(2) = 0.554, p = 0.76$; Paper: $\chi^2(2) = 0.761, p = 0.68$). All means of difficulty scores by question type
are shown in Figure 9. Further research may be required to answer why students use more scratch paper when answering conceptual questions on paper versus computer even when difficulty of conceptual questions does not differ from difficulty of other questions.

Table 5. Percentage of students who either did not use scratch paper or used it for answering 1 - 6 questions by the testing mode and question type. Most students used scratch paper to answer at least 3 algorithmic questions, up to 2 conceptual questions, and almost did not use it at all for definition questions.

<table>
<thead>
<tr>
<th>Question type/Mode</th>
<th>Number of questions for which students used scratch paper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Algorithmic</td>
<td></td>
</tr>
<tr>
<td>Computer</td>
<td>2.68%</td>
</tr>
<tr>
<td>Paper</td>
<td>1.83%</td>
</tr>
<tr>
<td>Conceptual</td>
<td></td>
</tr>
<tr>
<td>Computer</td>
<td>56.25%</td>
</tr>
<tr>
<td>Paper</td>
<td>41.28%</td>
</tr>
<tr>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>Computer</td>
<td>94.64%</td>
</tr>
<tr>
<td>Paper</td>
<td>85.32%</td>
</tr>
</tbody>
</table>

Figure 9. Mean difficulty level index of practice test questions by question type.

Lastly, we investigated two relationships associated with the use of scratch paper. First, we used a Pearson’s correlation to assess the relationship between mental effort and proportion
of questions for which students used scratch paper. For computer mode, we found no correlation, \( r = .08, n = 112, p = .40 \). However, for paper mode, we found a small positive correlation, \( r = .25, n = 109, p < .05 \). When taking a paper test, as students use more scratch paper, they report higher mean mental effort. Second, we examined the relationship between performance on the practice test and use of scratch paper. Using Spearman Rank Order Correlation, no relationship was found between these two variables (computer mode: \( \rho = .14; p\text{-value} = .15 \); paper mode: \( \rho = -.15; p\text{-value} = .12 \)). In sum, taking a chemistry test on paper shows (1) higher use of scratch paper than taking the same test online; (2) the more scratch paper students use the higher they rate their mental effort; but (3) the behavior of using scratch paper is not associated with higher chemistry practice test performance. These findings have significant practical implication. Our results suggest that changing from paper to computer testing would not impose a significant cognitive load on students, regardless of question type. However, as this is the only study to date to empirically measure cognitive load and use of scratch paper with two testing modes, our conclusions must remain tentative and represent an exciting opportunity for future research.

6.1. Limitations and future research

This study has some limitations. The study took place in a general chemistry course at a large Midwestern university, therefore the participants of this study represented a well-defined, selected group and our tests included only chemistry-related questions. These features restrict our findings to this subject group and material only. Consequently, future research needs to further investigate testing mode with more diverse participants and material.

Another limitation of this study was measuring cognitive load using subjective measures. Although we supplemented the subjective measure of item difficulty with the objective measure (i.e., item difficulty index), mental effort was investigated solely with subjective ratings. Future
studies on testing mode, may incorporate physiological measures to record changes in cognitive load in real time as students take tests in either testing mode. This step has become possible with a variety of physiological measures such as functional magnetic resonance imaging (fMRI, Brandt et al., 2015; for a review see Whelan, 2007), electroencephalography (EEG, Antonenko, Paas, Grabner, & van Gog, 2010; Krigolson, Hassall, Satel, & Klein, 2015; Thilaga et al., 2015; Vijayalakshmi et al., 2015), heart rate monitoring (Cranford, Tiettmeyer, Chuprinko, Jordan, & Grove, 2014; Durantin, Gagnon, Tremblay, & Dehais, 2014), and eye-tracking (Jarodzka, Janssen, Kirschner, & Erkens, 2015; Mcewen & Dubé, 2015; Park, Korbach, & Brünken, 2015; van Gog & Jarodzka, 2013).

Finally, in this study we only measured whether or not students used scratch paper. In the future studies, researchers may choose to measure how students used scratch paper (e.g., how much they write and what they write) when taking tests on paper versus computer. Such research could shed light on why students may use more scratch paper during a paper-based test than during an online test. The current study is only the first step in what we believe is a promising field of study in the area of testing mode.

Acknowledgments

We thank Dr. Appy and Logan Fischer who helped us to run the experiment. We also would like to thank Dr. Baluyut, Dr. Reed, and Gauri Ramasubramanian for helping create chemistry material, all student teaching assistants for proctoring and grading quizzes, and students for participating in this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
References


CHAPTER 5. GENERAL CONCLUSIONS

This dissertation has presented three separate peer-reviewed journal papers that explore testing mode and its effect on multiple testing in the context of general chemistry. Assessing testing mode is important because it is useful to know whether the mode of tests impacts test performance and represents an important difficulty to students, a difficulty that can be controlled by instructors. Although testing has been extensively studied, to date, there is no similar work on testing mode due to the difficulty of conducting an experimental design in an authentic setting. The current work filled this gap by conducting experiments in actual general chemistry classes with students from these classes and with material that represented either class assignments or practice tests for upcoming exam. The work of this dissertation adds to the growing literature that points to the utility of online testing in educational settings and makes empirical and practical contributions to the fields of chemical education, computers in education, and human computer interaction.

To better understand the effect of testing mode in educational settings, a series of studies were conducted. While one study represented a replication of previous work and used unbalanced chemistry material (Chapter 2), another study was conducted in a more authentic classroom environment and contained an equal number of algorithmic, conceptual, and definition questions (Chapter 3). In addition to student test performance, the testing mode was investigated with cognitive load and scratch paper use (Chapter 4).

