SUGGESTIONS FOR EBOOKS: A MULTIMETHOD CHARACTERIZATION OF UNDERGRADUATE STEM STUDENTS' INTERACTIONS WITH TEXTBOOK FEATURES

by

ELIZABETH L. DAY

(Under the Direction of Norbert J. Pienta)

ABSTRACT

This dissertation uses Interaction Theory to address the *learner-content* interactions before and after a format change to electronic textbook (ebook). This is significant because active learning pedagogies (such as flipped classrooms) increasingly push content-delivery outside of the classroom, and examination of the chemistry course resources is necessary to meet demands for interactive content-delivery devices.

To investigate this issue, a longitudinal survey of general chemistry students' perception of their use of course resources and textbook features before and after the adoption of an electronic textbook. As such, we can see that the main sources of information for students are the textbook, Google, friends, and occasionally peer or private tutors. As far as a format change, the students engage with the features in a similar manner despite the availability for more interactive choices. If the format does not engage them more, it begs the question, is this all we can do with an ebook?

Based on the responses of the survey, eye-tracking experiments were designed to investigate the *learner-content* interactions with animations and worked examples. In the animation investigation, students with higher post-test scores had shorter total fixation duration when ordering the events overall, especially planning and solving phases of the problem-solving process. In the planning phase, both groups of participants had patterns between adjacent choices (*e.g.* DCD or FED) without any jumps (*e.g.* ADE). In the problem-solving phase, the incorrect group's patterns were between adjacent choices or steps (*e.g.* ABC or 432) while the correct group's patterns were between choices and corresponding steps.

The final experiment investigated participants' visual attention to conceptual and algorithmic information in worked-out examples using eye tracking. While there was not a significant main effect for AOI type, there was a significant main effect for example type; participants responded differently to the various stimuli. One might infer from the traditional (algorithmic) worked-example that general chemistry students are only interested in the mathematical content to aid problem-solving. The interaction effect demonstrates that there is a significant change in fixation length for participants when the example type incorporates conceptual information or uses conceptual, particulate-level diagrams to explain.

INDEX WORDS: textbook, eye tracking, interaction, worked examples, animation, cognitive load, algorithmic, conceptual, ebook, UTAUT2,

SUGGESTIONS FOR EBOOKS: A MULTIMETHOD CHARACTERIZATION OF UNDERGRADUATE STEM STUDENTS' INTERACTIONS WITH TEXTBOOK FEATURES

by

ELIZABETH L. DAY

BS, University of South Dakota, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

SUGGESTIONS FOR EBOOKS: A MULTIMETHOD CHARACTERIZATION OF UNDERGRADUATE STEM STUDENTS' INTERACTIONS WITH TEXTBOOK FEATURES

by

ELIZABETH L. DAY

Major Professor:	Norbert Pienta
Committee:	Richard Morrison
	Gary Douberly

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

ACKNOWLEDGEMENTS

Dissertations are difficult to write; I thought I would write what I had done, the implications, and the limitations of these investigations, but that proved to be more challenging than I expected. I did not struggle for ideas, and with assistance I enjoyed devising the methods to explore those ideas. I learned the statistics and qualitative analytical techniques necessary to analyze and report my data. I believe the part with which I struggled was looking back at my work—seeing how far I have progressed as a researcher—and knowing I am capable of so much more now.

I would like to thank the contributions of my advisor, Dr. Nobert Pienta, for introducing me to the field of chemical education research and supporting my exploration of topics. This supervision has fostered independence and inquiry that will serve me well in my future research. I thank my committee members Dr. Richard Morrison and Dr. Gary Douberly for their patience and flexible scheduling as well as their insights, which greatly benefitted my research.

I appreciate the support of my friends and family. Dr. Hayley Broussard and Dr. Hunter Hickox were there to see me through the worst moments of impostor syndrome, and I am grateful for the push. My family—parents David and Lynne as well as siblings Daniel, Sarah, and David—was instrumental in providing love, support, good examples, and moving assistance from South Dakota all the way to the Deep South. I also attribute this dissertation's success to the mentorship of Dr. James Calhoun, Dr. Kristen Miller, and Judy Milton.

Finally, I thank the students for their (often too-enthusiastic) participation and willingness to provide feedback on the General Chemistry course; likewise, I appreciate our General Chemistry instructional team for helping my recruitment efforts. I especially appreciate the feedback and opportunities provided by collaborators, the University of Georgia's Scientists Engaged in Education Research (SEER), the chemical education community, and the Pienta research group members, especially Dr. Tang whose expertise in eye tracking and R made this dissertation possible.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

Page

CHAPTER 1

THE PEDAGOGICAL REVOLUTION

Active Learning and the Academy

The modern university was built on the pedagogy of the lecture.[1] However, based on insights from educational psychology and cognitive science, the lecture format has been classified as passive and passé. Educators in science, technology, engineering, and mathematics (STEM) have embraced constructivist learning theories and implemented active learning pedagogies across the STEM disciplines. Courses designed with active learning require student involvement beyond passively receiving the information from a lecturer.[2] In these pedagogies, students are required "to engage cognitively and meaningfully" in the course material and demonstrate this engagement through higher-order cognitive functions such as analyzing, evaluating, and synthesizing.[2]

Active learning pedagogies commonly create opportunities for inductive reasoning in a student-centered inquiry approach to complement (and in extreme cases supplant) the deductive method of lecture. A few examples of these active learning methods include: think-pair-share, cooperative learning, collaborative learning, problem-based learning, and inquiry. A simple, classic pedagogy is the "think-pair-share" discussion in which students individually contemplate a question posed by the instructor before pairing with one or more classmates to discuss their individual brainstorming sessions. The final stage in this model includes a report by the smaller groups to the entire class.[3] Think-pair-share is the least-structured application of the method of learning cycles; for instance, the STAR (Software Technology for Action and Reflection) Legacy model introduces a challenge problem, allows students to brainstorm, iteratively provides more context to be incorporated and assessments for an opportunity to test their explanations, and ends with a presentation of a final model or explanation.[4] These learning cycles rely on problem-based learning (PBL). In PBL pedagogy, a topic- or course-relevant problem is introduced at the start of the learning progression, and it is continually used to provide context for new information and motivation for the groups.[5] Popular applications of PBL are frequently paired with cooperative or collaborative learning. Cooperative learning maximizes instructional benefit by grouping students into teams to accomplish a learning activity through five key principles, including interdependence and accountability.[6] The incorporation of personal accountability into this pedagogy distinguishes it from another popular research topic, collaborative learning.[6] The implementation of this form of group learning activity may be formal (central to the course design) or informal.[6] Many versions of inquiry-based pedagogies fall underneath the umbrella term of inquiry, even those that incorporate cooperative group-based learning as most popularly demonstrated in Process-Oriented Guided Inquiry Learning or POGIL or learning cycles.

The vast literature on the implementation of active learning pedagogies provides enough statistical power to compare across differences in method, discipline, and population. In a 2014 meta-study, Freeman and colleagues explored the efficacy of active or constructivist course designs as compared to passive, lecture-based learning.[7] This yielded two impactful results. The first was that students in a passive environment who scored in the 50th percentile would benefit from an active-course design and would score in the 68th percentile under active learning conditions.[7] The second was that the change in introductory students' examination scores increased by 0.47 standard deviations, which would raise examination scores by six percent and produce a 0.3-point increase to final grades. In terms of the letter-based GPA system, the median score in active learning environments would rise from, for example, a B- to a B.[7] In the analysis of the risk of failing the course between the two conditions, Freeman *et. al.* calculated an odds ratio of 1.95; as an illustration of how to interpret such a ratio, they described how trials of medical interventions would respond to similar results: stop the control (lecture) and place all participants on the overwhelmingly more beneficial experimental treatment (active learning).[7]

This foundation of educational research has provided a space for second-generation educational research questions aimed at the implementation of these findings. There are several overarching challenges to changing the classroom culture to reflect evidence-based practices, even for faculty with the desire and organizational power to enact pedagogical change. First, there is a wealth of literature with unspecific, broad recommendations to "engage students cognitively" or to "encourage meaningful learning" with learning activities.[2] While thoughtful research is carried out with these research questions in mind, some publications give these recommendations without a strong underpinning of the theoretical frameworks that these terms were conceived in. This has led to an ambiguity in meaning and usage, especially as the popularity of a particular term increases, and it becomes more difficult to evaluate reported measures of these educational constructs.[8] Second, there are few criteria to differentiate how "active" a particular engagement is, as well as few criteria on the design and implementation of new cognitive engagements.[2] Furthermore, the Freeman paper invokes an analogy to medical research for a compelling interpretation of the odds ratio, although their meta-study focused specifically on course design rather than specific learning activities.[7] Mimicking this analogy, these findings do not answer the question of dosage: how much (in terms of interactivity) or how many (in terms of number of learning activities) is needed to see this benefit. Finally, there is limited information on how to modify pre-existing activities to optimize engagement.[2] These limitations can hinder faculty from applying evidence-based practices to their existing courses.

Pedagogical Revolution: Flipping the Classroom

Of the active learning pedagogies, the "flipped" classroom model has gained popularity since its first proposal in 2000,[9,10] and it has reached a critical mass in the past few years. Given the multiple names for the pedagogy—including inverted classroom/teaching, reverse teaching, backwards classroom, flip teaching, upside-down classroom, and variants—its popularity is understated.[11] While there is no official protocol for flipping the classroom,[11] formal class-time is used to provide active learning and problem-solving opportunities with the instructor acting as a guide.[12] In the flipped model, students are expected to prepare for class by reviewing the information that historically has been presented in direct instruction.[13] This content delivery is accomplished with hard-copy textbooks and multimedia presentations (which are often, but not exclusively, created by the instructor).[9] Some flipped courses include self-assessment opportunities in the form of homework or quizzes, which allows the dedication of synchronous course meetings to interactive problem-solving sessions or an opportunity to ask clarifying questions of the instructor.[14]

This redistribution of learning activities increases demands on students' time and effort, but in a study by Smith, the majority of students report that this reorganization made the course more engaging.[14] Because of these increased demands of the student, flipping the classroom establishes an andragogical approach to introductory courses; the locus of responsibility for a student's learning no longer lies solely with the instructor, but rather the instructor plays a role of encouragement and guidance of the learner's increasing self-directedness.[15] The expanding role of technology, such as learning management systems (LMS), to aid in this revised schedule of learning activities further blurs the line between flipped/inverted methods and hybrid course structures. A hybrid (or blended) course model involves some ratio of classroom interaction to online, asynchronous interaction.[16] At the college level, the use of personal-response devices (clickers) to ensure class attendance is one of few deterrents employed to prevent students from treating the flipped classroom as a hybrid course. Furthermore, these clickers provide opportunities to spark inquiry between instructors and students, and learning technology is necessary as a means of prompting instructional feedback for pre-class preparations as well.[16]

This pedagogical revolution has generated demands for adaptive, data-driven learning technologies in addition to complementary, interactive delivery formats for the traditional lecture materials, such as electronic textbooks (ebooks).[10] As the textbook becomes one of the main content-delivery tools, it becomes increasingly important to evaluate the function and value of a textbook (and its features), especially as the STEM community transitions to new teaching models and learning technologies. One study noted that sacrificing course meeting time to give students an opportunity to ask questions of their instructor on pre-prepared lecture material mitigates some of the benefit of the flipped model; however, this action was crucial "as it helps overcome the missing student-teacher interaction piece that is potentially present in the traditional lecture setting."[14]

Theoretical Frameworks

Active learning methods are grounded in the constructivist epistemological view. Constructivism is a learning theory that describes the process of learning as the construction of knowledge in the mind of the learner.[17] This theory is an instrumentalist view of the learner-centered search for knowledge.[17,18] One influence, Jean Piaget, conceptualized the key mechanisms for the construction of knowledge by a learner. The passive mechanism is assimilation—the incorporation of new knowledge from the environment into students' pre-existing conceptions.[18,19] This pathway for knowledge construction is the default mechanism until the cognitive demand of the task overwhelms the existing knowledge structure and triggers the reorganization of the cognitive structure to accommodate new information.[19] Although constructivism has its roots in Piaget's theory of intellectual development, constructivist learning theories recognize that an instructor's actions may influence the learner's experience and thus incorporate the central tenet of Vygotsky's social constructivism.[18]

Vygotsky emphasized that social interactions are crucial to the development of higher-order cognitive functions.[18,20] Unlike Piaget's focus on the learner's interaction with the environmental stimuli, Vygotsky hypothesizes that an individual is limited to a Zone of Proximal Development, a boundary phase on the ability to learn without the interaction with peers or experts.[20] Furthermore, while information learned passively is less memorable than knowledge constructed individually, effective direct instruction or interaction with peers may guide and support learning when the discovery learning process becomes lengthy or discouraging.[2,19] The learning process within the individual consists of adding information to and reorganizing cognitive structures,[19] which can be difficult to measure directly. Instead, pedagogical approaches have focused on the actions of the instructor, the instructor's learning activities, and student engagement with these activities.[2]

In an active learning classroom, the role of the instructor shifts from a central dispensary of knowledge to a manager;[12] the design of instructor's assistant manager has become online content-delivery systems (lecture notes, LMS, ebooks, online homework, etc.). This places responsibility on the student to prepare for class and assessments with minimal personal feedback, especially in large class sizes. In the process, the line between distance education (in which hybrid/blended course formats are abundant) and on-campus education has blurred. Therefore, a look at the guiding frameworks of distance education can inform how best to transition to a computer-mediated learning environment.

Although more complex and increasingly distance-learning specific models of interaction exist, Anderson's expansion of Moore's theory of three primary forms of interaction can serve as an organization of this work. Moore (1989) hypothesizes that deep and meaningful learning (although not explicitly in the Ausubelian sense)[21] is based on three types of interaction: *learner-content*, *learner-instructor*, and *learner-learner*.[22] Anderson's equivalency theorem (2003) suggests that although multiple types of interaction can lead to a better leaning experience, a high level of one type of interaction may compensate for missing or poor levels in the other two types of interaction,[23] although this principle of substitutivity does not appear to hold up to a meta-analysis.[24] Anderson (2003) also expanded Moore's model to include *instructor-content*, *instructor-instructor* (like professional development networks), and *content-content* interactions. This expanded model is visualized in Figure 1.

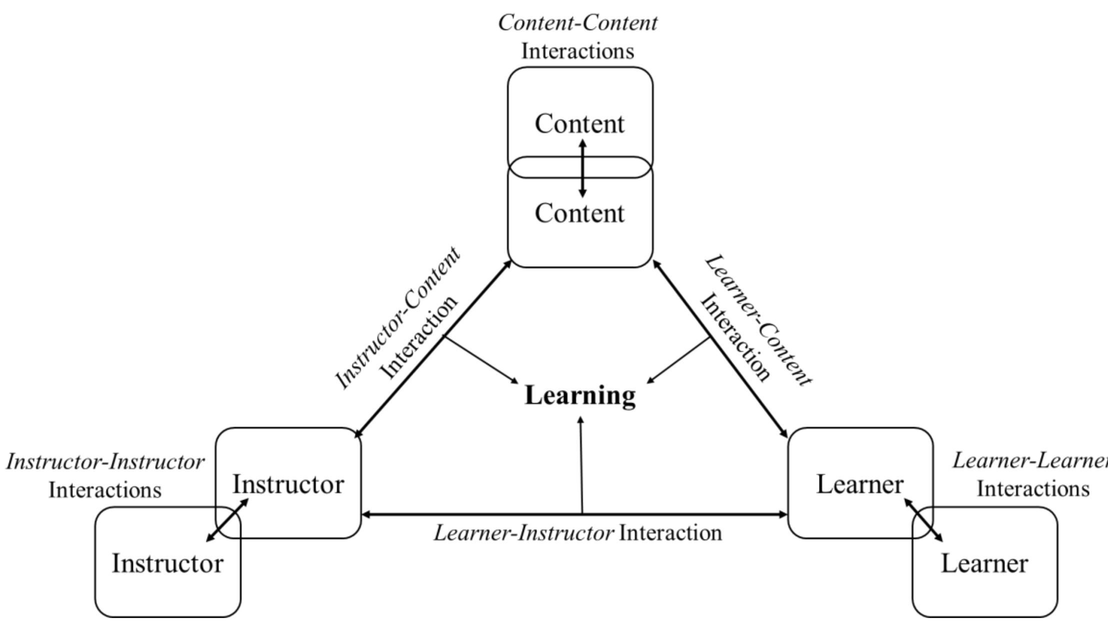


Figure 1: an illustration of Moore's Theory of Interaction with Anderson's expansions of the theory.[22,23] *Adapted from a review by Madland and Richards, but edited to reflect original terminology of theory.*[25]

In distance education (and increasingly in active learning environments), *learner-instructor* interaction is embedded into the choice of content-delivery device and therefore subsumed by the learner-content interaction. While this critical *learner-content* interaction is theorized to be the "defining characteristic of education" and responsible for restructuring conceptual organization within the learner,[22] *learner-instructor* interaction is an expected, highly-valued interaction in modern education.[24] Although consolidation of the *learner-instructor* interaction into the course content materials produces uniform delivery, the social and mentoring function of the *learner-instructor* interaction has not been appreciably replicated by the content-delivery systems. This deficiency stands in direct contrast to the observation by Smith (2013) about the value of sacrificing content coverage in a flipped classroom to devote synchronous course time to the *learner-instructor* interaction that can be difficult to incorporate in flipped instructional models.[14] As Xiao (2017) notes, "[u]nless reciprocal interpersonal interaction is built on learners' interaction with course content, its impact on learners' academic achievement and perceptions […] may not be as positive as desired."[24]

Education literature has studied *learner-learner* and *learner-instructor* interactions; evidence of the efficacy and implementation of POGIL, collaborative, and cooperative learning falls under *learner-learner* interactions, whereas development of classroom observation protocols seek to characterize *learner-instructor* interactions.[26–28] While there has been work on student understanding of the content, there has been fewer studies focusing on the interaction between learners and the assigned content-delivery devices. With the increased student responsibility on preparation of content, there is a vital interest in how learners make meaning of the content to inform instructor's and content developer's decisions about the structure and level of interactivity of that content.

Structure of Dissertation

This dissertation is comprised of three areas of focus: student use of textbooks and other course resources, the effect of animation or static images on problem-solving, and the visual perusal of worked-out example problems. Because each area has its own distinct swath of literature, the background and project-specific theoretical framework for each focus area will be included in the relevant chapter. A general discussion of the educational research methods used (surveys, eye-tracking, and think-aloud interviews) and data analyses (significance testing and coding qualitative data) will be discussed in Chapter 2. Specifics about each technique and the analyses used will be included in the relevant chapter for each project.

This project structure arose out of the work on students' use of course resources. A pilot study on textbooks and their perceived value relative to other course resources during a transition to an ebook revealed two further areas of inquiry: the increased value of visualizations and the decreased value of worked-out examples. The overarching goal of this dissertation work is to characterize *learner-content* interactions in terms of (i) which content-delivery devices do students prioritize or value and (ii) how can those features be optimized for an electronic textbook?

CHAPTER 2

AN OVERVIEW OF EDUCATIONAL RESEARCH METHODS

This chapter will begin with an overview of data analysis methods. Quantitative analysis, both string-edit methods and statistical, can be used to analyze surveys and eye-tracking data. The qualitative analysis methods will describe various approaches to coding responses to surveys and interview transcripts.

Quantitative Analysis

Many of the data collection methods listed below yield quantitative data, which can be summarized through descriptive statistics or condensed into character strings and analyzed statistically. Classical education research relies on statistical methods, particularly significance testing, to evaluate performance of different groups or an intervention. In this work, the various types of t-tests (parametric, non-parametric, and the paired t-test), Analysis of Variance (parametric or the non-parametric Kruskal-Wallis), and regression methods were used. The p-values of this type of null hypothesis significance testing (NHST) are reported with effect sizes as standardized measure of the magnitude of significance.[29]

In educational research, the most common tool in NHST is the independent-samples t-test, which is used to evaluate whether the means on a dependent variable of participants assigned to two experimental conditions are different at a specified criterion (typically an alpha-level of 0.05). As the observed difference between the sample means varies from the expected difference between the populations, there is a greater likelihood that the sample means are different because of the experimental conditions.[29] By contrast, paired t-tests compare the response variable means from a single group of participants who are subjected to two experimental treatments, such as a pre-/post-test measure. In this case, as the deviation of the difference in sample means increases from the expectation of "no difference" in population means, there is a higher likelihood that there is a significant difference in response variable scores.[29] These t-tests are parametric tests, which rely on the assumptions of a normal sampling distribution of independent observations.[29] If the dataset produces a non-normal distribution, a non-parametric t-test called the Mann-Whitney U uses rank scores (as opposed to sample means) to achieve test statistics and p-values.

NHST relies on p-values to indicate whether there is a significant effect at a specified level, but the magnitude of the p-value does not describe the magnitude of a significant effect. Instead, social, behavioral, and educational research groups report effect sizes as an objective description of how significant an observed effect is.[29] The effect size for a t-test is commonly reported as a Cohen's *d*, Glass's Δ, and Hedges's *g*. Cohen's *d* is a popular choice, and is most appropriately used when the two conditions produce datasets with similar variance because the use of standard deviation in its calculation:[30]

$$\hat{d} = \frac{\bar{X}_1 - \bar{X}_2}{s} \tag{1}$$

In which X_i represents the means and s represents the sample standard deviation.[29] This equation calculates d-hat, which is the sample's effect size and serves as an estimate of the true effect size on the population.[29] While this calculation assumes that the two samples' variances (and standard deviations) are similar, if there is a large difference in the two groups' standard deviations, a researcher can instead calculate Glass's delta using the control group's standard deviation ($\sigma_{control}$) (if applicable) or to calculate Hedges's *g* through the pooled standard deviation (σ_{pooled}) to be used in place of the standard deviation.[29,30] Interpretation of a Cohen's *d* effect size generally follows the prescription that a small effect is expressed as *d* less than 0.2, medium falls in the range of 0.21 to 0.79, and large effects sizes can range from 0.8 up to a value of 2.0.[30]

Analysis of Variance (ANOVA) returns an *F*-ratio of the variance of the response (typically sample means) explained by a linear regression model to the variance explained by unsystematic factors; in the context of comparison of group means, the magnitude of the ratio will determine whether groups belong to the same population (large *F*-ratio) or not (small *F*-ratio).[29] Unlike a t-test, an ANOVA can compare more than two sample means, although a large *F*-ratio amongst many experimental conditions only signifies that a significant difference exists and not between which groups there is the significant difference. Attempts to determine where the significance lies may be determined through careful application of post-hoc analyses.[29] Similar to a t-test, a one-way ANOVA assumes homogeneous variances and a normal distribution of scores *within a group*, although it is robust to violations of non-normality if the sample sizes are equivalent.[29] ANOVA has a non-parametric substitution, in the case of the violation of normality, called a Kruskal Wallis and alternatives in the Welch's F and Brown-Forsythe F tests to account for heterogeneity of variance and unequal sample sizes.[29] The effect size for an ANOVA is calculated commonly through η^2 or ω^2 values. In this dissertation, effect sizes are reported as ω^2 and are calculated as:

$$\omega^2 = \frac{SS_M - (df_M \times MS_R)}{SS_T + MS_R} \tag{2}$$

In which SS_M is the sum of squares of the model (also described as "between groups"), df_M is the degrees of freedom of the model, MS_R is the mean square of the residual (also described as "within groups"), and SS_T is the total sum of squares.[29] Omega-squared values of 0.01 are considered small, 0.06 are considered medium, and 0.14 and greater are considered large effect sizes.[29]

While one-way ANOVA is a robust method for comparing multiple response variables for the same group of participants, one of the eye-tracking studies produced a dataset in which all participants contributed multiple responses (and therefore a need for repeated measures design) as well as multiple predictor variables. For this study, a mixed design factorial ANOVA can compare the responses of each group on multiple repeated-measures dependent variables.[29] The assumptions for ANOVA are that the data in the dependent variables are normally distributed with a homogeneity of variance and sphericity (ε), a type of compound symmetry in which the variance of differences in-between groups is homogeneous.[29] If the Mauchly's test of sphericity is significant, using the Greenhouse-Geisser estimate of sphericity will reduce the degrees of freedom in the ANOVA and control the type I (false positive) error rate by reporting a more conservative p-value. The results of a factorial design ANOVA will report whether the main effects of each independent variable are significant, as well as the significance of any interaction term specified from combinations of independent variables. To elucidate further meaning from these results, within-subjects contrasts compare each level of the independent variables against each other, and effect sizes are reported with the contrasts (as opposed to the mixed-design factorial ANOVA itself, for clarity's sake). The effect size of contrasts is *r*, computed from the F-ratios of the ANOVA:

$$\mathrm{r} = \sqrt{\frac{F(1,\mathrm{df}_R)}{F(1,df_R) + df_R}} \qquad (3)$$

In which the df_R indicates the degrees of freedom from the residuals (error term) of the model.[29] An r-value of greater than 0.5 is considered a large effect.

The t-test and ANOVA statistical methods are applications of a regression model. Regression is a method of fitting the outcome variable to the identified predictor variables to return parameters (regression coefficients) that quantify the relationship between the outcome and predictor variables; these coefficients can be used to determine the existence of significant relationships between the outcome and one or more predictors.[29] Of the various types of regression, a logistic regression quantifies the relationship between a categorical outcome and predictor variables that are categorical or continuous in nature. In Chapter 4, an ordered logistic regression is used to predict the probability of an outcome with more than two levels given a set of predictor variables.[29,31]

Other quantitative measures used in this dissertation analyzed character strings—generated from eye-tracking or qualitative codes—through string-edit methods. String-edit analyses quantify how similar character strings are through the calculation of the minimum number of operations needed to align one string with a hypothesized or another character string. Most commonly-reported is the Levenshtein distance, which is accomplished through insertions, deletions, and substitutions, although some variants may only allow for certain operations or might include transposition of characters as an operator.[32]

Qualitative Analysis

Qualitative techniques (such as classroom observations, interviews, journal/writing assignments, and focus groups) provide insight on the motivation or reasoning behind behaviors and, as a result, produce datasets that are often non-numerical and require a different analytical approach.[33] Of these methods, the interview produces the richest dataset, but requires a rigorous approach to provide meaningful conclusions. The most common interview format is an open-ended approach, in which the interviewer has prepared a guide of questions that range on a spectrum from totally unstructured to a closed-answer style, as seen in Figure 2.[34]

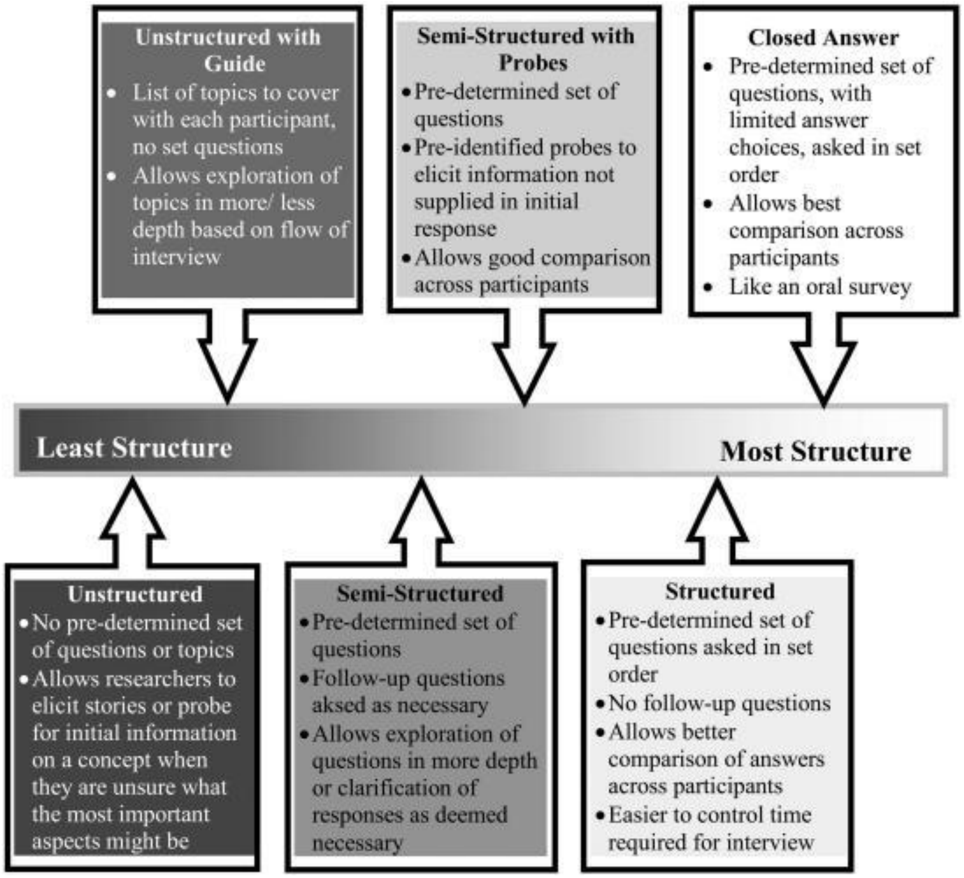


Figure 2: levels of structure of the interview guide for open-ended interviews, reproduced from Herrington and Daubenmire.[34]

The other interview format (as will be utilized in Chapter 5) is a think-aloud interview, which may vary in the level of structure as well. Rather than asking participants to respond to a prompt, the participant is instructed to verbalize their thoughts while performing a specific task. The interviewer interjects only to encourage articulation or to delve into a participants' description of their thought process. As a result, this method of interviewing focuses on research questions that address a participants' understanding, reasoning, and use of algorithms or heuristics to complete a task.[34] One potential pitfall of this approach is the tendency to describe the stimulus rather than one's problem-solving approach, resulting in a participants' reasoning after (not during) problem-solving. Some research protocols, such as viewing eye movements, may favor a retrospective think-aloud approach by encouraging participants to view their behavior and provide explanation after completion of the task.[35] In most problem-solving oriented research questions, a concurrent think-aloud approach offers the best vantage point as a participant's problem-solving approach unfolds.

Once a think-aloud interview approach is selected, careful consideration of the task or stimulus is as essential as a well-crafted interview guide. Ideally, the task is accessible to the participant and self-explanatory (because direct instruction would render the resulting dataset useless). With the goal of unprompted, uninterrupted articulation of task-performance, the task or stimulus must be challenging enough to elicit a participant's thinking without being burdensome or too time-consuming.[34]

Unlike quantitative methods, sample size in a qualitative study is not reliant on statistical power because the general purpose of the dataset is not to quantify the frequency of responses or utterances.[36] Instead, sampling techniques are based on the notion of saturation, the point at which novel responses to prompts or questions are no longer detected.[34] The initial selection of participants can fall into the categories of convenience, judgment, or theoretical sampling. A convenience sample recruits participants in a manner that is the least costly and time-intensive to the researcher, but the results of this less rigorous technique are sometimes a poor fit to the constructs under study.[37] A purposeful or judgment sample seeks participants who are most likely to provide insight on the underlying constructs; although demographic variables are a contributing factor, they are not the sole determination in a judgement sampling technique. Underneath this approach to sampling, researchers may recruit a broad swath of the population (termed a maximum variation sample), target the outliers (also known as a deviant sample), or focus on participants with specific experiences or expertise (the critical case and key informant samples, respectively).[37] Underneath a purposeful sampling technique, it is common to ask the participants to suggest other subjects who are eligible for study, commonly referred to as a snowball sample. This set of techniques provides the richest dataset with perspectives that confirm and disagree with the theory of the study. A theoretical sample, the basis of a grounded theory approach, iteratively recruits participants in response to the emergent themes from the dataset to examine newly-formed constructs; this approach is utilized to generate new theories, rather than test the validity of a previously-determined theoretical framework.[37]

The choice of analysis technique for a dataset is determined largely by the researcher's theoretical framework and research question. If the researcher is designing an experiment to explore the basis of a pre-existing theoretical framework, the constructs of that framework provide the codes that are assigned to the text-based data (such as the analysis in Chapter 3). In the studies that do not have a theoretical framework to provide codes, a grounded theory approach can be employed to generate a theory from the qualitative dataset. Based on pragmatism and symbolic interactionism, the two guiding principles of grounded theory are the dynamic nature of phenomena in response to sociocultural and environmental conditions as well as the deterministic view that the subjects of the study are actors who are capable of controlling outcomes in response to these conditions.[38] In such an approach, a theory is evolved through iterative processes of data collection and analysis to derive concepts from the data. Categories of codes that emerge from the data are reviewed and grouped until they are reduced to themes that characterize the content of the data.[39] This approach will be utilized in Chapter 5 to analyze think-aloud interview data.

Eye Tracking

A thorough discussion of eye-tracking would span multiple chapters; therefore, this section aims first to introduce the eye-mind hypothesis that connects the physiology of the eye to the foundations of working memory and cognition. Then, the language and schematic organization of an eye-tracking experiment will be briefly described. This section will end with a brief description of the eye-movement metrics and analyses, including scanpath pattern analysis.

Light reflected from objects in the physical world is converted and interpreted as visual sensory information through the optic nerve of the human visual system. As seen in Figure 3,[40] light passes through the cornea that covers the pupillary opening and is focused on the retina.[41]

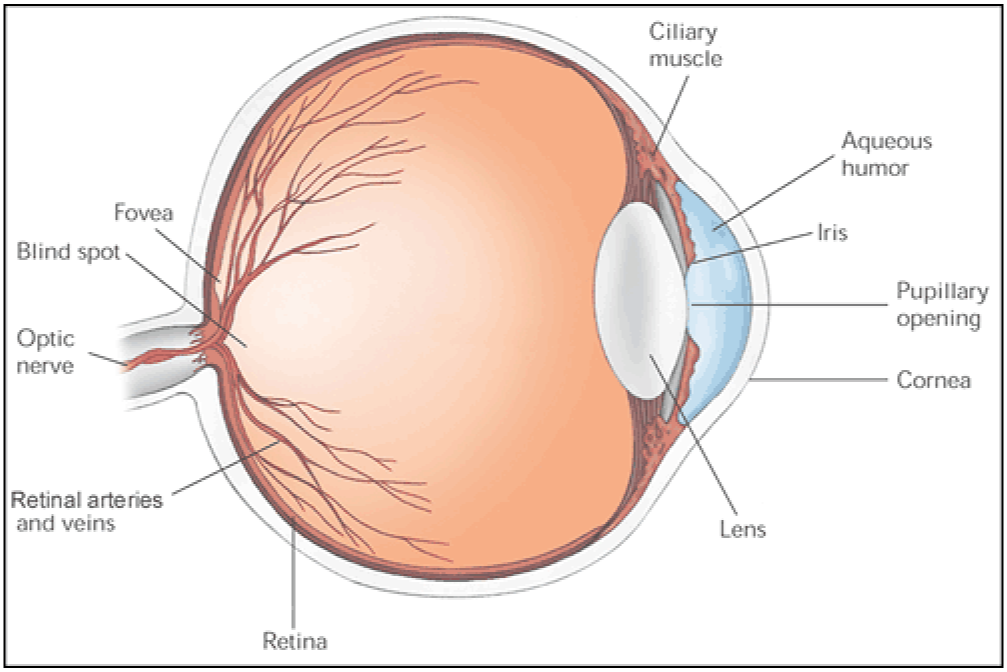


Figure 3: graphic depiction of the anatomy of the human eye.[40]

The light-sensitive nerve cells in this retina and fovea convert the projected image to electrical impulses; cones process bright light, perceive and discriminate between colors, and detect rapid movements in the scene, and rods resolve spatial relationships and shapes or serve as low-light or peripheral vision.[41,42] The relatively-narrow (1-5° angle of visual field) innermost region is the fovea centralis (commonly referred to as fovea), in which the vision is in sharpest focus due to the highest density of cone-type photoreceptors.[41] The visual architecture can attend to a visual stimulus through repositioning of the fovea through fast eye movements (10-100 milliseconds) called saccades.[41] Stabilization of the foveal region over a stationary visual stimulus is termed a fixation, and these pauses in the eye movements provide a metric by which to characterize a participant's visual interest. This is the eye-mind hypothesis—that the object under fixation is being processed visually and therefore the time spent fixating is a measure of the time spent processing. Similarly, there is an assumption of immediacy in which the participant's processes of the scene before the saccade to the next fixation.[41]

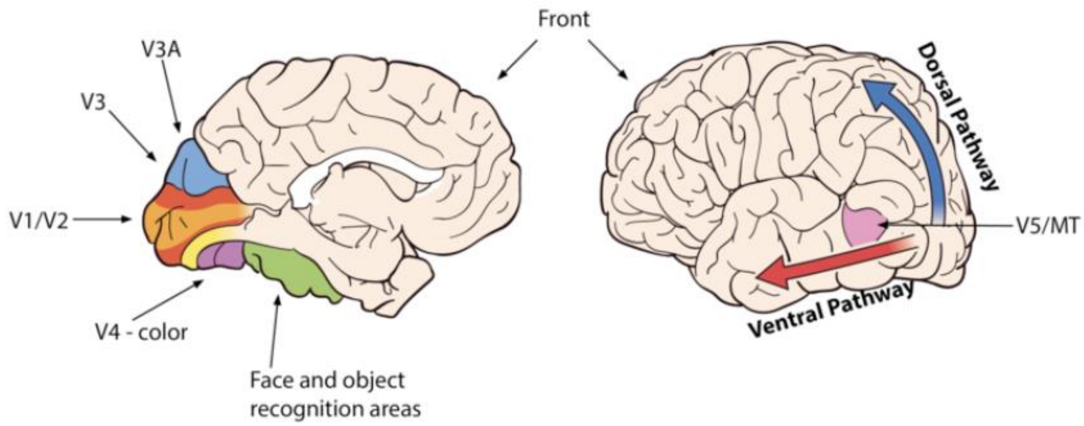


Figure 4: illustration of the visual pathways of the brain, reproduced from the NOBA project.[43]

As illustrated in Figure 4, the electrical impulses generated by the photoreceptors in the retinal and foveal regions of the eye are processed in the occipital lobe by the primary visual cortex V1 (in which orientation and some chromatic differentiation) occurs before being passed to areas V2 and V4 for further differentiation of form, color, and motion.[42] From the V1 area, impulses in the dorsal stream of information utilize the posterior parietal complex (PPC) to direct attention of sensory and location-based stimuli; the PPC disengages visual attention from a stimulus, the superior colliculus (SC) targets eye movements to a new stimulus, and the Pulvinar region re-engages (and sometimes enhances) attention to a visual stimulus.[42] At the same time, impulses from V1 are transmitted via the ventral stream of processing to the inferotemporal complex, which holds the cognitive processing; these streams of information interact and "cross-talk" significantly, creating a feedback system of visual attention.[42]

Although eye-tracking technology focuses on the eye movements generated by the dorsal stream, the implications of these metrics are attributed to the ventral stream of cognitive processing. Simplistic models of the cognitive architecture include information processing (IP) theory, which describes the cognitive states (called stores or memories) and the retention and discard processes of handling stimuli from the sensory nervous system. This theory describes a network of information that increases in size, inter-association, and complexity with more experience.[44] The inactive nodes serve as a long-term memory to permanently store information, although they can become activated and serve as a short-term store of memory. This activation allows these nodes to control the processing of information through rehearsal, elaboration, and encoding, but any information not encoded to long-term nodes is forgotten from these activated, short-term stores.[45]

More recent theories have revised this short-term memory into a multicomponent working memory. The working memory is an executive processor that controls attention to sensory information through the "subsidiary slave systems" dedicated to auditory stimuli and visual imagery.[46–48] The implications of these cognitive processing theories will be discussed in greater detail in Chapter 4.