The results of these studies were mixed. Chapter 2 consisted of 2 studies (i.e., Study 1 and Study 2). Although the effect of testing mode was detected, overall, in the first of these studies, that effect was not detected, overall, in the second study. Furthermore, while additional analysis in both studies showed that the Paper – Paper condition produced greater gains than one
or more other conditions (greater than Paper – Computer in the first study and greater than Computer – Paper in the second study), those differences were not consistent between the studies, and showed small effects when statistical significance was detected. The study described in Chapter 3 revealed no significant difference between online and paper-based testing in a more authentic learning context. In other words, there was no significant testing mode effect on the normalized gains (Hake, 1998).

Consistent with recent studies (Chua & Don, 2013; Inuwa, Taranikanti, Al-Rawahy, & Habbal, 2012; Meyer, Innes, Stomski, & Armson, 2015), the results of Chapter 2 (Study 2) and Chapter 3 suggest that the testing mode is not affecting student performance on a general chemistry test. The testing mode comparability finding represents a practical point for chemistry instructors; general chemistry instructors may format and deliver their quizzes and practice tests using paper-based mode, online mode, or any combination of both modes. It is also worth noting that this finding was attained with two different sample inclusion strategies. In Chapter 2, sample included only students for whom overall, algorithmic, and conceptual gains could be calculated, thus the sample size for overall, algorithmic, and conceptual gain analyses was consistent (i.e., Study 1 – 207 students and Study 2 – 215 students). Since many students answered correctly two definition questions, it was decided to omit definition gain analysis in order to avoid a larger sample cut. For example, if the sample size of Study 2 were based on students who had overall, algorithmic, conceptual, and definition gain, it would have decreased from 215 to 82. In Chapter 3, all students for whom an overall, algorithmic, conceptual, and definition gain could be calculated were included in the analyses, which resulted in a different sample size for each analysis (e.g., Quiz 1 overall gain – 207 students; Quiz 1 algorithmic gain – 112 students; Quiz 1 conceptual gain – 189 students; Quiz 1 definition gains – 149 students). This fluctuation in
sample size explains why trends in overall gain can appear different than trends in the other categories, and highlights the fact that figures of each gain analysis should be evaluated individually, not collectively.

Analyses of cognitive load (Chapter 4) lent further support for online testing. Figures 1–3 show that on each assessment (i.e., Quiz 1, Quiz 2, and the practice test) students reported at least one question to be easy, one question to be difficult, and most questions being in the middle, thus defining each assessment not too easy nor too hard. As expected, no significant difference was found in the cognitive load measures between the computer and paper modes. This means that one testing mode does not impose significantly different cognitive load on students than the other mode. A significant difference between the two modes was detected with use of scratch paper, but the effect was very small ($\eta^2 = .04$). This finding was not surprising; when Prisacari (2015) asked general chemistry students to select the preferred mode of their next chemistry test most students selected paper mode and 94% of them indicated the ability to work things out on paper as main reason. Although the results of this work demonstrated that students used scratch paper more when taking a chemistry test on paper than online, it is not clear whether the difference of approximately one question represents any practical significance.
Figure 1. Self-reported mean mental effort and level of difficulty for 10 Quiz 1 questions.

Figure 2. Self-reported mean mental effort and level of difficulty for 8 Quiz 2 questions.
Online testing offers several benefits. From an efficiency point of view, administering some or all class quizzes might optimize classroom time by reducing grading time for instructors and teaching assistants and spending more class time answering student questions or providing more information. Additionally, both instructors and students would benefit from fast feedback that online mode provides. According to Gikandi, Morrow, and Davis (2011) “online and blended settings offer the teacher more ongoing opportunities to monitor and identify patterns of students’ areas of weaknesses and provide feedback as concurrent scaffolded interventions (by being visible to all) that can meet the identified needs. Through ongoing monitoring of evidence of learning, the teacher can observe and identify patterns in students’ progress and achievements, interpret them, and make inferences about students’ progress, which in turn informs the appropriate formative feedback to serve common needs.” Receiving scores, instructors could make some adjustments to their upcoming lectures. For example, the instructor could include and work out the problem from a quiz that most students solved incorrectly. Additionally, online
testing could permit instructors to test students with a more versatile material such as interactive visualizations and media (DeBoer et al., 2014; for a review on computer-based science assessments see Kuo & Wu, 2013). For students, their quiz scores could provide feedback on what material they need to review or practice more.

In summary, this dissertation has shown the potential of online testing and its practical importance to pedagogical methods within the undergraduate general chemistry courses, using class material and four testing mode conditions. The present dissertation gives support for online testing and provides evidence that suggests test performance is unrelated to the mode in which chemistry tests are administered. Instructors don’t need to be concerned about testing mode when designing chemistry quizzes or practice tests.

Recommendations for Future Research

This work is limited by the contextual characteristics of the studies, including the nature of the participants, material used, and general chemistry context. Although in Article 2 and Article 3, a subset of tests included two required class quizzes, the final practice test was optional. Thus, the sample included students who volunteered to take a practice test in addition to all in-class quizzes (i.e., in Article 1, all practice tests were optional). Furthermore, these studies were conducted at a large Midwestern research university. Although the general chemistry course is very common and taught at many institutions, student population may vary from one university to another. Therefore, the student sample of these studies may not accurately represent the entire general chemistry student population.

Another suggestion for future work would be to repeat these experiments with different content, either in another chemistry class (e.g., organic chemistry) or discipline (e.g., general
physics, history, psychology). This material included algorithmic, conceptual, and definition question types that are common to general chemistry. It is possible that other disciplines may contain different questions types and show different testing mode effects. Of particular interest would be disciplines in which students might be predisposed to use one mode more than the other (e.g., art, computer science, engineering, mathematics). Another area of future work that relates to general chemistry as a course is the setting of the experiments. Castillo-Manzano, Castro-Nu, Opez-Valpuesta, and Teresa Sanz-Díaz (2017) conducted a study with economics college students and found that about 30% of students take either a laptop or tablet to class. Anticipating a similar effect, chemistry students were asked and reminded several times to bring their own laptops to class or practice sessions that represented data collection sessions of these studies. Despite these precautions, a few students still forgot to bring their laptops or to sufficiently charge them. As a result, data from these students were eliminated from data analyses. To address this design feature, future studies could be conducted in classes that take place in computer labs or encourage students to bring their laptops to every class.

Thus, it remains for future work to discover whether the results of this work would generalize to other students, material, and courses. Knowing how testing mode behaves in other educational settings can help improve future curricular assessment practices.

References


APPENDIX A. IRB APPROVAL FORM

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Date: 8/18/2015
To: 
CC: 
From: Office for Responsible Research
Title: Test Effect as a Function of Test Format and Delivery
IRB ID: 12-520
Study Review Date: 8/18/2015

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

1. Research conducted in established or commonly accepted education settings involving normal education practices, such as:
   - Research on regular and special education instructional strategies; or
   - Research on the effectiveness of, or the comparison among, instructional techniques, curricula, or classroom management methods.

2. Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey or interview procedures with adults or observation of public behavior where
   - Information obtained is recorded in such a manner that human subjects cannot be identified directly or through identifiers linked to the subjects; or
   - Any disclosure of the human subjects' responses outside the research could not reasonably place the subject at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

3. Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified directly or through identifiers linked to the subjects.

The determination of exemption means that:

You do not need to submit an application for annual continuing review.

You must carry out the research as described in the IRB application. Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. Changes to key personnel must also be approved. The purpose of review is to determine if the project still meets the federal criteria for exemption.

Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

Detailed information about requirements for submission of modifications can be found on the Exempt Study Modification Form. A Personnel Change Form may be submitted when the only modification involves changes in study staff. If it is determined that exemption is no longer warranted, then an Application for Approval of Research Involving Humans Form will need to be submitted and approved before proceeding with data collection.
Please note that you must submit all research involving human participants for review. Only the IRB or designees may make the determination of exemption, even if you conduct a study in the future that is exactly like this study.

Please be aware that approval from other entities may also be needed. For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted.

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.
APPENDIX B. TEST ITEMS (ARTICLE 1)

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

Q 1: Chapter 1, Question 1, MC
Answer: B
Which quantum number gives the shape of the atomic orbital?

A. \( n \)  B. \( \ell \)  C. \( m_\ell \)  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. \( \ell \)  C. \( m_\ell \)  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: Definitions

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?
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?
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. Cr₃CO₃  B. Cr(CO₃)₃  C. Cr₂(CO₃)₃  D. Cr₃(CO₃)₂

Q 18: Chapter 2, Question 1, OE
Exact Answer: Cr₂(CO₃)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. Co₂PO₄  B. Co(PO₄)₂  C. Co₂(PO₄)₃  D. Co₃(PO₄)₂

Q 20: Chapter 2, Question 1a, OE
Exact Answer: Co₃(PO₄)₂
What is the formula of cobalt(II) phosphate?
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?
\[ \text{BaCl}_2 + \text{K}_2\text{SO}_4 \rightarrow 2\text{KCl} + \text{BaSO}_4 \]

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

Q 30: Chapter 2, Question 4, OE
Exact Answer: \( \text{Ba}^{2+}(aq) + \text{SO}_4^{2-}(aq) \rightarrow \text{BaSO}_4(s) \)
Acceptable Answer: \( \text{Ba}^{2+} + \text{SO}_4^{2-} \rightarrow \text{BaSO}_4 \)
What is the net ionic equation for the reaction?
\[ \text{BaCl}_2 + \text{K}_2\text{SO}_4 \rightarrow 2\text{KCl} + \text{BaSO}_4 \]

Q 31: Chapter 2, Question 4a, MC
Answer: D
What is the net ionic equation for the reaction?
\[ \text{NaCl} + \text{AgNO}_3 \rightarrow \text{NaNO}_3 + \text{AgCl} \]

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

Q 32: Chapter 2, Question 4a, OE
Exact Answer: \( \text{Ag}^+(aq) + \text{Cl}^-(aq) \rightarrow \text{AgCl(s)} \)
Acceptable Answer: \( \text{Ag}^+ + \text{Cl}^- \rightarrow \text{AgCl} \)
What is the net ionic equation for the reaction?
\[ \text{NaCl} + \text{AgNO}_3 \rightarrow \text{NaNO}_3 + \text{AgCl} \]
Q 33: Chapter 3, Question 1, MC
Answer: C
How many sigma (σ) and pi (π) bonds are in one molecule of butyric acid (shown in the figure)?

(A) 12 σ bonds and 1 π bonds
(B) 12 σ bonds and 2 π bonds
(C) 13 σ bonds and 1 π bonds
(D) 13 σ bonds and 2 π bonds

Q 34: Chapter 3, Question 1, OE
Exact Answer: 13 σ bonds and 1 π bonds
How many sigma (σ) and pi (π) bonds are in one molecule of butyric acid (shown in the figure)?

Q 35: Chapter 3, Question 1a, MC
Answer: C
How many sigma (σ) and pi (π) bonds are in one molecule of 3-hydroxy butanal (shown in the figure)?