Eye tracking has been widely used in usability, literacy, psychology, education, neurology, visual search, and numerous sub-disciplines.[35,41,42] As a consequence of the technique's wide-ranging experimental applications, a diverse range of eye-tracking apparatuses are available commercially.[42] These instruments can vary on apparatus (head-mounted, table-mounted) and incorporation of tracking head position relative to eye position.[41] The hardware consists of a (typically infrared) light source, a video camera to detect reflections, and software that can map the change in position of the reflections to the image.[41] In this work, a table-mounted video-based eye tracker utilized the corneal reflection of infrared light as a point of reference for image-processing algorithms to map a participant's gaze to the stimulus.[41]

Eye-tracking data consists of various eye movements, as outlined by the non-exhaustive list in Table 1.[35,41] The majority of data in this dissertation will consist of fixations and saccades. As demonstrated in previous studies, eye-movement metrics on the same area-of-interest (AOI) and stimulus are often highly correlated ($r \geq 0.9$).[31,49–51]

Table 1: Eye Movement Metrics

Movement	*Metric*	*Unit*	*Definition*
Fixation	Fixation Duration	s	Fovea stabilizes and focuses attention on an area of interest on a stimulus; total amount of time spent on an area of interest before resuming movement
	Total Fixation Duration	s	Total amount of time spent on an area of interest, including return visits
Saccade	-	s	Rapid movement to move fovea to a new position on the stimulus
Smooth Pursuit	-		Slow, steady movement of fovea when visually tracking a moving stimulus
Nystagmus	-	-	A mixture of smooth pursuits and saccades resulting in a saw-tooth pattern
Microsaccade	-		Minute eye movements that comprise fixations as the fovea attempts to stabilize
Tremor	-		
Drift	-		

From the output of the eye tracker, statistical tests or string-edit manipulations can be performed on the eye-movement metrics to determine relationships between participants, between AOIs, or patterns of visual behavior. Significant relationship between metrics and conditions is often attributed to burdensome cognitive load.

Fixation duration and fixation count are commonly-reported metrics that can indicate cognitive load. Fixation duration and fixation count are correlated to the difficulty and complexity of the visual stimuli and can reflect cognitive load;[52,53] duration or higher count implies more complex material under visual perusal.[54] Similarly, experts' shorter fixation durations have confirmed the assumption that expertise yields faster and more efficient encoding and retrieval of information than novices.[55] A meta-analysis of eye-tracking studies also confirmed the theory that expert-like schemata optimize information processing; experts had shorter fixations on task-irrelevant information than novices, and experts spent longer fixating on information that was relevant to the task (as compared to novices).[55]

Scanpath Analysis

A scanpath (or gaze sequence) is a pattern of fixations and saccades constructed from the path of eye movements over a certain timespan.[35,41] Multiple visits to an area of interest (AOI) in these sequences may indicate features that require more cognitive processing from the viewer.[56] Therefore, experts tend to have shorter, more efficient/focused scanpaths than novices.[57] From the scanpaths, shorter sub-sequences called transitions may be defined and used to measure cognitive load. In the design of educational materials, the number of transitions between text and picture is related to complexity of stimulus.[35]

Unlike singular eye-movement metrics, scanpaths reveal the spatial and sequential allocation of attention to a stimulus.[58] Scanpaths can be compared via a dissimilarity calculation, which employs the string-edit method. Fewer edits indicate strings with more similarity and would yield a smaller Levenshtein distance.[59] Because the strings in an analysis are frequently of different lengths, a calculation of the normalized string edit distance ($\hat{d}$) from the distance (d) and the maximum string length, as seen in Equation 4.[35]

$$\hat{d} = 1 - \frac{d}{\text{maximum string length }(m, n)} \tag{4}$$

A transition is most often defined as a gaze shift from one AOI to another and is modelled as a 2-character string. Within the group of AOI strings, transitions between any combination of AOIs can also be modelled through a transition matrix.[35,59] From this matrix of transitions, a frequency-ordered list of transitions can be used to calculate the entropy of transitions as shown in Equation 5.

$$H = -\sum_{i=1}^{n} freqs_i \times \log_2 \left(freqs_i \right) \tag{5}$$

In which $freqs_i$ is the *ith* frequency of a particular transition in a string.[35] This calculation can be performed on an individual scanpath using an ordered list of transitions within a single string or within a group using an ordered list of transitions from a group of strings. The entropy value is an indication of the dispersion of fixation distributions on a stimulus; a larger entropy value indicates equal probabilities of transitions between any two AOIs.[60]

Entropy is appropriate to calculate when a research question calls for a quantification of a participants' eye movements to report as a single value. Commonly interpreted using the eye-mind hypothesis, smaller values of entropy indicate more targeted scanpaths, such as those displayed by experts. Large values of transition entropy imply a greater dispersity of transitions and reveal search behaviors for AOIs on the stimuli that are far apart physically. Therefore, a large entropy value is an indication of higher cognitive load and/or novice-like visual behavior.[58]

CHAPTER 3

LONGITUDINAL SURVEY OF STUDENT RESOURCE USE

From the instructor's point-of-view, the purpose of the textbook is to serve as a multipurpose reference to deliver in-depth conceptual explanation of the lecture content, and it is commonly used by students to find problem-solving algorithms—and (ideally) to gain conceptual understanding of the lecture content.[61] The instructor-chosen textbook often drives the curriculum, especially for instructors with heavy teaching loads. Over the past 90 years, the traditional hard-copy text that has been used to train future chemists has undergone a curricular change from purely descriptive chemistry to a "theory-first" presentation.[62,63] Even with the recent popularity of the flipped model, an adaptation of the textbook to meet the current pedagogical revolution has been slow.[64]

Despite the central role of the textbook, the prices of traditional college textbooks can be prohibitive. A Government Accountability Office (GAO) report in 2005 noted that between December 1986 and December 2004, the prices of textbooks increased 186 percent (twice the rate of inflation).[65] The increase in textbook prices mirrors the overall trend of rising tuition costs, as shown in Figure 5.[65] Another GAO report in 2013 noted that—from 2002 to 2012—textbook prices have further increased by 82 percent;.[66] this increase is shown in Figure 6. One potential cost-saving option for students is an electronic book (ebook),[10,67,68] especially with the availability of open-source ebooks.[12,64]

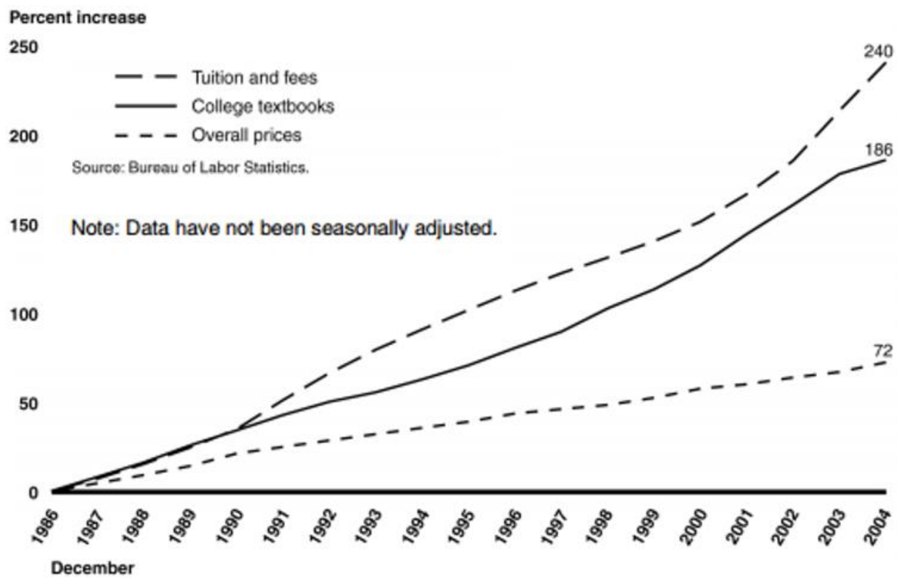


Figure 5: The percent increase in the price of textbooks, tuition/fees, and overall consumer spending from December 1986 to December 2004.[65] *Over this period, textbook prices increased 186%, double the rate of the 72% increase in overall prices. This increase in textbook prices mirrors the trend for the increase in tuition and fees for higher learning.*

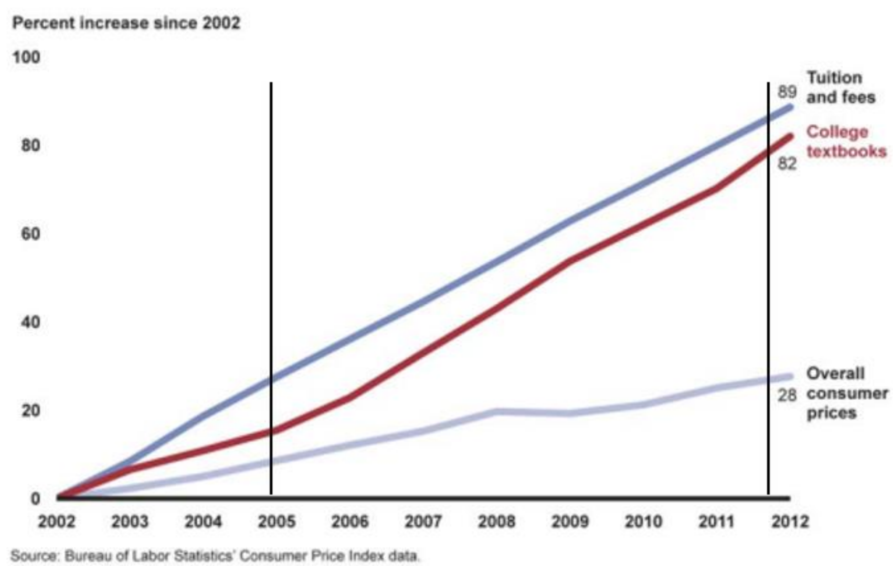


Figure 6: The percent increase in prices of college textbooks, tuition/fees, and overall consumer prices from 2002 to 2012.[66] *During this period of 2002 to 2012, textbook prices increase 82% overall, at an average rate of 6% per year. The vertical black lines are annotation to indicate visually how prices have increased since the first GAO report in 2005.*

Since the early 2000s, several studies have documented the adoption of ebooks into classrooms.[10,64,69,70] The 2013 GAO report notes that the financial discount of an ebook is similar to that of a used textbook or renting a textbook for a semester; ebooks are often discounted up to 50 percent of the new textbook price.[66] However, the price of "renting" a digital file from the publisher benefits the publisher far more than the student.[69] The constraints of digital rights management include expiring access after a certain amount of time, limits on the number of pages that can be printed, or a limit on the number times the book can be downloaded to any device; these issues present some challenges to the student consumer.[69] Other disadvantages include the tendency of the reader to skim material when presented electronically—rather than read deeply—as well as technical access concerns.[10,70] The slow adoption of ebooks into college chemistry classrooms might be attributed to electronic platform concerns,[10]. Despite these disadvantages, ebooks are cheaper, more environmentally-friendly, and offer more interactive possibilities. This interaction can be mathemagenic (in the form of embedded questions, videos, or intuitive "tutoring" feedback) or generative (such as concept maps or study guides).[10,24]

The hefty price tag and the role of the textbook imply an instructor's expectation that students utilize the text frequently and effectively. Although students and instructors in one study both perceived the function of the textbook to be primarily for assigning homework, their perceptions differed on the extent of the textbook's role in the classroom.[71] This study by Tulip and Cook found that instructors viewed the textbook as a guide of how deeply and in what order to structure the course, a common practice that may not align with sound pedagogy.[71] The textbook market is driven by the profit of cost-effective updates, but the role of textbook author is not valued as genuine authorship in the way that literature or scientific research is.[72] This removal of the author's perspective and incentive to innovate facilitates a reputation as an empirical source of knowledge, which is an implicit statement of the values, culture, as well as the ontology and epistemology of the scientific discipline.[72] This prevailing approach to textbook generation has instilled a "passive consumption" pedagogical model that places a learner in a constrained, subordinate learning environment in which the target is knowledge that reflects the information on the final exam.[72]

Theoretical Frameworks

Given that the present population of students are part of the "digital-native" generation, an instructor may expect that the ebook be easily adopted.[10] The Unified Theory of the Acceptance and Use of Technology (UTAUT2) describes factors that predict consumers' technology use.[73] Although this theory was conceptualized originally around technology adoption in organizational contexts (UTAUT) and consumer contexts (UTAUT2), the factors that contribute to the Behavioral Intention to adopt technology can be applied to the open-ended survey items regarding students' format preference. The original theory (UTAUT) only identified external factors of performance expectancy, effort expectancy, facilitating conditions, and social influence as constructs, but the latest iteration in the UTAUT2 has further identified hedonic motivation, habit, and price value as co-contributors to the behavioral intention that determines a consumer's use behavior.[73] This is illustrated in Figure 7.[73]

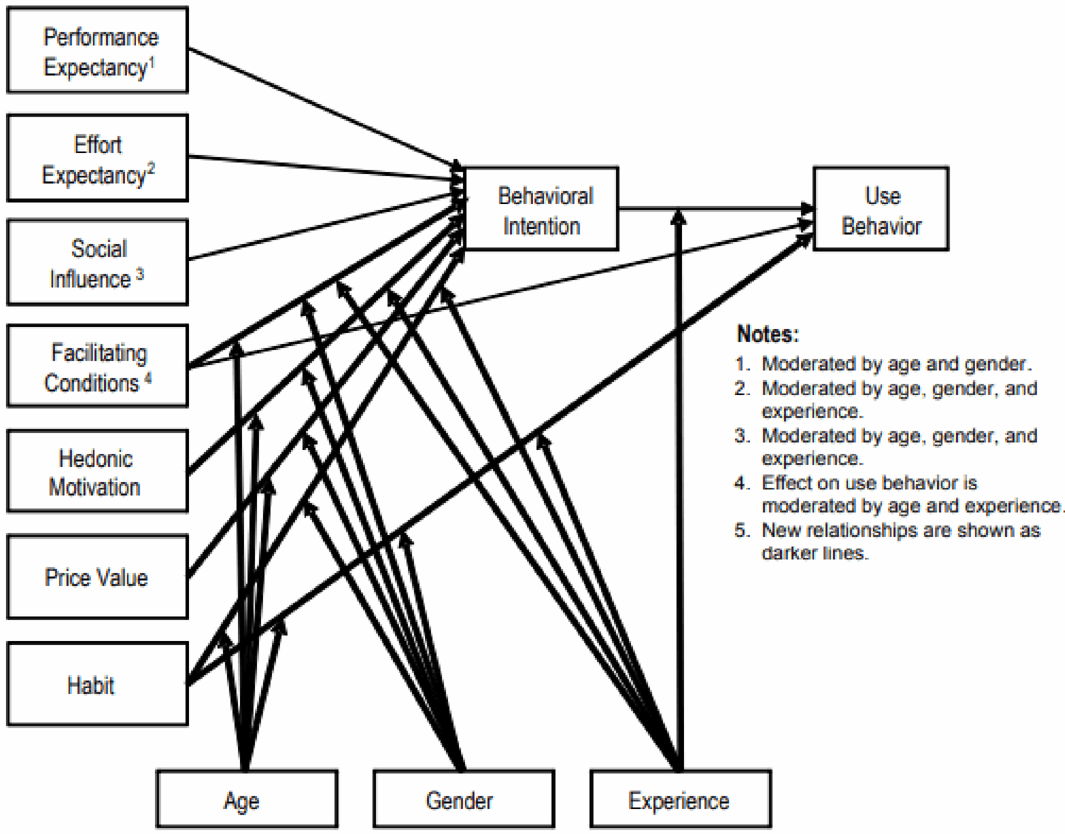


Figure 7: Venkatesh et. al.'s Unified Theory of Acceptance and Use of Technology 2 model.[73]

Of the constructs in this model, performance expectancy is the strongest predictor of technology acceptance; it is described as "the degree to which using a technology will provide benefits to consumers in performing certain activities."[73] Effort expectancy is another utilitarian predictor of behavior and defines the "degree of ease associated with consumers' use of technology."[73] The opinions of important personal contacts (friends, family, mentors, and professors) are incorporated in the construct of social influence, or how consumers perceive others emphasize the use of a particular technology. The consumers' perception of the resources and support available to them to perform a behavior with adopted technology is termed facilitating conditions in this model, and this construct can be influenced by the age, gender, and experience of the consumer. Other constructs that are mediated by age, gender, and experience are hedonic motivation (the perceived enjoyment in using the technology), price value (the consumer's cost-benefit analysis), and habit (the "extent to which an individual believes the behavior to be automatic").[73] These constructs inform a consumer's intention to adopt and therefore use a technology.

Literature Review

The evolution of the modern general chemistry textbook is evident in the changes in page number as well as the complexity of the theoretical chemistry invoked to describe core principles. It represents a paradigm shift from small books of descriptive chemistry and tables of experimentally-derived values to massive encyclopedias of theoretical principles and algorithmic-driven problem-solving. The effect on the student, as Lloyd (1992) notes, is a belief that solving a mathematical word problem is equivalent to "learning what chemistry the problems represent."[62] The modern textbook's common ancestor with descriptive handbooks of the past is Sienko and Plane's 1961 "manifesto" for communicating general chemistry.[74] This popular theory-first textbook has a heavy influence on the field of introductory-level textbooks in chemistry.[63]

Despite the instructor's assumption of wide-spread student use, there are few studies on textbooks to validate this premise. The topics of publications on textbooks in chemistry have focused on suggestions for reform.[62,75–79] Chemical education research (CER) on textbooks has focused on discourse analysis of the text,[80] the use of analogies within the text,[81,82] a content analysis of diagrams,[83] classifying end-of-chapter problems based on Bloom's taxonomy,[84] problematic explanation styles,[85] issues of representation,[86,87] misconceptions/alternative conceptions in general chemistry textbook,[88–94] as well as an analysis of visual representations of particulate matter in the text.[95–98] Even fewer have demonstrated how students use their textbooks or how valuable students find the various features of these ever-expanding textbooks.[61,99–101]

The now-defunct TextRev national textbook survey was designed in 2002 to study the usage of textbooks by first-year general and second-year organic chemistry students.[99] This survey collected data from 3200 general and organic chemistry students and 23 instructors from nine higher-learning institutions across the United States regarding the time spent on various textbook features and the "helpfulness" of those textbook features. The TextRev survey found that general chemistry students on average used their textbooks less than organic chemistry students (4.1 ± 0.1 hours and 5.8 ± 0.2 hours, respectively).[99] As seen in Figure 8, when the self-reported average hours of textbook use are plotted against the students' anticipated letter grade, only the general chemistry students' use has a statistically significant relationship to the anticipated letter grade (ANOVA: $F = 10.9$, $p = 10^{-7}$).[99] While it is possible that students who expected a lower course grade consciously inflated the self-reported hours of use, the lack of a significant relationship between anticipated grade and organic students' reported use suggests that hours spent with the study materials alone does not guarantee a better grade. In addition, researchers reported that students overestimated their final grades by 0.45 grade-point average (GPA) points for the fall semester of their first year, 0.25 points of their spring first year, and by 0.2 points in their second year.[99] Even though the students are initially poor at estimating their final grade (likely due to poor metacognitive skills), the underestimation of their final grades does not change the conclusion that textbook use alone is not the sole predictor of course performance.

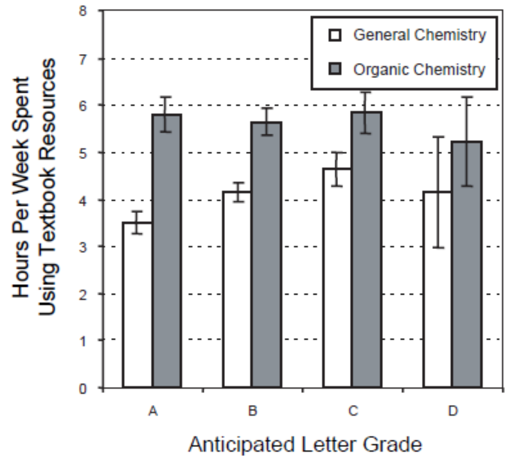


Figure 8: The self-reported hours spent weekly using the textbook's resources (text, study guide, solutions manual, website, textbook's CD) arranged according to students' anticipated grade in the course. Error bars correspond to 95 percent confidence interval.[99]

Another item on the TextRev survey asked students to evaluate how helpful[1] various print textbook features were on a 0-10 Likert scale. The results, as seen in Figure 9, demonstrate that images, end-of-chapter problems, and the solutions manual were significantly more helpful to organic chemistry students, while real-world applications and animations were more helpful to general chemistry students.[99] The two general chemistry courses were taken from a single student population over two semesters, with the same textbook and the same instructor. This figure also notes that between the two semesters of general chemistry, various textbook features are more helpful to the first-semester general chemistry students, but by the second semester students consider the in-

---

[1] Here, helpfulness was not defined in the Smith & Jacobs paper. Given that the Technology Acceptance Model (TAM) and initial UTAUT theories were developed around the same time of publication, it is unlikely that the researchers intended that precise definition.

chapter example problems (commonly called worked-example problems) and the text itself to be the most helpful features.

Table 3. Average Student Ratings of Textbook Features[a]

Textbook Feature	First-Year versus Second-Year Courses			General Chemistry Class over Two Semesters[b]		
	General Chemistry (n = 2084)	Organic Chemistry (n =1137)	p value (two-tailed probability)[c]	1st Semester (n = 250, 48% of class)	2nd Semester (n = 168, 36% of class)	p value (two-tailed probability)[c]
Written text	6.8 ± 0.1	7.0 ± 0.1	.110	7.2 ± 0.3	7.7 ± 0.3	.025
Images or photographs	6.4 ± 0.1	7.1 ± 0.1	10^{-13}	6.8 ± 0.3	5.9 ± 0.4	.001
Examples of real world applications or situations	5.3 ± 0.1	4.5 ± 0.1	10^{-12}	4.7 ± 0.3	3.6 ± 0.4	.001
In-chapter example problems	7.4 ± 0.1	7.3 ± 0.1	.76	7.0 ± 0.3	7.7 ± 0.4	.008
End-of-chapter problems	7.1 ± 0.1	7.3 ± 0.1	.010	7.4 ± 0.3	7.7 ± 0.4	.19
Chapter summaries or glossaries	6.1 ± 0.1	6.2 ± 0.1	.14	6.2 ± 0.3	5.5 ± 0.4	.011
Solutions manual	6.4 ± 0.1	7.4 ± 0.1	10^{-21}	6.1 ± 0.4	5.9 ± 0.5	.043
Study guide	5.1 ± 0.1	5.0 ± 0.2	.25	4.8 ± 0.4[d]	3.9 ± 0.5[d]	.023
Animations or simulations	5.2 ± 0.1	4.8 ± 0.2	.004	5.3 ± 0.4	4.1 ± 0.4	.001

Figure 9: Organic and general chemistry students' rating (on 0-10 Likert scale) of the helpfulness of textbook features. Uncertainties represent 95 percent confidence limits.[99]

Within CER on textbooks, fewer studies have focused on ebooks. In terms of performance, one study found that an open-educational resource ebook was a viable alternative to a traditional textbook via a non-inferiority multiple regression.[64] There was no significant difference between groups of students who used an ebook and those who used a print textbook in performance on most assessments; the learning gains by the ebook group were not inferior to the gains by the print book group.[64] Similar results in non-inferior learning gains have been reported for students using ebooks in statistics courses.[102,103] In a more qualitative study, Salami and Omiretu used a non-random purposeful sampling technique to survey general chemistry students who voluntarily chose to use an ebook over a traditional print textbook for the course; their participants

(N = 46) rated statements on a 7-point Likert scale to elucidate how the constructs of the Unified Theory of Acceptance and Use of Technology (UTAUT) and Technology Acceptance Model (TAM) why explained these participants chose to use the ebook over the traditional print book.[10,104] Salami and Omiretu found that performance expectancy and effort expectancy significantly correlated to the behavioral intention to adopt the new technology.[10,73]

With regards to course materials generally, a few studies give context to the results presented in this work. One study collected data from general chemistry students for ten years, asking participants to rate the value of formal class meetings, course materials, and other resources, such as a paid tutor.[61] Figure 10 shows the change in percent rank score for selected course materials over the ten year period.[61] The textbook is consistently rated as highly valuable to students, only to be surpassed by the electronic homework towards the end of that decade. While the author noted that further study was necessary to demonstrate the efficacy of the textbook,[61] the sharp rise in perceived value of the electronic homework suggests an opportunity to combine the students' two most valuable resources—textbook and electronic homework—into a single source of knowledge.

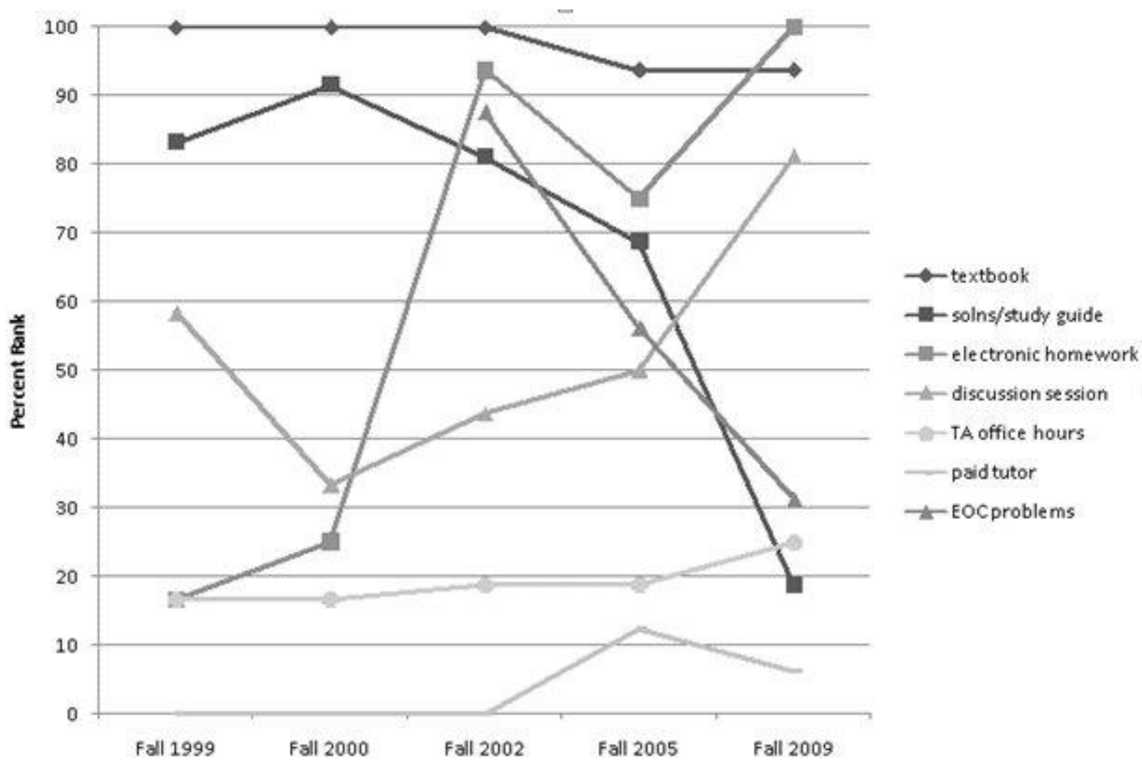


Figure 10: Percent rank score by year for selected course items. The textbook was ranked consistently as one of the most used course components/study resources available to students. Also notable is the prevalence of electronic homework rose, the perceived value of the solutions manual fell rather sharply.[61]

Online homework systems have been the focus of study recently. These tools provide immediate and detailed—although not specific to the student—feedback on problem-solving.[101] In comparing two electronic homework systems, researchers evaluated the students' preference, attitudes, and performance with the different formats.[101] The MindTap system (which is also the system used at the University of Georgia and a topic of the presented work) has questions embedded within the ebook text that can be assessed for points; each ebook section is followed by "mastery" questions designed to quiz the student. This system also embeds interactive figures, interactive video tutorials, and tutorial solutions to problems. Williamson and Zumalt compared this MindTap system to the Online Web Learning (OWL) system, in which triads of mastery questions were assigned and the electronic text was accessible through a link to a separate browser tab.[101] Previous work by these researchers concluded that students perceived that they learned more with MindTap and preferred the embedded arrangement of homework and text.[100] Researchers compared the average grades for hourly exams, final exam grades, and overall course grade between groups that used each homework system.[101] They noted that the two groups had a similar number of withdrawals from the course and similar grade distributions.[101] Although this comparison was between two electronic texts, the rationale of similarity between educational interventions will be relevant in the conclusions of the study presented in this document. In light of the results of the TextRev national textbook survey, Smith and Jacobs found that students (N = 3200) were slightly underestimating their grades (by as much as 0.45 points).[99] For the presented work, accepting a slight underestimation in anticipated course grade (especially in light of the similar grade distributions between ebooks in the previous studies) allowed for a larger sample size.

Research Questions

It is important to make the distinction between what students perceive as valuable to their learning and what course components help students learn.[61] While the latter—what influences students' learning—is a highly variable and complex web of resources, prior knowledge, and student affective characteristics, the former can be demonstrated in the following work. The justification is that the resource has no opportunity to influence learning if the students do not perceive it as helpful or valuable. Thus, to explore the viability of an ebook as a primary content-delivery medium for an active-learning course structure, it is important to survey students about their attitudes towards and their perceived usage of the textbook, particularly in relation to other course resources.

During the implementation of this survey, the University of Georgia Chemistry Department transitioned from a traditional general chemistry print textbook (in fall 2015) to an ebook (in fall 2016). These books were from the same publisher (Cengage) with different authors (Kotz *et. al* for the print and Vinings *et. al*. for the ebook)[105,106] and were uniformly adopted across multiple sections with different instructors of the general chemistry course. This transition provided a unique opportunity to measure students' reactions to a format change. The goal of this research is to inform general trends in student textbook use, and to answer the questions:

1. How often do students perceive that they use their textbooks, and how does this frequency compare to other course resources?
2. Which features of their textbooks do students consider helpful, and does the helpfulness of a feature change during the format transition?

<u>Methods</u>

An Institutional Review Board (IRB)-approved 21-item Qualtrics survey was distributed to general chemistry students at the end of the semester. This data collection began in fall 2015 (in which the textbook was a print book with an available PDF format) and continued until May 2018. In spring 2016, the instructional team for the chemistry department uniformly adopted a required, primarily electronic book by the same publisher (with a print copy available) for the General Chemistry I. This same ebook was adopted for General Chemistry II in fall 2016. This transition provided a unique opportunity to measure students' reactions to a format change in the textbook. The longitudinal data is from the General Chemistry I population.

Modeled after previous work on students' use of resources and textbooks,[61,99] the initial survey items are listed in Appendix A. These questions included demographics (age, sex, and anticipated course grade), the average use of resources per week, and the helpfulness of textbook features. Some questions were included to probe students' communication preferences. After the switch to the ebook in the General Chemistry I course, the survey was revised to include a branch that reflected the features in the ebook.

Some of the items generated responses that were quantitative. When appropriate, the data were organized initially in SPSS and later data organization and statistical analyses were carried out in JMP by SAS.[107,108] Table 2 lists these questions and the statistical test performed. For the qualitative data, responses were analyzed in Atlas.ti with a mixture of *a priori* (from UTAUT2) and emergent codes.[109]

Table 2: List of Quantitative Survey Items

Survey Item	*Statistical Analysis*
Hours/Week on Chemistry Resources	ANOVA, Welch's t-test
Dates/Times Most Likely to Study	n/a
Helpfulness of Textbook Features	Mann-Whitney (nonparametric) t-test

For the items that asked participants to estimate their weekly use of course resources, each semester's data was plotted to inspect for outliers. Outliers were defined as data points that were 1.5 times beyond the interquartile range (IQR) and were eliminated, which was approximately a 10 percent trim of each dataset.[29] After removing outliers in JMP, cumulative density plots were generated for each course resource; resources with non-parametric distributions were subjected to tests that variances are equal (such as Levene's test). Normally-distributed data that had significantly unequal variances are analyzed with a Welch t-test, whereas non-parametric data with variances that were not significantly unequal were treated to a Kruskal Wallis. Internal validation checks were used on the items about weekly use of course resources and helpfulness of textbook features.

Table 3 describes the average age, numbers of males and females within each semester, as well as the number of participants and completed response rate. The sample size is the number of participants that completed the survey at the end of the semester, and the population size is the number of students still enrolled in the course at the end of the semester. The completed response rate reflects the number of participants who passed internal validation measures within the survey.

Table 3: Demographics and Unit Response Rate

Semester	*Age*		*Sex*		*Total Number of Responses*	*Course Enrollment for GC1*	*Completed Response Rate (%)*
	Avg.	*Median*	*F*	*M*			
Fall 2015	18.52	18	363	231	595	1144	50.0
Spring 2016	19.09	19	206	101	326	683	45.2
Fall 2016	18.54	18	558	278	845	981	85.9
Spring 2017	18.99	19	424	188	620	929	65.9
Fall 2017	18.41	18	478	260	751	1256	58.8
Spring 2018	18.90	19	468	152	625	969	64.1

Student Use of the Print Textbook

In the initial semester (fall 2015), a print textbook was used in the general chemistry sequence. The co-requisite for the course was a one-credit, three-hour laboratory course, and the typical lecture met three times weekly for 50 minutes each. Students in general chemistry also had access to lecturer’s presentation slides and lecture videos through the classroom management system (LMS), and they were assigned online homework through a separate product. The textbook and its solutions manual, lecture slides and videos, and homework problems constitute the *learner-content* interaction in the model in Figure 1. Each instructor offered office hours, and teaching assistants (TAs) were required to hold weekly office hours in addition to teaching sections of the laboratory. These professor and TA office hours provided students in the course an opportunity to interact with a content-expert with any questions they may have; this fits the Interaction Theory model under *learner-teacher* interaction. Furthermore, students in the course had a variety of *learner-learner* interactions available, including the use of informal studying with friends, formalized learning communities/study groups, university-sponsored free peer tutoring, and students' paid use of tutoring from local private tutoring companies. Each of these interactions (and the constituent course resources) represent sources of information from which general chemistry students may use to help construct meaning (or simply to memorize algorithms) to aid their chemistry studying.

To ascertain what resources students used in the general chemistry course, participants were asked "To study for chemistry, how many hours per week do you use the following resources?" The provided resources each had a slider-bar allowing participants to estimate their use on a scale with a maximum of 168 hours per week; participants also had an opportunity to suggest other resources in a free-response format. A full list of resources that were presented is available in the Appendix A. This dataset was analyzed for outliers (roughly equivalent to a ten percent trim of the dataset) to remove participants whose total for all responses exceeded the maximum hours in a week.

While students (especially first-semester students) are notoriously poor at estimating the amount of time or effort expended (related to their poor metacognition skills), the dataset reveals what the participants perceive to be the resources they utilized the most. Responses from the fall 2015 semester demonstrate that students use the textbook (5.278 ± 4.151 hours per week), homework (4.053 ± 1.894 hours per week), formal class "lecture" meetings (3.017 ± 0.183 hours per week), laboratory period (3.000 ± 0.000 hours per week), and professor's PowerPoint lecture slides (2.141 ± 1.604 hours per week) most often. Participants' reported use of peer, instructor, and content interactions for this print-book semester are shown in Figure 11. Of the resources that involve *instructor-learner* interaction, respondents only spent 0.646 ± 0.694 hours per week in the instructor's office hours and 0.302 ± 0.513 hours per week in the teaching assistants' (TA) office hours. While it is unclear the quality or extent of *instructor-learner* or *learner-learner* interaction occurred in the large lecture halls of the formal class meeting or laboratory periods, participants reported spending an average of 2.665 ± 2.127 hours per week meeting with friends and 1.798 ± 1.889 hours per week study groups. The average weekly use of these social interactions findings is also presented in Figure 11.

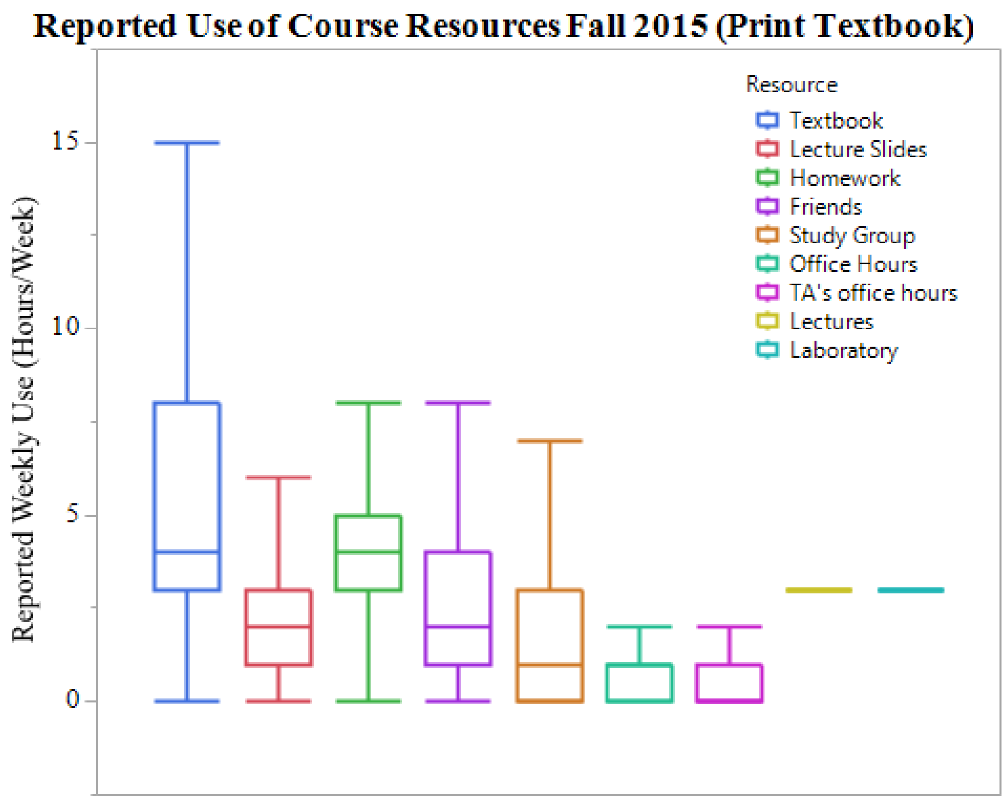


Figure 11: boxplot representation of General Chemistry I students' reported weekly use of course resources for the fall 2015 semester (print textbook).

Given the averages, it appears that most of the learners' time is spent interacting with the course content (*learner-content* interaction) and peers (*learner-learner* interaction) rather than interacting with the instructor directly. This reticence to engage with the instructor on an individual basis is manifested in the preferential use of peer (1.103 ± 1.373 hours per week) and private tutoring (1.810 ± 2.320 hours per week). Who are these students? The boxplots in Figure 12 demonstrate who the respondents (of various anticipated letter grades) seek out for clarification. The most striking observation is that respondents who anticipated receiving an F did not report using any outside social interaction. ANOVA of anticipated course grade against the use of private tutor (F =

2.0279, p = 0.1197), professor's office hours (*F* = 0.6541, p = 0.5235), TA office hours (*F* = 0.5365, p = 0.5889), and peer tutoring (*F* = 1.8832, p = 0.1434) was not significant. Although the results are not statistically significantly different, Figure 12 demonstrates that high-achieving students favor study groups and friends, whereas lower-performing students increasingly use private and peer tutoring in addition. While students who anticipate a letter grade of A, B, or C report some use of the professor's or TA's office hours, this use disappears in the students who anticipate not passing the class (D and F).

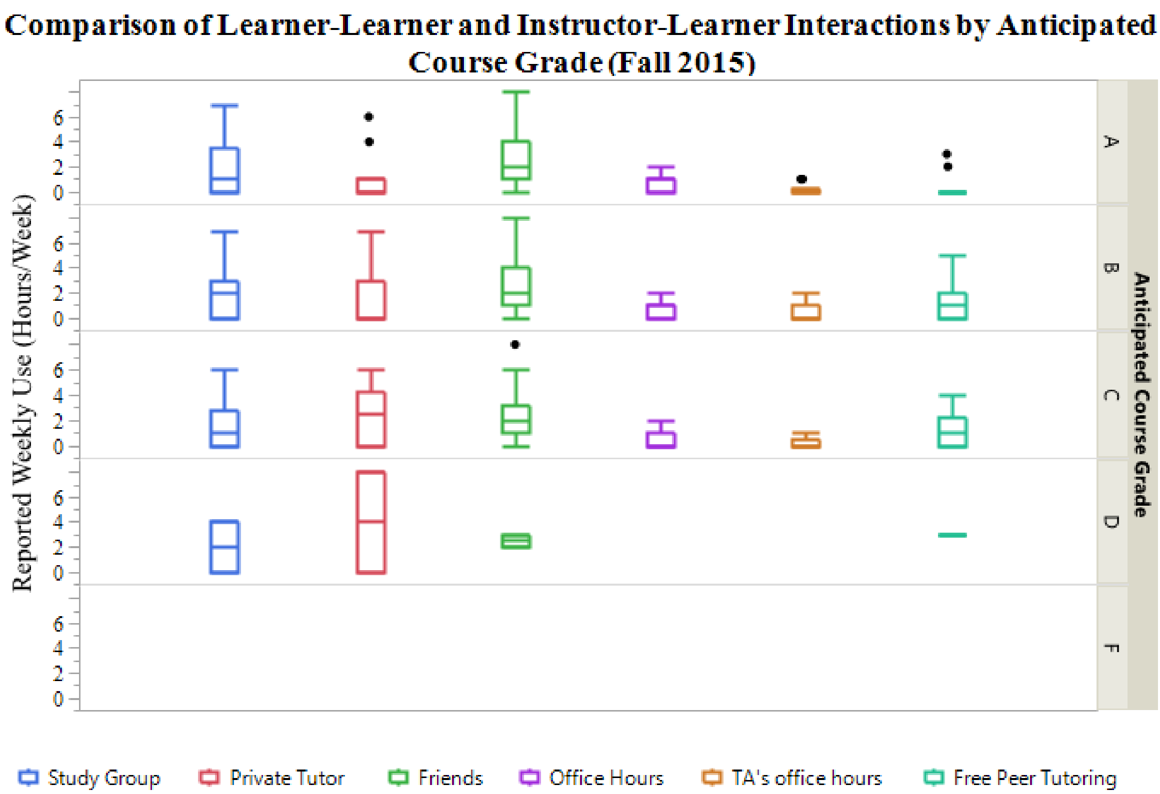


Figure 12: a stratified set of boxplots based on anticipated course grade for fall 2015 (print) General Chemistry I participants; these selected course resources represent student-student or instructor-student interactions.

Participants were also asked about their motivation and the resources that contributed to achieving that goal. Participants were asked to select which resources they believed helped them *understand the course content* and which resources helped them *earn a good grade*. Understandably, 69.8 percent of fall 2015 General Chemistry I students said that *earning a good grade* was more important, as compared to 27.5 percent that valued *understanding the course content*. Figure 13 demonstrates the fall 2015 General Chemistry I respondents' perception of course resources toward helping them achieve the goals of *understanding the course content* (gray) or *earning a good grade* (red).

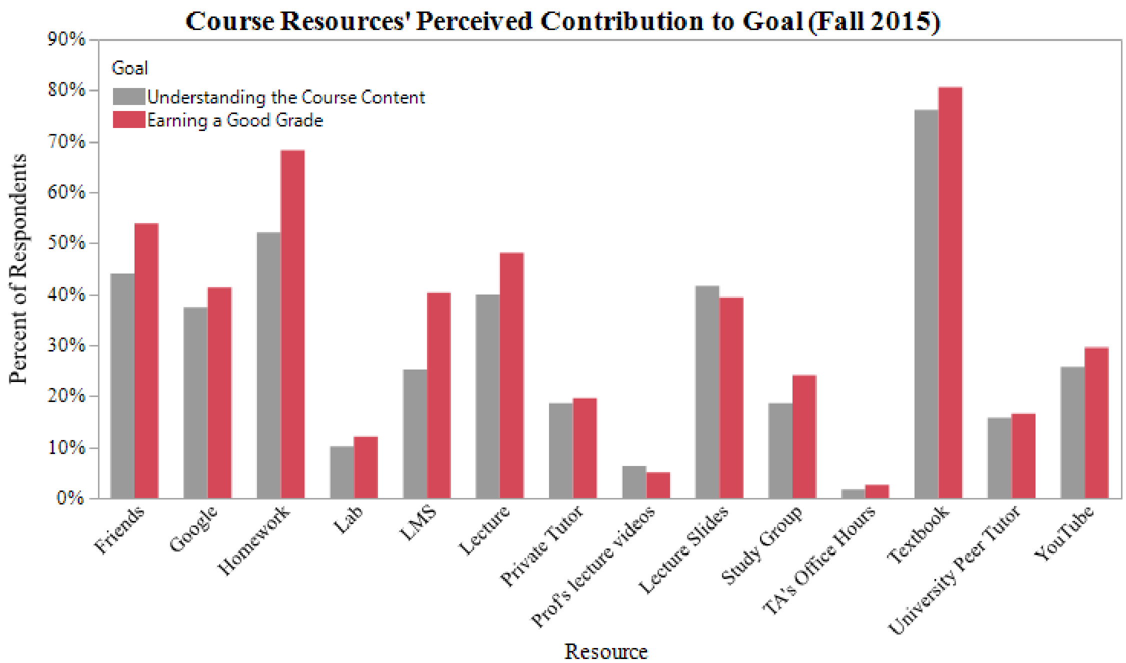


Figure 13: the percent of participants who selected each resource (multiple selection allowed) that contributed to each goal, understanding the course content (gray) and earn a good grade (red).

Given that most participants valued *earning a good grade* (red) more, it is apparent that students in General Chemistry I in this semester relied mostly on the print textbook (80.56 percent), online homework (68.30 percent), friends (53.94 percent), and lecture (48.16 percent). These resources are also valuable towards the goal of *understanding the course content* (gray), but the percent of each resource is smaller than the percentage reported for the *earning a good grade* goal. This underscores that participants valued *learner-content* interaction with the textbook and homework, *learner-learner* interaction with their friends, and *learner-instructor* interaction through lecture.

Considering the students' interaction with the content-delivery devices, an ANOVA on anticipated letter grade for the use of textbook (F = 2.0099, p = 0.1148), homework (F = 2.2389, p = 0.0670), and lecture slides (F = 1.1309, p = 0.3395) were not significant. Although these results are not significant, the comparison is illustrated in Figure 14. Students who anticipated receiving an F did not report interacting with the textbook or course lecture slides, and students who expected to receive a D report spending more time with the homework than the students who did well. This illustrates a pattern of reading for understanding rather than extensive homework practice to memorize algorithms, although the association is not significant.

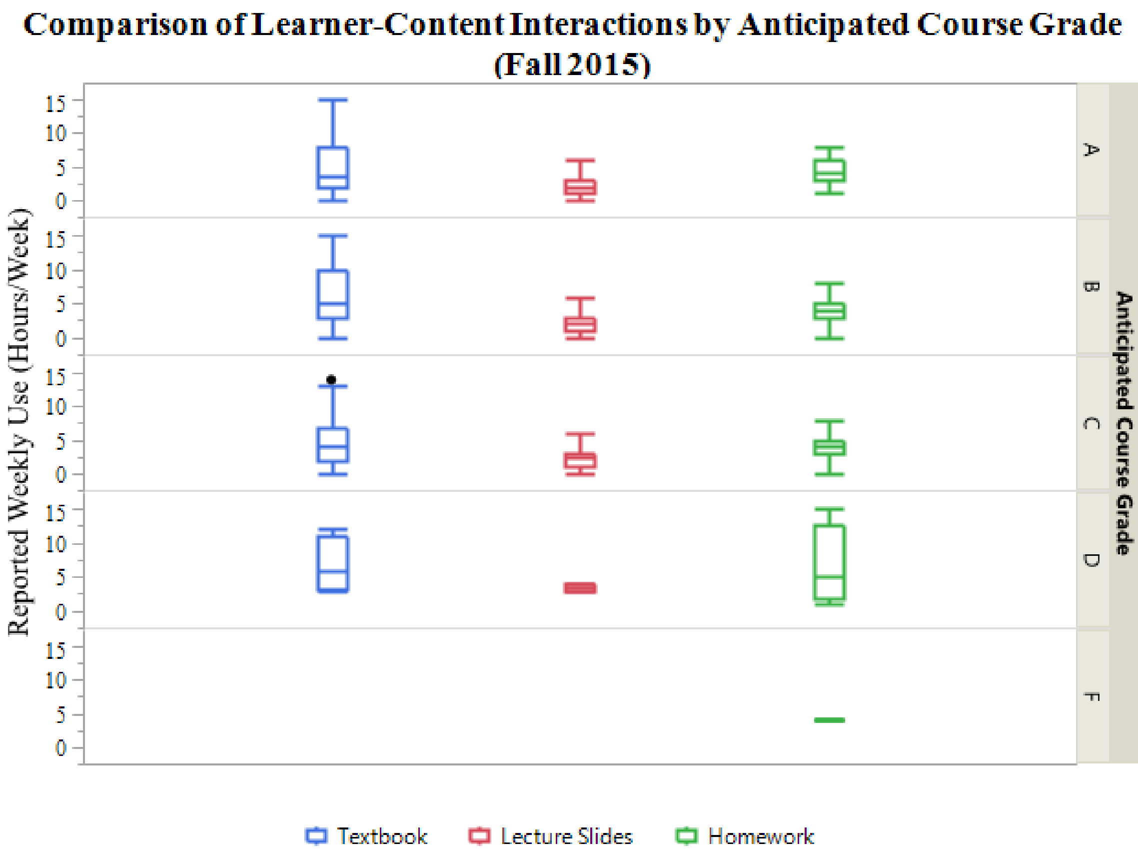


Figure 14: a stratified set of boxplots to illustrate how General Chemistry I participants of various expected final grades interacted with content-delivery resources.

If the respondents, regardless of anticipated grade, are favoring the textbook as their most often-used content-delivery source, what feature of that interaction is the most helpful to them? Figure 15 visualizes the features of the print textbook (as well as the online homework) on a Likert scale that ranges from “Very helpful” to “Very unhelpful” with an “I don’t use” option; percentages of each rating for each feature described in Appendix A. Participants ranked the written text, the worked examples, the end-of-chapter problems, and the tutorials on WebAssign as the most helpful features of this print book. The features with the highest rates of “I don’t use” reported include the “Key Experiment” sections (40.3 percent), solutions manual (30.0 percent), chapter goals (28.2

percent), strategy maps (28.5 percent), and chapter outline (24.8 percent), among others. This visualization attempts to probe specifically what parts of the content-delivery that the students find valuable and helpful in their studies. Students favor mathemagenic features that either demonstrate or explain algorithmic problem-solving, especially those in a stepwise manner.

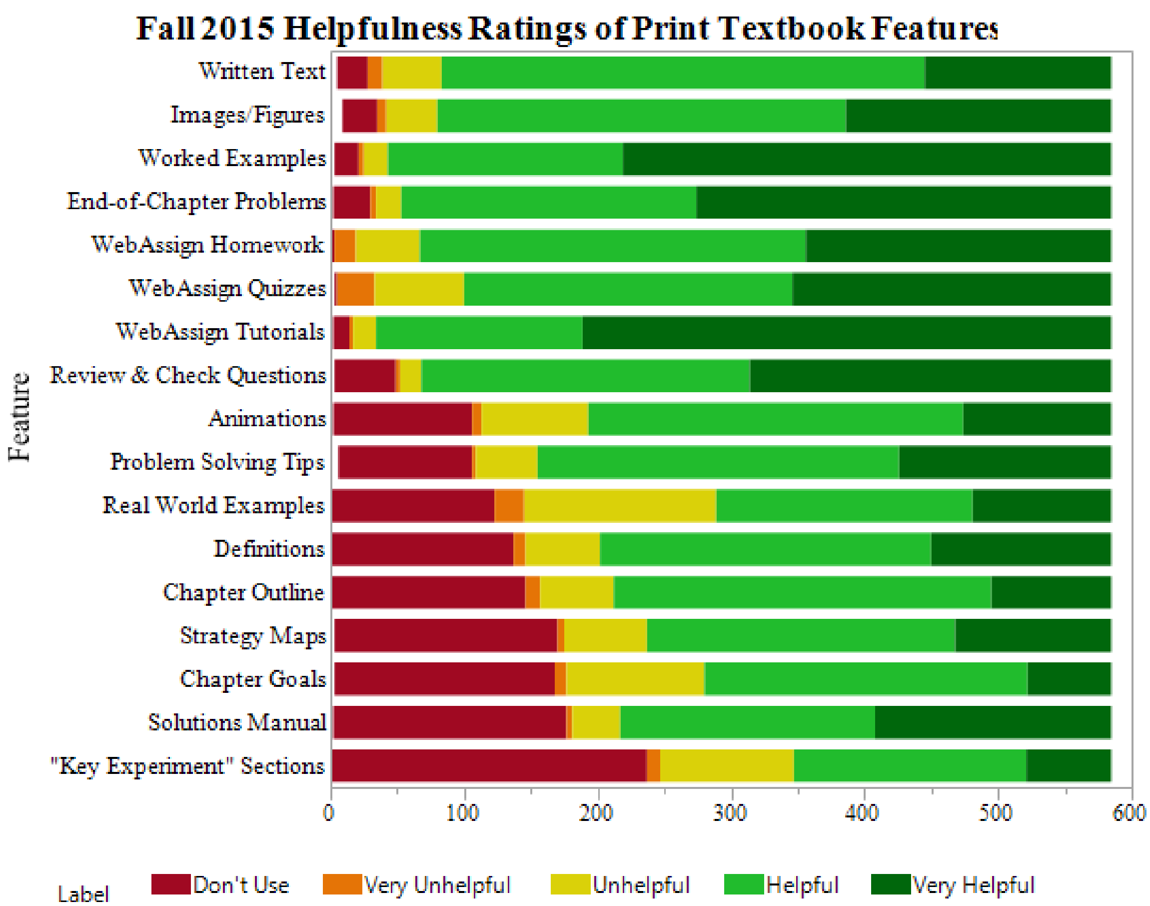


Figure 15: the Likert-style ratings of each textbook feature of the print textbook; participants were asked to rate the features on a scale from "I don't use/NA" to "very helpful."

The results visualized in Figure 15 suggest that authors and instructors should focus on the quality of the text, figures, problems, and examples. General Chemistry I participants do not value the generative features, *i.e.*, "key experiment" sections, strategy maps, and other features that tend to reside in the margins of the text. The highly-valued mathemagenic features will be compared to the equivalent features in the electronic textbook.

The Transition to an ebook

Following the Department's transition to the MindTap ebook in spring 2016, revisions were made to the survey to reflect the current resources and textbook features available to general chemistry students. Figure 16 visualizes General Chemistry I respondents' use of course resources in the semester following a transition to an ebook. While the textbook is still a commonly-reported resource, students now report spending more time on the homework. Study group use and reliance on friends still outpace use of professor or TA office hours.

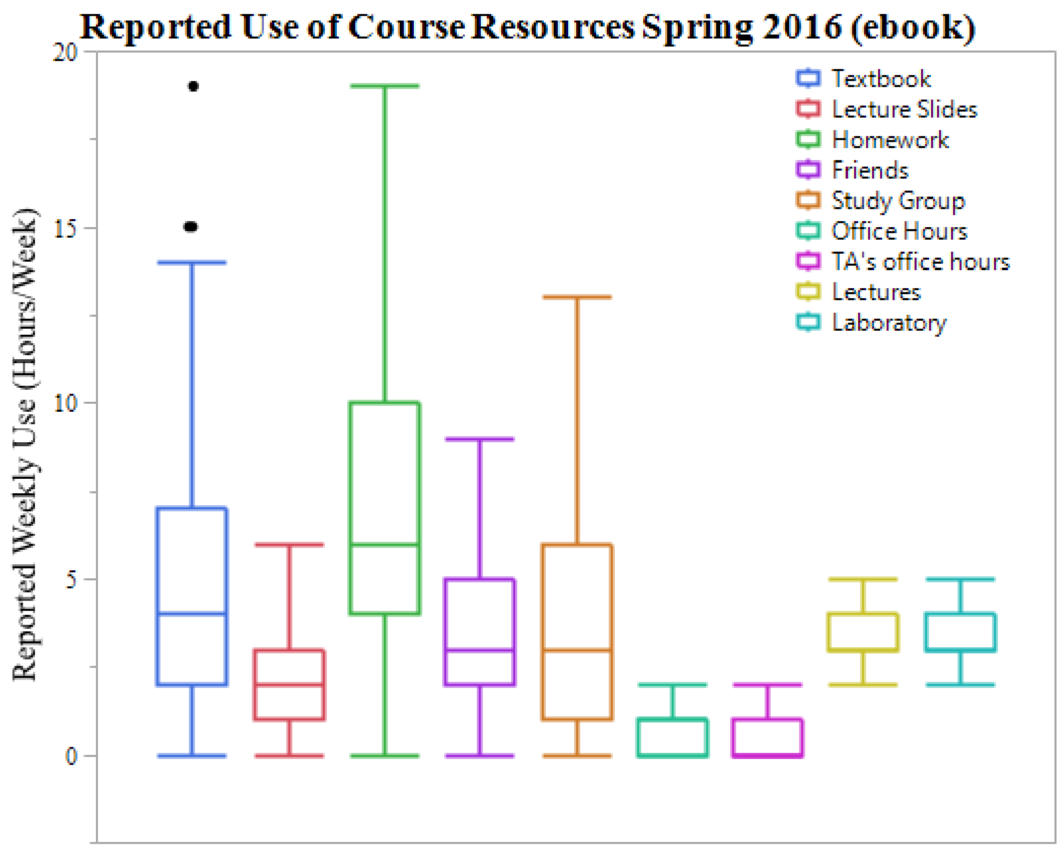


Figure 16: General Chemistry I students' reported weekly use of course resources for the spring 2016 semester (ebook)

Although this sample is a different population of students compared to fall's population in terms of preparation, this data represents the immediate transition to a new textbook format. Therefore, a statistical comparison of the reported use of resources between fall 2015 (print) and spring 2016 (ebook) is shown in Figure 17. Fall 2015 represents a traditional, hard copy textbook, and spring 2016 represents an ebook. An asterisk (*) indicates that a Welch's t-test returned a significant p-value ($p < 0.05$), and a double asterisk (**) indicates $p < 10^{-7}$; a single dagger (†) indicates a Cohen's *d* in the small to medium range, whereas a double dagger (‡) indicates a large effect size.

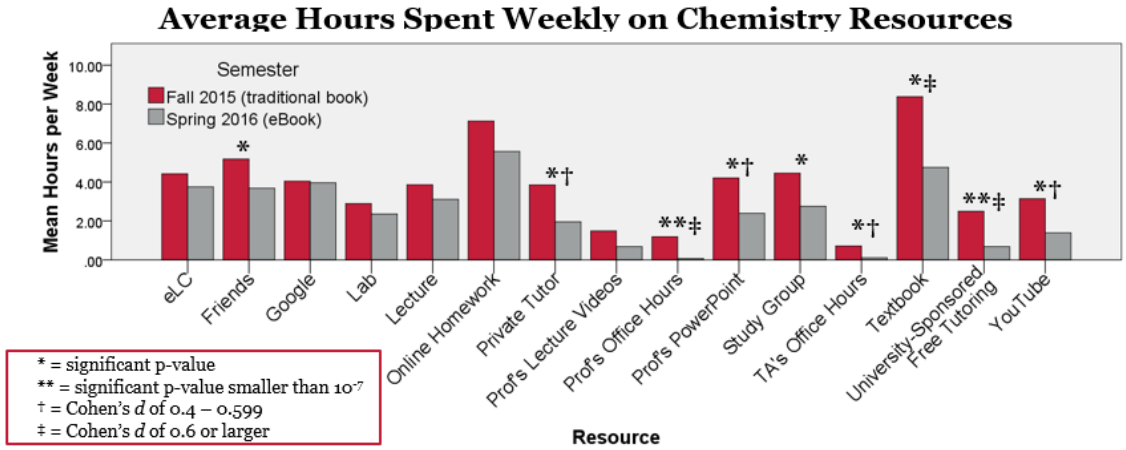


Figure 17: a t-test comparison of the average weekly reported hours spent with course resources between a print book semester (fall 2015, red) and the first semester with the ebook (spring 2016, gray); asterisks denote different magnitudes of significant p-values, the single dagger indicates a medium effect size, and a double dagger indicates a large effect size.

From the print book to the ebook, there were significant decreases in average reported weekly time spent with friends ($p = 0.024$, $d = 0.366$), private tutors ($p = 0.014$, $d = 0.410$), professor's office hours ($p < 0.001$, $d = 0.935$), professor's lecture slides ($p = 0.020$, $d = 0.535$), study groups ($p = 0.017$, $d = 0.399$), TA office hours ($p < 0.001$, $d = 0.585$), textbook ($p = 0.003$, $d = 0.602$), and university-sponsored peer tutoring ($p < 0.001$, $d = 0.761$). These results suggest that during the initial ebook implementation, respondents engaged for less time with approved content-sources (textbook, lecture slides and videos), spent less time interacting with experts (professor and TA office hours, lecture, lab), and sought out less peer interaction outside of class (friends, peer tutoring, study groups). The immediate effect of the format transition was significantly decreased use, although the difference in constituent population is a confounding factor in this analysis. This difference in population is suggested by the decreased use of all resources from fall (red) to spring (gray) semesters. Therefore, a comparison between fall semesters is necessary.

A more specific look at how much time respondents of different anticipated letter grades devoted to interaction with the instructor-approved content sources is demonstrated in Figure 18. The well-performing learners devoted the most time to the homework, and roughly an equal amount of time with the textbook. Students who expected to fail did not use these content-delivery resources at all; students who expected to receive a D devoted more time to homework and textbook use than the well-performing students, indicating that a student's time spent working problems or reading the textbook is not a guarantee of performance.

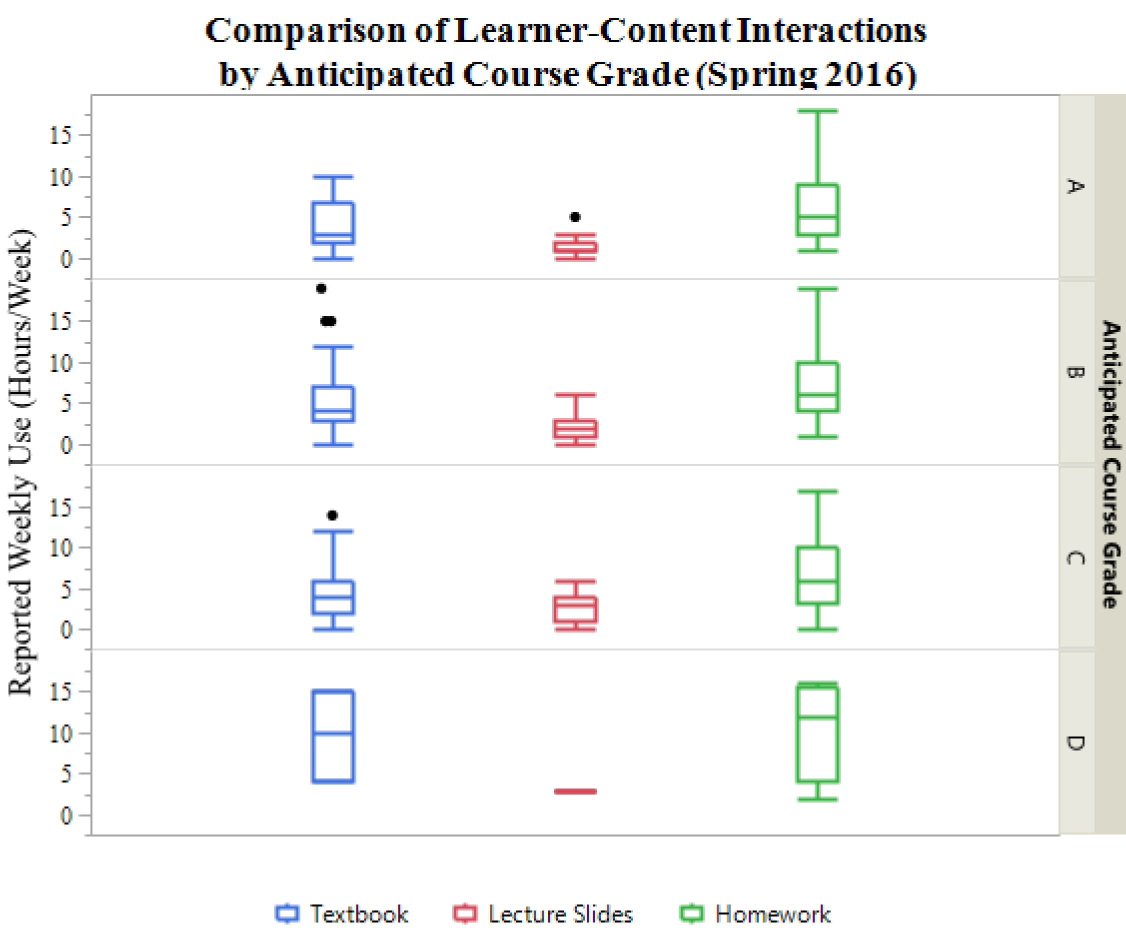


Figure 18: boxplots to demonstrate how spring 2016 participants of various anticipated course grades report using content-delivery devices on a weekly basis.

With the increased amount of *learner-content* interaction for the poor-performing students, the question becomes where are these students going for supplemental explanations of the content? Has the interactivity of an eBook influenced their use of peer or instructor interaction? Figure 19 demonstrates how much time students at different anticipated course grades spend with study groups, private tutors, friends, professor and TA office hours, and free university-sponsored free tutoring. The reliance on private tutors has decreased for poor-performing students, but the well-performing students are increasing the amount of interaction with these private, paid tutors. The learner-instructor interactions have not been affected appreciably at any letter-grade, which suggests that the learner-content interaction does not further subsume the *learner-instructor* interactions. One encouraging trend is an increase in reported *learner-learner* interactions in the form of university-sponsored free tutoring for learners at each anticipated grade level.

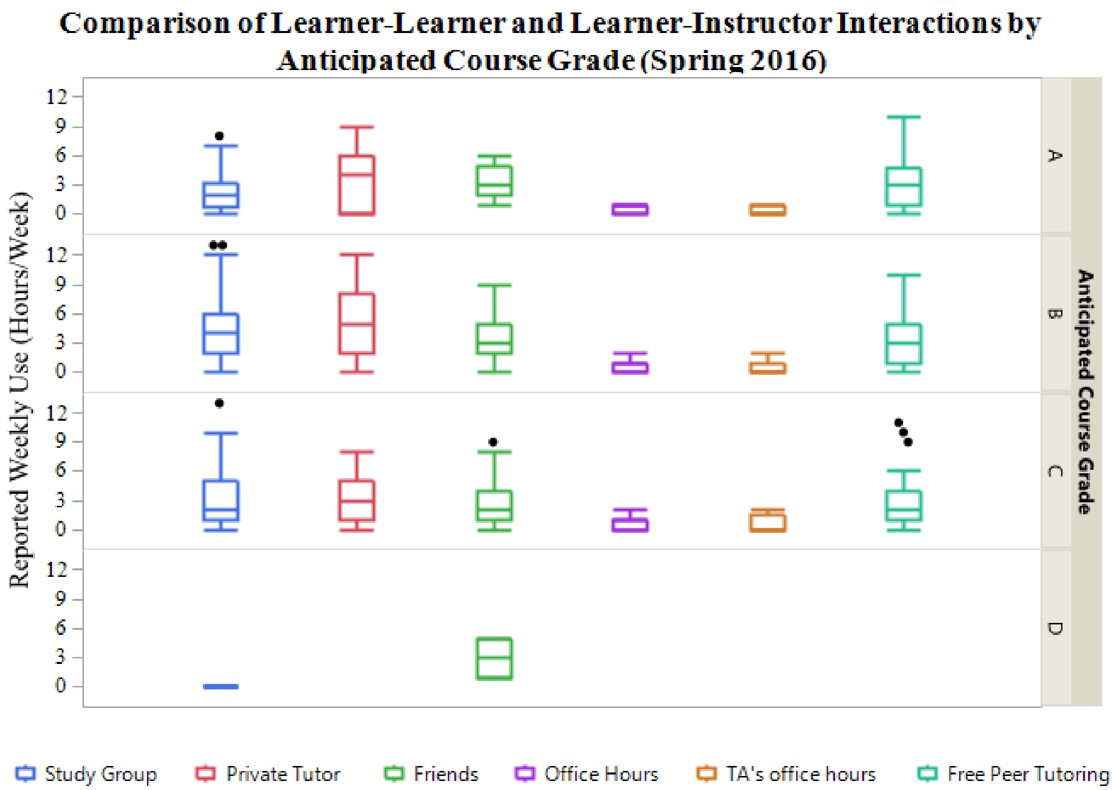


Figure 19: stratified box-plots to describe learner-learner and learner-instructor interactions for General Chemistry I students in the first ebook (spring 2016) semester across anticipated course grade.

These trends are also demonstrated in the set of questions regarding goals and resource use. Figure 20 demonstrates the spring 2016 General Chemistry I respondents' perception of course resources toward helping them achieve the goals of *understanding the course content* (gray) or *earning a good grade* (red). Predictably, 70.3 percent of spring 2016 General Chemistry I students said that *earning a good grade* was a more important goal, as compared to 24.5 percent that valued *understanding the course content*.

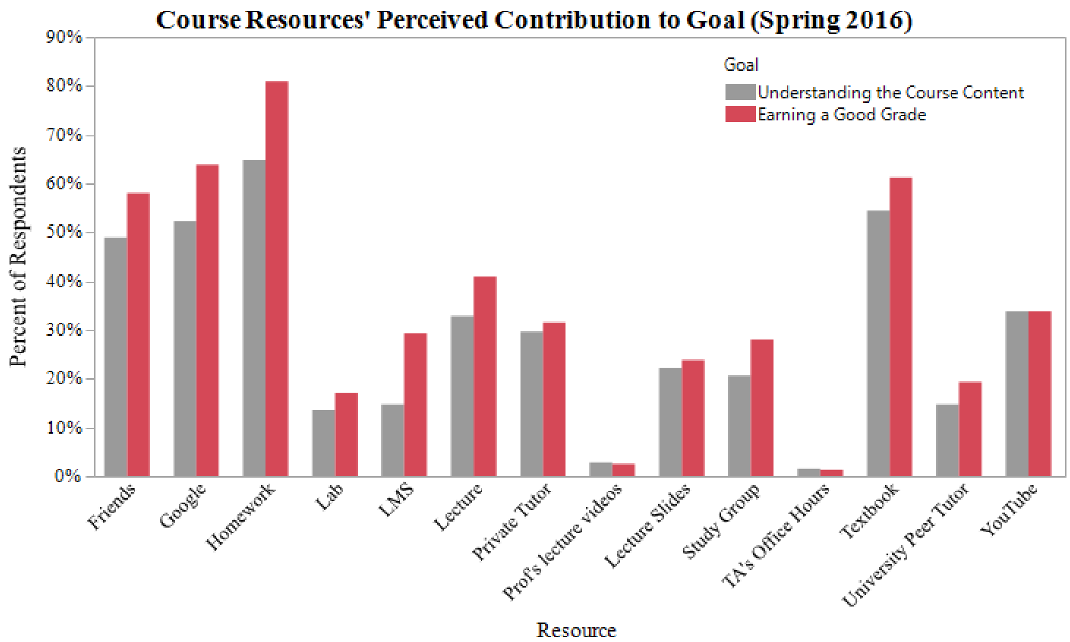


Figure 20: number of responses for each resource (multiple selections allowed) that contributed to each goal, understanding the course content (gray) and earning a good grade (red).

Once again, the resources that most participants said helped them achieve the goal of *earning a good grade* was online homework (80.97 percent), Google (63.87 percent), the now-electronic textbook (61.29 percent), and friends (58.06 percent). The *learner-instructor* interaction of lecture (which was deemed helpful towards achieving a desirable grade) dropped to only 40.97 percent of respondents and instead are supplanted by using Google to find answers. When asked what resources helped participants *understand the course content*, online homework (64.84 percent), the electronic textbook (54.52 percent), Google (52.26 percent), and friends (49.03 percent) were the most selected. For achieving an understanding of the conceptual material of the course, using Google to find the answer was selected nearly as often as the instructor-approved homework and ebook. Typing a query into a search engine would take the same effort to use the search feature of the ebook and these choices were selected nearly the same amount. The change in response patterns from the print-book semester indicates that participants found the textbook and lectures less helpful for achieving either goal.

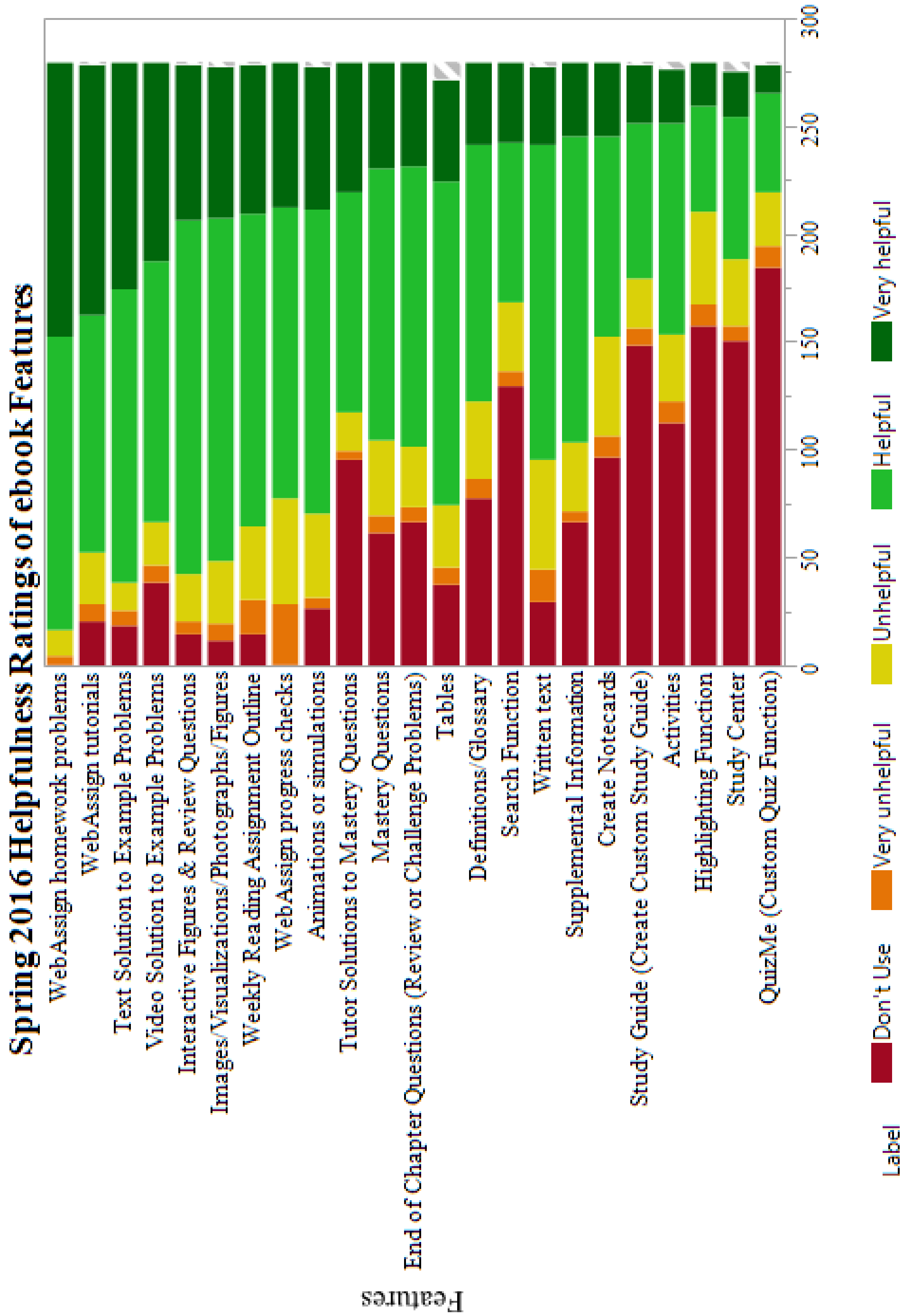


Figure 21: Likert-rating of ebook features from the spring 2016 sample.

In Figure 21, the features of the ebook and the General Chemistry I spring 2016 (N = 280) participants' Likert-scale ratings (from "I don't use/NA" to "Very helpful") are colored coded. The percentages of ratings for each feature are contained in Appendix A. The features rated most helpful were not the features of the ebook, but instead are the online homework problems and tutorials for that homework system. Participants rated the ebook features of text solution to example problems, interactive figures and review questions, images/visualizations/photographs/figures, reading assignment outline, video solution to example problems, and animations/simulations highly. General Chemistry I participants frequently reported not using the generative ebook features that are meant to help aid studying, such as the search function (46.43 percent), create notecards (34.64 percent), custom study guide (53.41 percent), activities (40.79 percent), highlighting function (56.43 percent), study center (54.71 percent), and custom quiz function (66.31 percent). Participants regarded the text of the ebook as mostly positive, but out of all the features included in the textbook, the text itself garnered the largest percentages of unhelpful opinions (18.35 percent unhelpful, 5.40 percent very unhelpful, and 10.79 percent "I don't use").

Fall 2016 ebook Adoption

The final set of descriptions is the data from the fall 2016 semester, the first fall in which the eBook was used. Like the spring 2016 transition semester, the reported use of homework (6.962 ± 4.923 hours) has outpaced the use of the textbook (5.680 ± 4.418 hours). Students favor interactions with their peers through friends (3.227 ± 2.731 hours), study groups (3.046 ± 3.224 hours), and tutoring (2.025 ± 2.777 hours and 1.785 ± 1.901

hours for private and peer, respectively). While participants still sought out their instructor during lecture and office hours (0.644 ± 0.718 hours), the students reported not using their TA's office hours (0 hours). These results are depicted in Figure 22.

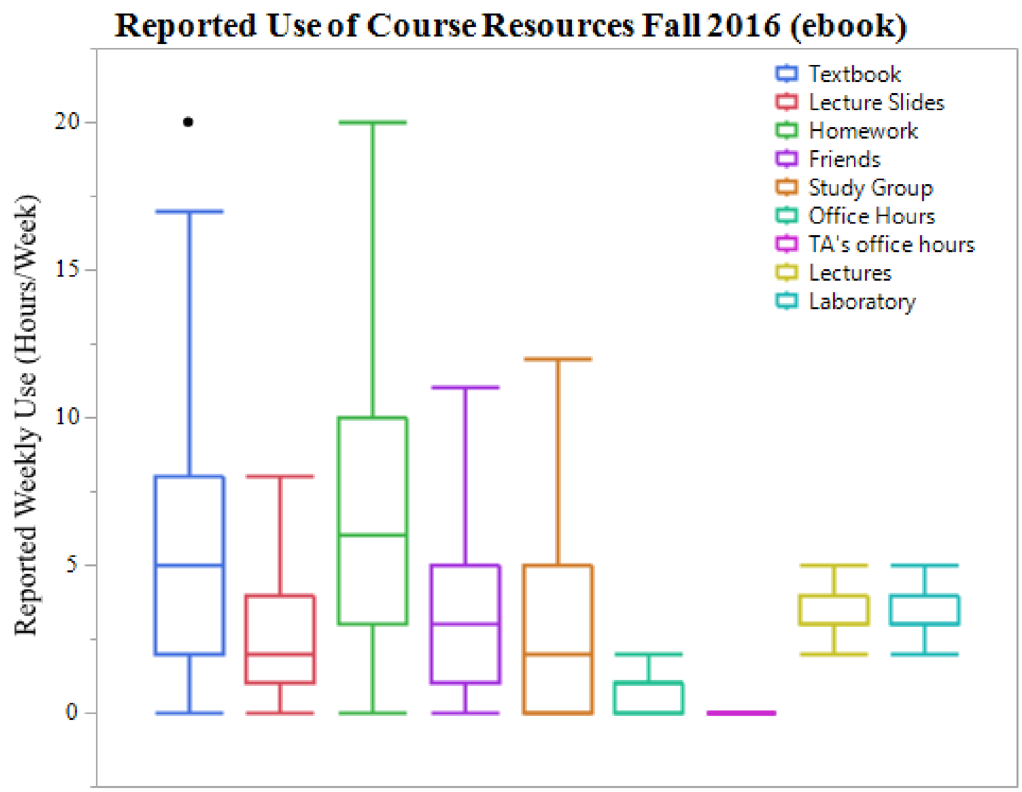


Figure 22: reported weekly use of course resources from fall 2016 General Chemistry I participants; this represents the first fall semester of ebook adoption

To explore the use of content-delivery devices, Figure 23 details how participants of various anticipated course grades used the textbook, lecture slides, and homework. While the print semester (fall 2015) participants who were not doing well in the course (and anticipated an F) reported not using the (print) textbook and lecture slides, failing students in the 2016 fall semester reported using the ebooks, lecture slides, and homework. These students reported using the textbook and homework more often than students who anticipated an A, although this could be attributed to either students' poor metacognition or a study strategy based on rote learning and memorization of algorithms.