(A) 9 σ bonds and 1 π bonds
(B) 9 σ bonds and 2 π bonds
(C) 10 σ bonds and 1 π bonds
(D) 10 σ bonds and 2 π bonds

Q 36: Chapter 3, Question 1a, OE
Exact Answer: 10 σ bonds and 1 π bonds
How many sigma (σ) and pi (π) bonds are in one molecule of 3-hydroxy butanal (shown in the figure)?
Chapter 3: Structure & 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?
Q 45: Chapter 3, Question 4, MC
Answer: D
What is the molecular shape of dichloromethane?

\[
\begin{align*}
\text{H} & \quad \text{C} & \quad \text{Cl} \\
\text{H} & \quad \quad & \quad \text{Cl} \\
\quad & \quad & \quad \text{Cl}
\end{align*}
\]

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?

\[
\begin{align*}
\text{H} & \quad \text{C} & \quad \text{Cl} \\
\text{H} & \quad \quad & \quad \text{Cl} \\
\quad & \quad & \quad \text{Cl}
\end{align*}
\]

Q 47: Chapter 3, Question 4a, MC
Answer: B
What is the molecular shape of trichloromethane?

\[
\begin{align*}
\text{H} & \quad \text{C} & \quad \text{Cl} \\
\quad & \quad & \quad \text{Cl} \\
\quad & \quad & \quad \text{Cl}
\end{align*}
\]

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?

\begin{center}
\begin{tikzpicture}
  \node at (0,0) (H) {H};
  \node at (1,0) (C) {C};
  \node at (2,0) (Cl) {Cl};
  \node at (3,0) (Cl1) {Cl};
  \node at (2,1) (Cl2) {Cl};
  \draw (H) -- (C);
  \draw (C) -- (Cl);
  \draw (C) -- (Cl1);
  \draw (C) -- (Cl2);
\end{tikzpicture}
\end{center}
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 H₂?

\[ 4\text{H}_2\text{O}(g) + 3\text{Fe}(s) \rightarrow \text{Fe}_3\text{O}_4(s) + 4\text{H}_2(g) \]

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 H₂?

\[ 4\text{H}_2\text{O}(g) + 3\text{Fe}(s) \rightarrow \text{Fe}_3\text{O}_4(s) + 4\text{H}_2(g) \]

Q 51: Chapter 4, Question 1a, MC
Answer: B
How many moles of H₂ are produced by 10.0 mol of Fe?

\[ 4\text{H}_2\text{O}(g) + 3\text{Fe}(s) \rightarrow \text{Fe}_3\text{O}_4(s) + 4\text{H}_2(g) \]

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 H₂ are produced by 10.0 mol of Fe?

\[ 4\text{H}_2\text{O}(g) + 3\text{Fe}(s) \rightarrow \text{Fe}_3\text{O}_4(s) + 4\text{H}_2(g) \]
Q 53: Chapter 4, Question 2, MC
Answer: B
How many moles of Al₂O₃ can be produced from 27 g of aluminum and 16 g of oxygen?

<table>
<thead>
<tr>
<th>Atomic molar masses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
</tr>
<tr>
<td>O</td>
</tr>
</tbody>
</table>

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 Al₂O₃ can be produced from 27 g of aluminum and 16 g of oxygen?

<table>
<thead>
<tr>
<th>Atomic molar masses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
</tr>
<tr>
<td>O</td>
</tr>
</tbody>
</table>

Q 55: Chapter 4, Question 2a, MC
Answer: C
How many moles of Fe₂O₃ can be produced from 56 g of iron and 32 g of oxygen?

<table>
<thead>
<tr>
<th>Atomic molar masses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe</td>
</tr>
<tr>
<td>O</td>
</tr>
</tbody>
</table>

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; 0.49 mol
How many moles of Fe₂O₃ can be produced from 56 g of iron and 32 g of oxygen?
Chapter 4: Stoichiometry
Topic 3: Molarity
Question Type: Algorithmic

Q 57: Chapter 4, Question 3, MC
Answer: D
What mass of MgCl\textsubscript{2} is required to prepare 2.00 L of 0.550 M solution?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Molar mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>1.10 g</td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td>26.2 g</td>
<td></td>
</tr>
<tr>
<td>C.</td>
<td>28.9 g</td>
<td>MgCl\textsubscript{2} 95.2 g/mol</td>
</tr>
<tr>
<td>D.</td>
<td>105 g</td>
<td></td>
</tr>
</tbody>
</table>

Q 58: Chapter 4, Question 3, OE
Exact Answer: 105 g
Acceptable Answer: 104.7 – 105 g
What mass of MgCl\textsubscript{2} is required to prepare 2.00 L of 0.550 M solution?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Molar mass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MgCl\textsubscript{2} 95.2 g/mol</td>
</tr>
</tbody>
</table>

Q 59: Chapter 4, Question 3a, MC
Answer: D
What mass of anhydrous Na\textsubscript{2}S is required to prepare 4.00 L of 0.270 M solution?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Molar mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>1.08 g</td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td>5.27 g</td>
<td></td>
</tr>
<tr>
<td>C.</td>
<td>72.2 g</td>
<td>Na\textsubscript{2}S 78.0 g/mol</td>
</tr>
<tr>
<td>D.</td>
<td>84.2 g</td>
<td></td>
</tr>
</tbody>
</table>

Q 60: Chapter 4, Question 3a, OE
Exact Answer: 84.2 g
Acceptable Answer: 84.24 g; 84 g
What mass of anhydrous Na\textsubscript{2}S is required to prepare 4.00 L of 0.270 M solution?

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Molar mass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Na\textsubscript{2}S 78.0 g/mol</td>
</tr>
</tbody>
</table>


Q 61: Chapter 4, Question 4, MC
Answer: B
Balance this equation and then answer the following question:

\[ ? \text{S} + ? \text{HNO}_3 \rightarrow ? \text{H}_2\text{SO}_4 + ? \text{NO}_2 + ? \text{H}_2\text{O} \]

How many moles of H\(_2\)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:

\[ ? \text{S} + ? \text{HNO}_3 \rightarrow ? \text{H}_2\text{SO}_4 + ? \text{NO}_2 + ? \text{H}_2\text{O} \]

How many moles of H\(_2\)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:

\[ ? \text{Na}_2\text{S} + ? \text{O}_2 + ? \text{CO}_2 \rightarrow ? \text{Na}_2\text{CO}_3 + ? \text{SO}_2 \]

How many moles of SO\(_2\) are formed per mole of sodium sulfide, Na\(_2\)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:

\[ ? \text{Na}_2\text{S} + ? \text{O}_2 + ? \text{CO}_2 \rightarrow ? \text{Na}_2\text{CO}_3 + ? \text{SO}_2 \]

How many moles of SO\(_2\) are formed per mole of sodium sulfide, Na\(_2\)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?