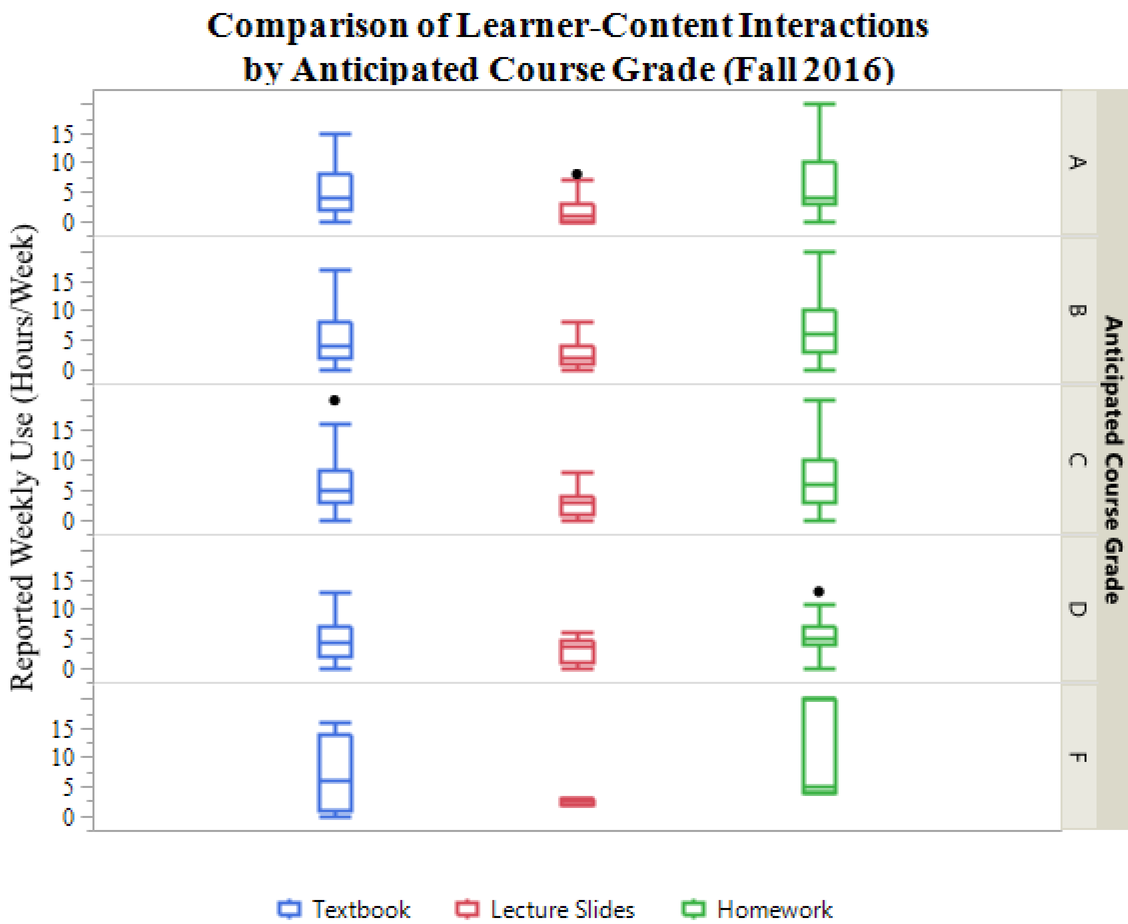


Figure 23: set of stratified boxplots to demonstrate how participants of different anticipated course grades use content-delivery resources on a weekly basis.

The evidence for a brute-force, shallow memorization study strategy is reflected in the selections featured in Figure 24. This figure demonstrates the fall 2016 General Chemistry I respondents' perception of course resources toward helping them achieve the goals of *understanding the course content* (gray) or *earning a good grade* (red). Consistent with previous semesters, 74.8 percent of fall 2016 General Chemistry I students said that *earning a good grade* was more important, as compared to 21.6 percent that valued *understanding the course content*.

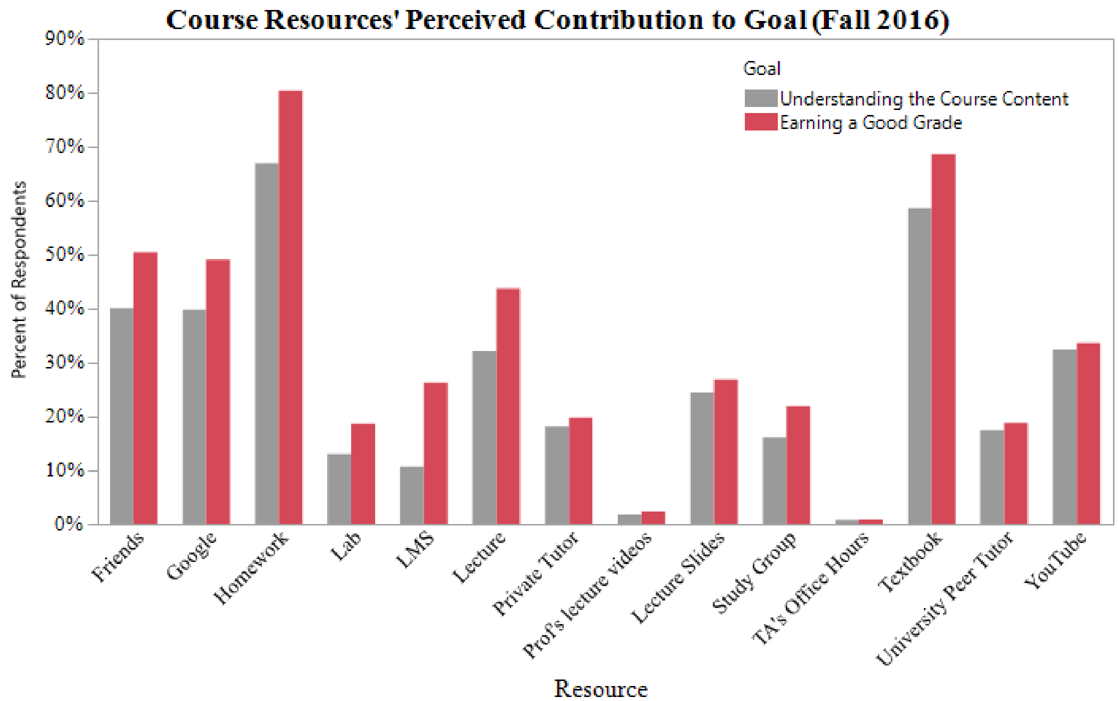
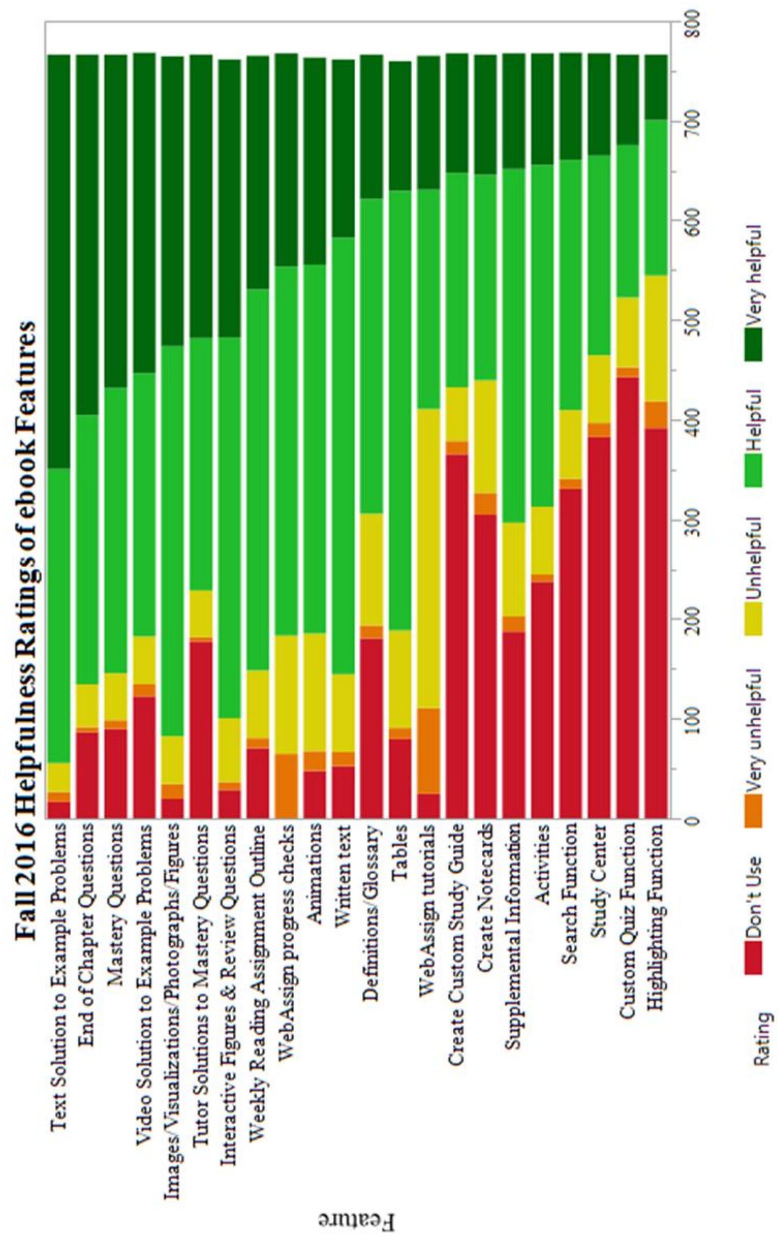


Figure 24: General Chemistry I participants were asked to select all the resources that contributed to each goal: understanding the course content (gray) and earning a good grade (red).

Toward the goal of *earning a good grade*, many participants perceived the online homework (80.40 percent), ebook (68.65 percent), friends (50.48 percent), Google (49.05 percent), and lectures (43.71 percent) as helpful. These results suggest that *learner-content* interaction with homework and ebook are still highly helpful, but *learner-learner* and *learner-instructor* interactions are roughly as helpful as using Google to find an answer. Toward a goal of understanding the content, online homework (66.98 percent), ebook (58.67 percent), friends (40.14 percent), and Google (39.79 percent) are considered by participants to help generate understanding, albeit at much lower percentages. There seems to be a general downward trend in the perception of these features toward achieving either goal.

To explore their perceptions of the ebook, participants were also asked to rate the helpfulness of features of the text on a Likert scale. Figure 25 summarizes the fall 2016 responses. Percentages of each feature's ratings are in Appendix A.

Figure 25: Likert rating of helpfulness of ebook features from fall 2016 participants



Fall semester participants rated the worked examples ("text solution to example problems") the most valuable feature, followed by end-of-chapter questions and mastery questions. The end-of-chapter questions were supplemental to the ebook, provided by the instructors to match the difficulty of the quiz and exam questions. The mastery questions were the questions embedded at the end of each section.

Students also rated the tutor solutions to these mastery questions highly. Static and interactive figures were also rated highly to similar magnitudes, suggesting that the level of interactivity of the figures does not contribute to the helpfulness of the feature. The text is viewed as helpful, although it is ranked lower than the weekly reading assignment outline. A comparison of helpfulness ratings of equivalent features between the two formats will illustrate which features' helpfulness is format-dependent.

Comparison of the Two Formats

Because the first semester that piloted the ebook was a spring semester (in which there is lower enrollment and different constituencies in the General Chemistry I course), a comparison of fall 2015 (print) to fall 2016 (eBook) can be used to determine significant differences in use of resources. Table 4 summarizes the responses from the General Chemistry I fall 2015 (print) and fall 2016 (ebook) participants. This includes the resources available to general chemistry students at the University of Georgia, each fall semester's reported average use per week (in hours per week) with the number of responses retained for each resource, as well as the test statistic, p-value, and effect size (if p-value is significant at alpha-level of 0.05). In two cases—TA office hours and laboratory—the uniformity of response in one group yielded unequal variances, and therefore the lack of variance meant the Welch's t-test could not return a value; F-test statistic and p-value were reported instead.

Table 4: Comparison of Reported Weekly Use of Course Resources between Textbook Formats

	Print (Fall 2015)		*eBook (Fall 2016)*		*Significance Tests*		
Resource	*Average Use (Hours/Week)*	*N*	*Average Use (Hours/Week)*	*N*	*Test statistic*	*p-value*	*Effect Size*
Textbook	5.278 ± 4.151	162	5.659 ± 4.383	686	0.0336*	0.855	
Lecture Slides	2.141 ± 1.604	121	2.558 ± 2.176	468	2.355**	**0.0193**	0.201‡
Homework	4.053 ± 1.894	170	6.944 ± 3.995	731	12.396**	**<0.0001**	0.783‡
Google	1.866 ± 1.554	142	3.205 ± 2.907	606	7.608**	**<0.0001**	0.495‡
LMS	1.959 ± 1.549	123	3.142 ± 2.869	599	6.485**	**<0.0001**	0.440‡
Study Group	1.798 ± 1.889	109	3.074 ± 3.234	445	5.380**	**<0.0001**	0.423‡
Private Tutor	1.810 ± 2.320	63	2.061 ± 2.794	277	0.442*	0.507	
Friends	2.664 ± 2.127	128	3.214 ± 2.718	548	2.487**	**0.0136**	0.210‡
YouTube	1.215 ± 0.883	93	1.855 ± 1.606	441	5.364**	**<0.0001**	0.425‡
Professor's Office Hours	0.646 ± 0.694	65	0.656 ± 0.719	221	0.0098*	0.921	
TA Office Hours°	0.302 ± 0.513	43	0.000 ± 0.000	119	41.7284*	**<0.0001**	0.201†
University Peer Tutoring	1.103 ± 1.373	58	1.808 ± 1.906	292	3.325**	**0.0012**	0.385‡
Lecture Videos	0.442 ± 0.639	52	0.431 ± 0.669	160	0.0109*	0.917	
Lecture	3.017 ± 0.183	120	3.420 ± 0.778	526	10.670**	**<0.0001**	0.570‡
Laboratory°	3.000 ± 0.000	69	3.315 ± 0.777	606	11.353*	**0.0008**	0.015†

* indicates F-statistic ** indicates Welch's t-test † indicates ω^2 ‡ indicates Hedges' *g*

° indicates that the uniformity of responses in one group yielded unequal variances, but the lack of variance meant the Welch's t-test could not return a value; F-test statistic and p-value reported instead

Average weekly reported use of professor's lecture slides, online homework, Google, the LMS, study groups, friends, YouTube, university-sponsored free tutoring, lecture attendance, and laboratory all increased significantly between these semesters. While there are multiple factors influencing a student's use of course resources, it is confusing to see participants reporting more time spent in lecture because the standard three hours of lecture had not changed. The most striking result is that—despite a change to a more "interactive" textbook format —participants' reported weekly use of the textbook was not significantly different between formats. The only significant decrease in reported use was the TA office hours, which dropped from an average of 0.302 ± 0.513 hours per week to zero hours per week. These trends mirror the results shown in Figures 26 and 27, in which *learner-instructor* interaction is less utilized than *learner-learner* or *learner-content* interactions. The format switch did not significantly affect students' use of the textbook and did not appear to have a discouraging effect on less-desirable sources of understanding (namely search engines like Google and Yahoo Answers or private local tutoring companies). The significant increase (with large effect size) in reported weekly use of homework is promising; both semesters utilized online homework, either in a separate product or embedded within the textbook. It appears that students in General Chemistry I are spending more time interacting with their peers and engaging in the homework.

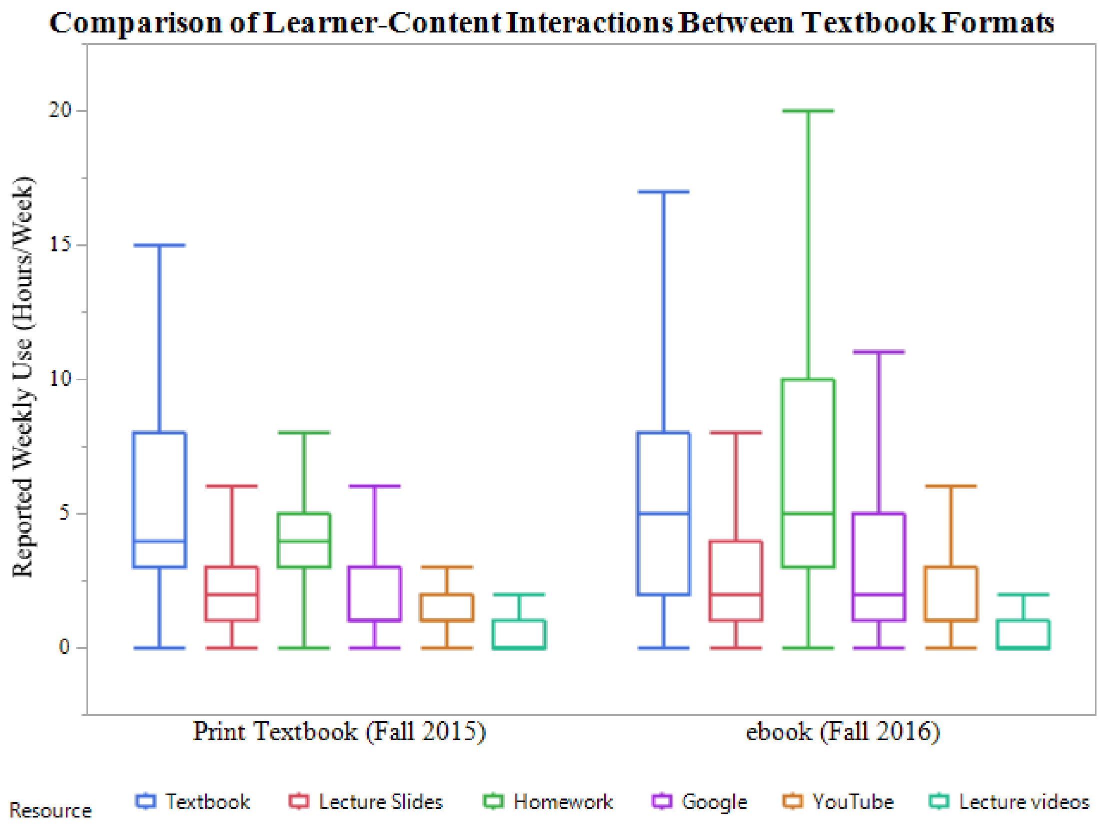


Figure 26: The average self-reported hours of content-delivery resources used by General Chemistry I students between a print semester (fall 2015, left) and ebook semester (fall 2016, right).

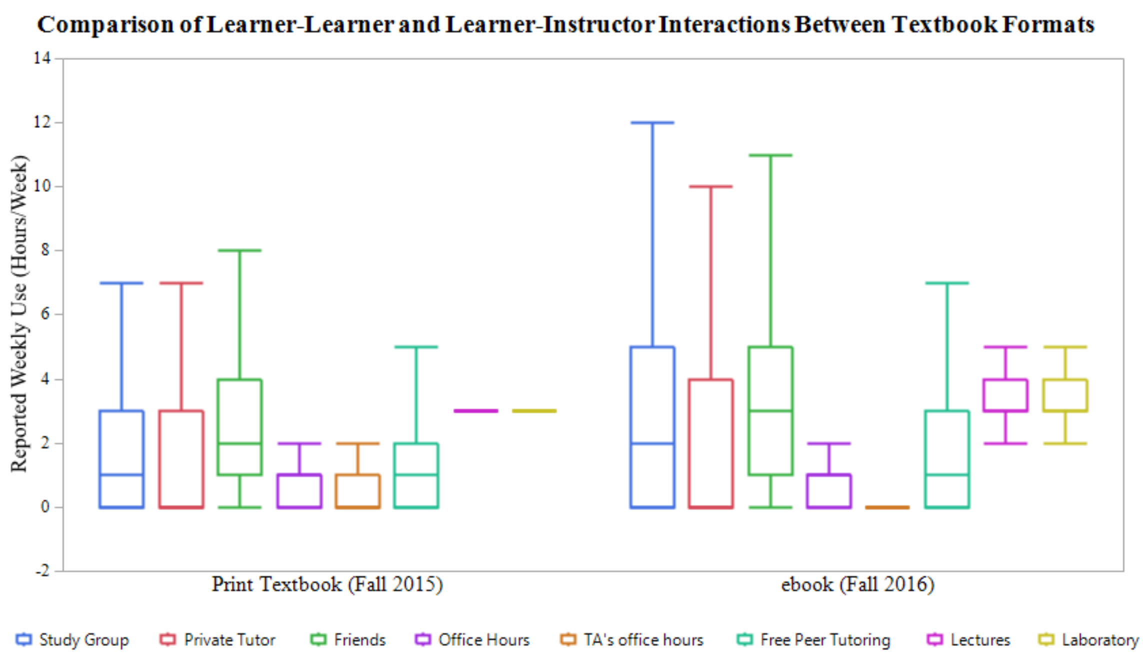


Figure 27: The average self-reported hours of student-student and instructor-student interactions used by General Chemistry I students between a print semester (fall 2015, left) and ebook semester (fall 2016, right).

The *learner-instructor* and *learner-learner* interactions depicted in Figure 27 can be broken down further in Figure 28. Which group of students is driving the increase in the use of study groups, tutoring, and friends?

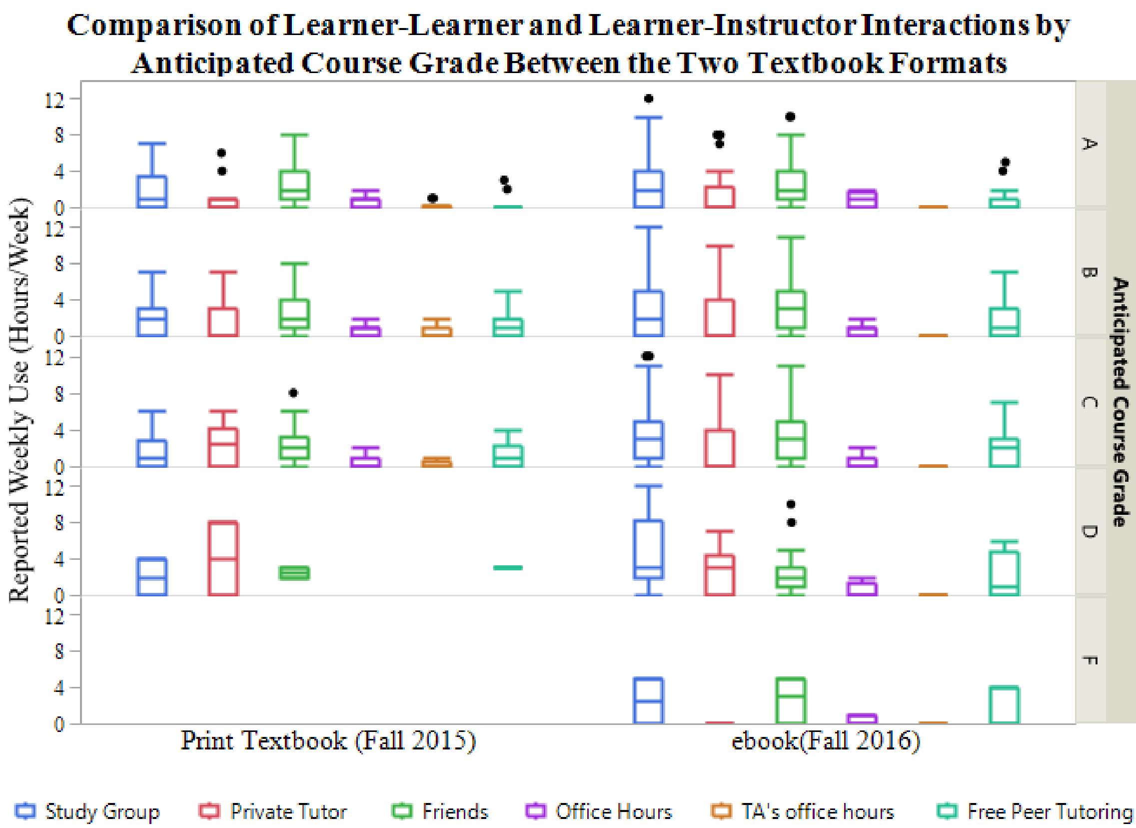


Figure 28: boxplots stratified by anticipated course grade showing the use of learner-learner and learner-instructor interactions for the print (left) and ebook (right) semesters

The biggest change is in the response of the lowest-achieving students; in the print book semester, students who anticipated an F reported not using peer interactions or seeking out an expert for help. In the ebook semester, students who expect an F reported using study groups, friends, and free peer tutoring frequently. It is unclear what effect this help has on their course performance, because students at all achievement levels in the fall 2016 semester reported similar frequency of visiting professor's office hours.

The results of the significance tests in Table 4 and results in Figures 25, 26, and 27 between print and ebook semesters can be further exemplified by arranging the use according to the participants' anticipated course grade, as seen in Figure 29. An ANOVA of these course resources had significant differences in use within the semester between participants of different anticipated grades. The ω^2 effect sizes were small for fall 2015 textbook, fall 2015 lecture slides, fall 2016 search engine use, and fall 2015 university-sponsored peer tutoring. Fall 2015 private tutor and YouTube videos had medium effect sizes.

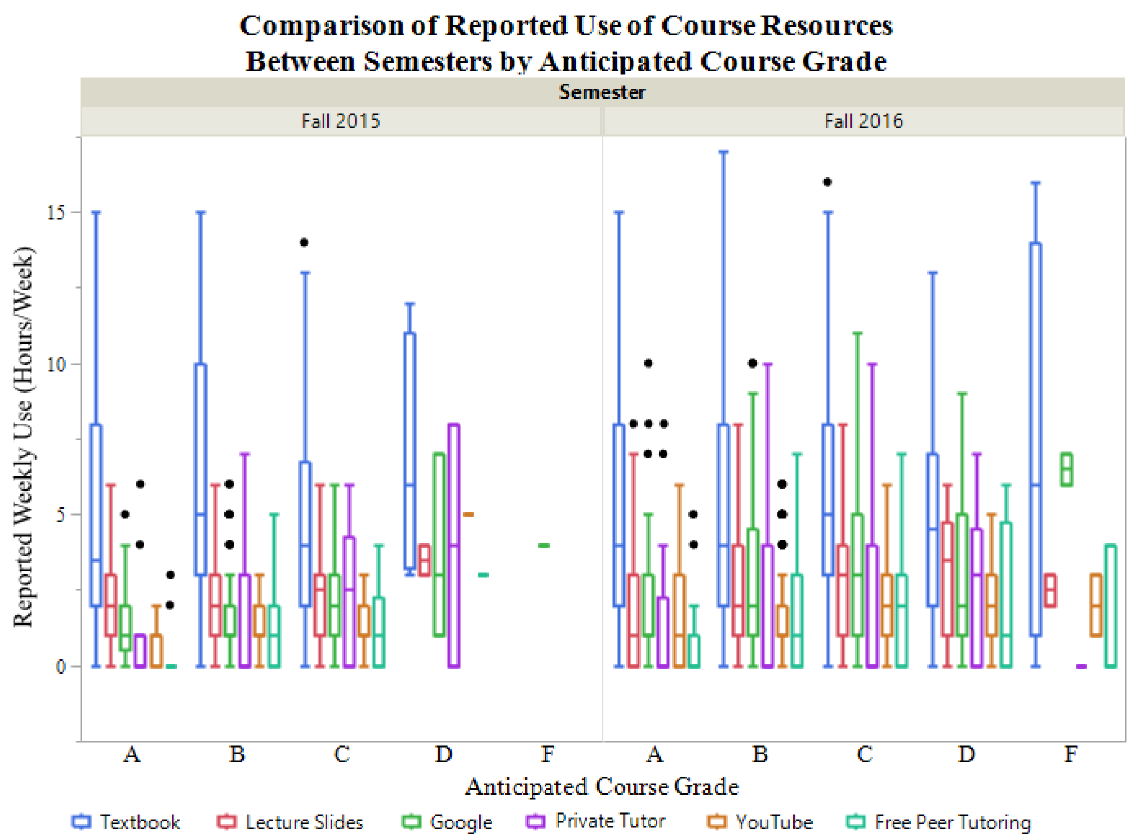


Figure 29: boxplots to demonstrate the spread of reported weekly use of resources that had a significant difference between formats (print on the left, ebook on the right); each resource is split by anticipated course grade.

Figure 29 illustrates that in the print book semester the failing participants reported almost no use of resources, but in the ebook semester the failing participants reported using the textbook more than their better-performing peers. The amount of time spent (or at least perceived to be spent) with the textbook is not a direct link to successful outcomes. In addition, there is an increase from print to ebook in the amount of time that participants reported using search engines, private and university sponsored peer tutoring to aid their studies. This suggests that the more-interactive textbook medium is not answering the students' concerns as it once did.

Between the two formats, seven equivalent features were identified. The reading assignment in the electronic format is roughly equivalent to the reading outline in the print textbook, whereas the written text, definitions, and images/figures in either format were roughly equivalent (different authors, but same curriculum). The end-of-chapter exercises (or "chapter exercises") had slightly different placement as the mastery questions (mastery questions were embedded within as opposed to grouped at the end of the chapter), but the function of these features was equivalent. The worked-out examples in the print format are similar in function to the "text solutions to the example problems." The online quizzes were administered through the outside platform and drew from the same question banks that had been generated by the instructors.

As seen in Table 5, a Mann-Whitney nonparametric analysis of the ordinal-type Likert ratings can elucidate whether there was a significant difference in the participants' perceived helpfulness of that textbook feature. For measures of central tendency, the median rating of each feature is also listed. These results demonstrate that students do perceive the equivalent features of the two mediums to differing degrees of helpfulness, although a qualitative interview would be needed to discover the likely reasons for the differing perceptions.

Table 5: Comparison of Perceived Helpfulness Ratings for Print and ebook

	Print (Fall 2015)		*ebook (Fall 2016)*		*Significance Testing*	
Feature	*Median*	*N*	*Median*	*N*	*p-value*	*Effect Size (r)*
Reading Outline	3	577	3	767	**<0.0001**	0.23
Images/Features	3	569	3	766	0.1774	
Animations	3	576	3	765	**<0.0001**	0.13
Written Text	3	573	3	763	0.0679	
Worked-Out Examples	4	575	4	768	**0.0018**	0.085
End-of-Chapter Questions	4	576	3	768	**<0.0001**	0.13
Definitions/Glossary	3	577	3	768	**0.0481**	0.054
Online Quizzes	3	575	3	769	**<0.0001**	0.15
Participants asked to rate the "helpfulness" of the feature on a scale of 0= "I don't use," 1 = "very unhelpful," 2 = "Unhelpful," 3 = "helpful," and 4 = "very helpful"						

Although not reflected in the median values in Table 5, reading outline, animations, worked-out examples, end-of-chapter questions, definitions, and online quizzes were all significantly different with small effects ($r < 0.3$). The perceived helpfulness of the reading outline increased significantly ($p < 0.0001$, $r = 0.23$) as did the animations ($p < 0.0001$, $r = 0.13$). Surprisingly, the helpfulness rating of the worked-out example decreased significantly ($p = 0.0018$, $r = 0.085$); this represents an opportunity to explore and innovate how a worked-out example functions in an interactive, electronic textbook. The end-of-chapter questions also significantly decreased ($p < 0.0001$, $r = 0.13$); free responses and instructor feedback noted that the level of difficulty of the questions between the two textbooks differed. The helpfulness of online quizzes and definitions also decreased significantly, although further inquiry is needed to discern what the reason might be for these significant changes, because the quizzes were instructor-generated and hosted in a separate product from either textbook.

Longitudinal Data

Six semesters' worth of data has been collected with this survey instrument. This provides an opportunity to measure trends in students' opinions in the resources available to them for general chemistry. Some questions were added in response to students' open-ended responses; early on, participants' comments indicated that students find professors' and TA's office hours inconvenient to their study schedules. This likely contributes to their reported lack of use of *instructor-learner* interactions. To probe this sentiment further, questions about students' study timetables were implemented in the spring 2017 semester. The first question of this pair asked participants to select all the times they would typically study for General Chemistry I. Figure 30 shows three semesters' worth of data (spring 2017 through spring 2018); each sample (N = 641, 739, 621, respectively) selected the hours in which they were most likely to study for chemistry. Students overwhelmingly favor studying later at night.

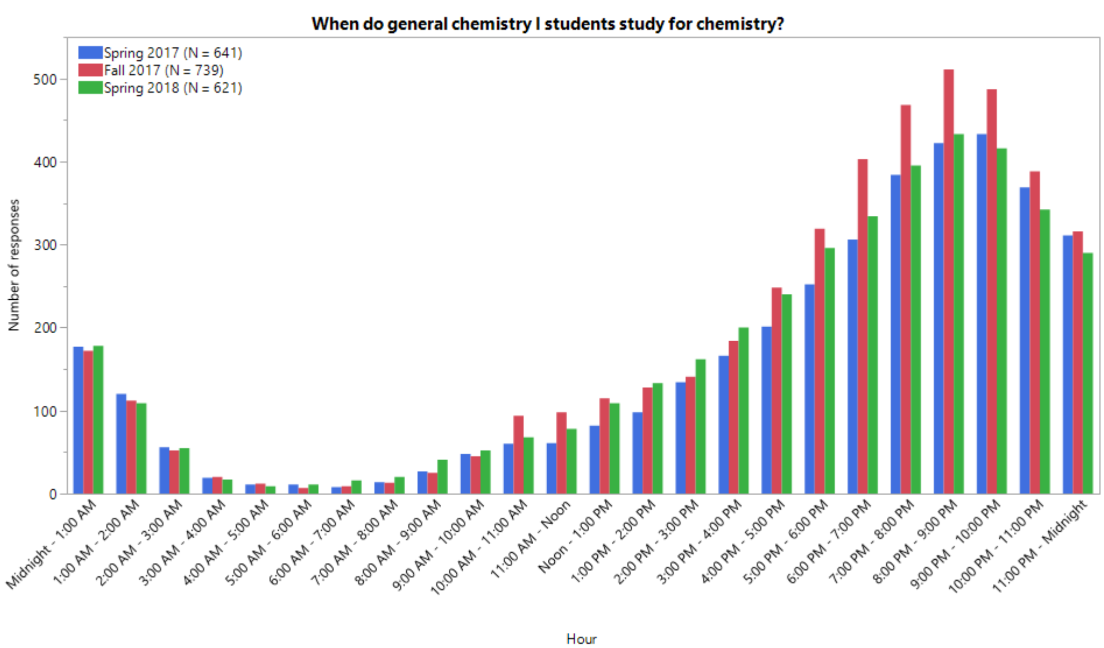


Figure 30: General Chemistry I students' typical study hours for the spring 2017 (blue), fall 2017 (red), and spring 2018 (green) semesters.

While the students tend to study in the afternoon/evening hours (3 PM to midnight), instructors and TAs are not always available during these hours. At the beginning of the implementation of this survey, professor's office hours were limited to a few times a week and TAs were available from 9 AM to 5 or 6 PM most days. By the time this item was added to the survey, professors had moved to larger classrooms to hold more open and accessible office hours, but they were still held during business hours. TAs were only available from 9 AM until 9 PM, although it is clear from other survey items that the TAs were not a highly-regarded source of information. The new set of survey items also asked participants which days they were most likely to study, independent of study hours, and Figure 31 summarizes their responses. Participants were most likely to study—and therefore ask questions or need assistance—at the beginning of the week, and this may conflict with experts' availability.

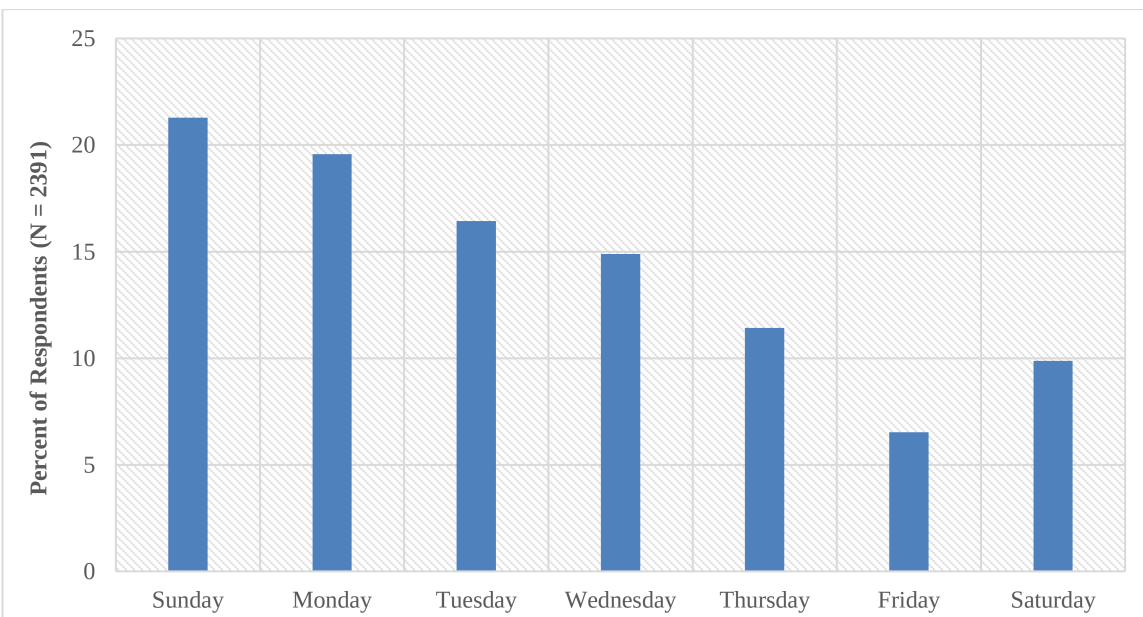


Figure 31: in response to a question about which days General Chemistry I students are most likely to study for chemistry; this plot shows the percent of respondents who each day (multiple selections were possible).

If the instructors' office hours are misaligned with students' study schedules, how do students communicate with each other, and would they be open to a moderated message board? In a separate set of questions, participants from all six semesters (N = 3646) were asked whether they would be willing to participate in a moderated message board, as demonstrated in Figure 32. The two most popular choices for the forum include the learning management system (eLC) and the app GroupMe, a popular group messaging app. These are avenues that are already used by most students. Piazza is another message board option with a LaTeX equation editor and LMS-integration that has been utilized by the mathematics department at UGA.[110]

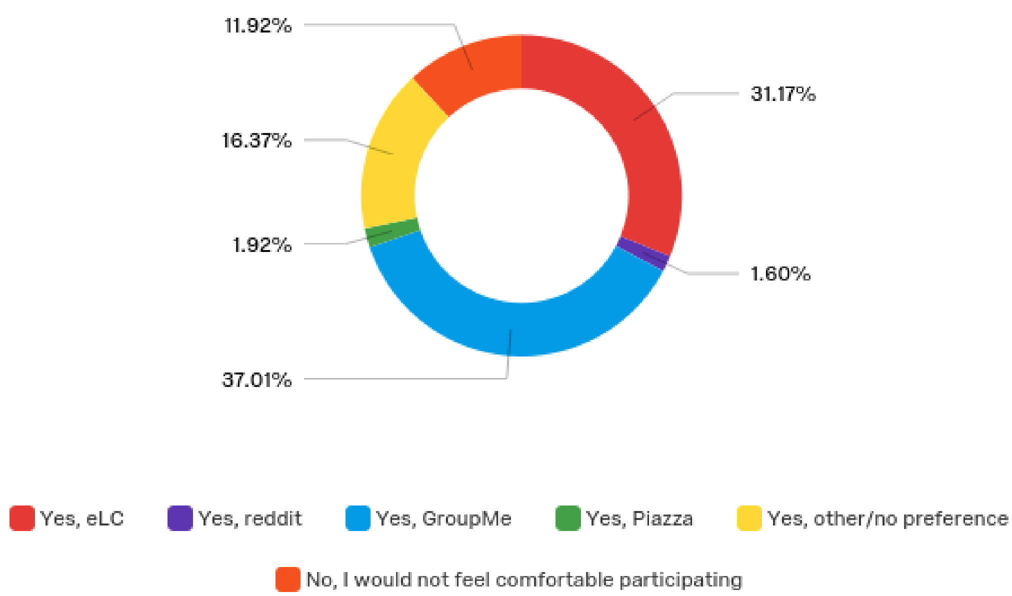


Figure 32: summary of General Chemistry I students' willingness to participate in a moderated message board; results include GC1 participants from all six semesters.

The final set of survey items addressed the students' format preference and allowed for an open-ended response on their preferred format. Participants from all six semesters were split by their stated preference (print/ebook), and their reasoning was compiled in a dataset that was coded using *a priori* codes from the UTAUT2. These codes as well as the emergent themes from the dataset are summarized in Table 6.

Table 6: List of Format Preference Codes

Code	*Definition*
Accessibility*	Ergonomic or designed for use regardless of disability
Availability*	Convenience or reliability of the textbook format
Concentration*	Ability to focus on textbook material without distraction
Effort Expectancy	Degree of ease associated with use
Environmentally Friendly*	Effect that technology eBook adoption has on environment
Facilitating Conditions	Availability of technical support
Habit (Automatic)	Extent to which individual believes preference or use is automatic
Habit (Prior Behavior)	Prior experience with preferred format textbook format
Hedonistic Motivation	Fun or enjoyment derived from use (perceived enjoyment)
Interactive Features*	Receive feedback on understanding
Mark the Physical Copy*	Tendency to mark, highlight, or annotate the textbook
Performance Expectancy	Degree of utility or benefit in using preferred format
Price Value	Trade-off between benefits and cost
Social Influence	Extent to which individual believes important people advocates for technology
Tactile*	Desire to hold or view a physical book
Task Management*	Preferred format encouraged respondent to stay on task
Up to Date*	Preferred format has the most current and up-to-date information
*indicates an emergent theme	

Each semester, survey respondents were asked to indicate whether they preferred a print or electronic format. As seen in Figure 33, prior to the adoption of the electronic textbook (to the left of the vertical line), most participants (88 percent) preferred the print format. Once the eBook was adopted, the polarity of preference dropped sharply until spring 2017 (three semesters after adoption) when the preference for each format was roughly equal. The codes are represented on the pie charts, describing why each group of respondents chose that format; this illustrates how the reasoning behind students' preference evolved over time and exposure to the electronic format.

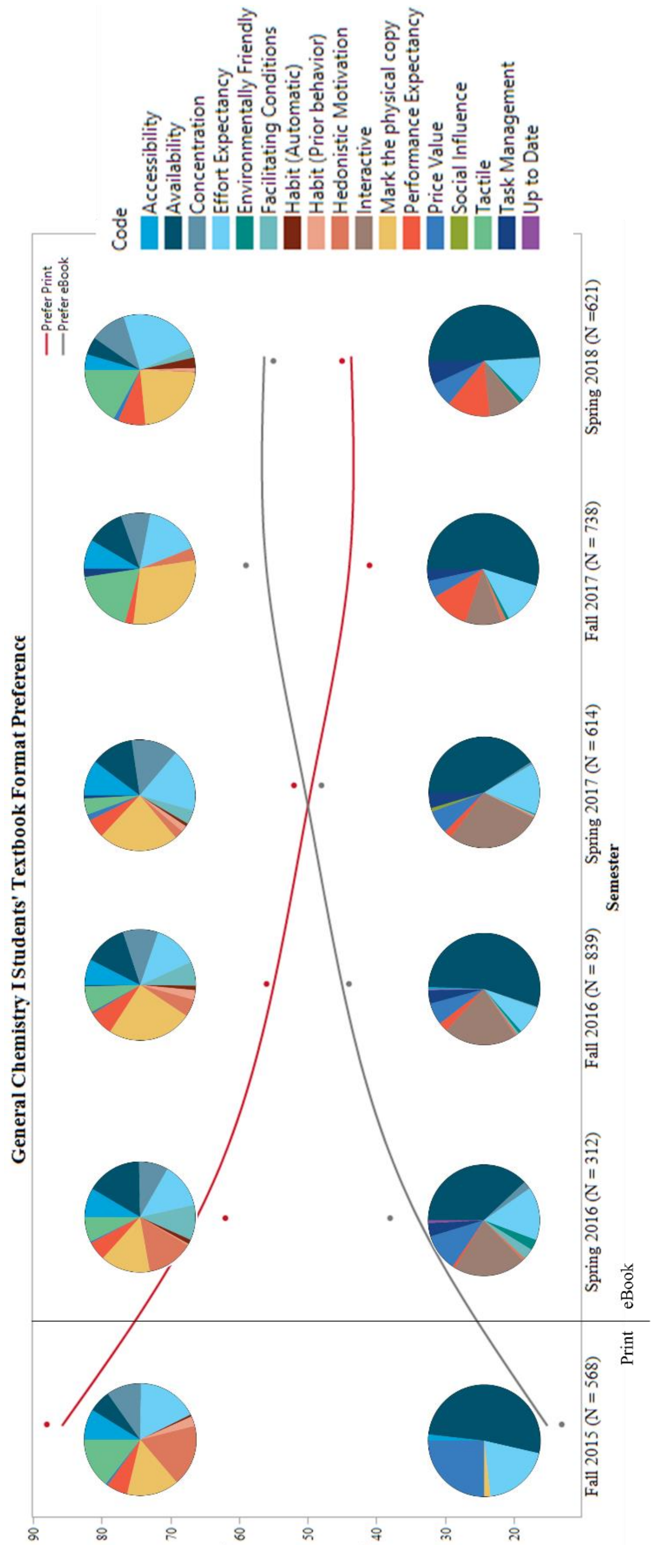


Figure 33: the base plot shows the change in the format preference over the six semesters; the red line connects the percentage of General Chemistry I participants who preferred the print format, the black represents the percentage who preferred the electronic format. The vertical line demarcates the plot to represent a course-wide adoption of a new textbook format. Each pie chart represents the codes for that group of participants in that semester.

To summarize the general trend in responses, those who preferred the print format had more reasons that contributed to their preference. Initially, the print textbook was "easier to use" (effort expectancy), easier on the eyes (accessibility), less distracting (concentration), and that the desire to hold a physical copy (tactile) and mark that copy contributed to an enjoyment of use (hedonistic motivation). With exposure to the ebook, the enjoyment of using a physical copy lessened, but a desire to hold and mark a physical textbook remained, along with an increasing expectation that using a print textbook would be require less effort. The participants who preferred an ebook had fewer codes and more homogeneous and consistent reasoning; the main reason in every semester was availability, but the lower cost (price value) of the ebook was almost as attractive as the interactive features. With increased use of the ebook and its "interactive" features, the perceived benefit of use (performance expectancy) and ease of use (effort expectancy) increased.

Both groups of students referred to their preferred format's availability, or the convenience or reliability of the textbook format. This emergent theme manifested in print textbooks as ability to use regardless of internet connection. For the ebook, this code described the sentiment of being able to access[2] the textbook on their computer any time and avoiding the burden of another heavy textbook. For instance, one respondent in fall 2017 described the availability of the ebook as:

> "The eBook is easier to access and carry around as you only need your laptop. As a college student lugging around heavy textbooks is impractical

---

[2] "Access" here is distinct from "accessibility" whose operational definition is included in Table 6; codes of this nature commented on eye strain from excessive reading from a screen.

as carrying your laptop is easy and needed for other classes other than chem as the chem book is only applicable to one class."

Unlike the print textbook, respondents often commented on the "interactive" features of the electronic format. One fall 2016 student described it as, "The electronic copy included interactive activities, extra practice with immediate feedback, [and] tutorials for problems." This highlights an important advantage for ebooks, the ability to deliver immediate (and potentially customized) feedback on problem-solving and more opportunities to practice. Comments in this theme suggested a strong openness to a textbook that embeds interpersonal interaction (such as a message board), as students respond positively to this type of innovation.

While the electronic copy provided tools to highlight and annotate, those who preferred print did not care for these tools. As a spring 2018 student said, "I like to write in the book and see my notes. I do not like the comment part or the highlighting in the textbook. I like to underline the important information instead of highlight." Some responses were unaware of these tools in the ebook, especially in the early adoption semesters.

In terms of the price value construct, the participants who preferred ebooks commented on the smaller price tag of the electronic version. However, one respondent who preferred the print book remarked, "[print textbooks] can be resold at the end of the semester to help recoup some of the horrendous costs of buying the books. These companies are essentially renting the information to me at the same cost as buying the book, it's despicably greedy." This echoes one of the concerns from the literature,[69] that ebooks that are not open-source are charging students for renting a digital file. This dataset does not contain information about the wisdom of purchasing a more expensive print copy with the goal of reselling at the end of the course to recover the investment; the quote illustrates the college student mindset of trying to minimize costs and feeling squeezed by the bookstore, the publisher, and the higher education system.

Conclusions

In fall 2015, the print textbook and homework were tied for most widely used on average. When broken out by anticipated course grade, students who performed well spent more time with the textbook, whereas students who performed poorly favored the homework over the textbook. There was minimal use of *learner-instructor* interactions, *learner-learner* interactions were more used; however, use of these two types of interactions was not significantly different across anticipated course grade. The most helpful features of the print textbook (tutorials, worked examples, end-of-chapter problems) are mathemagenic features that provided an explanation or opportunity to practice algorithmic problem-solving. The least popular features were the real-world examples or key experiment sections that provided real-world context to the information. General Chemistry I students favor resources and textbook features that help them earn a good grade (69.9 percent), and conceptual understanding is a secondary concern.

In the immediate transition to an ebook, reported homework use increased dramatically, overtaking the textbook as the most-used course resource. However, the most realistic comparison of formats is between the fall semesters. In fall 2016, homework use was still more frequently reported than textbook use, and the increase in homework use per week was significant with a large effect. After the transition to the electronic platform, *learner-learner* interactions (study groups, tutoring, and friends) increased significantly with small to medium effect sizes. While the use of professor's office hours remained steady but infrequent, TA office hour use dropped to zero, limiting the interaction that students have with content-experts as opposed to peer or private tutors. The biggest shift in interactions with peers and experts is seen in the poor-performing students in the fall 2016 semester, who report interacting with both groups unlike the fall 2015 semester. With the print book, students who anticipating receiving high (A's and B's) and low (C's and D's) grades reported similar amounts of textbook use, but with the eBook the students who anticipate doing well in the course tended to use their textbooks more than the homework.

For equivalent textbook features, a comparison between formats revealed that the helpfulness ratings of reading outline and animations increased significantly whereas worked-examples and end-of-chapter questions decreased significantly. These results provide an opportunity to employ eye-tracking methodology to probe how students interact with these features in their current form.

The longitudinal data reveals unsurprising trends. General Chemistry I students tend to study later at night, even though there are fewer content-experts (instructors or TA's) available to them at those hours. Predictably, this population also favors studying at the beginning of the week ahead of immediate deadlines. These trends are important for instructors to pay attention to when scheduling office hours. When an expert is unavailable while the students are studying (especially if under a deadline), students will fall back on Google, peers, and private tutoring options. Almost 90 percent of all General

Chemistry I respondents favor a moderated, online message board option to alleviate this concern, and their open-ended responses suggested that they are creating these messaging opportunities amongst themselves. For the best outcomes, instructors might consider creating spaces and opportunities to meet with students virtually. An online option for teaching assistant hours could provide more flexibility in busy graduate student schedules, as well. Textbook publishers should consider this opportunity to innovate and embed a truly interactive feature in their ebooks, such as a moderated message board.

The qualitative data on students' format preference is valuable feedback for instructors interested in adopting an electronic textbook. Students who were enthusiastic about the ebook loved the convenience and availability, and with increased exposure they credited the interactive features as one aspect that would increase their performance expectancy. The students who were reticent about the electronic format were concerned about being able to mark and hold a tactile copy of the textbook, attributing that as essential to their study performance. After exposure, the sample exhibited less polar preferences in textbook format; at the end of data collection, the split of opinions was nearly half for each format.

The limitation of this work is the self-report nature of this data collection method. The researchers were unable (due the sample size) to verify the participants' demographic data or confirm the accuracy of their perceived use; therefore, these results demonstrate how the first-semester General Chemistry students believe they utilized the course resources and textbook features.

Future research in this area includes longitudinal studies, analysis of data from other introductory courses, and studying how an ebook utilizing a different curriculum can impact student interaction patterns. There are indications that prior experience with a particular technology predicts future behavioral intention and use behaviors;[73] accordingly, the predictive value of experience could be evaluated by following participants who used the ebook in the first semester of General Chemistry and answered the survey in the second semester of the course to compare their answers on a repeated-measures model.

CHAPTER 4

EFFECT OF DYNAMIC OR STATIC VISUALIZATIONS ON STUDENT GAZE PATTERNS DURING PROBLEM-SOLVING

One benefit of ebooks is the features that allow interaction between the learner and content in a more dynamic and constructive fashion. In the longitudinal study on students' perceptions of their textbook, one of the highly-valued features of the ebook was the animations. To more deeply explore this *learner-content* interaction, an in-depth investigation was devised to explore how students utilize the information presented in either a static series of images or a dynamic animation to aid problem-solving. The goal of this work was to measure and characterize students' visual attention patterns post-test "ordering" items to the two presentation formats—a set of static images and an animation of those images—that are almost exactly visually equivalent and share the same audio overlay.

Literature Review

While eye-tracking studies on problem-solving have been well-documented,[50,111–115] there is little eye-tracking research pertaining to solving "ordering questions" or on "temporal sequence" problems. These terms reflect the available literature on educational research databases. One of the few relevant studies assessed difficulties in causal reasoning in biology; researchers tested eighth and tenth grade students on their ability to organize steps in a temporal sequence after a short, written introduction to a topic.[116] The researchers found that most students could not place the steps in the proper temporal sequence, likely due the complexity of the material.[116]

Despite the availability of psychological research into animations for education,[48,117–120] the research on animations versus static graphics yields conflicting results as to whether the number of user-controlled views was important to animation effectiveness, as well as whether "decorative" animations were useful in an online course.[121] Instances in which animations succeeded over static images in teaching a concept could be attributed to differences in presentation of the information; more detail provided in the information or a higher element of interactivity with animations.[119] Animations may also be more difficult to understand by novices and may be distracting or harmful to the construction of knowledge. One study compared the effectiveness of animations and static images as a tutorial for medical students and found that the static tutorial was just as effective as the animation (as determined by a test score); furthermore, the static images produced the same amount of cognitive load burden as the animation.[122]

In studies on problem-solving, researchers commonly divide the process into multiple problem-solving phases. Some studies on mathematical word problems have described four problem-solving phases;[54] (1) crafting mental representations from the components of the prompt of the problem, (2) integrating the mental representations into a cohesive understanding, (3) planning an approach, and (4) executing the approach. A more straight-forward definition of problem-solving is demonstrated in studies that used two problem-solving phases: an initial systematic reading phase and a solution phase.[123]

In the present work, we divided the students' actions into three problem phases: planning, problem-solving, and checking.[31]

Theories of Working Memory and Eye Tracking

Eye tracking has been used as a tool in scientific education research. Eye-tracking techniques rely on the theory of Information Processing (IP theory), which focuses on cognitive states (memories) and the processes by which individuals retain or decay information.[44] This theory describes the structure of various cognitive processing centers (memories) as well as the schematic movement of information from sensory memory to short-term/working memory, where it might be encoded and stored in long-term memory. The work of Baddeley and Hitch gave rise to the term "working" memory, rather than short-term, because they emphasized the processing capacity of this cognitive state.[46]

Eye-tracking instruments also rely on the eye-mind hypothesis to derive meaning from the eye-movement measures.[124] This hypothesis considers eye movements to be the observable measure of visual attention that is linked to the cognitive processing of information.[125] Although there are some limits to this hypothesis,[126] researchers can relate certain eye-tracking measures to the level of processing.

Cognitive load theory (CLT) was first proposed in 1988,[127] and in its current form, it provides guiding information in the design of instructional materials.[117] CLT builds on the foundation of IP theory and assumes a working memory that is limited to holding 5-9 elements (or "bits") of novel information and can only actively process 2-4 bits simultaneously.[117] The construction and automation of schemas can treat multiple bits of information as one to reduce the strain (or cognitive load) on the working memory.[117]

Schema is a term to describe a participants' domain knowledge. For an expert, this knowledge is stored (and can be recalled) from long-term memory in larger, more coherent chunks based on domain principles. Novices tend to organize their unconnected fragments of information around surface features. Sometimes these superficial, self-explanatory interpretations of novices are termed "phenomenological primitives" or p-prims. As the number of encounters within a domain increases, the novice's p-prims cluster and organize into more complex, expert-like schema.[128]

Mayer's cognitive theory of multimedia learning also relies on IP theory to inform the design of educational materials.[48] The working memory is composed of a visual sketchpad processing center and phonological/verbal loop processing center. While these centers are separate within working memory, their processing capabilities and speeds are interdependent in an additive fashion. Information-processing by both centers (via multiple modalities of input) is more effective than relying on one processor alone. This strategy allows extraneous cognitive load to be reduced and has led to several instructional design principles.[129]

Methods

To analyze students' visual attention on the stimuli and the post-test, we used a Tobii T120 eye tracker with a sampling rate of 60 Hz on an LCD screen (1280 by 1024 pixels) running Tobii 2.0.4 software.[130] By default, fixations are defined as 100 milliseconds on a radius of 30 pixels.[31] Eighty-nine anatomy/physiology students served as participants and were randomly assigned to either the animation or the static presentation; only 81 participants had sufficient calibration to the eye tracker (greater than 80 percent of fixations) were included in the analysis.[131] Using the first hourly exam scores, an independent-samples t-test found that differences in scores between the two experimental groups on the first course exam were not significant; this indicated that the overall membership of each assigned condition was not significantly different in terms of course achievement. The groups watched the assigned stimulus twice, with a short break in the middle to prevent attentional and gaze drifts during the six-minute presentations.

After the two viewings, all participants answered the same post-test questions under eye tracking. The 17-item post-test was devised of fact-recall, conceptual evaluation questions, and "ordering" questions; an example of these questions is in Appendix B. These ordering questions were included to test student recall of the steps of the biochemical pathway presented in the multimedia materials. The language of each step was taken directly from the text on the screen of the nearly-identical presentations, and this stimulus was designed such that the student could "drag-and-drop" the words of each step into the appropriate blank, as seen in Figure 34. The grading key of the post-test was reviewed by two post-doctoral physiology fellows and two physiology graduate students for validation.

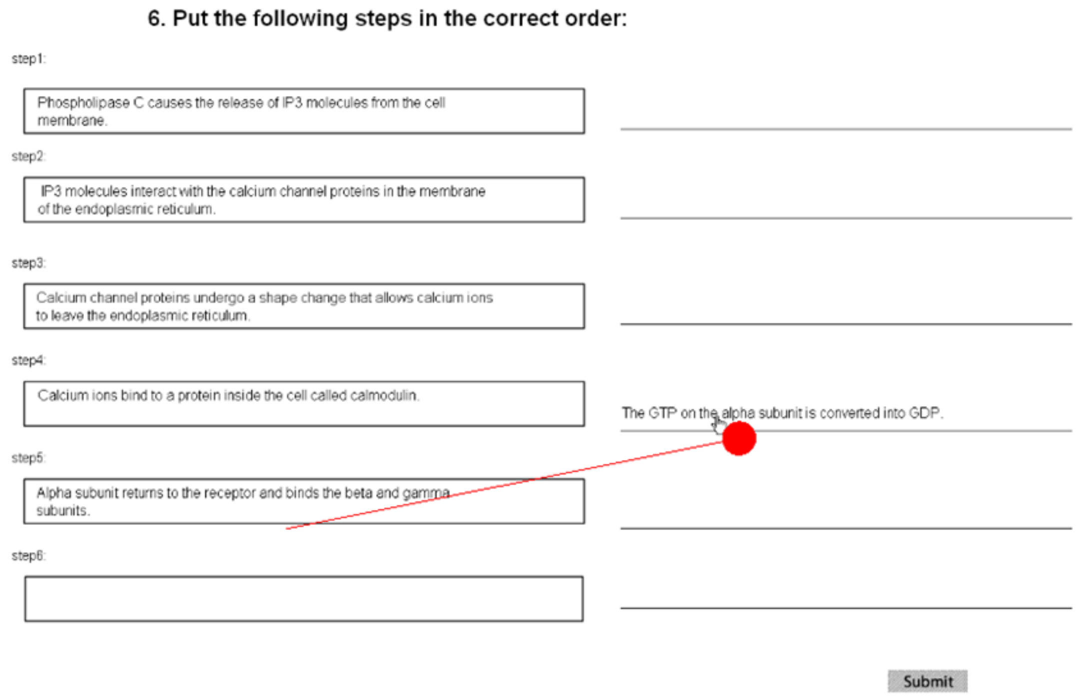


Figure 34: An example of the ordering items on the post-test. These items were designed to test the participants' recall of the steps in the biochemical pathway. The steps—taken directly from the narration that was identical in both presentations—were presented on the right in a random order to participants. The participant was asked to drag-and-drop the steps into the correct sequence on the left.

In this experiment, the dependent variable was the post-test score on the three ordering questions (for a total of 0 to 3 points). The independent variables include presentation type (animation or static), sex, time spent in each problem-solving phase, total fixation duration, fixation count, and visit count.[31] Statistical analyses were executed using R 3.1.2.[132] For this study, we had three hypotheses. First, eye-tracking measures (such as fixation duration and fixation count) are correlated. Second, that participants assigned to the animation condition would perform better than students assigned to the static condition on the "ordering" questions. Third, between students who correctly answered the ordering questions and those who answered incorrectly, there would be a significant difference in (i) amount of time spent answering the questions, (ii) the fixation durations and fixation counts, (iii) the time spent in each phase of the problem solving, and (iv) the scanpath patterns.

Results and Discussion

The presentation styles appeared equivalent at first glance, but there were several minute background details in the animated version that were not present in the slides, demonstrated in Figure 35. These differences in the visuals made it difficult to find equivalent scenes on which to compare eye-movement metrics.

Figure 35: a side-by-side comparison of the static (left) and dynamic (right) presentation formats of the alpha-receptor pathway.[133]

The post-test results were not significant overall between the two conditions, likely due to the participants' pre-exposure to the subject (see Appendix B). However, using multiple variables in a logistic regression can reveal relationships between score and eye-movement metrics and other independent variables. Because there were 18 independent variables, dimension reduction was necessary to simplify the regression model. Based on previous research,[49,51,123] it is likely that the eye-movement metrics are correlated. This work collected information on the participants' total fixation duration (the total of all fixation durations within an AOI), fixation count, visit count, time-on-task, and number of mouse click events (from moving answer choices). Based on a Pearson's r correlation matrix, these metrics were correlated ($r = 0.56$ or higher for each) within each ordering item, and therefore any significant difference in one of the eye-tracking metrics is likely to be significant for another metric. Total fixation duration, fixation count, and time-on-task were highly correlated ($r \geq 0.95$), and so the commonly-reported total fixation duration was chosen as the metric for the logistic regression models.

To elucidate problem-solving strategies on these ordering questions, participants were split into two groups—those who answered correctly and who answered incorrectly. Forty-six participants (56.8 percent) answered the first item correctly, nine participants (11.1 percent) answered item two correctly, and 47 (58.0 percent) answered the final item correctly. For each ordered logistic regression, the possible score for each six- or seven-step ordering questions was 0, 1, 2, and possibly 3 (in the case of the third ordering item). To determine the relationship between post-test score on ordering item and eye-movement metrics within each item, sex, and movie type, the results are summarized in Table 7.[31] Negative coefficients indicate that participants who answered incorrectly (and had lower scores overall) fixated longer on that item, which is the first suggestion of cognitive load. Of the ordering items, this difference in total fixation duration was only significant for the third ordering item, which corresponds to steps at the end of the biochemical pathway. This significant result provided the basis for further analyses on the third ordering item. Contrary to the expectation that the media type (static versus dynamic) would influence their performance and serve as a distinguishing factor between groups, there was not a significant difference in score on the ordering items because of visualization type.

Table 7: ordered logistic regression of participant scores and total fixation duration on each ordering item, presentation format, and sex

Independent Variable	*Coefficient*	*t(81)*	*p-value*
Visualization Type (Static or Dynamic)	0.186	0.438	0.661
Sex	0.564	1.317	0.188
Ordering item #1	-0.015	-1.881	0.060
Ordering item #2	0.014	1.274	0.203
Ordering item #3	-0.023	-2.181	**0.029***
Significant p-values are bolded and starred.			

The results of Table 7 further reduce the dataset to total fixation duration on the third ordering item of the post-test. As previously mentioned, some studies on problem-solving define different phases to the solving process. Given the wealth of information from the eye tracker, mouse click events provided a rationale for defining three problem-solving phases. The *reading-and-planning* (or *planning*) phase is defined from the time when the participant first views and begins to read the ordering item up to the first click event that ended in a completed drop of a choice into a blank (choices that were clicked but not moved did not count). The average total fixation duration in this phase was 17.40 ± 0.45 seconds. The *problem-solving* phase is defined as the end of the first completed answer-choice drop until the final answer choice has been dropped and there are no further adjustments to the order. The average total fixation duration was 40.47 seconds, which is considerably longer. The third phase is the *checking* phase, which is from the last completed answer drop until the "submit" button was clicked to advance the test. This was the shortest phase on each ordering item, with an average total fixation duration of 4.98 seconds.[31]

These problem-solving phases were defined for the third ordering item, which had a significant difference in total fixation duration between “correct” and “incorrect” groups in Table 7. The results of the ordered logistic regression between participant scores (0-3 points possible) and the visualization type, sex, and problem-solving phase on the third question are summarized in Table 8.[31] There was a significant difference between correct and incorrect groups for problem-solving phase two and three (*problem-solving* and *checking*). In phase two, the negative coefficient indicates that participants who scored incorrectly fixated longer, again suggesting a greater cognitive load. In phase three, the positive coefficient indicates that the correct group fixated longer. Since this does not fit the trend of incorrect participants experiencing high cognitive load and therefore fixating longer, an investigation of the scanpath patterns of each problem-solving phase of the third ordering item may explain the results of the ordered logistic regressions.[31]

Table 8: ordered logistic regression of participant scores and total fixation duration on each phase on third ordering item

Independent Variable	*Coefficient*	*t(81)*	*p-value*
Visualization Type (Static or Dynamic)	0.046	0.0105	0.916
Sex	0.750	1.692	0.091
Planning phase	0.000	0.011	0.991
Problem solving phase	-0.020	-3.795	< **0.001***
Checking phase	0.037	2.297	**0.022***
Significant p-values are bolded and starred.			

Finally, to investigate the significant difference in total fixation duration on the third ordering item in two of the problem-solving phases, a scanpath analysis was used to describe patterns of participants’ visual attention during those phases. Figure 36 shows how the areas-of-interest (AOIs) for the third ordering item (which was the eleventh question of the post-test). The AOIs labeled with letters are the steps in the pathway taken directly from the audio that was common to both presentations. The AOIs labeled with letters represent the blanks into which the participants were to drag the text of each step. The correct sequence of answer choices was DFCEBGA.

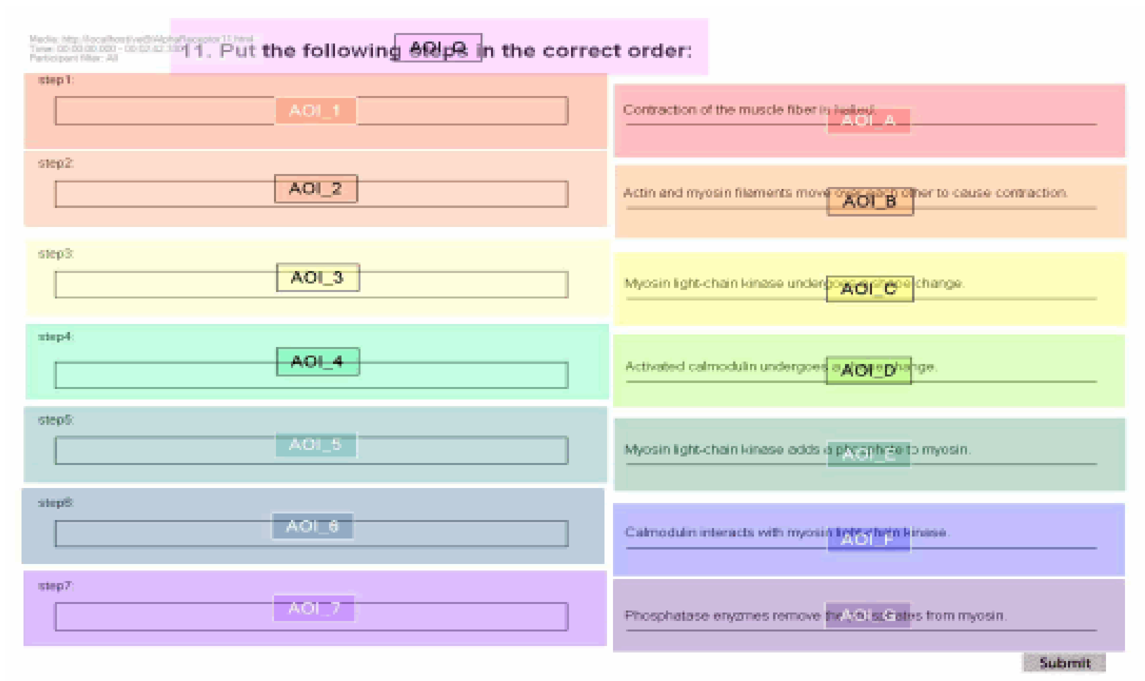


Figure 36: a screenshot of the third ordering item with the area-of-interest (AOI) assignments superimposed. The right column contained the initial position and order of the pathway steps (AOIs A-G), and the left column contained the blanks (AOIs 1-7) for participants to order the steps. AOI_Q indicates the question stem.[31]

As described in Chapter 2, a scanpath is the sequence of fixations and saccades that describes a participant's visual perusal of a stimulus. This is qualitatively and visually described in Figure 37.[31] When completing a quantitative or statistical pattern analysis, a scanpath (or group of scanpaths) are represented as a character strings. Scanpaths may contain any (or not contain any) subsequence of minimum length of three characters. Common patterns (shared by a specified percent of the group) can be elucidated through packages in the R statistical program **GrpString**.[134] Once common patterns to each group of scanpaths are identified, featured patterns from each group can be discovered.

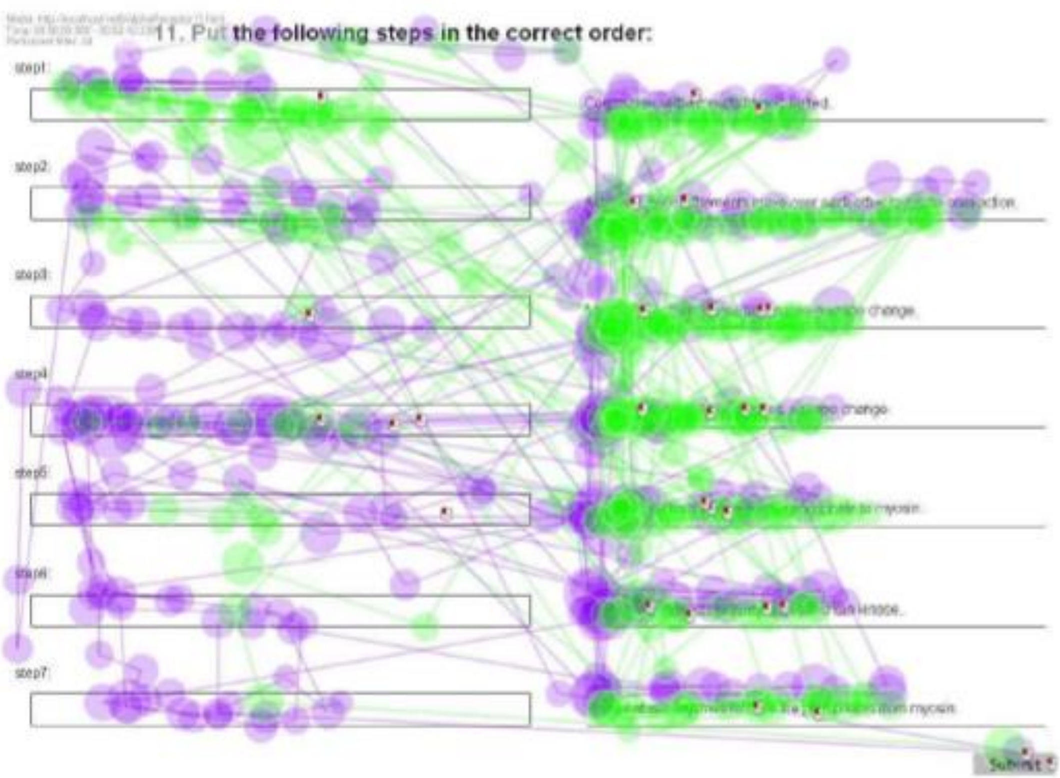


Figure 37: a visualization of an example of a scanpath of a participant who answered correctly (in green) and an example of a scanpath of a participant who answered incorrectly (purple).[31]

First, a look at the patterns within each problem-solving phase between correct and incorrect groups illuminates interesting trends. Table 9 shows the common patterns for correct and incorrect groups for phase one, the *reading-and-planning* phase, with a 20 percent criterion.[31] This cut-off dictates that a subsequence belongs to that group of scanpaths if it is present in at least 20 percent of the group's sequences.

Table 9: Common patterns returned for the planning phase on the third ordering item (with a cutoff of ≥ 20%)

Participants who answered correctly		*Participants who answered incorrectly*	
Pattern	*Percentage*	*Pattern*	*Percentage*
ABC	63.83	ABC	73.53
CDE	57.45	BCD	73.53
BCD	57.45	CDE	67.65
EFG	55.32	DEF	64.71
DEF	53.19	EFG	58.82
BCDE	51.06	BCDE	55.88
ABCD	44.58	CDEF	52.94
DEFG	42.55	ABCD	50.00
ABCDE	40.43	DEFG	50.00
CDEF	40.43	BCDEF	41.18
BCDEF	36.17	CDEFG	41.18
EDC	36.17	ABCDE	38.24
CDEFG	31.91	QAB	35.29
QAB	31.91	BCDEFG	29.41
ABCDEF	31.91	QABC	29.41
BCDEFG	31.91	1AB	29.41
ABCDEFG	29.79	ABCDEF	26.47
FGF	29.79	BAB	26.47
QABC	27.66	ABA	26.47
BAB	27.66	CBC	26.47
GFE	27.66	FGF	26.47
FED	27.66	1ABC	23.53
1AB	25.53	QABCD	20.59
1ABC	23.40	1QA	20.59
EFGF	23.40	AQA	20.59
DCD	23.40	BCB	20.59
DED	23.40	BCBC	20.59
1QA	23.40		
BABC	21.28		
EDE	21.28		

The list of letters indicates reading through the answer choices, often sequentially (no answer choices were skipped over). The correct group has more sequential reading patterns and less commonly, regressions (re-visits) to previous answer choices. The incorrect group, on the other hand, had fewer common patterns but more regressions in the scanpath and more frequently glanced back at the question stem. This could suggest the planning phase was more orderly and straight-forward for the correct group.

The designation as a common pattern does not preclude the pattern from appearing for both groups, such as 1AB in the planning phase. Instead, the **GrpString** package was used to find featured patterns, or patterns that have a high probability of being a member of one group, but a low probability of membership to the other group.[58] Table 10 shows the featured patterns for each group that are unique to that group of participants.

Table 10: Featured patterns returned for the planning phase on the third ordering item (generated with a criterion of ≥ 20 percent).

Participants who answered correctly	*Participants who answered incorrectly*
ABCDEFG	QABCD
ABCDEF	BCBCA
BABC	ABA
EDC	CBC
DED	AQA
GFE	BCB
FED	
DCD	
EDE	

The participants who answered correctly spent the planning phase reading through the answer text (AOIs A-G), generally sequentially. In contrast, those participants who ordered these steps incorrectly spent their planning phase deliberating over the first few answer choices (AOIs A, B, and C) and referring back to the question stem (AOI Q) which contained no salient information.[58]

In phase two, the actual problem-solving "drag-and-drop" phase, the function returned fewer common patterns for the correct group, as demonstrated in Table 11.[31] Participants who answered correctly tended to deliberate between starting points (such as in pattern BAB or 212), but generally the patterns were placing the steps into the correct order (for instance G6A represents looking at choice G, correctly placing it in blank 6, and then fixating on choice A). The wealth of patterns common to the incorrect participants share the common theme of deliberation between adjacent choices (such as CBC) or blanks (as in 212); some incorrect participants are still reading through the choices (as evidenced in the pattern CDE), while others are attempting to place the steps in the correct blanks (as seen in the patterns 3C3 and C3C).

Table 11: Common patterns returned for problem-solving phase on the third ordering item (with a cutoff of ≥ 20 percent)

Participants who answered correctly		*Participants who answered incorrectly*	
Pattern	*Percentage*	*Pattern*	*Percentage*
212	36.17	212	35.29
BAB	31.91	3C3	32.35
G6A	29.79	323	29.41
5G6	27.66	CDE	29.41
4BA	25.53	CBC	26.47
F21	23.40	BAB	23.53
2C3	23.40	545	23.53
D21	21.28	434	20.59
		234	20.59
		454	20.59
		543	20.59
		21A	20.59
		23D	20.59
		BCD	20.59
		C3C	20.59

By focusing on the featured patterns of the problem-solving phase, Table 12 demonstrates that participants who answered incorrectly continued the habit from the planning phase of deliberating between adjacent AOIs (such as in CBC, 323, or 434) and belatedly reviewing the choices sequentially (*e.g.* BCD). The correct group, on the other hand, demonstrates the placement of an answer choice between the correct position and the adjacent position (such as in D21, in which D is the correct first step).[58]

Table 12: Featured patterns returned for problem-solving phase on the third ordering item (with a cutoff of ≥ 20 percent).

Participants who answered correctly	*Participants who answered incorrectly*
G6A	323
5G6	CBC
F21	CDE
2C3	434
D21	3C3
4BA	454
	543
	23D
	C3C
	545
	234
	21A
	BCD

Finally, Table 13 details the common patterns for the correct and incorrect groups' scanpaths for phase 3 (the checking phase); no common patterns were detected for the incorrect group, even when the criterion was lowered to ten percent. Effectively, these patterns listed in Table 13 are the featured patterns for the correct group, and the incorrect group had no discernable patterns. This could be another indication of cognitive load; incorrect patterns seem less orderly.

Table 13: Common patterns returned for checking phase on the third ordering item (with a cutoff of ≥ 10 percent)

Participants who answered correctly		*Participants who answered incorrectly*	
Pattern	*Percentage*	*Pattern*	*Percentage*
456	19.15	No data detected	
345	17.02		
234	14.89		
123	14.89		
567	14.89		
67G	12.77		

With the TransEntropy() function in the **GrpString** R package, the average Shannon entropy value for each list of normalized transition frequencies within each phase of each group was calculated.[58] As suggested in Chapter 2, these entropy values are an expression of the participants' dispersity of fixations and can be related to the cognitive load of a participant. A statistical comparison of the entropy values is shown in Table 14.[58] In phases 1 and 2, the participants who answered incorrectly had higher entropy values, suggesting less targeted or more scattered fixations; the difference in these entropy values was only significant in the problem-solving phase, $t(81) = -3.515$, $p < 0.001$, $d = 0.80$, with a large effect. In phase 3, the incorrect participants had a smaller (although not significantly) entropy value; when viewed in conjunction with the lack of discernable patterns for this group in phase 3, this is likely the result of a lack of time spent in the checking phase.

Table 14: t-test of the average entropy values of correct and incorrect groups within each problem-solving phase on item three.

	Average Entropy		*Significance Tests*		
	Correctly Answered	*Incorrectly Answered*	*t(81)*	*p-value*	*d*
Planning phase	3.549	3.596	-0.383	0.703	
Problem solving phase	4.799	5.118	-3.515	**< 0.001***	0.80
Checking phase	1.672	1.352	1.160	0.251	

Based on the lack of expertise of all participants, the difference in entropy values is unlikely due to difference in prior knowledge structure. The significantly higher entropy value for the incorrect group in problem-solving (phase 2) indicates less-purposeful fixations, and the wealth of featured patterns for this group in phase 2 suggests a higher cognitive load.