<table>
<thead>
<tr>
<th>Partial Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
</tr>
<tr>
<td>CO₂</td>
</tr>
<tr>
<td>CH₄</td>
</tr>
<tr>
<td>C₂H₄</td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Partial Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
</tr>
<tr>
<td>CO₂</td>
</tr>
<tr>
<td>CH₄</td>
</tr>
<tr>
<td>C₂H₄</td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Partial Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
</tr>
<tr>
<td>CO₂</td>
</tr>
<tr>
<td>CH₄</td>
</tr>
<tr>
<td>C₂H₄</td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Partial Pressures</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
</tr>
<tr>
<td>CO₂</td>
</tr>
<tr>
<td>CH₄</td>
</tr>
<tr>
<td>C₂H₄</td>
</tr>
</tbody>
</table>
Q 77: Chapter 5, Question 4, MC
Answer: C
A student collected 40.0 mL of N₂ gas when the temperature was 20.0°C and the pressure was 720. torr. The next day the temperature was still 20.0°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 N₂ gas when the temperature was 20.0°C and the pressure was 720. torr. The next day the temperature was still 20.0°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 O₂ gas when the temperature was 25.0°C and the pressure was 750. torr. The next day the temperature was still 25.0°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 O₂ gas when the temperature was 25.0°C and the pressure was 750. torr. The next day the temperature was still 25.0°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 J·g⁻¹·°C⁻¹) emits 233 J of heat when it cools from 125°C?

A. –84°C  B. 41°C  C. 84°C  D. 209°C

Q 82: Chapter 6, Question 1, OE  
Exact Answer: 41°C or 314.15 Kelvin  
Acceptable Answer: 314 Kelvin; 40.6 – 41.4 °C

What is the final temperature when a 21.5 g sample of gold (specific heat = 0.129 J·g⁻¹·°C⁻¹) emits 233 J of heat when it cools from 125°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 J·g⁻¹·°C⁻¹) absorbs 421 J of heat when it is heated from 22°C?

A. 34°C  B. 50°C  C. 56°C  D. 608°C

Q 84: Chapter 6, Question 1a, OE  
Exact Answer: 56°C or 329.15 Kelvin  
Acceptable Answer: 329 - 330 Kelvin  
Accept: 56.0 – 56.4 °C

What is the final temperature when a 13.7 g sample of aluminum (specific heat = 0.900 J·g⁻¹·°C⁻¹) absorbs 421 J of heat when it is heated from 22°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

$\text{H}_2(g) + \text{Cl}_2(g) \rightarrow 2\text{HCl}(g)$ \hspace{1cm} $\Delta H^\circ = -185$ kJ

$\text{H}_2(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{H}_2\text{O}(g)$ \hspace{1cm} $\Delta H^\circ = -242$ kJ

A. –57 kJ \hspace{1cm} B. –185 kJ \hspace{1cm} C. –306 kJ \hspace{1cm} 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

$\text{H}_2(g) + \text{Cl}_2(g) \rightarrow 2\text{HCl}(g)$ \hspace{1cm} $\Delta H^\circ = -185$ kJ

$\text{H}_2(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{H}_2\text{O}(g)$ \hspace{1cm} $\Delta H^\circ = -242$ kJ

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

$\text{H}_2(g) + \text{Br}_2(l) \rightarrow 2\text{HBr}(g)$ \hspace{1cm} $\Delta H^\circ = -72$ kJ

$\text{H}_2(g) + \frac{1}{2}\text{O}_2(g) \rightarrow \text{H}_2\text{O}(g)$ \hspace{1cm} $\Delta H^\circ = -242$ kJ

A. –98 kJ \hspace{1cm} B. –170 kJ \hspace{1cm} C. –278 kJ \hspace{1cm} D. –314 kJ
Q 88: Chapter 6, Question 2a, OE
Exact Answer: $-170 \text{ kJ}$

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

$\text{H}_2(g) + \text{Br}_2(l) \rightarrow 2\text{HBr}(g) \quad \Delta H^\circ = -72 \text{ kJ}$

$\text{H}_2(g) + \frac{1}{2} \text{O}_2(g) \rightarrow \text{H}_2\text{O}(g) \quad \Delta H^\circ = -242 \text{ kJ}$
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, \( \text{Cl}_2(g) + \text{F}_2(g) \rightarrow 2\text{ClF}(g) \)?

<table>
<thead>
<tr>
<th>Bond Energies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl–Cl</td>
<td>243 kJ/mol</td>
</tr>
<tr>
<td>F–F</td>
<td>159 kJ/mol</td>
</tr>
<tr>
<td>Cl–F</td>
<td>255 kJ/mol</td>
</tr>
</tbody>
</table>

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, \( \text{Cl}_2(g) + \text{F}_2(g) \rightarrow 2\text{ClF}(g) \)?

<table>
<thead>
<tr>
<th>Bond Energies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl–Cl</td>
<td>243 kJ/mol</td>
</tr>
<tr>
<td>F–F</td>
<td>159 kJ/mol</td>
</tr>
<tr>
<td>Cl–F</td>
<td>255 kJ/mol</td>
</tr>
</tbody>
</table>

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, \( 2\text{ClF}(g) \rightarrow \text{Cl}_2(g) + \text{F}_2(g) \)?

<table>
<thead>
<tr>
<th>Bond Energies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl–Cl</td>
<td>243 kJ/mol</td>
</tr>
<tr>
<td>F–F</td>
<td>159 kJ/mol</td>
</tr>
<tr>
<td>Cl–F</td>
<td>255 kJ/mol</td>
</tr>
</tbody>
</table>

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, \( 2\text{ClF}(g) \rightarrow \text{Cl}_2(g) + \text{F}_2(g) \)?