Conclusions

In this study, we compared eye movement patterns for students who correctly ordered the events of a biochemical pathway responsible for vasoconstriction against those students who ordered them incorrectly. Similarly, we compared the patterns of students' eye movements from two groups—those who viewed the material using animations previously associated with improved student understanding of this pathway against those who viewed the material using static images extracted from those animations. For two of the three ordering problems, students with higher scores had shorter total fixation duration when ordering the events and spent less time (fixating) in the planning and solving phases of the problem-solving process (phases 1 and 2). This finding was supported by the scanpath patterns that demonstrated that students who correctly solved the problems used more efficient problem-solving strategies.

Total fixation duration, fixation count, mouse click (mostly moving choices), visit count, and problem-solving time are always highly correlated in the same phase. Only in phase two of item three, participants with incorrect score had significantly longer duration ($p < 0.001$). Sex and movie type did not have significant effect on post-test ordering item score.

In an investigation of scanpath patterns between participants who answered correct compared to incorrectly, in the planning phase both groups had patterns between adjacent choices (e.g. DCD or FED) without any jumps (e.g. ADE). In the problem-solving phase, the incorrect group's patterns were between adjacent choices or steps (e.g. ABC or 432) while the correct group's patterns were between choices and corresponding steps. In phase three, the correct group's patterns indicated that more than ten percent of participants checked answers sequentially. For the checking phase, the incorrect group had no checking patterns detected with ten percent cutoff, indicating that this group either did not check answers or that the fixations were so scattered as to not form patterns.

The major limitation to this study was that AOIs represent fixations on the original position of an answer choice or step, regardless of whether the text had been moved. Pairing this method of eye-tracking ordering problems while simultaneously recording their screen and choreographing the timing of drag-and-drop choices to fixation sequence would provide a more detailed account of the problem-solving phase.

The scanpath analysis technique employed in this study serves as a method for characterizing (through patterns and entropy values) the *learner-content* interactions that STEM students have with the features of their textbooks. Future studies should consider employing this technique to mathematical problem-solving or assessment items that involve visuospatial reasoning to elucidate patterns of visual behavior.

CHAPTER 5

EYE-TRACKING MEASURES OF STUDENTS' VISUAL ATTENTION TO CONCEPTUAL AND ALGORITHMIC INFORMATION IN WORKED EXAMPLES

As demonstrated by the results of the survey of General Chemistry students' ratings of helpful textbook features, the worked-out example is regarded as a very helpful feature of their textbook. Even with the confirmation that students value this feature, this information does not describe how these students interact with the content and form of the worked-out example. Previous work in the domain of educational psychology has characterized the *learner-content* interaction of this feature from the perspective of student behaviors, but this research perspective has been studied exclusively for problems requiring an algorithmic solution. The information on the interplay between instructional goal and corresponding student problem-solving approach would further detail how students interact with this highly-rated textbook feature.

Literature Review

In chemical education, the long-standing instructional paradigm was based on the notion that if students could solve algorithmic problems, they had also achieved conceptual understanding; this assumption was first addressed in 1987, when Nurrenbern and Pickering asked students multiple-choice items that needed a traditional algorithmic approach and paired these with a multiple choice conceptual items on the same topic.[135]

While the algorithmic questions were "plug-and-chug," the conceptual questions contained no mathematical content. The topics chosen were a mixture of gas laws and stoichiometry. They found that students were more successful at answering algorithmic items, and this result was determined to be statistically significant ($p < 0.05$) via a Chi-squared analysis. They found that 65 percent of students could solve the algorithmic problem, but only 35 percent of students could solve the conceptual problem. The authors attributed this result to instructional strategies because algorithmic problem solving has typically dominated exam items, and that this performance discrepancy exists despite efforts to emphasize conceptual understanding.[135]

In 1990, Sawrey repeated the Nurrenbern study with a larger, more uniform group of students.[136] In order to see if the effect shown in Nurrenbern would also affect the top students in the class, she separated the student success on these paired questions for the top and bottom 27 percent of the class. The results showed a statistically significant difference ($p << 0.001$) between traditional and conceptual items for all students, the top 27 percent, and the bottom 27 percent on the topics of stoichiometry and gas laws.[136] Also in 1990, Pickering administered the paired gas laws items to second-semester general chemistry students, and then followed the performance of these students into organic chemistry I. He analyzed their performance in organic chemistry as a function of their performance on the paired items.[137] Although the sample size decreased (not all of the participants went on to organic), if students were matched for their grade in general chemistry, there was no difference in their ability to perform well in organic regardless of performance on the conceptual gas law item. Pickering argues that the difficulty with the conceptual gas law item is because of a "lack of some specific factual knowledge about gases, not some arcane ability difference."[137] Furthermore, he suggested that there are not two types of students, but rather two instructional goals that emphasize one question type over the other.

Despite the conclusions of these two studies, Nakhleh (1993) performed a study to determine if students were conceptual thinkers or algorithmic problem-solvers using the paired conceptual and algorithmic questions.[138] Figure 38 shows the possible categories of differential performance on conceptual or algorithmic items that students may exhibit. She expanded the topics from the original stoichiometry and gas laws to include chemical equations, limiting reagents, empirical formulas, and density. She tested 1,000 students over four freshmen chemistry courses (remedial, science/engineering, chemistry majors, and honors), hypothesizing that the remedial course would have the highest population of conceptual thinkers, that the honors students would demonstrate both conceptual and algorithmic skills, and that the engineering students would be more algorithmic rather than conceptual. She found that 85 percent of the students could answer an algorithmic gas law item, but only 49 percent could answer the paired conceptual item. The hypothesis that remedial students had poor algorithmic skills was not supported by the data. Significant differences across courses were present for gas laws ($p = 0.016$), equations, and density ($p = 0.006$). Referring to the quadrant in Figure 38, 49 percent of students were high algorithmic/high conceptual, 31 percent were high algorithmic/low conceptual, ten percent were low algorithmic/high conceptual, and ten percent were low algorithmic/low conceptual.[138]

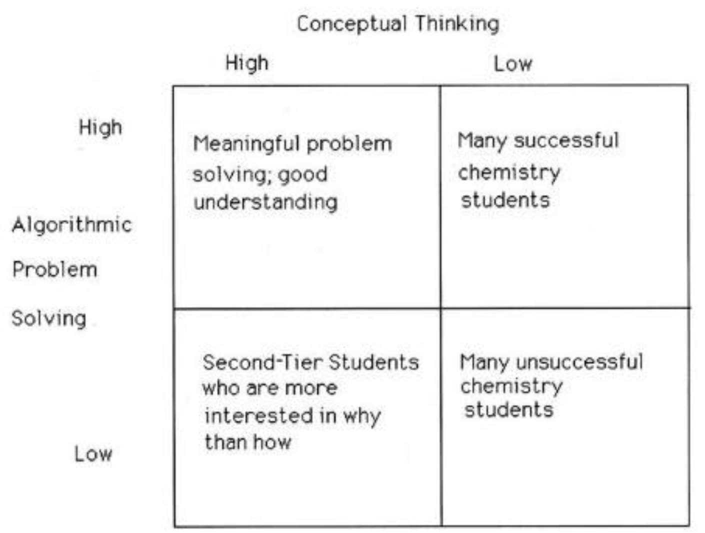


Figure 38: Designation of the four categories of conceptual-algorithmic thinkers typically found in a general chemistry classroom.[138]

In 1993, Nakhleh and Mitchell took the study of conceptual thinkers and algorithmic problem-solvers one step further. Students were categorized using the paired gas law items on an exam, and then two students from three of the four categories were interviewed.[138,139] In each 50-minute interview, students worked through the same two paired items and an extra paired pair of items on stoichiometry while vocalizing their thought processes. Participants were also asked to compare the two types of problems on their perceived level of difficulty and which problem type the student would prefer. Researchers also noted whether students attempted to use an algorithm on the conceptual item. Half of the students solved the conceptual gas law item with an algorithm, and two students used the elimination approach to answer the problem (an algorithmic test-taking strategy). In half of the attempts at the stoichiometry conceptual item, an unnecessary algorithm was used. The results suggest that high algorithmic students use algorithms to solve conceptual items, possibly because they do not trust their conceptual ability. The preferences for problem type often came with stipulations; most students would prefer conceptual homework problems, but not conceptual exam problems. Furthermore, students were more confident in algorithm-derived answers, but success on conceptually-driven answers was seen as more gratifying.[139]

Worked examples are the instructional devices that consist of a problem-stem and a set of problem-solving, algorithmic steps.[140] Worked examples often do not contain all information necessary for learners to solve the problem.[140,141] These tools illustrate an expert's problem solving method, but because experts may skip or combine steps, the step-by-step algorithm may not accurately reflect the conceptual structure of the problem.[140,141] In a review article, Atkinson notes that worked examples often fail to contain justifications or explanations for step-by-step solutions.[140]

Since the 1950s, worked examples have been studied as a means to provide example-based learning, in which a worked example problem is paired with an isomorphic set of practice problems.[140] This instructional method was found to be effective, especially for learners with low domain knowledge. Since the 1970s, educational researchers have been aware of the discrepancy between expert and novice interpretations of the worked example. Experts tended to focus on deeper structural/conceptual aspects, while novices tended to be distracted by surface/contextual features.[140] In the late 1980s and 1990s, attention in this area of educational research focused on the observation (and later the fostering of) the learner's self-explanative behaviors.[140,141] Studies had shown that students who were successful at learning from worked examples could be classified as either anticipative reasoners or principle-based explainers, whereas unsuccessful students were either passively or superficially interacting with the worked example.[140,141] Anticipative reasoners predicted the next step and checked themselves. Principle-based explainers tried to come up with a conceptual reason for each step. The conclusion was that learners who self-explained were more successful. The goal of this body of research was to either explicitly train students to self-explain, or to design the worked example to either prompt an explanation from the learner or fade out the algorithmic steps in order to prompt anticipative reasoning.[140,141] These design principles have not been studied under the chemistry domain, specifically in regards to how students use the conceptual and algorithmic elements of the worked example problem in their textbooks.

Methods

This IRB-approved quasi-experimental eye-tracking study recruited from the General Chemistry II student population, from all quadrants of the model in Figure 39 as determined by the score on a pre-test. The pre-test consisted of paired conceptual and algorithmic multiple-choice items on kinetics, and it was administered as clicker questions in class during the coverage of the relevant chapter. These pre-test paired items are listed in Appendix C.

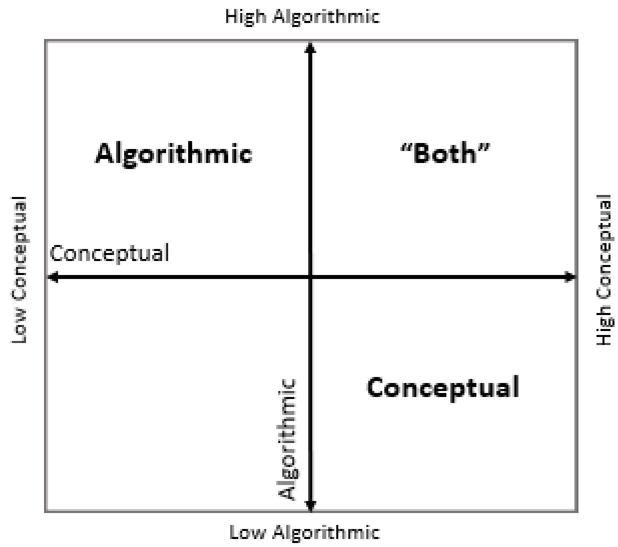


Figure 39: classification system for general chemistry II students' performance on paired algorithmic-conceptual questions.

Once students were classified, the groups identified in Figure 39 emerged. All students in the course were invited to participate, and 26 students agreed to participate. These 26 participants were majority "both conceptual and algorithmic" (N = 13) who answered equal numbers of conceptual and algorithmic items correctly or "more algorithmic" (N = 12) who answered more algorithmic items correctly; there was one student who was initially classified as "more conceptual" for answering both pre-test conceptual items correctly but only one algorithmic item.

The participants viewed three worked-out examples on a Tobii T120 eye tracker with a sampling rate of 60 Hz on an LCD screen (1280 by 1024 pixels) running Tobii 2.0.4 software.[130] By default, fixations are defined as 100 milliseconds on a radius of 30 pixels.[31] Participants were instructed to read through the example as they would an example in their textbook. These examples of acid-base equilibria varied in their presentation of the information (and are displayed in Appendix C). One example—

termed the “algorithmic” example—represents the typical, basic worked-out examples found in textbooks; this example outlined the algorithmic approach to solving for the pH of a weak acid solution using “plug-and-chug” math and stepwise instructions. By contrast, the “conceptual” example demonstrated how to order weak acids by relative acid strength/percent ionization; this example provided no mathematical content, but instead communicated the relative acid strength with particulate level drawings and text-descriptions of the principles of ionization. The “integrated” example mirrored the algorithmic steps and mathematical content of the algorithmic example, but each algorithmic step was preceded by a short conceptual justification of the reasoning for the next algorithmic step.

These examples were provided in varying order to the participants. The average viewing time on each example is illustrated in Table 15 below, in which each row represents a different viewing order. The average viewing time is calculated by summing each participant’s total fixation length (on and off AOIs) and then taking the average of total fixation lengths within a viewing track. One participant (originally classified as “both” and assigned to track 2) was excluded on all statistical analyses of eye-tracking metrics because of poor calibration with the eye tracker.

Table 15: average viewing time for each example in each of the possible viewing orders

		Average Viewing Time (seconds)			
Viewing Order	*Number of Participants*	*Algorithmic (A) Example*	*Conceptual (C) Example*	*Integrated (I) Example*	*Overall*
A, C, I	5	26.684	44.815	61.810	133.309
A, I, C	4	24.575	54.786	63.467	142.828
C, A, I	4	21.362	54.339	58.484	134.185
C, I, A	4	19.737	67.115	74.230	161.081
I, C, A	4	36.425	60.991	57.812	155.228
I, A, C	4	24.086	44.659	52.918	121.664

After viewing the worked-out examples on the eye tracker, participants completed a concurrent think-aloud post-test consisting of paired algorithmic-conceptual items on the topic of general equilibria. These post-test items are included in the Appendix C. The interviews were transcribed and analyzed in Atlas.ti using emergent themes.

Given the open-ended structure of the post-test items, it is possible to assign partial credit; while the binary correct/incorrect scoring was useful for statistical analyses, partial credit allows for calculation of item difficulty and discrimination. For each item, a point was awarded for each component of the correct answer, and then the overall possible score was scaled to a value between 0 and 1. The first item on the post-test was an algorithmic item that asked participants to (i) determine whether the system was at equilibrium and (ii) to determine the shift in the net reaction if the system was not at equilibrium. The second item was a paired conceptual item which presented three reaction mixtures and asked participants to evaluate (i) whether each reaction mixture was at equilibrium and (ii) the direction of the shift of net reaction for mixture C if not at equilibrium. Each component was assigned a point (for a total of four) and scaled to a value between 0 and 1. The third item was another algorithmic item which asked participants to calculate the equilibrium pressures; the fourth item provided a gaseous reaction mixture at equilibrium and asked participants to draw the effect of increasing the volume of the system, and for full points participants needed to (i) predict the correct shift in the net reaction and (ii) draw an equilibrium state that obeyed the Law of Conservation of Mass.

Research questions initially relied heavily on quadrant classification as seen in Figure 39. Based on the eye-tracking data, does participant visual behavior have a relationship with the classification? After the post-test, does the classification in the quadrant system change?

<u>Analysis of Item Scores</u>

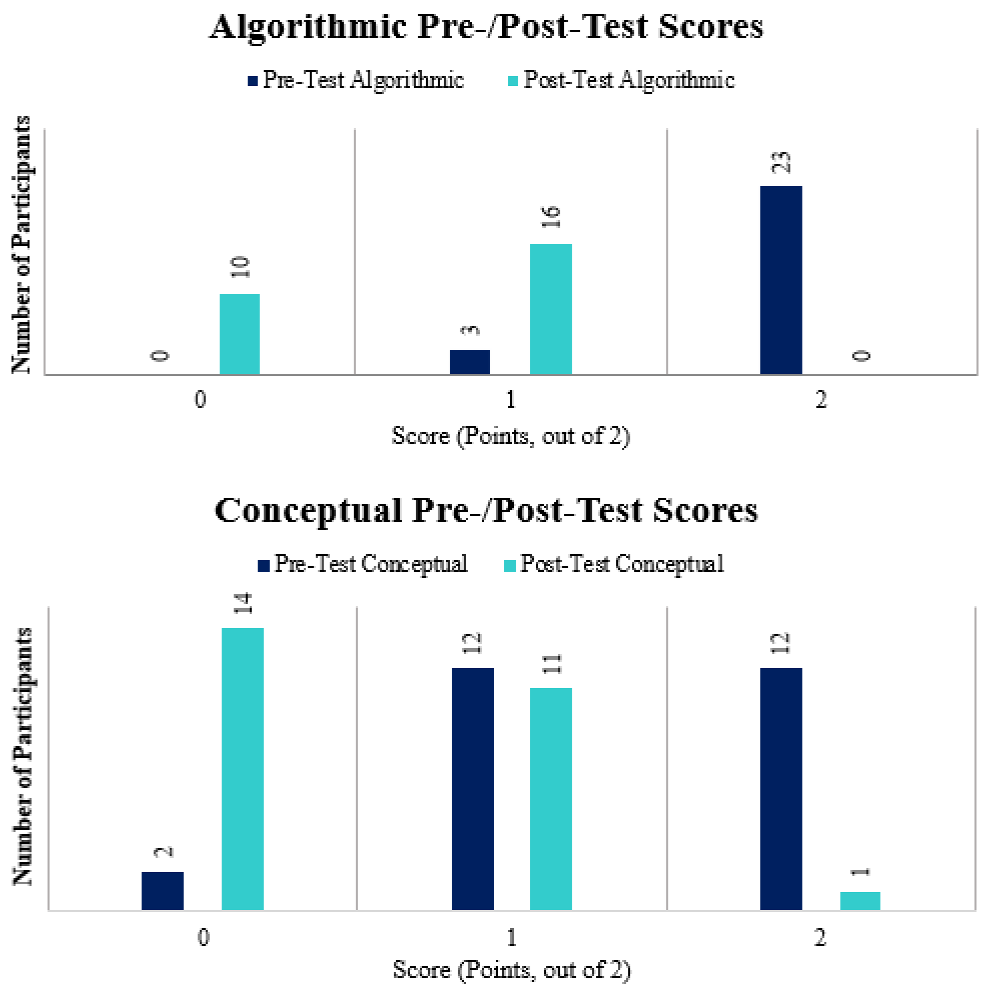


Figure 40: (top) participants' pre- (navy) and post-test (teal) scores on algorithmic items; (bottom) participants' pre- (navy) and post-test (teal) scores on conceptual items.

Initially, pre- and post-test items were scored on a binary correct/incorrect scale; each set of pre- and post-test scores were out of a possible four points (two conceptual and two algorithmic). In Figure 40, the top portion shows the distribution in scores (out of a total of 2) for the pre- and post-test algorithmic items, whereas the lower half shows the distribution of pre- and post-test conceptual item scores. The pre-test scores in navy were used to assign classifications to participants. Thirteen participants scored equally well on conceptual and algorithmic items, and this group was called “both algorithmic and conceptual.” Twelve participants had higher scores on algorithmic items are were classified “algorithmic,” and only one participant scored higher on conceptual items than algorithmic items and was classified “conceptual.”

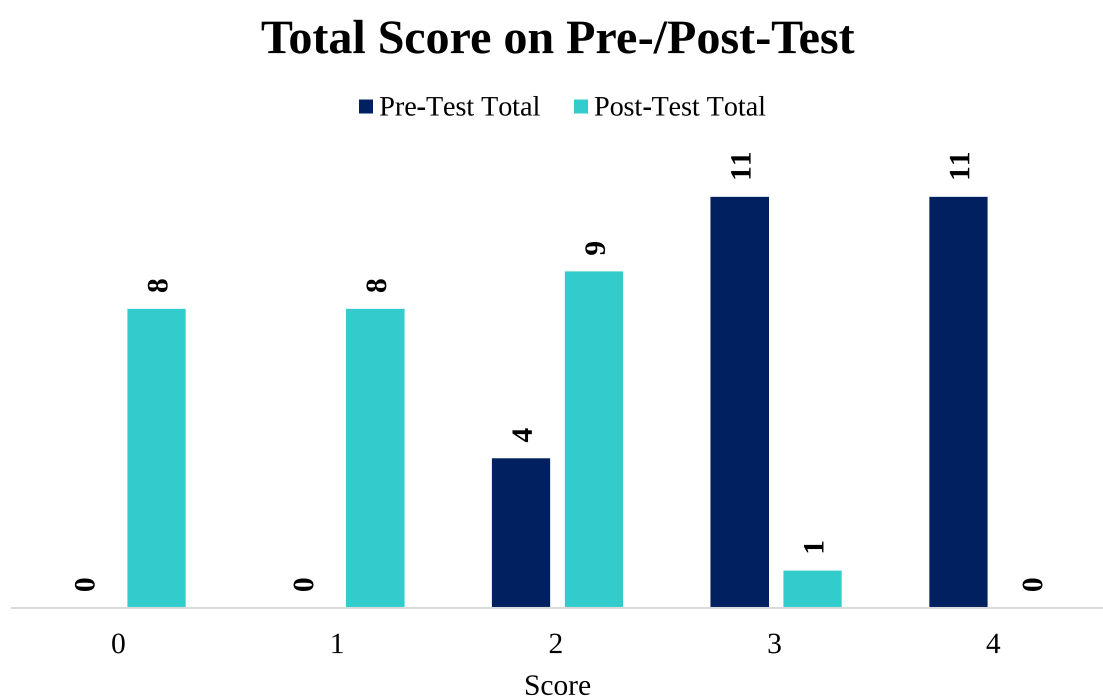


Figure 41: distribution of participants' pre- (navy) and post-test (teal) scores on all the paired items (for a total score out of 4).

In Figure 41, the overall distribution of pre-test scores is shown in navy (M = 3.27, SE = 0.142) and the distribution of post-test scores is shown in teal (M = 1.12, SE = 0.178). This difference in mean scores (-2.15) was significant, $t(26) = 8.981$, $p < 0.0001$, with an extremely large effect size (Cohen’s d = 2.616). After the post-test, 15 participants were reclassified; the one participant initially classified as conceptual was re-classified to "neither", as were four initially-algorithmic participants and three originally "both" participants. Three participants were able to be re-classified from algorithmic to "both." Three participants initially classified as "both" were moved to "conceptual" due to post-test scores, and one participant was re-classified from "both" to just algorithmic.

Using a more nuanced grading scale (as described in the methods section), partial credit was assigned and normalized to a total of one point for each item. From these scores, assessment metrics can be calculated to explain the significant and extremely large effect size reported in Figure 41. Based on Classical Test Theory, item difficulty and discrimination metrics can be calculated as demonstrated in Equations 6 and 7, respectively.[142,143]

$$p = \frac{respondents\ who\ answered\ correctly}{total\ number\ of\ respondents} \tag{6}$$

$$d = top\ performing\ quartile - low\ performing\ quartile \tag{7}$$

Item difficulty values range from zero to one, indicating the proportion of students who answered an item correctly. Within the context of the assessment, benchmarks for interpreting these values are proposed as challenging if $p < 0.25$ and too difficult if the $p > 0.75$.[143] Item discrimination measures how well an assessment item discriminates between high- and low-performing respondents. Discrimination values can range from negative one to one, negative values indicating a flaw in the exam item. Discrimination values below 0.20 suggest that the assessment item needs to be revised due to poor discrimination and the target value is 0.40 or higher for excellent discrimination.[143] Table 16 lists each item, its subject, the item difficulty score (p), and the item discrimination score (d).

Table 16: Item difficulty and discrimination scores for pre- and post-test items

Item	*Subject*	*Difficulty (p)*	*Discrimination (d)*
Pre-Test #1	Conceptual: describing reaction order with a linear best-fit	.70	.727
Pre-Test #2	Algorithmic: solving for partial order of a reactant given multiple experimental trials	.95	.170
Pre-Test #3	Conceptual: estimating the number of half-lives that pass to decompose a certain percent of the sample	.47	.861
Pre-Test #4	Algorithmic: calculating the remaining number of atoms after a first-order decay	.86	.474
Post-Test #1	Algorithmic: calculating reaction quotient to determine whether a system is at equilibrium and the direction of the net reaction	.83	0.286
Post-Test #2	Conceptual: evaluating whether three visualizations depict equilibrium and prediction of the direction of the net reaction	.67	0.714
Post-Test #3	Algorithmic: calculating the equilibrium pressures given K_C	0	0
Post-Test #4	Conceptual: predicting and drawing the result of an increase in volume of a gaseous equilibrium	.40	0.429

The pre-test algorithmic items were not difficult enough ($p > 0.75$) and the second item did not discriminate well ($d = 0.170$). The ease of the algorithmic items suggests that the classification scheme did not yield valid results. Furthermore, the first post-test item was rated an easy algorithmic item ($p = 0.83$) that did not discriminate well ($d = 0.286$) between the top and lower quartiles of participants. On the other hand, the third post-test item was a particularly difficult ($p = 0$) algorithmic item on a general, gaseous equilibrium which did not discriminate well ($d = 0$).

Analysis of Eye-Tracking Metrics

In previous work,[31,58] eye-tracking metrics are typically highly correlated on the same stimulus. For this set of stimuli, the correlation between metrics as very high (r >

0.9), and thus dimensionality of the dataset was reduced by selecting fixation length as the dependent variable for future analyses. The size of each AOI (contained in Appendix C) was used to create weighted sum of fixation lengths for each participant on text-based AOIs and math/image-based AOIs. The fixation length metric on each AOI was scaled to the percent of stimulus and summed for each type of AOI as shown in Equation 8.

$$Scaled\ FL = \sum\left(FL\ on\ AOI \times \frac{AOI\ \%\ of\ Screen}{100\%}\right) \tag{8}$$

In Figure 42, the average weighted-sum fixation lengths of algorithmic learners (N = 12) and "both" learners (N = 12) on text-based AOIs for each example type are plotted. Algorithmic learners spent an average of 0.1602 ± 0.0562 seconds on text-based AOIs on the algorithmic example, 2.0507 ± 0.6478 seconds on the text in the conceptual example, and 1.8778 ± 0.7197 seconds on the text-based AOIs in the integrated example. These average weighted-sum fixation lengths of algorithmic learners are consistently longer than the averages for learners classified as "both" high algorithmic and high conceptual. The "both" group fixated on text-based AOIs for an average of 0.1344 ± 0.0932 seconds on the algorithmic example, 1.8140 ± 0.7388 seconds on the conceptual example, and 1.7423 ± 0.6929 seconds on the text in the integrated example. The annotations on Figure 42 describe the result of one-way independent samples ANOVA of learner-classification on average weighted-sum fixation length on text-based AOIs. There was no significant effect of classification on fixation length of text-based AOIs for the algorithmic example, $F(1,22) = 0.673$, $p = 0.421$, the conceptual example, $F(1,22) = 0.696$, $p = 0.413$, or the integrated example, $F(1,22) = 0.221$, $p = 0.643$.

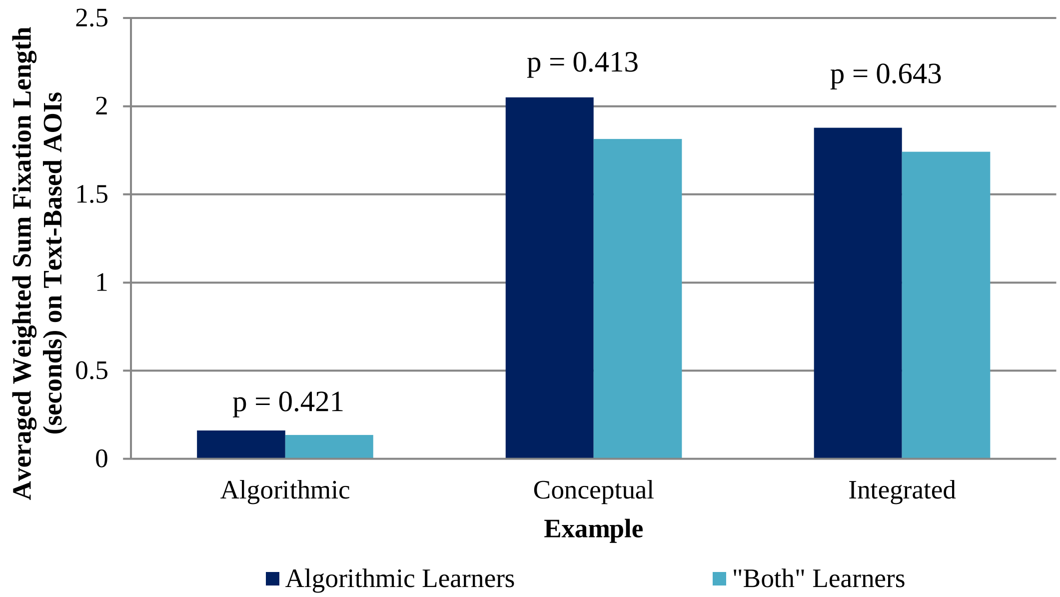


Figure 42: average weighted-sum fixation lengths (in seconds) spent on text-based AOIs on each example type; navy indicates learners classified as "algorithmic" and teal indicates "both" classified learners. Annotations detail the results of a one-way ANOVA of average weighted-sum fixation length on text-based AOIs between the two classifications of learners for each example type.

Figure 43 demonstrates a similar analysis of the participants' average weighted-sum fixation lengths on math/image-based AOIs. Here, learners classified as "both" consistently fixated longer on math/image-based AOIs for the algorithmic example (0.4272 ± 0.2494 seconds), conceptual example (3.3370 ± 1.7271 seconds), and integrated example (0.5127 ± 0.1683 seconds). The algorithmic-classified learners fixated for less time on average on these types of AOIs; on the algorithmic example they spent 0.3715 ± 0.1993 seconds, as compared to 2.2542 ± 1.4923 seconds on the images of the conceptual example or 0.5035±0.2082 seconds on the math of the integrated example. The results of the one-way independent samples ANOVA on these weighted-sum fixation lengths is designated in Figure 43. There was no significant effect of classification type on how long participants fixated on math/image-based AOIs for the algorithmic, $F(1,22) = 0.365$, $p = 0.552$, conceptual, $F(1,22) = 2.700$, $p = 0.115$, or the integrated example $F(1, 22) = 0.014$, $p = 0.906$.

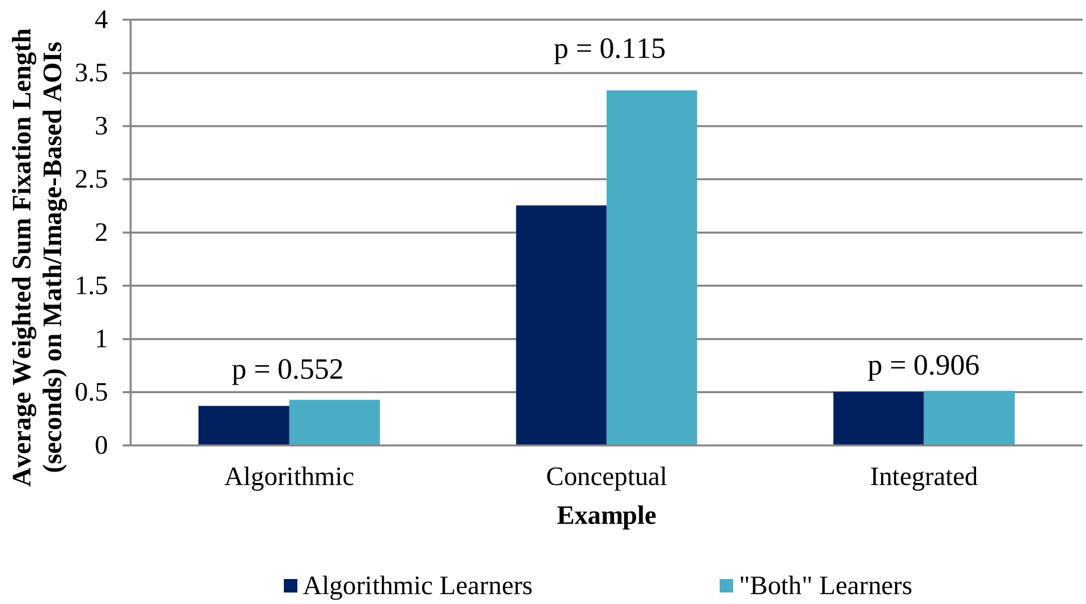


Figure 43: average weighted-sum fixation lengths (in seconds) spent on math/image-based AOIs on each example type; navy indicates learners classified as "algorithmic" and teal indicates "both" classified learners. Annotations detail the results of a one-way ANOVA of average weighted-sum fixation length on math/image-based AOIs between the two classifications of learners for each example type.

While the one-way ANOVAs on weighted-sum fixation lengths on each example type for each classification were not significant, a final attempt to justify the classifications was made using scanpath patterns. By reducing each participants' scanpath to a character string of AOIs, the permutation test in the R package **GrpString** was used to calculate the dissimilarity between the algorithmic and "both" groups of strings; again, the conceptual group was excluded due to small sample size (N = 1). For the conceptual example, the difference in scanpaths between algorithmic and "both" participants was not significant, p = 0.80200. For the algorithmic example, there was a significant difference in scanpath strings, p = 0.04700. Finally, for the first page of the integrated example there was a significant difference, $p = 0.01200$, but the difference on the second page was not significant, $p = 0.07200$, likely due to the familiarity of that problem type.

After disregarding the classification system, a mixed factorial design for all 25 participants was devised. Based on Q-Q plots as well as Kolmogorov-Smirnov and Shapiro-Wilk tests of normality, the weighted-sum fixation length of text- and math-based AOIs on the algorithmic example differed significantly from normal; after trimming the outliers that fell 1.5 times outside the interquartile range (IQR), the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality did not indicate a significant deviation from a normal distribution. This trim resulted in losing two cases from the algorithmic text-based AOI fixation lengths, one case from the algorithmic math-based AOI fixation lengths, and one case from the conceptual math-based AOI fixation lengths.

After removal of the outliers, the mean weighted fixation length sum for all AOIs on the algorithmic example ($M = 0.241$, $SD = 0.024$) was shorter than for the conceptual example ($M = 2.195$, $SD = 0.197$) or the integrated example ($M = 1.097$, $SD = 0.073$). While two of the example types featured math, the conceptual example type offered an image to explain the concept. On math/image based AOIs, there were math/image-based AOIs had a shorter weighted-sum average fixation length for algorithmic examples ($M = 0.365$, $SE = 0.045$) than conceptual ($M = 2.462$, $SE = 0.372$) or integrated examples ($M = 0.514$, $SE = 0.036$). The fixations on text-based AOIs were typically shorter; the weighted-sum fixation length on algorithmic ($M = 0.117$, $SE = 0.010$) was much shorter than conceptual ($M = 1.928$, $SE = 0.107$) and integrated examples ($M = 1.680$, $SE = 0.136$).

Using a mixed factorial ANOVA allowed for the study of not only the significance of each main factor (example type and AOI type), but also the interaction between these two factors. First, Mauchly's test indicated that the assumption of sphericity was violated for the main effect of example type, $\chi^2(2) = 13.060$, $p < 0.001$, as well as the interaction between example and AOI type, $\chi^2(2) = 9.218$, $p < 0.010$. To account for this violation, SPSS used the corrected degrees of freedom of the Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.566$ for the main effect of example type and $\varepsilon = 0.609$ for the interaction between the main effects).

From the within-subjects effects, there was a significant main effect for the example type on participants' weighted-sum fixation length, $F(1.133, 11.327) = 63.124$, $p < 0.0001$. This indicates that there is a significance difference in weighted-sum fixation lengths on each example type, even without considering other variables. For the main effect of example type, the average of the weighted-sum fixation lengths on the algorithmic example was 0.241 seconds (SE = 0.024), 2.195 seconds (SE = 0.197) for the conceptual example, and 1.097 seconds (SE = 0.073) for the integrated example. In the Bonferroni-adjusted pairwise comparison the algorithmic and conceptual example types, the mean difference (-1.954 seconds) in participant weighted fixation lengths was significantly ($p < 0.0001$) shorter on the algorithmic type of example. A similar comparison between conceptual and integrated example types yielded a mean difference (1.098 seconds) that was significantly ($p < 0.0001$) longer fixation lengths on conceptual example. Finally, the third pairwise comparison between integrated and algorithmic examples resulted in a mean difference (0.856 seconds) that was significantly ($p < 0.0001$) longer fixations on the integrated example type. Based on the within-subjects contrasts, the estimated mean weighted-sum fixation length on the algorithmic example compared to the integrated example was significant with a large effect size, $F(1, 10) = 178.843$, $p < 0.0001$, $r = 0.973$; similarly, the contrast between the conceptual example and the integrated example was significant with a large effect size, $F(1, 10) = 29.701$, $p < 0.0001$, $r = 0.865$.

For the main effect of AOI type, the assumption of sphericity was not violated; the average of the weighted-sum fixation lengths for text-based AOIs was 1.241 seconds (SE = 0.072) and 1.114 seconds (SE = 0.112) for math/image-based AOIs. There was not a significant main effect for AOI type, $F(1, 24) = 0.148$, $p = 0.549$, and the Bonferroni-adjusted pairwise comparison between these AOI types yielded a mean difference of 0.128 seconds (SE = 0.122) in average weighted-sum fixation length and was also not significant.

There was a significant interaction effect between the example type and the AOI type on the weighted-sum fixation lengths, $F(1.219, 12.188) = 13.786$, $p < 0.002$. This indicates that AOI type was fixated upon for different lengths of time depending on which worked example was presented. Figure 44 details the results of the interaction term for the mixed-design factorial repeated-measures ANOVA between example and AOI type. The navy bar represents weighted-sum fixation lengths on text-based AOIs and the teal bar represents math or image-based AOIs, regardless of group classification or viewing order of the participants.

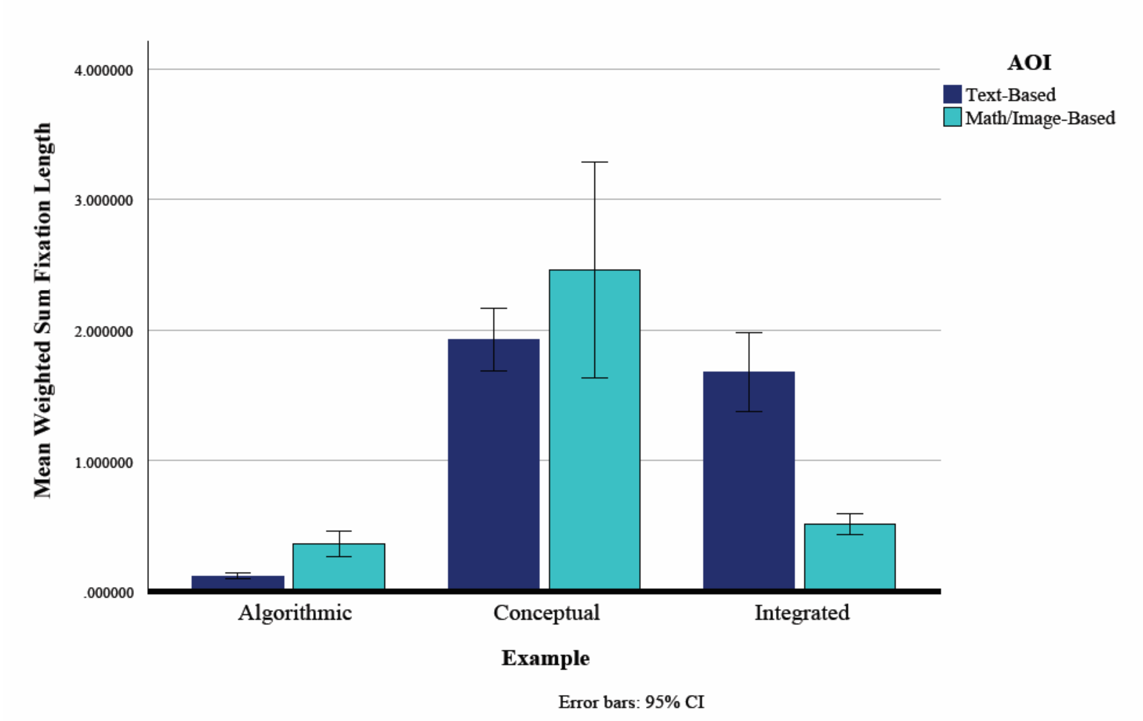


Figure 44: an interaction plot depicting the mean weighted-sum fixation length on text-based (navy) and math/image-based (teal) AOIs for each example type for all participants; error bars indicate 95% confidence interval.

From Figure 44, we can interpret the results of this mixed factorial ANOVA to indicate that participants generally spent less time fixating on text-based AOIs in the traditional algorithmic or the conceptual style worked-example. For the algorithmic example, participants spent an average of 0.117 seconds (SE = 0.010) fixating on text-based AOIs and 0.365 seconds (SE = 0.045) fixating on math-based AOIs. These fixations are much shorter than those on the conceptual example, in which the participants fixated on text-based AOIs for an average of 1.928 seconds (SE = 0.107) and 2.462 seconds (SE = 0.372) on math/image-based AOIs. The average fixation length on the AOIs of the integrated example were 1.680 seconds (SE = 0.136) for text-based AOIs and 0.514 seconds (SE = 0.036) for math-based AOIs. The plot in Figure 44 illustrates the effect of example-organization on participants' visual attention: when conceptual information is included, participants fixate longer on the text-based AOIs, and participants fixate longest on particulate-level conceptual diagrams.

This significant interaction term for example and AOI type can be broken down by within-subjects contrasts. There was a significant contrast between the algorithmic and integrated examples between the two types of AOIs with a large effect size, $F(1, 10) = 77.006$, $p < 0.0001$, $r = 0.941$. The contrast between the conceptual and integrated examples on the two AOI types was also significant with a large effect size, $F(1, 10) = 17.718$, $p = 0.002$, $r = 0.800$.

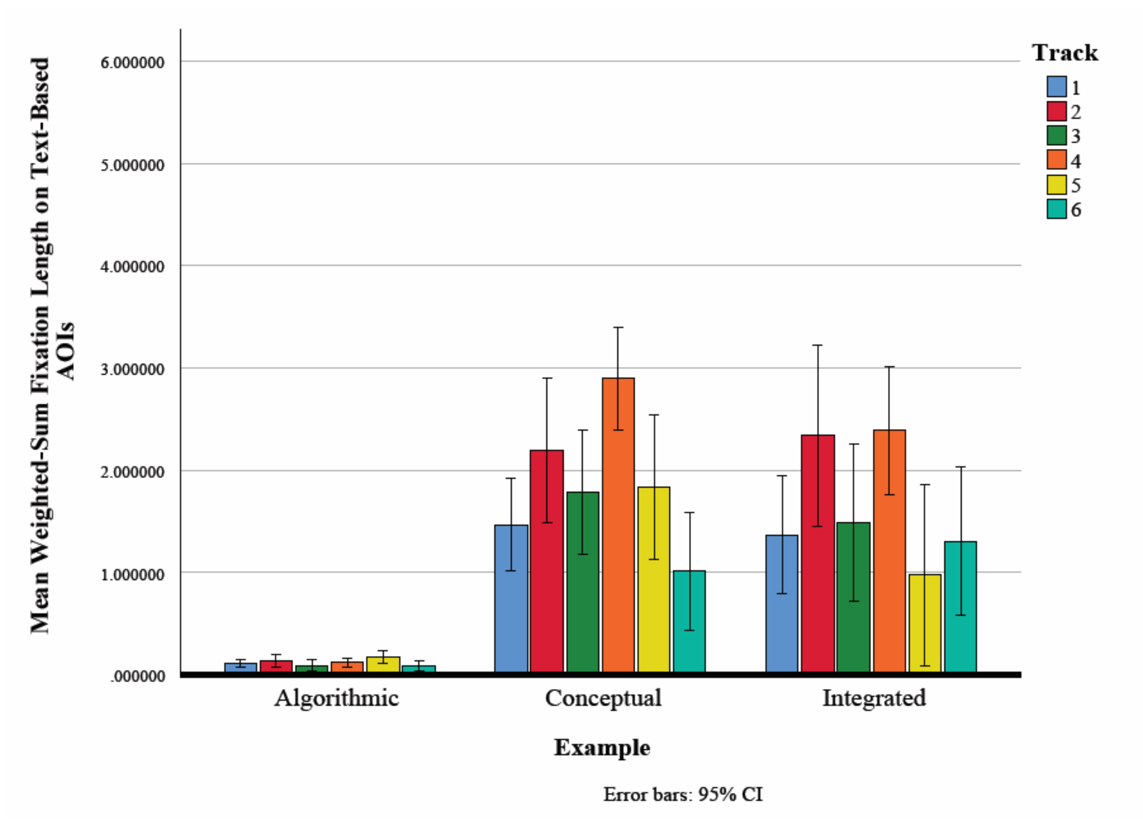


Figure 45: plot of the estimated marginal means of weighted-sum fixation length on text-based AOIs for each example type for each viewing track; error bars indicate 95% confidence intervals.

Figure 45 breaks down the weighted-sum fixation length on text-based AOIs for each example type by the viewing order ("track") as defined by Table 15. Despite the difference in viewing order, the main interaction and within-subjects contrasts of example by AOI by viewing track were not significant at the 0.05 alpha level. Similarly, in Figure 46, the weighted-sum fixation length on math/image-based AOIs for each example type by viewing order is shown, despite the lack of significant main effect or within-subjects contrasts.

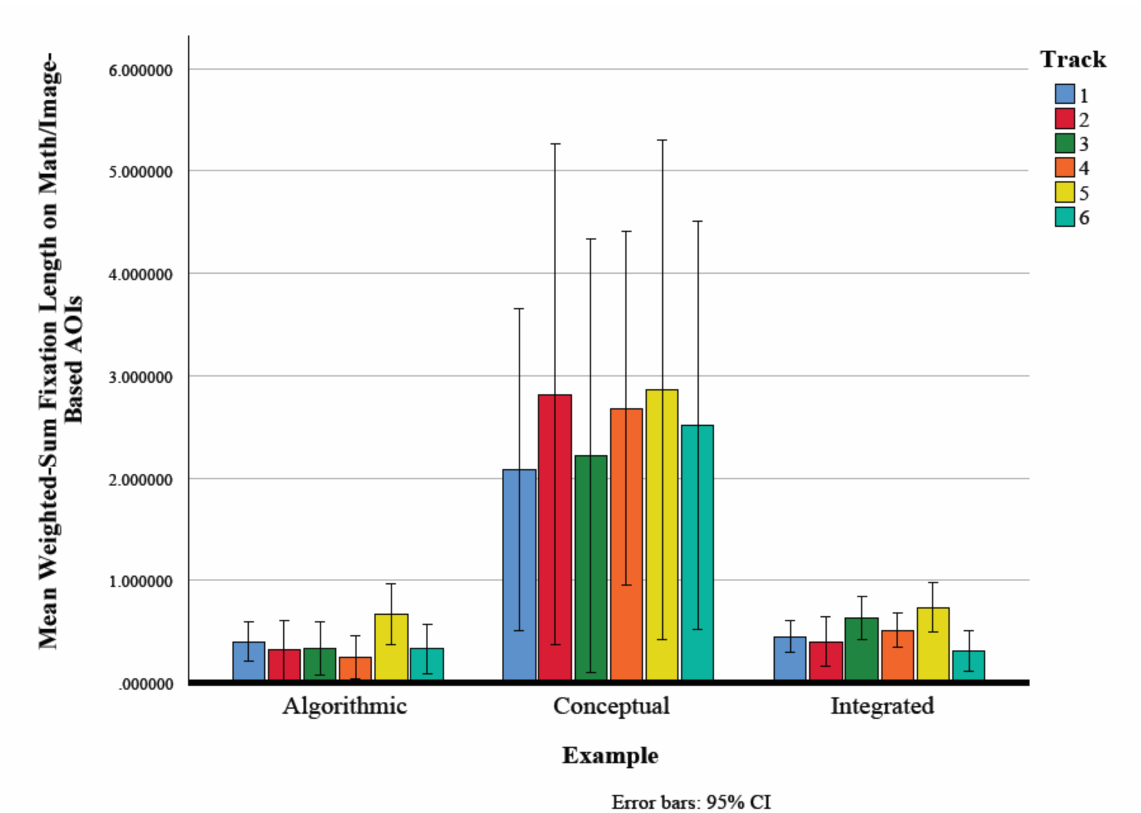


Figure 46: plot of the estimated marginal means of weighted-sum fixation length on text-based AOIs for each example type for each viewing track; error bars indicate 95% confidence intervals.

Not only were the interactions between example, AOI, and track that are featured in Figures 45 and 46 not significant, but the Bonferroni-adjusted pairwise comparisons were also not significant. The average difference in weighted-sum fixation length on each

AOI type on each example type was not significantly different because of viewing order. Furthermore, any interaction term involving the group classification (algorithmic-learners vs “both” learners) previously studied in Figures 41 and 42 was not significant. The significant main effect of example type and the significant interaction effect of example by AOI type are not significant due to viewing order or the group classification assigned to participants.

Analysis of Post-Test Interview Data

A concurrent think-aloud post-test of paired algorithmic/conceptual items was performed to investigate a relationship between participants’ viewing behaviors on the worked examples and the participants’ problem-solving behaviors. As previously demonstrated in the eye-tracking analysis, there was not a significant difference in fixation length on different elements of the worked example as a function of a participants’ pre-test classification (“algorithmic” or “both”). Despite the difference in pre- and post-test item difficulties and discriminations (Table 16), the initial hypothesis was that the classification system in Figure 38 could be used to differentiate between participants in terms of their visual attention as well as their problem-solving strategies. As such, the first analysis of the interview data is summarized by Table 17, in which the total number of quotes of major classification of participants was characterized as either algorithmic or conceptual problem-solving approaches. Regardless of classification, many participants employed an algorithmic approach to problem-solving.

Table 17: characterization of each group of participants' quotes as either an algorithmic or conceptual problem-solving method

Classification	*Problem-Solving Method*		*Total Number of Quotes*
	Algorithmic	Conceptual	
Algorithmic	68%	32%	231
"Both"	78%	22%	272

As demonstrated in Figures 40 and 41, there was a significant difference in pre- and post-test scores when the items were scored on the binary correct/incorrect fashion. The tables of codes below reflect the same binary grading system; each table lists the codes common to the participants who answered that question either completely correctly or at least partially incorrectly. Next to each code is an illustrative quote of that code.

For the first post-test item, most participants solved the entire problem (both parts) correctly. Predictably, all participants had some algorithmic problem-solving; it would be difficult to discuss the solution to an algorithmic item without utilizing algorithmic problem-solving methods.

Table 18: Codes for Post-Test Item #1

Participants who answered correctly (N =13)		*Participants who answered incorrectly (N =4)*	
Code	Quote	Code	Quote
Algorithmic	A8600165: "And I got 3.834 × 10^3. And this is essentially your Q value, which is the value that you used to compare to the actual K_C value that it gives you. And because the Q value is greater than the K_C value, that it's not in equilibrium, and that the direction has to go towards the reactants, so towards the K_C."	Algorithmic	B1553020: "So I have about 1533, which if I wanted to put into similar format as given value, it would be one to the 1.5 × 10^3. Okay. And since it's lower than the K_C value, the equilibrium is going to shift right, to the product, because this means there's less product than there should be when equilibrium has been reached.
Conceptual	A8481633: "Equilibrium is way farther to the left, so there's way more products than reactants at this point in time than there would be at equilibrium."	Confusion	B1553020: "Alright 0.041A so that's a 1 to 2 ratio... 0.152 B 0.70 C in a five liter at six hundred K... Wow, why would they give me that?"
Erroneous Use of Kinetics Terminology	A4052558: "Uh, the coefficient two tells me that it is, um, a second order reaction"	Confusing Q and K_C	B8046467: "This is K_C at equilibrium and this is K_C for now."
Rote Learning	A9755546: "And since 3.7 × 10^3 is 3,700, it's much greater than the value that I have here. And is that the Q value? And what was the rule for that? I think if it's... I think it's moving to the left, because it's less on the reactant side. No shortcuts, right? I feel like it's wrong. I just forgot the rule. But if I knew it, I think I'd be able to answer that. Yeah, that's what I'm thinking for that one."	Gaseous not Aqueous	B1553020: "Okay. Let's see, Here we have... Wow these are gaseous. I thought I would be dealing with aqueous. Well I guess not. Alright so 0.041 moles... I can write on here, right?"

Superfluous Information	B1924479: "Okay. So first, what I'm going to do is things that don't matter is the temperature because... you already have the Kc for the reaction, so the temperature, you're not going to use that number into the equation. So I see that it's 5 liters, so I'm going to convert the moles to molarity, so I'm going to do... Well, actually, at first I'm going to write the equation down"	Superfluous Information	A9968443: "I guess I wouldn't need the Kc or the temperature itself because the question just asks if there's an [net] flow in terms of... It didn't really change any of that. It already should be at equilibrium if anything"
Realized Mistake	B3611841: "I might not need an ICE table for this"	Rote Learning	A9968443: "Well, the… I just know it. [chuckle]"

Interestingly, the participants who answered the item correctly invoked conceptual descriptions of equilibrium, even if the attempt ended in erroneous use of terminology. Correct participants also realized mistakes, such as employing an unnecessary ICE table. Incorrect participants also utilized incorrect terminology, but they confused the position of equilibrium (K_C) and the reaction quotient (Q). One incorrect participant was also confused by the gaseous nature of the generic elements in the equilibrium, which is a theme seen in the third post-test item as well. Both sets of participants noted the superfluous information of reaction temperature, as it was not needed to solve for the reaction quotient, and in certain cases incorrect participants were confused by its inclusion. Furthermore, both groups describe rote learning as the source of the algorithmic knowledge on display. In general, most participants were able to identify that the system was not at equilibrium, but those who answered incorrectly predicted the wrong shift in the net reaction. As shown by the participant identifiers in Table 18, the classification of students was distributed between correct and incorrect answers and all participants utilized some algorithmic problem-solving.

The second post-test item had four parts, and most participants did not predict the correct shift of the net reaction. This problem had visualizations depicting the particulate-level of three different reaction mixtures, and it was possible to calculate a reaction quotient by using the equilibrium expression and the number of molecules shown as a concentration. Surprisingly, participants who solved the problem conceptually compared the ratio of molecules in the visualization to the stoichiometric ratio in the equilibrium. Since this approach did not expressly calculate a reaction quotient value, it was deemed a more conceptual problem-solving approach.

Table 19: Codes for Post-Test Item #2

Participants who answered correctly (N = 7)		*Participants who answered incorrectly (N=11)*	
Code	Quote	Code	Quote
Algorithmic	B8057201: "[C]ounting each [molecule] as a concentration… And so it's telling me the equilibrium constant is one, so that must mean if the K_c equals one then the concentration of... Well the quotient of the concentration of products, which in this case is X over... X_2 over Y, over X_2 multiplied by Y_2, has to equal one"	Algorithmic	A4052448: "I'll just go ahead and write the expression, which is going to be, uh, K_c equals, um, the concentration of X_2Y squared because it's a second degree […] And then divide it by X_2 squared because of the coefficient and, uh, times the concentration of Y_2."
Conceptual	A5668900: "I'm gonna use a ratio to see if, like, that works? And that'll give me the answer I'm looking for? So it looks like it's a 2 to 1 to 2 ratio at equilibrium… and in [reaction mixture] C, there's 1 to 3 to 2. That['s] not at equilibrium [because] the numbers aren't matching, like this is half of what it should be and this is three times what it should be so that just means the whole thing is off."	Conceptual	A8600165: "Well I know that for it to be in equilibrium, it has to have the same ratios of the reaction that's given. And I know that if it is in an equilibrium or if it is in equilibrium, but more products is added then equilibrium would shift or the reaction would go towards the products because you have more reactants. But if it was the other way around and you had more products then the reaction would shift towards the reactants. And there doesn't seem to be any with the right ratio because box A has $2X_2$ but it has $2Y_2$ and $2X_2Y$,

			whereas B has $2X_2$, $3X_2Y$, and $1Y_2$, and C has $1X_2$, $2X_2Y$, and $3Y_2$, so none of them seem to be in equilibrium. So for reaction mixture C, because there's more Y's, there is three Y_2's and one X_2 and two X_2Y because there is an increase in Y_2, obviously that's a net reaction will perceived in a forward direction…”
Dissociation	B5599778: “So, I know based on reaction we're looking... I'm assuming if it's at equilibrium then that means there is this same amount of disassociation on both sides […]Because, I don't know... Or is it supposed to be the less? I'm pretty sure it's the most moles, because that was that one concept that I was just like, "I don't know, that doesn't make any sense." But you're just supposed to go towards the side with the most moles. Because it's based on like the disassociation of it.”	Conflating Stability with Equilibrium	A9755546: “I would guess... I wanna say yes, because if you look at the elements like X_2 and Y_2, that's their natural state where they're stable, so they could be at equilibrium.”
Erroneous Use of Acid/Base Terminology	B8244738: “Okay so, if the question is, for the reaction, made sure was [reaction mixture] C, well then that reaction occurred... Proceed in the forward or reverse direction. And since your Y_2 would be Y_2 to your conjugate base, judging by the pictures would be, and then the X_2 would be your... So then your X_2 would have to, that's your conjugate base, and that would mean the acid would have to be X_2Y, so then how many H_3O^+ would be the X_2…”	Erroneous Use of Acid/Base Terminology	B9322110: “Because it has more base…”
Unfamiliar with Problem Type	B5599778: “We’re gonna look at the example [from the eye tracker] with all the confusing bubble things because I’ve never really seen any of these before. [On what makes it confusing] I guess because I’ve never seen a problem like this in a textbook like I did all the	Erroneous Use of Kinetics Terminology	A4052448: “…concentration of X_2Y-squared because it’s a second degree…”

	reading and stuff… or in lecture."		
Wants Math	A5668900: "I feel like this is the wrong, method-" [laughs] "-to use to find this […] 'cause [sic] this isn't really helping me. Like it tells me that it's not at equilibrium but I don't… [On what would help] Probably some type of math."	Law of Conservation of Mass	B5683065: "They wouldn't disappear at equilibrium would they? No I don't think that they'd disappear, I think they'd all be here at the same time… Like when the reaction occurs. I don't know how the box works, but I don't know if they would... If once the reaction comes to equilibrium, the box changes. You can't... There's the law of concentration of matter. But, that's not what it's called. But... The conservation of matter. So it wouldn't be gone, but I don't know why none of these add up to the reaction."
Rote Learning	B5599778: "I'm kind of confused about […] the rules I'm supposed to have already memorized, because I guess that was chapter 15, so that was like two chapters ago."	Confusion Interpreting the Visualization	A9755546: "Yeah, basically the part about an element being in its natural state, reading the reaction. I'm looking at the diagrams, and I'm trying to figure out if they're different, and what that would tell me. I know the equation is equilibrium, but I don't know how to tell if each one is in equilibrium. I wanna say yes, because the equation is, but I can't explain to you which one or if all of them are at equilibrium, if that makes any sense."
References	B8244738: "I used the [conceptual example from the eye tracker] for […] the conceptual more so on how acids and… the equation works, consuming and products products, and the relationship between those two, from prior knowledge, to solve that problem."	Confusion with $K_C = 1$	A9849278: "I don't really know what the question's asking. I think it's asking me which way this goes. But I don't know where the one constant is coming from."
		Confusion with Generic Elements	B1553020: "Yeah. I guess it's because instead of actual elements, like what's on the periodic table, it's just given as

			X and Y."
		References	B9322110: "I was looking at [the conceptual example from the eye tracker] because they have a similar picture."
		Rote Learning	A9755546: "I think I started using logic to understand that after a while, when I felt like I was forgetting some of the chemistry rules, I did try to make sense of it all."

For instance, the participant A8600165 is quoted in Table 19 solving the problem correctly using a conceptual approach; however, the participant did not answer the second portion of the item correctly because this participant incorrectly suggested Reaction Mixture C would shift to the right "[b]ecause the number of Y_2 molecules increased." On the other hand, participant A5668900 solved the problem correctly using a conceptual approach by comparing the ratio of molecules to the stoichiometric ratio. Despite this success, participant A5668900 also comments that they are unsure of this approach and would prefer math instead. Similarly, although participant B3611841 solved the problem correctly, they remarked, "I'm thinking, how do you find equilibrium if you don't have numbers? Can you just look at it?" As demonstrated in Table 19, participants did find an algorithmic approach to solving a problem that did not provide concentrations; as participant B8057201 phrased it, "If I can utilize putting random number, random concentrations to see if I can get to [a K_C of] one[,]" although this participant (as shown in Table 19) did count the number of molecules as a form on concentration to calculate reaction quotient values.

Another interesting trend in both groups was the use of inappropriate terminology to describe this generic equilibrium. Both groups erroneously ascribed acid/base labels to species in the equilibrium, due to the relatively recent coverage of the acid/base equilibria chapter as well as the participants who referenced the conceptual worked-out example from the eye tracker. Participant B8244738 even drew a parallel between the hydronium ion and the generic X_2Y product; despite this wrong description, this participant did solve the problem correctly. One incorrect participant also used kinetics terminology incorrectly to describe the powers in the equilibrium expression as “orders” but this quirk was not wide-spread.

Within the groups of correct and incorrect participants, some chose to refer to the conceptual worked-out example that they had seen on the eye tracker, largely because of the similarity in the visualizations. The confusion of generic elements may have contributed to the use of this reference, as one incorrect participant noted. The solution to the conceptual example also prompted participant B5599778 to comment on the amount of dissociation, although this participant could not predict how much dissociation was required on each “side” of the equilibrium arrows.

Whether the participant answered the problem correctly or incorrectly using either a conceptual or algorithmic approach, a common theme from both groups was a reliance on rote learning. Participants referred to nebulous “rules” of chemistry that would help them solve the problem, and others admitting that they were supposed to have “memorized” this information several chapters ago. Just as in post-test item #1, the algorithmic instructional view influences how the students conceive of problem-solving approaches.

None of the participants were able to solve the third post-test item, which was an algorithmic problem that asked for equilibrium pressures of two chemical species. For this equilibrium, there were two possible algorithmic approaches: using initial concentrations and the given K_C value to find equilibrium concentrations and convert to equilibrium pressures or use the formula to convert K_C to K_P as well as convert initial concentration to a pressure and solve for equilibrium pressures directly.

Table 20: Codes for Post-Test Item #3

Participants who answered correctly (N = 0)		*Participants who answered incorrectly (N=26)*	
Code	Quote	Code	Quote
No Participants Answered Correctly		Algorithmic Approach Using K_C	A4052448: "So not sure how to calculate equilibrium pressures, but I'm going to guess it has to do with, um, the Kc equaling the concentration of N2O4 divided by the concentration of NO2 squared because its second order. Um, so I need to find the concentrations so I'm given grams so I need to convert that to moles, and then I need to convert that to, um, molarity"
		Algorithmic Approach Using K_P	A9968443: "And the other information is given I could use to determine the pressure first and then just use the ICE chart to [find equilibrium]."
		Incorrect Use of K_P	A8481633: "And instead of concentrations, there would be pressures. So it would be the pressure of N_2O_4 over the pressure of NO_2-squared. So K_p equals X over 0.00913 [M] squared […] But at the same time I don't have the right units for this either. Which is another reason why I don't think that this is right, 'cause I need a unit of pressure and not a unit of… Concentration."
		Checking the Assumption of Negligible Dissociation	A9755546: "Okay, now I have to do the quadratic formula… I was just checking to make sure that the exponent of K_C wasn't less than -3 because, if it was, I could just cancel out the X and I would skip… It would cause me a lot less trouble."
		Rote Learning	A9755546: "I kind of just followed the steps, the ICE steps."
		Conflating K_P with K_C	B5683065: "Okay, if K_P equals K_C it's at equilibrium. [When asked why] 'Cause it's just a fact."
		Erroneous Use of Kinetics Terminology	B5599778: "Because we're trying to find partial pressures, and since I don't really have any other formulas, I'm assuming that that's the one that we're

		supposed to use. Deductive reasoning tells me that this is my only option. It would kind of make sense because this is such a lower value than the KC constant. And maybe the KP is supposed to make up for it. Because I know K does have units, depending on the orders of reactions, and usually involves molarity and seconds and stuff like that. So hopefully it maybe crosses out somehow, even though I'm not really gonna mess with any of the units at the [22:53] ____. But, maybe it does cross out."
	Superfluous Information	A9755546: [On why they said the temperature is not useful] I don't know where I would include it in the problem, so…"
	Gaseous not Aqueous	B5599778: "Because I can't think of any formulas, but it makes sense. You gave me a K_C constant, but equilibrium pressure you'd need a K_P constant. So I'm trying to think of ways that you can convert... I know none of these examples talked about pressures, so it gets a little bit [14:40] ____ this shows you how to do disassociation, ICE, blah, blah, blah, ICE. So these are all disassociation, but none of them mention equilibrium of gases, or the KP. Huh"

As demonstrated by the algorithmic codes in Table 20, some participants recognized the approach that utilizes K_P; many participants attempted a solution from K_C and concentrations, because of the similarity to problems emphasized in-class. Of those who attempted to use K_P, there were issues with using K_P correctly. Some failed to calculate K_P correctly, even erroneously assuming that K_P and K_C were the same value (not true in this problem, because the stoichiometric coefficients were not equal). The majority who used K_P incorrectly used concentrations in the expression instead of pressures, as seen in the quote from participant A8481633.

The most common reason that participants did not answer the problem correctly is that they could not finish the problem. Participants worked through the solution until they reached the portion which required using the quadratic formula to solve for an x-value; at this point, a few participants gave up and outlined what they would do instead of attempting to solve, while others attempted to solve and were unsuccessful. One participant (most participants never made it to that step) made an error in the set-up of the ICE table leading to a K_C expression that did not have a squared term.

The fourth post-test item was a conceptual question that asked participants to draw the result of increasing the volume of the container on the equilibrium that was shown. An equilibrium constant was not explicitly given (although it could be inferred from the visualization) and generic elements (A and A_2) were used. A correct solution would predict a shift towards the reactants, drawing more monoatomic A's than diatomic A_2's. Participants were fully correct if the answer obeyed the Law of Conservation of Mass, *i.e.* did not just add more A's without a logical decrease in the number of A_2's. Because the problem does not contain information about the amount the volume is increased by, there were multiple potential ratios that were possible. The equilibrium state started with 7 A_2 and 5 A's drawn, and as such a correct answer needed to have a ratio of A_2 to A that reflect the dissociation of A_2 to create two A atoms. These solutions are illustrated in Table 21.

Table 21: Codes for Post-Test Item #4

Participants who answered correctly (N = 3)		*Participants who answered incorrectly (N=15)*	
Code	Quote	Code	Quote
Rote Learning	A5668900: "If you increase the volume it's gonna go towards the side with the most moles. [Why?] That's like a rule or something."	Algorithmic	A9968443: "So if I decrease the concentration... If I decrease the concentration why does it go towards the left? […]I don't know how to explain it I just know that it usually goes... It's because, well it has something to do with how concentration, it's related to moles over liters and since I'm increasing the liters of it, I'm reducing my concentration so it would move towards the more number of moles to make it

			equilibrium 'cause of the reduced concentration”
Conceptual	B3611841: “Yeah, I guess it's not, I'm gonna do... So I'm just gonna make... So there's five A in this one, I'll just do more than five. [pause] So I made nine. […] just because. I wanted to dissociate two of these things. So we have five A_2's, five single As. Then when we dissociate two, we have four more. So we're left with one, two, three, four, five of the A_2”	Conceptual	A8481633: “When you have more room, the pressure is gonna decrease and if they're further away from each other, then the chance of them colliding to react is less so therefore... Yeah, so there's gonna be less number of collisions that cause reactions at a greater number of volume and so there's going to be more unreacted reactants than there are products.”
Extent of Equilibrium	B3611841: “I'm wondering if depending on how much you increase the volume, if that has an effect on how the equilibrium shifts. Should you know an amount, increase the volume by [blank] liters or something? I always knew it just goes to the one with the most moles, moles but I don't know by how many, or by how much. It's weird.”	Extent of Equilibrium	A4052448: “Um… yeah so I'm—there's five A atoms in the left so I'll just draw like seven because I can't really say how many there is going to be because I don't know by how much the volume is increasing so…”
		Law of Conservation of Mass	A8600165: “I would have drawn less A_2 and more A's. Unless, the number of molecules that you have don't [sic] change, just shifts. So if you have 7 A_2's and 5 A's together would be 12 molecules… So I'm thinking that in the second box, there should still be 12… Because you're not making any more new molecules. That the number of molecules would still stay the same. You would just have a higher ratio of one more than the other.

			And because it's now shifting towards the reactants, you would have more A's than A_2's out of the 12. And if you are still keeping the 2 to 1 ratio consistent, then it could be 4 A_2's and 8 A's. So that would still add up to 12. But you would have more A's than A_2's."
		Picturing Gaseous Behavior	A9755546: "I'm trying to picture a container where it shows pressure, and you push the down, so you're losing your volume. When that happens, I guess it's less space for the molecules to bounce back, so maybe you need more... No, then they'd be more crowded. I know this is... I feel this is a rule, though. So less volume, it goes to the side with more moles."

Most participants predicted the correct direction of the net reaction, but were counted as incorrect because they did not obey the Law of Conservation of Mass. Most drew random ratios of each species, and as participant A8481633 phrased it, "As long as the number of the A molecules increases and the number of A_2 molecules decreases, other than that, there's not enough information there to tell you exactly what happens." However, the most common error for the incorrect group was violating the law, such as participant B5631673: "…I just don't know what to do with this [the species A_2], but I don't think it matters so I kept it at five for... That made it go up one. Oh, this is bad"

This violation of Law of Conservation of Mass is unsurprising, given the lack of information about the extent of equilibrium. The question prompt did not tell participant by how much the volume would increase, and participants who answered correctly and incorrectly noticed this discrepancy. A truly conceptually-derived correct answer would have incorporated this, although most correct conceptual codes discussed the action of dissociation without specifically mention that dissociation of A_2 as the source of the increased number of A. Although some participants attempted to picture the gaseous behavior and the collisions that would fuel the shift to the reactant side, there was still a gap in understanding. Most participants (regardless of correct or incorrect) ultimately relied on a rote or algorithmic approach to solve this problem.

Conclusions

Initially the goal was to classify student reasoning and measure visual attention as a function of that classification. Results from Figures 42 and 43 indicate that there was not a significant difference in average time spent fixating on text or math/image AOIs within each example, even after controlling for the relative size of each AOI. The extremely large Cohen's d effect size for the paired-samples t-test on the pre- and post-test scores suggests that the relative item complexity and difficulty (as seen in Table 16) could be confounding the classification of participants. The two major groups of this classification system favored similar algorithmic problem-solving approaches, as noted in Table 17. As such, any changes in a participants' classification after the study could be due to a wealth of confounding factors.

When the classification system is disregarded, interesting visual search behaviors emerge. While there was not a significant main effect for AOI type, there was a significant main effect for example type; participants responded differently to the various stimuli. A detailed look at the interaction effect suggests that when text is integrated into the problem structure, general chemistry students will have attended to it significantly longer than the mathematical statements. This finding is intriguing; one might infer from the traditional (algorithmic) worked-example that general chemistry students are only interested in the mathematical content to aid problem-solving. The interaction effect detailed in Figure 44 demonstrates that there is a significant change in fixation length for participants when the problem type is altered to incorporate conceptual information or uses conceptual, particulate-level diagrams to explain. This finding suggests that Atkinson's suggestion of incorporating conceptual justifications is an effect use of worked examples.[140]

Limitations of this study include a small sample size derived from a purposeful sampling method based on student problem-solving classification. While the results of this study cannot definitively prove Pickering's original assumption that there are not two types of students, only two instructional approaches[137] nor disprove Nakhleh's assertion that there is a quadrant to map students' problem-solving ability[138], the qualitative analysis of the post-test think-aloud interview suggests that participants' problem-solving approach depended on problem type; the references to the instructors' algorithmic problem-solving methods suggest an influence in how the instructor approaches these problems initially. The sample size in this study was limited based on student classification, which is one factor that limits the ability to conclude on the mixed effect ANOVA model as seen in Figures 44, 45 and 46. The item difficulty and complexity of the pre- and post-test items also limit the accuracy of any student classification. The final limitation was the assumption that the viewing/passive reading behavior of participants as captured by the eye tracker is linked to participants' problem-solving behaviors. While both the eye-tracking stimuli and post-test items were equilibria problems, one interviewee noted that these problems were not the same in their view; the question of transfer between different types of equilibria is another confounding factor.

Future work in this area should address the assumptions of validity of student classification, correlation between passive reading/viewing behavior and problem-solving behavior, and the transfer of student understanding of general to acid/base equilibria. These issues are best tackled in separate projects with sample sizes more than the current sample size.

CHAPTER 6

CONCLUSIONS

As the academy is adapting to technological and pedagogical advances, the structure of the higher education classroom is evolving. The prototypical model of higher education was self-directed study in advance of a lecture by an expert and rigorous practice and application of practice problems; if such a model was ever truly attained, it has been jettisoned in favor of collaborative, student-centered methods in which the sage steps off the stage and actively participates in student problem-solving.

Although the meta-analyses of this active, flipped pedagogy are conclusive—engaging students directly is more beneficial than lecturing at them—the manner of implementing this pedagogy has focused on the instructor. Technological aids such as clickers and published educational activities are developed to assist the instructor in managing personal *learner-instructor* interaction in classroom meeting spaces that are not well-suited for an active pedagogy. Recent investigations into *instructor-content* and *instructor-instructor* interactions, such as content knowledge and pedagogical practices, have designed professional development to equip instructors for this challenge and hold them accountable for student outcomes.[26,28,144] Investigations into STEM pedagogy have also extensively studied *learner-learner* interactions in order to design active pedagogies. This wealth of literature has armed instructors with a variety of ways to achieve collaborative, interactive classrooms.

The area in most need of attention is the *learner-content* interaction, which largely takes place outside of the classroom and is shielded from the researcher's eye. Ultimately, the desirable outcome for higher education classrooms is a self-directed student who engages with the content directly for the broad understanding, and then seeks out an expert or peer for clarification on finer points. By exploring what hierarchy the various course resources have in the students' perspective (*content-content* interaction), it is possible to explore the highly-valued content-delivery resources and make suggestions for future evolution of the current resources.

Summary of Study Results

In the first three semesters' worth of data analyzed, General Chemistry I students report relying on their textbook for multiple hours of studying per week; when the ebook was adopted, the homework resource was more frequently used. In the semester with the print textbook, students' use of textbook, YouTube, professor's lecture slides, private and peer tutors were all significantly different across anticipated grade. With an eBook, only their use of Google was significantly different across anticipated grade. In addition to mathemagenic features, ebook features that help students stay on-task or visualize reactions are significantly highly rated.

Out of all the resources available to students, the textbook is the one with which they report spending the most time. How helpful they find the textbook in achieving their goals or what specifically it is about the textbook that they find so helpful is also available to us in this dataset, but what this study is lacking is the instructors' choices. Textbook authorship has been hypothesized to not be as valued as a scholarly pursuit as other types of writing, and textbooks often lack the expert's voice.[72] In this sense, the textbook is interchangeable with others of a similar curriculum. In the students' perspective, the lack of voice and identical content does not justify the price. Furthermore, when it comes to selection of a textbook, there is no data on what factors or concerns the faculty take into consideration when choosing a textbook to reflect a specific format, curriculum, or emphasis. There is a prevailing attitude that one textbook is as good as another. If the textbook truly does not matter as this instructor view may suggest, this dataset shows that students engage with their textbooks far more often than they ever engage with their instructor. As such, we can see that the main sources of information for students are the textbook, Google, friends, and occasionally peer or private tutors. If an instructor hopes to influence students to utilize a resource, further study is necessary.

As far as a format change, the students engage with the features in a similar manner despite the availability for more interactive choices. If the format does not engage them more, it begs the question, is this all we can do with an ebook? What is truly the difference if the ebook is essentially a glorified PDF of the print book. All the innovations, educational interventions, and implementations of cognitive science are unappreciated and underutilized by today's student. Given that they do not use the written text as designed and that they often seek out interpersonal interaction for clarity, it is time to reinvent what a textbook—especially in the electronic age—truly can be.

One feature of the textbook that students valued was the embedded animations of the ebook. This result led to a study of the educational efficacy of an animation or a static set of images. This type of *learner-content* interaction was assessed using an ordering question, which is under-used in chemistry courses. Investigation of the effect of instructional format (learning the material using animations versus learning the material using static images extracted from those animations) via regression was not significant; the patterns of students' eye movements between correct and incorrect group demonstrated an ability to detect and quantify cognitive load. For two of the three ordering problems, students with higher scores had significantly ($p < 0.001$) shorter total fixation duration when ordering the events and fixated for less time in the planning and solving phases of the problem-solving process. Sex and visualization type did not have significant effect on post-test ordering item score.

This finding was supported by the scanpath patterns that demonstrated that students who correctly solved the problems used more efficient problem-solving strategies. Between participants who answered correct vs incorrectly, in the planning phase (phase 1), both groups had patterns between adjacent choices (e.g. DCD or FED) without any jumps (e.g. ADE). In the problem-solving phase (phase two), the incorrect group's patterns were between adjacent choices or steps (e.g. ABC or 432) while the correct group's patterns were between choices and corresponding steps. In the checking phase (phase three), the correct group's patterns indicated that more than ten percent of participants checked answers sequentially. For this phase, the incorrect group had no checking patterns detected with ten percent cutoff, indicating that this group either did not check answers or that the fixations were so scattered as to not form patterns.

The survey data also revealed that general chemistry students value mathemagenic textbook features that aid with algorithmic problem-solving, in line with the focus of "earning a good grade" that the majority of survey participants expressed.

Given the consistently high helpfulness rating of the worked-out example and the wealth of literature on the efficacy of worked-examples as an instructional tool, attention to how learners interact with this textbook feature is a valuable line of inquiry. The ebook adopted by the Department offered multiple versions of the worked example ("tutor solutions to example problem," "video solutions to example problem," "tutor solutions to mastery questions") which offer more than one modality for delivering content. The common theme to these worked-examples is an underlying algorithmic approach to assessing student understanding.

The attempt to classify students by performance on algorithmic-conceptual paired items could not be supported by assessment performance or eye-tracking metrics; however, the interview results demonstrated that participants favored an algorithmic approach to problem-solving (regardless of whether the problem was algorithmic or conceptual). The reliance on rote-learning and memorization of algorithms indicates an unintended consequence from algorithmic examples and assessments. The eye-tracking analyses revealed that when conceptual information was available, participants (regardless of classification) fixated longer on that information. When the problem type was completely conceptual, all participants fixated on the information longer than the traditional style of worked examples. This finding is intriguing; one might infer from the traditional (algorithmic) worked-example that general chemistry students are only interested in the mathematical content to aid problem-solving; rather than two types of students, there is a possibility that the instructional goal of the content-delivery device has a greater effect on students' visual perusal.

Limitations

This dissertation work has several limitations in each project. Although the textbook/resource survey achieved high response rates each semester (for a total sample of 3,646 General Chemistry I students), this self-selection sampling was achieved at the expense of more precise data collection methods. Participants were asked to reflect on their study habits, which is unreliable in this age group without explicit metacognitive training.[145]s Furthermore, this recollection of the semester's study habits was collected at the end of the semester; this depiction represents the best approximate of average use, without an initial dataset for comparison or individual follow-up during the semester for external validation. Measures were taken to ensure the validity of the survey findings. Biases to internal validity were addressed by not forcing participants to respond to each question and embedding requests to "mark this as N/A" within large questions (to prevent students from answering the high value for each answer); participants who failed these measures (did not respond to questions or did not respond correctly to internal validation measures) were removed from the dataset prior to analysis. The longitudinal results show external validity in the consistency of responses over the six semesters. A potential avenue to confirm the validity of these results and measure the effects demonstrated at a finer grain would be to track usage of textbook and course resources for a smaller group of students at multiple points throughout the semester; this study could be accomplished through short interviews or questionnaires.