<table>
<thead>
<tr>
<th>Bond Energies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl–Cl</td>
<td>243 kJ/mol</td>
</tr>
<tr>
<td>F–F</td>
<td>159 kJ/mol</td>
</tr>
<tr>
<td>Cl–F</td>
<td>255 kJ/mol</td>
</tr>
</tbody>
</table>
### Pair 1

**Topic:** Matter  
**Question type:** Definition

#### Quiz 1

**Answer:** b

Which are condensed phases of matter?  
i) solid  
ii) liquid  
iii) gas

- a) i only  
- b) i and ii  
- c) ii and iii  
- d) All are condensed phases  
- e) None are condensed phases

#### Practice Test

**Answer:** d

In which phase(s) are particles able to move around (i.e., not locked in place)?

- a) Gas only  
- b) Liquid only  
- c) Solid only  
- d) Gas and liquid  
- e) Liquid and solid
<table>
<thead>
<tr>
<th>Quiz 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answer:</strong> a</td>
</tr>
</tbody>
</table>

Which intermolecular force is present in all substances?

a) London forces  
b) Dipole-dipole interactions  
c) Ion-dipole interactions  
d) Hydrogen bonding  
e) None of the above

<table>
<thead>
<tr>
<th>Practice Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Answer:</strong> d</td>
</tr>
</tbody>
</table>

Hydrogen bonding is…

a) Present in all substances  
b) Present in any molecule with hydrogen atoms  
c) Present in all polar molecules  
d) Present in molecules with hydrogen atoms attached to oxygen atoms  
e) Present in any molecule containing both hydrogen and oxygen atoms
Pair 3
Topic: Intermolecular forces
Question type: Conceptual

Quiz 1

Answer: b

Which intermolecular forces are present in the substance formaldehyde?

:O\(\text{C}\)\(\text{H}\)

i) London forces ii) dipole-dipole forces iii) hydrogen bonding

a) i only
b) i and ii
c) ii and iii
d) i, ii, and iii
e) None of the above

Practice Test

Answer: d

Which intermolecular forces are present in C\(_2\)H\(_5\)OH?

i) London forces ii) dipole-dipole forces iii) hydrogen bonding

a) i only
b) i and ii
c) ii and iii
d) i, ii, and iii
e) None of the above
Pair 4
Topic: Intermolecular forces
Question type: Conceptual

Quiz 1
Answer: b

Considering all forces involved, which substance would you expect to have the lowest boiling point?

a) H₂O  
b) H₂S  
c) H₂Se  
d) H₂Te  
e) All of the above substances have similar boiling point

Practice Test
Answer: a

Which substance would you expect to have highest boiling point?

a) HF  
b) HCl  
c) HBr  
d) HI  
e) All of the above substances have similar boiling point
Quiz 1
Answer: b

Which covalent bond is the most polar?

<table>
<thead>
<tr>
<th>Electronegativities</th>
</tr>
</thead>
<tbody>
<tr>
<td>C 2.5</td>
</tr>
<tr>
<td>N 3.0</td>
</tr>
<tr>
<td>O 3.5</td>
</tr>
<tr>
<td>F 4.0</td>
</tr>
</tbody>
</table>

a) N-F
b) C-F
c) O-F
d) F-F
e) All of the above are of similar polarity

Practice Test
Answer: a

Which covalent bond is the least polar?

<table>
<thead>
<tr>
<th>Electronegativities</th>
</tr>
</thead>
<tbody>
<tr>
<td>H 2.1</td>
</tr>
<tr>
<td>N 3.0</td>
</tr>
<tr>
<td>O 3.5</td>
</tr>
<tr>
<td>F 4.0</td>
</tr>
</tbody>
</table>

a) H-H
b) H-N
c) H-O
d) H-F
e) All of the above are of similar polarity
Quiz 1
Answer: a

Rank the following substances in order of increasing boiling point: He, H₂O, CH₄

a) He < CH₄ < H₂O  
b) CH₄ < He < H₂O  
c) CH₄ < H₂O < He  
d) H₂O < CH₄ < He  
e) None of the above statements are correct

Practice Test
Answer: e

Rank the following substances in order of increasing boiling point: CH₄, C₂H₅OH, CH₃OCH₃, C₂H₆.

a) C₂H₅OH < CH₃OCH₃ < C₂H₆ < CH₄ 
b) CH₄ < C₂H₆ < C₂H₅OH < CH₃OCH₃ 
c) C₂H₆ < CH₄ < CH₃OCH₃ < C₂H₅OH 
d) CH₄ < CH₃OCH₃ < C₂H₆ < C₂H₅OH 
e) CH₄ < C₂H₆ < CH₃OCH₃ < C₂H₅OH
Quiz 1
Answer: d

Which substance would you expect to exhibit the greatest London forces?

a) H₂O
b) H₂S
c) H₂Se
d) H₂Te
e) All of the above substances exhibit similar degree of London forces

Practice Test
Answer: c

Which substance would you expect to exhibit the weakest dispersion forces?

a) CH₃Cl
b) CH₃OH
c) CH₄
d) CH₃F
e) CH₃Br
Quiz 1
Answer: d

Which is a network covalent substance?

a) Water  
b) Carbon dioxide  
c) NaCl  
d) Quartz  
e) Gold

Practice Test
Answer: c

Which exhibits metallic bonding?

a) Diamond  
b) Ice  
c) Magnesium  
d) Table salt  
e) Sugar
Quiz 1
Answer: e

When the following chemical equation is balanced with whole number coefficients, what is the sum of the coefficients? (include “1”s in your calculation)
$SO_2 + O_2 \rightarrow SO_3$

a) 3  
b) 7  
c) 10  
d) 12  
e) None of the above

Practice Test
Answer: c

$N_2 + H_2 \rightarrow NH_3$
When the above reaction is completed and balanced, what is the sum of the coefficients? (include “1”s in your calculation)

a) 4  
b) 5  
c) 6  
d) 8  
e) None of the above
Quiz 1
Answer: a

Classify the following reaction type:
Fe₂O₃(s) + 2Al(s) \rightarrow Al₂O₃(s) + 2Fe(s)

a) Single replacement
b) Double replacement
c) Decomposition
d) Synthesis
e) Combustion

Practice Test
Answer: a

The general form of a chemical synthesis is:

a) A + B \rightarrow AB
b) AB \rightarrow A + B
c) AB + C \rightarrow AC + B
d) AB + CD \rightarrow AD + CB
e) None of the above
Pair 11
Topic: Stoichiometry
Question type: Conceptual

Quiz 2
Answer: d

Which of the following statements accurately describes what the equation below represents?

\[ 2\text{H}_2 + \text{O}_2 \rightarrow 2\text{H}_2\text{O} \]

a) 2 moles of H\(_2\) react with 1 mole of O\(_2\) to produce 2 moles of H\(_2\)O
b) 2 molecules of H\(_2\) react with 1 molecule of O\(_2\) to produce 2 molecules of H\(_2\)O
c) 2 grams of H\(_2\) react with 1 gram of O\(_2\) to produce 2 grams of H\(_2\)O
d) Both a and b are correct
e) a, b, and c are all correct

Practice Test
Answer: b

Which of the following statements regarding the reaction below is true?

\[ \text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O} \]

i. 1 mole of CH\(_4\) will produce 2 moles of H\(_2\)O
ii. 1 mole of O\(_2\) will produce 1 mole of H\(_2\)O
iii. 1 gram of CH\(_4\) will produce 1 gram of CO\(_2\)

a) i
b) i and ii
c) iii
d) i, ii, and iii
e) None of the above
**Pair 12**  
**Topic:** Solubility and precipitation reactions  
**Question type:** Algorithmic

**Quiz 2**  
**Answer:** b

Which of the following substances are soluble in aqueous solution?

i) NaBr  
ii) HgBr  
iii) Fe(NO$_3$)$_2$  
iv) Zn(OH)$_2$

---

**Table 8.1 Solubility Rules**

<table>
<thead>
<tr>
<th>Soluble in water</th>
<th>Insoluble in water</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sodium, potassium, and ammonium salts</td>
<td>Most phosphates, carbonates, and sulfides</td>
</tr>
<tr>
<td>All acetates and nitrates</td>
<td>Most sulfates</td>
</tr>
<tr>
<td>Most halides (chlorides, bromides, iodides)</td>
<td><strong>except</strong></td>
</tr>
<tr>
<td>Most sulfates</td>
<td><strong>except</strong></td>
</tr>
<tr>
<td></td>
<td>Halides of lead(II), silver(I), and mercury(I)</td>
</tr>
<tr>
<td></td>
<td>Sulfates of calcium, barium, lead(II), and strontium</td>
</tr>
</tbody>
</table>

a) i only  
b) i and iii  
c) i, ii, and iii  
d) All are soluble  
e) None are soluble
Practice Test
Answer: d

Select the precipitate of the reaction (if it is formed) when the following chemicals are allowed to react.
Zn(NO$_3$)$_2$(aq) + NaOH(aq) \rightarrow

**Table 8.1 Solubility Rules**

<table>
<thead>
<tr>
<th>Soluble in water</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sodium, potassium, and ammonium salts</td>
</tr>
<tr>
<td>All acetates and nitrates</td>
</tr>
<tr>
<td>Most halides (chlorides, bromides, iodides)</td>
</tr>
<tr>
<td>Most sulfates</td>
</tr>
<tr>
<td>except Halides of lead(II), silver(I), and mercury(I)</td>
</tr>
<tr>
<td>except Sulfates of calcium, barium, lead(II), and strontium</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insoluble in water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most phosphates, carbonates, and sulfides</td>
</tr>
<tr>
<td>Most hydroxides</td>
</tr>
<tr>
<td>except Sodium, potassium, and ammonium salts; calcium sulfide</td>
</tr>
<tr>
<td>except Sodium, potassium, calcium, and barium hydroxides</td>
</tr>
</tbody>
</table>

a) NaNO$_3$
b) Na(NO$_3$)$_2$
c) ZnOH
d) Zn(OH)$_2$
e) A precipitate is not formed
What is the net ionic equation for the following reaction?
\[ \text{Zn(NO}_3\text{)}_2(\text{aq}) + 2\text{NaOH(aq)} \rightarrow \text{Zn(OH)}_2(\text{s}) + 2\text{NaNO}_3(\text{aq}) \]

a) \[ \text{Zn}^{2+}(\text{aq}) + \text{Na}^+(\text{aq}) \rightarrow \text{NO}_3^-(\text{aq}) + \text{OH}^-(\text{aq}) \]
b) \[ \text{Zn}^{2+}(\text{aq}) + \text{NO}_3^-(\text{aq}) \rightarrow \text{Zn(NO}_3\text{)}_2(\text{s}) \]
c) \[ 2\text{NO}_3^-(\text{aq}) + 2\text{OH}^-(\text{aq}) \rightarrow 2\text{NO}_3\text{OH(aq)} \]
d) \[ \text{Zn}^+(\text{aq}) + \text{OH}^-(\text{aq}) \rightarrow \text{ZnOH(s)} \]
e) None of the above

What is the net ionic equation for the following reaction?
\[ \text{HCl(aq)} + \text{NaOH(aq)} \rightarrow \text{NaCl(aq)} + \text{H}_2\text{O(l)} \]

a) \[ \text{H}^+(\text{aq}) + \text{OH}^-(\text{aq}) \rightarrow \text{H}_2\text{O(l)} \]
b) \[ \text{Cl}^-(\text{aq}) + \text{Na}^+(\text{aq}) \rightarrow \text{NaCl(aq)} \]
c) \[ \text{H}^+(\text{aq}) + \text{H}^-(\text{aq}) \rightarrow \text{H}_2(\text{g}) \]
d) \[ \text{H}^+(\text{aq}) + \text{Cl}^-(\text{aq}) + \text{Na}^+(\text{aq}) + \text{OH}^-(\text{aq}) \rightarrow \text{Na}^+(\text{aq}) + \text{Cl}^-(\text{aq}) + \text{H}_2\text{O(l)} \]
e) Acid-base reactions do not have net ionic equations
Pair 14
Topic: Acid-Base reactions
Question type: Definition

Quiz 2
Answer: c

Which substance acts as an acid in aqueous solution?
 a) NaOH
 b) CH₃CH₂OH
 c) CH₃COOH
 d) All of the above
 e) None of the above

Practice Test
Answer: c

Which substance acts as a base in aqueous solution?
 a) CH₃COOH
 b) CH₃CH₂OH
 c) NaOH
 d) AlCl₃
 e) None of the above
### Quiz 2

**Answer:** b

How much solid NaOH, in grams, must you weigh out in order to have 1.50 mole of NaOH? (periodic table is provided)

- a) 40.0 g
- b) 60.0 g
- c) $3.75 \times 10^{-2}$ g
- d) 26.7 g
- e) $9.97 \times 10^{-23}$ g

### Practice Test

**Answer:** d

How many moles of Cl$^{-}$ ions are there in 194.64 g of FeCl$_3$?

- a) 1.2 moles
- b) 0.4 moles
- c) 2.5 moles
- d) 3.6 moles
- e) 4.8 moles
How many chlorine atoms are in 2.5 moles of AlCl₃?

a) 5.0 x 10²³
b) 1.5 x 10²⁴
c) 3.0 x 10²⁴
d) 4.5 x 10²⁴
e) 6.0 x 10²⁶

How many oxygen atoms are in 27.4 grams of Fe(NO₃)₂?
(F.W. 179.9 g/mol)

a) 5.50 x 10²³ atoms
b) 9.17 x 10²³ atoms
c) 6.02 x 10²³ atoms
d) 2.37 x 10²⁵ atoms
e) 3.95 x 10²⁴ atoms
Quiz 2
Answer: c

One mole of O$_2$ molecules contains the same number of molecules as
a) 0.667 moles of O$_3$
   b) 2.00 moles of CH$_3$COOH
   c) 1.00 mole of CH$_3$CH$_2$OH
   d) All of above
   e) None of the above

Practice Test
Answer: d

One mole of NH$_3$ molecules contains the same number of hydrogen atoms as
a) 3.00 moles of HCl
   b) 1.50 moles of H$_2$O
   c) 1.00 mole of CH$_3$Cl
   d) All of above
   e) None of the above
**Pair 18**

Topic: Stoichiometry  
Question type: Definition

**Quiz 2**

Answer: a

When balancing a chemical equation, one may change:
- a) Coefficients only
- b) Subscripts only
- c) Both a and b
- d) Neither a nor b
- e) Depends on the type of reaction

**Practice Test**

Answer: b

A chemical reaction is balanced when…
- a) The number of atoms on both sides of the equation are equal to each other
- b) There is the same number of each kind of atoms on both sides of the equation
- c) The number of molecules on both sides of the equation are equal to each other
- d) The number of substances on both sides of the equation are equal to each other
- e) None of the above
APPENDIX D. INSTRUCTIONS FOR COMPUTER QUIZ (ARTICLES 2 & 3)

Time: 15 minutes

Time:
Section #:
Location:
TA:
# of students:

Before quiz
1. Erase everything on the blackboard/whiteboard
2. Distribute the periodic table to all students
3. Remind students to answer all questions on the quiz and double check everything before submitting the quiz online.

How to access online quiz
1. Click on the link that was emailed.
2. Before telling students the password (below), make sure that everyone is on the same page that requests password. We want all students to start the online quiz at the same time.
3. Password is: [TYPE PASSWORD]. Password contains only lowercase letters.
4. In case a student forgot to bring a laptop or student has problems with his/her computer (e.g., low battery), give him/her a printed version of the quiz.

During the quiz
1. Walk around and make sure students have only the quiz page open.

Questions? Contact Anna.
APPENDIX E. INSTRUCTIONS FOR PAPER QUIZ (ARTICLES 2 & 3)

Time: 15 minutes

Time:
Section #:
Location:
TA:
# of students:

Before quiz
4. Erase everything on the blackboard/whiteboard
5. Distribute the periodic table to all students
6. Remind students to answer all questions on the quiz and double check everything before giving you the quiz.

During the quiz
2. Walk around and make sure students don’t cheat.

After the quiz
1. When student is done with his/her quiz, quickly scan for completion. Has the student answered all questions? If you see some questions unanswered, suggest student to finish his/her quiz, if time permits.

Questions? Contact Anna
Question 15

How many moles of Cl\textsuperscript{−} ions are there in 194.64 g of FeCl\textsubscript{3}\textsuperscript{−}?

a) 1.2 moles  
b) 0.4 moles  
c) 2.5 moles  
d) 3.6 moles  
e) 4.8 moles

\[ \frac{194.64 \text{ g FeCl}_3}{126.951 \text{ g FeCl}_3} \times \frac{3 \text{ mol Cl}^-}{1 \text{ mol FeCl}_3} \]

Fe - 55.845  
Cl - 35.453

Rate how much mental effort you invested in answering Question 15.

1 Very low  
2  
3  
4  
5  
6 Very high

Rate how difficult you perceived Question 15.

1 Very easy  
2  
3  
4  
5  
6 Very difficult  
7
Indicates item number for which student used scratch paper