In the study intended to compare student visual attention to static and dynamic visualizations of a biochemical pathway, the analysis of student problem-solving patterns on ordering questions had three limitations. The first was that both groups of students had already been previously instructed on the material in the visualization, and therefore the educational efficacy of the two formats could not be inferred from the difference in the post-test scores. Although the static version of the visualization was created from the dynamic visualization, small changes in post-production (before the study was proposed) interfered with finding equivalent frames between the two formats; as such, participants' visual attention to the instructional visualizations could not be directly compared. The final limitation to this study was that the scanpath analysis used the fixations on AOIs with static position; these AOIs remained in the original position of an answer choice or step, regardless of whether the text had or had not been moved. To overcome this limitation, pairing this method of analysis with simultaneous screen-recording and choreographing the timing of drag-and-drop choices to fixation sequence would provide a more detailed account of the problem-solving phase.

As addressed in Chapter 5, the eye-tracking study on worked examples was predicated on the ability to classify students on the basis of the paired conceptual-algorithmic assessment item performance. While this study could not conclusively disprove this classification scheme, the results suggest that this classification does not extend to visual attention behaviors or problem-solving approaches. This purposeful sampling strategy limited the overall sample size, and thus the power of the mixed-effect factorial ANOVA employed in the study. Future work could utilize item response theory (IRT) to examine how the item difficulty and discrimination values contribute to performance, providing a revised classification for evidence of whether this classification exists. The final limitation was the assumption that the viewing/passive reading behavior of participants as captured by the eye tracker is linked to participants' problem-solving behaviors. While both the eye-tracking stimuli and post-test items were equilibria problems, one interviewee noted that these problems were not the same in their view; the question of transfer between different types of equilibria is another confounding factor. Future studies should limit the scope of content of the stimuli (focusing on just one type of equilibria, for example) to limit transferability effects.

Suggestions for Publishers and Authors

The ratings of helpfulness of textbook features indicate that students find value in mathemagenic features that aid problem-solving (worked examples and practice problems) or explain information with multiple modalities (text, figures, and animations). It does not appear that they find the contextual pieces in the margins (sections that explain experiments, deeper or historical contexts, or provide real world examples) as “helpful.” As such, one suggestion is to minimize those features in favor of the aforementioned “helpful” features. Furthermore, participants expressed a willingness to participate a moderated message board, and this type of embedded feature with a LaTeX equation editor would simplify implementation for instructors and learners.

As demonstrated in the eye-tracking study on worked examples, participants directed their visual attention to conceptual information when it was embedded in algorithmic problems and spent more time engaged with conceptual style example problems. The prevalence for algorithmic problem-solving strategies in the post-test interviews reflected an instructional paradigm, as evidenced by their statements. One innovation would be to iteratively provide the “carrot” of math/algorithmic steps as a reward to a prompt for the conceptual justification for that algorithmic step. These responses could be open-ended (if a novel problem type is used) or allow for a multiple-choice response (which could contribute to the students' course grade). Approaches to worked examples like this reflect the extensive literature on worked examples and the results of the longitudinal survey in Chapter 3, while incentivizing conceptual understanding over rote memorization of algorithms.

Suggestions for Instructors

Instructors interested in adopting an ebook are primarily focused on outcomes: did the students do better? In the current climate of education, this question is the natural first impulse, and yet all of the factors that contribute to an individual students' final grade (time management, affective behaviors, intellectual interests, math background), the contribution of the format of the textbook alone is difficult to discern. While previous research has shown that open-educational ebooks are non-inferior, the goal of this work was to explore how students perceive and use the textbook in relation to other course resources.

This study added to literature which suggests that students of differing achievement levels utilize the course resources differently. One suggestion for instructors is to seek out an electronic text that offers an individual, tailored experience for students or at least a textbook resource that is free or low-cost. In semesters with a print book, students reported using the book more than other resources (although low-performing students reported not using the textbook); in the ebook semesters, students at all anticipated grades reported using the textbook, although they increasingly relied on the homework and *learner-learner* interactions (whether study groups, friends, or peer/private tutoring). A textbook that incorporates these *learner-learner* interactions would best suit the modern student; with the moderation of an expert or TA, the discussion can extend outside the classroom space.

A final suggestion to instructors would be to reflect on their own expectations and values in a textbook. Despite being one of the most expensive pieces of a college student's education, its value has long been assumed and therefore not discussed in the literature. Given each instructor's goals for their students and this dataset on what students find helpful, the gap in the literature is which concerns and considerations a textbook selection committee has.

Suggestions for Researchers

As summarized at the end of Chapter 3, future research in the area of textbooks includes longitudinal studies, analysis of data from other introductory courses, and studying how an ebook utilizing a non-traditional curriculum can impact student interaction patterns. There are indications that prior experience with a particular technology predicts future behavioral intention and use behaviors;[73] accordingly, the predictive value of experience could be evaluated by following participants who used the ebook in the first semester of General Chemistry into the second semester of the course and compare their answers on a repeated-measures model. Finally, the missing piece from this study was the instructor perspective; while instructors want the best outcomes for their students, a qualitative study to investigate how this population conceives of the textbook and the factors that go into selection is necessary. Ultimately, it is not the student who chooses the textbook.

The scanpath analysis technique employed in Chapter 4 serves as a method for characterizing (through patterns and entropy values) the *learner-content* interactions that STEM students have with the features of their textbooks. Future studies should consider employing this technique to mathematical problem-solving or assessment items that involve visuospatial reasoning to elucidate patterns of visual behavior. When paired with observation protocol (such as COPUS codes), patterns of *learner-instructor* interactions could be quantified as entropy values and visualized in patterns.

The results of Chapter 5 recommend a study of the validity of classifying students on the conceptual-algorithmic quadrant system; this assumption needs to be addressed with a larger sample size recruited across multiple instructors and examined on several topics. Future work from this project should also investigate students' interaction with a more interactive form of the worked example, such as the one described in the suggestions to publishers/authors section.

REFERENCES

(1) Brockliss, L. Curricula. In *A History of the University in Europe*; Cambridge University Press: Cambridge, UK, 1996; pp 565–620.

(2) Chi, M. T. H.; Wylie, R. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol* **2014**, *49* (4), 219–243 DOI: 10.1080/00461520.2014.965823.

(3) Andrews, T. M.; Leonard, M. J.; Colgrove, C. A.; Kalinowski, S. T. Active Learning Not Associated with Student Learning in a Random Sample of College Biology Courses. *CBE Life Sci. Educ.* **2011**, *10* (4), 394–405 DOI: 10.1187/cbe.11-07-0061.

(4) Prince, M. J.; Felder, R. M. Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. *J. Engr. Educ.* **2006**, *95* (2), 123–138.

(5) Prince, M. J. Does Active Learning Work? A Review of the Research. *J. Engr. Educ.* **2004**, *93* (3), 223–231.

(6) Smith, K. A.; Sheppard, S. D.; Johnson, D.; Johnson, R. T. Pedagogies of Engagement: Classroom-Based Practices - ProQuest https://search.proquest.com/docview/217964349?pq-origsite=gscholar.

(7) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafora, N.; Jordt, H.; Wenderoth, M. P. Active Learning Increases Student Performance Inscience, Engineering, and Mathematics. *PNAS* **2014**, *111* (23), 8410–8415 DOI:

10.1073/pnas.1319030111.

(8) Cooper, M. M. It Is Time To Say What We Mean. *J. Chem. Educ.* **2016**, *93* (5), 799–800 DOI: 10.1021/acs.jchemed.6b00227.

(9) Ryan, M. D.; Reid, S. A. Impact of the Flipped Classroom on Student Performance and Retention: A Parallel Controlled Study in General Chemistry. *J. Chem. Ed.* **2016**, *93*, 13–23 DOI: 10.1021/acs.jchemed.5b00717.

(10) Salami, T. O.; Omiteru, E. O. E-Textbooks and the Digital Natives: A Study of First-Year Chemistry Students' Attitudes toward E-Textbooks; 2015; pp 45–59.

(11) Christiansen, M. A. Inverted Teaching: Applying a New Pedagogy to a University Organic Chemistry Class. *J. Chem. Educ.* **2014**, *91* (11), 1845–1850 DOI: 10.1021/ed400530z.

(12) Eubanks, I. D. Challenges for the Next Generation. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; Orna, M. V., Ed.; American Chemical Society: Washington, DC, 2015; pp 339–355.

(13) Seery, M. K. Flipped Learning in Higher Education Chemistry: Emerging Trends and Potential Directions. *Chem. Educ. Res. Pr.* **2015**, *16*, 758–768 DOI: 10.1039/c5rp00136f.

(14) Smith, J. D. Student Attitudes toward Flipping the Chemistry Classroom. *Chem. Educ. Res. Pr.* **2013**, *14*, 607–614 DOI: 10.1039/c3rp00083d.

(15) Knowles, M. S. *The Modern Practice of Adult Education*; Follett: Chicago, IL, 1980.

(16) Ealy, J. B. Development and Implementation of a First-Semester Hybrid Organic Chemistry Course: Yielding Advantages for Educators and Students. *J. Chem.*

Educ. **2013**, *90* (3), 303–307 DOI: 10.1021/ed200858p.

(17) Bodner, G. M. Constructivism: A Theory of Knowledge. *J. Chem. Ed.* **1986**, *63*, 873–878.

(18) Bunce, D. M. Does Piaget Still Have Anything to Say to Chemists? *J. Chem. Educ.* **2001**, *78* (8), 1107 DOI: 10.1021/ed078p1107.2.

(19) Anderson, J. R.; Reder, L. M.; Simon, H. A. Applications and Misapplications of Cognitive Psychology to Mathematics Education http://act-r.psy.cmu.edu/papers/misapplied.html (accessed Feb 24, 2018).

(20) Ferguson, R. L. Constructivism and Social Constructivism. In *Theoretical Frameworks for Research in Chemistry/Science Education*; Bodner, G. M., Orgill, M., Eds.; Pearson: Upper Saddle River, NJ, 2007; pp 28–49.

(21) Novak, J. D. Meaningful Learning: The Essential Factor for Conceptual Change in Limited or Inappropriate Propositional Hierarchies Leading to Empowerment of Learners. *Sci. Ed.* **2002**, *86*, 548–571 DOI: 10.1002/sce.10032.

(22) Moore, M. G. Editorial: Three Types of Interaction. *Am. J. Distance Educ.* **1989**, *3* (2), 1–7 DOI: 10.1080/08923648909526659.

(23) Anderson, T. Getting the Mix Right Again: An Updated and Theoretical Rationale for Interaction. *Int. Rev. Res. Open Distrib. Learn.* **2003**, *4* (2).

(24) Xiao, J. Learner-Content Interaction in Distance Education: The Weakest Link in Interaction Research. *Distance Educ.* **2017**, *38* (1), 123–135 DOI: 10.1080/01587919.2017.1298982.

(25) Madland, C.; Richards, G. Enhancing Student-Student Online Interaction: Exploring the Study Buddy Peer Review Activity. *IRRODL* **2016**, *17* (3), 157–175.

(26) Smith, M. K.; Jones, F. H. M.; Gilbert, S. L.; Wieman, C. E. The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE Life Sci. Educ.* **2013**, *12* (4), 618–627 DOI: 10.1187/cbe.13-08-0154.

(27) Wieman, C.; Gilbert, S. The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science. *CBE—Life Sci. Educ.* **2014**, *13* (3), 552–569 DOI: 10.1187/cbe.14-02-0023.

(28) MacIsaac, D. L.; Falconer, K. A. Reforming Physics Education via RTOP. *Phys. Teach.* **2002**, *40* (8), 479–485.

(29) Field, A. *Discovering Statistics Using IBM SPSS Statistical Software*, 4th ed.; Carmichael, M., Ed.; SAGE Publications: London, 2013.

(30) Becker, L. A. Effect Size (ES) - Effect Size Calculators https://www.uccs.edu/lbecker/effect-size.html (accessed Mar 1, 2018).

(31) Tang, H.; Day, E.; Kendhammer, L. K.; Moore, J. N.; Brown, S. A.; Pienta, N. J. Eye Movement Patterns in Solving Science Ordering Problems. *J. Eye Mov. Res.* **2016**, *9* (3) DOI: 10.16910/JEMR.9.3.6.

(32) Le Meur, O.; Baccino, T. Methods for Comparing Scanpaths and Saliency Maps: Strengths and Weaknesses. *Behav. Res. Methods* **2013**, *45* (1), 251–266 DOI: 10.3758/s13428-012-0226-9.

(33) Talanquer, V. Using Qualitative Analysis Software to Facilitate Qualitative Data Analysis. In *Tools of Chemistry Education Research*; Bunce, D. M., Cole, R., Eds.; American Chemical Society: Washington, DC, 2014; pp 83–95.

(34) Herrington, D. G.; Daubenmire, P. L. Using Interviews in CER Projects: Options,

Considerations, and Limitations. In *Tools of Chemistry Education Research*; Bunce, D. M., Cole, R., Eds.; American Chemical Society: Washington, DC, 2014; pp 31–59.

(35) Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oxford University Press, 2011.

(36) Chi, M. T. H. Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *J. Learn. Sci.* **1997**, *6* (3), 271–315 DOI: 10.1207/s15327809jls0603_1.

(37) Marshall, M. N. Sampling for Qualitative Research. *Fam. Pract.* **1996**, *13* (6), 522–525.

(38) Corbin, J.; Strauss, A. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Osterr Z Soziol.* **1990**, *19*, 418–427.

(39) Taylor-Powell, E.; Renner, M. *Analyzing Qualitative Data*; Madison, WI, 2003.

(40) Cengage Learning. CogLab: Blind Spot https://coglab.cengage.com/labs/blind_spot.shtml (accessed Mar 2, 2018).

(41) Havanki, K. L.; VandenPlas, J. R. Eye Tracking Methodology for Chemistry Education Research; 2014; pp 191–218.

(42) Duchowski, A. T. *Eye Tracking Methodology*, 2nd ed.; Springer: London, UK, 2007.

(43) Buetti, S.; Lleras, A. *Areas of the Brain*; University of Illinois Urbana-Champaign, 2018.

(44) Atkinson, R. C.; Shiffrin, R. M. Human Memory: A Proposed System and Its Control Processes. *Psychol. Learn. Motiv.* **1968**, *2*, 89–195 DOI: 10.1016/S0079-

7421(08)60422-3.

(45) Shiffrin, R. M.; Schneider, W. Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. **1977**, *84* (2).

(46) Baddeley, A. D.; Hitch, G. Working Memory. *Psychol. Learn. Motiv.* **1974**, *8*, 47–89 DOI: 10.1016/S0079-7421(08)60452-1.

(47) Baddeley, A. Working Memory. *Science (80-.).* **1992**, *255*, 556–559.

(48) Mayer, R. E.; Moreno, R. Aids to Computer-Based Multimedia Learning. *Learn. Instr.* **2002**, *12*, 107–119.

(49) Stieff, M.; Hegarty, M.; Deslongchamps, G. Identifying Representational Competence With Multi-Representational Displays. *Cogn. Instr.* **2011**, *29* (1), 123–145 DOI: 10.1080/07370008.2010.507318.

(50) Schuttlefield, J. D.; Kirk, J.; Pienta, N. J.; Tang, H. Investigating the Effect of Complexity Factors in Gas Law Problems. *J. Chem. Educ.* **2012**, *89* (5), 586–591 DOI: 10.1021/ed100865y.

(51) Tang, H.; Kirk, J.; Pienta, N. J. Investigating the Effect of Complexity Factors in Stoichiometry Problems Using Logistic Regression and Eye Tracking. *J. Chem. Educ.* **2014**, *91* (7), 969–975 DOI: 10.1021/ed4004113.

(52) Slykhuis, D. A.; Wiebe, E. N.; Annetta, L. A. Eye-Tracking Students’ Attention to PowerPoint Photographs in a Science Education Setting. *J. Sci. Educ. Technol.* **2005**, *14* (5–6), 509–520 DOI: 10.1007/s10956-005-0225-z.

(53) de Jong, T. Cognitive Load Theory, Educational Research, and Instructional Design: Some Food for Thought. *Instr. Sci.* **2010**, *38* (2), 105–134 DOI:

10.1007/s11251-009-9110-0.

(54) Hegarty, M.; Mayer, R. E.; Green, C. E. Comprehension of Arithmetic Word Problems: Evidence from Students' Eye Fixations. *J. Educ. Psychol.* **1992**, *84* (1), 76–84 DOI: 10.1037/0022-0663.84.1.76.

(55) Gegenfurtner, A.; Lehtinen, E.; Saljo, R. Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educ Psychol Rev* **2011**, *23*, 523–552 DOI: 10.1007/s10648-011-9174-7.

(56) West, J. M.; Haake, A. R.; Rozanski, E. P.; Karn, K. S. eyePatterns. In *Proceedings of the 2006 symposium on Eye tracking research & applications - ETRA '06*; ACM Press: New York, New York, USA, 2006; p 149.

(57) Tang, H.; Topczewski, J. J.; Topczewski, A. M.; Pienta, N. J. Permutation Test for Groups of Scanpaths Using Normalized Levenshtein Distances and Application in NMR Questions. In *Proceedings of the Symposium on Eye Tracking Research and Applications*; ACM, 2012; pp 169–172.

(58) Day, E. L.; Tang, H.; Kendhammer, L. K.; Pienta, N. J. Sequence Analysis of Eye Movement Patterns in Solving Science Ordering Problems. In *Eye Tracking for the Chemistry Education Researcher*; VandenPlas, J., Hansen, S., Cullipher, S., Eds.; American Chemical Society, 2018.

(59) Eraslan, S.; Yesilada, Y.; Harper, S. Eye Tracking Scanpath Analysis Techniques on Web Pages: A Survey, Evaluation, and Comparison. *JEMR* **2016**, *9* (1), 1–19 DOI: 10.16910/jemr.9.1.2.

(60) Krejtz, K.; Duchowski, A. T.; Krejtz, I.; Kopacz, A.; Chrząstowski-Wachtel, P.

Gaze Transitions When Learning with Multimedia; 2016; Vol. 9.

(61) Pienta, N. J. General Chemistry Student Surveys: Longitudinal Data about Which Factors Helped Students Learn. In *Investigating Classroom Myths through Research on Teaching and Learning*; Bunce, D. M., Ed.; Oxford University Press, 2011; pp 121–135.

(62) Lloyd, B. W. A Review of Curricular Changes in the General Chemistry Course during the Twentieth Century. *J. Chem. Educ.* **1992**, *69* (8), 633 DOI: 10.1021/ed069p633.

(63) Bell, J. A. Getting It Right: A Paradigm for the Education of Chemists. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; Orna, M. V., Ed.; American Chemical Society: Washington, DC, 2015; pp 25–43.

(64) Allen, G.; Guzman-Alvarez, A.; Smith, A.; Gamage, A.; Molinaro, M.; Larsen, D. S. Evaluating the Effectiveness of the Open-Access ChemWiki Resource as a Replacement for Traditional General Chemistry Textbooks. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 939–948 DOI: 10.1039/C5RP00084J.

(65) GAO. *College Textbooks: Enhanced Offerings Appear to Drive Recent Price Increases*; 2005.

(66) Office, U. S. G. A. College Textbooks: Students Have Greater Access to Textbook Information. June 6, 2013.

(67) Lubliner, D. J. Integrated Learning Environments: From eBooks to E ^{2} Books: From E-Learning 3.0 to E-Knowledge and Beyond; 2010; pp 149–166.

(68) Shorb, J. M.; Moore, J. W. The ChemPaths Student Portal: Making an Online Textbook More than a Book Online; 2010; pp 283–308.

(69) Butler, D. Technology: The Textbook of the Future. *Nature* **2009**, *458* (7238), 568–570 DOI: 10.1038/458568a.

(70) Shepperd, J. A.; Grace, J. L.; Koch, E. J. Evaluating the Electronic Textbook: Is It Time to Dispense with the Paper Text? *Teach. Psychol.* **2008**, *35*, 2–5 DOI: 10.1080/00986280701818532.

(71) Tulip, D.; Cook, A. Teacher and Student Usage of Science Textbooks. *Res. Sci. Educ.* **1993**, *23* (1), 302–307 DOI: 10.1007/BF02357074.

(72) Issitt, J. Reflections on the Study of Textbooks. *Hist. Educ.* **2004**, *33* (6), 683–696 DOI: 10.1080/0046760042000277834.

(73) Venkatesh, V.; Thong, J. Y. L.; Xu, X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Q.* **2012**, *36* (1), 157–178.

(74) Orna, M. V. Introduction: The Evolution and Practice of Chemical Education. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; Orna, M. V., Ed.; American Chemical Society: Washington, DC, 2015; pp 1–24.

(75) Barrow, G. M. General Chemistry and the Basis for Change. *J. Chem. Educ.* **1994**, *71* (10), 874 DOI: 10.1021/ed071p874.

(76) Gillespie, R. G. Commentary: Reforming the General Chemistry Textbook. *J. Chem. Educ.* **1997**, *74* (5), 484 DOI: 10.1021/ed074p484.

(77) Russo, R. The Perfect Text. *J. Chem. Educ.* **1998**, *75* (2), 141 DOI: 10.1021/ed075p141.

(78) Trapp, D. W. Reforming High School Chemistry Textbooks. *J. Chem. Educ.* **1999**, *76* (2), 165 DOI: 10.1021/ed076p165.5.

(79) Jensen, W. B. Logic, History, and the Chemistry Textbook: II. Can We Unmuddle the Chemistry Textbook? *J. Chem. Educ.* **1998**, *75* (7), 817 DOI: 10.1021/ed075p817.

(80) Pyburn, D. T.; Pazicni, S. Applying the Multilevel Framework of Discourse Comprehension to Evaluate the Text Characteristics of General Chemistry Textbooks. *J. Chem. Ed.* **2014**, *91*, 778–783 DOI: 10.1021/ed500006u.

(81) Thiele, R. B.; Treagust, D. F. The Nature and Extent of Analogies in Secondary Chemistry Textbooks. *Instr. Sci.* **1994**, *22* (1), 61–74 DOI: 10.1007/BF00889523.

(82) Thiele, R. B.; Treagust, D. F. Analogies in Chemistry Textbooks. *Int. J. Sci. Educ.* **1995**, *17* (6), 783–795 DOI: 10.1080/0950069950170609.

(83) Liu, Y.; Treagust, D. F. Content Analysis of Diagrams in Secondary School Science Textbooks. In *Critical Analysis of Science Textbooks*; Springer Netherlands: Dordrecht, 2013; pp 287–300.

(84) Dávila, K.; Talanquer, V. Classifying End-of-Chapter Questions and Problems for Selected General Chemistry Textbooks Used in the United States. *J. Chem. Educ.* **2010**, *87* (1), 97–101 DOI: 10.1021/ed8000232.

(85) Talanquer, V. Explanations and Teleology in Chemistry Education. *Int. J. Sci. Educ.* **2007**, *29* (7), 853–870 DOI: 10.1080/09500690601087632.

(86) King, D.; Domin, D. S. The Representation of People of Color in Undergraduate General Chemistry Textbooks. *J. Chem. Educ.* **2007**, *84* (2), 342 DOI: 10.1021/ed084p342.

(87) Bazler, J. A.; Simonis, D. A. Are High School Chemistry Textbooks Gender Fair? *J. Res. Sci. Teach.* **1991**, *28* (4), 353–362 DOI: 10.1002/tea.3660280408.

(88) Seethaler, S.; Czworkowski, J.; Wynn, L. Analyzing General Chemistry Texts' Treatment of Rates of Change Concepts in Reaction Kinetics Reveals Missing Conceptual Links. *J. Chem. Educ.* **2018**, *95* (1), 28–36 DOI: 10.1021/acs.jchemed.7b00238.

(89) Gillette, G.; Sanger, M. J. Analysing the Distribution of Questions in the Gas Law Chapters of Secondary and Introductory College Chemistry Textbooks from the United States. *Chem. Educ. Res. Pr.* **2014**, *15* (4), 787–799 DOI: 10.1039/C4RP00115J.

(90) Sanger, M. J.; Greenbowe, T. J. An Analysis of College Chemistry Textbooks As Sources of Misconceptions and Errors in Electrochemistry. *J. Chem. Educ.* **1999**, *76* (6), 853 DOI: 10.1021/ed076p853.

(91) Niaz, M. A Rational Reconstruction of the Origin of the Covalent Bond and Its Implications for General Chemistry Textbooks. *Int. J. Sci. Educ.* **2001**, *23* (6), 623–641 DOI: 10.1080/09500690010006491.

(92) Furió-Más, C.; Luisa Calatayud, M.; Guisasola, J.; Furió-Gómez, C. How Are the Concepts and Theories of Acid–Base Reactions Presented? Chemistry in Textbooks and as Presented by Teachers. *Int. J. Sci. Educ.* **2005**, *27* (11), 1337–1358 DOI: 10.1080/09500690500102896.

(93) De Posada, J. M. The Presentation of Metallic Bonding in High School Science Textbooks during Three Decades: Science Educational Reforms and Substantive Changes of Tendencies. *Sci. Educ.* **1999**, *83* (4), 423–447 DOI: 10.1002/(SICI)1098-237X(199907)83:4<423::AID-SCE3>3.0.CO;2-9.

(94) Naughton, W.; Schreck, J.; Heikkinen, H. Seeking Evidence for "curricular

Relevancy” within Undergraduate, Liberal Arts Chemistry Textbooks. *J. Res. Sci. Teach.* **2008**, *45* (2), 174–196 DOI: 10.1002/tea.20205.

(95) Gkitzia, V.; Salta, K.; Tzougraki, C. Development and Application of Suitable Criteria for the Evaluation of Chemical Representations in School Textbooks. *Chem. Educ. Res. Pr.* **2011**, *12*, 5–14 DOI: 10.1039/C1RP90003J.

(96) Nyachwaya, J. M.; Gillaspie, M.; M. Nyachwaya, J.; Gillaspie, M. Features of Representations in General Chemistry Textbooks: A Peek through the Lens of Cognitive Load Theory. *Chem. Educ. Res. Pr.* **2015**, *17* (1), 58–71 DOI: 10.1039/c5rp00140d.

(97) Nyachwaya, J. M.; Wood, N. B. Evaluation of Chemical Representations in Physical Chemistry Textbooks. *Chem. Educ. Res. Pr.* **2014**, *15* (4), 720–728 DOI: 10.1039/C4RP00113C.

(98) Han, J.; Roth, W.-M. Chemical Inscriptions in Korean Textbooks: Semiotics of Macro- and Microworld. *Sci. Educ.* **2006**, *90* (2), 173–201 DOI: 10.1002/sce.20091.

(99) Smith, B. D.; Jacobs, D. C. TextRev: A Window into How General and Organic Chemistry Students Use Textbook Resources. *J. Chem. Educ.* **2003**, *80* (1), 99 DOI: 10.1021/ed080p99.

(100) Zumalt, C. J.; Williamson, V. M. Does the Arrangement of Embedded Text Versus Linked Text in Homework Systems Make a Difference in Students Impressions, Attitudes, and Perceived Learning? *J. Sci. Educ. Technol.* **2016**, *25* (5), 704–714 DOI: 10.1007/s10956-016-9625-5.

(101) Williamson, V. M.; Zumalt, C. J. How Do General Chemistry Students’

Impressions, Attitudes, Perceived Learning, and Course Performance Vary with the Arrangement of Homework Questions and E-Text? *Chem. Educ. Res. Pract.* **2017**, *18* (4), 785–797 DOI: 10.1039/C7RP00052A.

(102) Lovett, M.; Meyer, O.; Thille, C. JIME - The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning. *J. Interact. Media Educ.* **2008**, *2008* (1), 13 DOI: 10.5334/2008-14.

(103) Bowen, W.; Lack, K.; Chingos, M.; Nygren, T. *Interactive Learning Online at Public Universities*; 2012.

(104) Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Manage. Sci.* **1989**, *35* (8), 982–1003.

(105) Kotz, J. C.; Treichel, P. M.; Townsend, J. R. *Chemistry & Chemical Reactivity*, 8th ed.; Cengage: Belmont, CA, 2012.

(106) Vining, W. J.; Young, S. M.; Day, R.; Botch, B. *General Chemistry*; Cengage: Boston, MA, 2016.

(107) SPSS. IBM Corp.: Armonk, NY 2017.

(108) JMP. SAS Institute, Inc.: Cary, NC 2014.

(109) Atlas.ti. ATLAS.ti Scientific Software Development 2017.

(110) Piazza Information for Instructors https://piazza.com/ (accessed Jun 23, 2018).

(111) Knoblich, G.; Ohlsson, S.; Raney, G. E. An Eye Movement Study of Insight Problem Solving. *Mem. Cognit.* **2001**, *29* (7), 1000–1009 DOI: 10.3758/BF03195762.

(112) Green, H. J.; Lemaire, P.; Dufau, S. Eye Movement Correlates of Younger and

Older Adults' Strategies for Complex Addition. *Acta Psychol. (Amst).* **2007**, *125*, 257–278 DOI: 10.1016/j.actpsy.2006.08.001.

(113) Suppes, P.; Cohen, M.; Laddaga, R.; Anliker, J.; Floyd, R. A Procedural Theory of Eye Movements in Doing Arithmetic. *J. Math. Psychol.* **1983**, *27* (4), 341–369 DOI: 10.1016/0022-2496(83)90033-0.

(114) Grant, E. R.; Spivey, M. J. Eye Movements and Problem Solving. *Psychol. Sci.* **2003**, *14* (5), 462–466 DOI: 10.1111/1467-9280.02454.

(115) Topczewski, J. J.; Topczewski, A. M.; Tang, H.; Kendhammer, L. K.; Pienta, N. J. NMR Spectra through the Eyes of a Student: Eye Tracking Applied to NMR Items. *J. Chem. Educ.* **2017**, *94* (1) DOI: 10.1021/acs.jchemed.6b00528.

(116) Zohar, A.; Tamir, P. Assessing Students' Difficulties in Causal Reasoning in Biology—a Diagnostic Instrument. *J. Biol. Educ.* **1991**, *25* (4), 302–307 DOI: 10.1080/00219266.1991.9655229.

(117) van Merrienboer, J. J. G.; Sweller, J. Cognitive Load Theory in Health Professional Education: Design Principles and Strategies. *Med. Educ.* **2010**, *44* (1), 85–93 DOI: 10.1111/j.1365-2923.2009.03498.x.

(118) O'Day, D. H. The Value of Animations in Biology Teaching: A Study of Long-Term Memory Retention. *CBE Life Sci. Educ.* **2007**, *6* (3), 217–223 DOI: 10.1187/cbe.07-01-0002.

(119) Tversky, B.; Morrison, J. B.; Betrancourt, M. Animation: Can It Facilitate? *Int. J. Human-Computer Stud.* **2002**, *57*, 247 DOI: 10.1006/ijhc.2002.1017.

(120) Pollock, E.; Chandler, P.; Sweller, J. Assimilating Complex Information. *Fac. Educ. - Pap.* **2002**.

(121) Ruiz, J. G.; Cook, D. A.; Levinson, A. J. Computer Animations in Medical Education: A Critical Literature Review. *Med. Educ.* **2009**, *43* (9), 838–846 DOI: 10.1111/j.1365-2923.2009.03429.x.

(122) Tunuguntla, R.; Rodriguez, O.; Ruiz, J. G.; Qadri, S. S.; Mintzer, M. J.; Van Zuilen, M. H.; Roos, B. A. Computer-Based Animations and Static Graphics as Medical Student Aids in Learning Home Safety Assessment: A Randomized Controlled Trial. *Med. Teach.* **2008**, *30* (8), 815–817 DOI: 10.1080/01421590802263508.

(123) Tang, H.; Topczewski, J. J.; Topczewski, A. M.; Pienta, N. J. Permutation Test for Groups of Scanpaths Using Normalized Levenshtein Distances and Application in NMR Questions. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*; ACM Press: New York, New York, USA, 2012; p 169.

(124) Eivazi, S.; Bednarik, R. Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data. In *Proc. Workshop on Eye Gaze in Intelligent Human*; 2011; p 9.

(125) Just, M. A.; Carpenter, P. A. A Theory of Reading: From Eye Fixations to Comprehension. *Psychol. Rev.* **1980**, *87* (4), 329–354.

(126) Anderson, J. R.; Bothell, D.; Douglass, S. Eye Movements Do Not Reflect Retrieval Processes: Limits of the Eye-Mind Hypothesis. *Psychol. Sci.* **2004**, *15* (4), 225–231 DOI: 10.1111/j.0956-7976.2004.00656.x.

(127) Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn. Sci.* **1988**, *12* (2), 257–285 DOI: 10.1207/s15516709cog1202_4.

(128) Kozma, R. B.; Russell, J. Multimedia and Understanding: Expert and Novice

Responses to Different Representations of Chemical Phenomena. *J. Res. Sci. Teach.* **1997**, *34* (9), 949–968.

(129) Mayer, R. E. *The Cambridge Handbook of Multimedia Learning*; Cambridge University Press: New York, NY, 2005.

(130) Tobii. *Tobii Studio 2.X User Manual*; Stockholm, Sweden, 2011.

(131) Kruger, J.; Hefer, E.; Matthew, G. Attention Distribution and Cognitive Load in a Subtitled Academic Lecture: L1 vs L2. *J. Eye Mov. Res.* **2014**, *7* (5), 4.

(132) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria 2015.

(133) Buchanan, M. F.; Carter, W. C.; Cowgill, L. M.; Hurley, D. J.; Lewis, S. J.; MacLeod, J. N.; Robertson, T. P. Using 3D Animations to Teach Intracellular Signal Transduction Mechanisms: Taking the Arrows out of Cells. *J. Vet. Med. Educ.* **2005**, *32* (1), 72–78.

(134) Tang, H.; Day, E.; Atkinson, M. B.; Pienta, N. J. GrpString: An R Package for Analysis of Groups of Strings. *R J.* **2018**.

(135) Nurrenbern, S. C.; Pickering, M. Concept Learning versus Problem Solving: Is There a Difference? *J. Chem. Educ.* **1987**, *64* (6), 508 DOI: 10.1021/ed064p508.

(136) Sawrey, B. A. Concept Learning versus Problem Solving: Revisited. *J. Chem. Educ.* **1990**, *67* (3), 253 DOI: 10.1021/ed067p253.

(137) Pickering, M. Further Studies on Concept Learning versus Problem Solving. *J. Chem. Educ.* **1990**, *67* (3), 254 DOI: 10.1021/ed067p254.

(138) Nakhleh, M. B. Are Our Students Conceptual Thinkers or Algorithmic Problem Solvers? Identifying Conceptual Students in General Chemistry. *J. Chem. Educ.*

1993, *70* (1), 52 DOI: 10.1021/ed070p52.

(139) Nakhleh, M. B.; Mitchell, R. C. Concept Learning versus Problem Solving: There Is a Difference. *J. Chem. Educ.* **1993**, *70* (3), 190 DOI: 10.1021/ed070p190.

(140) Atkinson, R. K.; Derry, S. J.; Renkl, A.; Wortham, D. Learning from Examples: Instructional Principles from the Worked Examples Research. *Rev. Educ. Res.* **2000**, *70* (2), 181–214 DOI: 10.3102/00346543070002181.

(141) Renkl, A.; Atkinson, R. K. Interactive Learning Environments: Contemporary Issues and Trends. An Introduction to the Special Issue. *Educ Psychol Rev* **2007**, *19*, 235–238 DOI: 10.1007/s10648-007-9052-5.

(142) Barbera, J. A Psychometric Analysis of the Chemical Concepts Inventory. *J. Chem. Educ.* **2013**, *90* (5), 546–553 DOI: 10.1021/ed3004353.

(143) Towns, M. H. Guide To Developing High-Quality, Reliable, and Valid Multiple-Choice Assessments. *J. Chem. Educ.* **2014**, *91* (9), 1426–1431 DOI: 10.1021/ed500076x.

(144) Ebert-May, D.; Derting, T. L.; Hodder, J.; Momsen, J. L.; Long, T. M.; Jardeleza, S. E. What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *Bioscience* **2011**, *61* (7), 550–558 DOI: 10.1525/bio.2011.61.7.9.

(145) Casselman, B. L.; Atwood, C. H. Improving General Chemistry Course Performance through Online Homework-Based Metacognitive Training. *J. Chem. Educ.* **2017**, *94* (12), 1811–1821 DOI: 10.1021/acs.jchemed.7b00298.

(146) McMurray, J. E.; Fay, R. C. *General Chemistry: Atoms First*, 2nd ed.; Pearson: Boston, MA, 2010.

APPENDIX A: LONGITUDINAL SURVEY DATA

Table A1: Initial survey items (FALL 2015) for the print textbook

Survey Item	*Response (Scale)*	*Evaluating*
"To study for chemistry, how many hours per week do you use the following resources:"	Average Hours Used Per Week (0-168 hours)	Textbook PowerPoint Online Homework (WebAssign) Google Solutions manual eLC Study group/Learning Community Paid, private tutor Friends YouTube Videos Professor's Office Hours TA's office hours University-sponsored free tutoring Professor's Lecture Videos Lectures Laboratory
"Were you unaware of any of these resources?"	Multiple select	Solutions manual Paid, private tutoring YouTube Videos Professor's Office Hours TA's Office Hours University-sponsored free tutoring Professor's lecture videos
"How many hours per week do you spend on the following activities:"	Average Hours Per Week (0-168 hours)	In class (all courses, not just chemistry) Studying outside of class (all courses) Studying outside of class (for chemistry) Using WebAssign to complete homework Using WebAssign to complete progress checks (quizzes) Working at a job Extracurricular activities (sports/music practice or performances) Entertainment "Living" (eating, sleeping, personal hygiene, laundry, going to/from activities)

"How helpful are the following textbook features?"	Likert Scale: 0 = I don't use/NA 1 = very unhelpful 2 = unhelpful 3 = helpful 4 = very unhelpful	Chapter outline Chapter goals Written text Images/visualizations/photographs/figures Examples of real world applications or situations In-chapter, worked-examples problems End-of-chapter problems Animations or simulations Solutions manual Strategy Maps Problem solving tips Review & Check question for each section Definitions/glossary WebAssign homework problems WebAssign tutorials WebAssign progress checks "Key Experiment" sections
"Rank these features of WebAssign (1 = most helpful):"	4 possible positions	Tutorials Practice another problem Detailed answers to homework problems Progress checks
"Rank these features of your textbook (1 = most helpful):"	14 positions	Chapter outline Chapter goals Text Pictures/figures Molecular visualizations/representations Strategy map Problem solving tips Real-World Application Worked-Out example problems in the chapter Review and Check questions for each section End-of-Chapter problems Appendices Answers to Odd Problems in the back of the book Solutions manual
"Which version would you prefer?"		Print copy of textbook Electronic copy of textbook
"Please explain your preference:"	Free response	
"Which resources do you believe helped you earn the grade you received?"	Multiple select	Textbook PowerPoint Online Homework (WebAssign) Google Solutions manual eLC

		Study group/Learning Community Paid, private tutor Friends YouTube Videos Professor's Office Hours TA's office hours University-sponsored free tutoring Professor's Lecture Videos Lectures Laboratory
"Which resources do you believed helped you understand the content presented in the course?"	Multiple select	Textbook PowerPoint Online Homework (WebAssign) Google Solutions manual eLC Study group/Learning Community Paid, private tutor Friends YouTube Videos Professor's Office Hours TA's office hours University-sponsored free tutoring Professor's Lecture Videos Lectures Laboratory
"Which goal was more important to you:"		Earning a good grade Understanding the course content
"How do you communicate with your friends/lab partners/fellow classmates about chemistry?"	Multiple select	Facebook Texting UGA email GroupMe Yik Yak Personal email eLC Twitter Reddit Snapchat
"Do you use other methods to communicate about your chemistry course? If so, please describe:"	Free response	

Table A2: Fall 2015 Participants' Rating of Helpfulness of Textbook Features

Feature	*I don't use this*	*Very unhelpful*	*Unhelpful*	*Helpful*	*Very helpful*
Chapter Outline	24.79%	1.88%	9.40%	48.38%	15.38%
Written Text	3.93%	1.88%	7.52%	61.88%	23.93%
Images/Figures	4.44%	1.20%	6.50%	52.31%	34.02%
Worked Examples	3.08%	0.68%	3.08%	30.09%	62.56%
"Key Experiment" Sections	40.34%	1.71%	17.09%	29.74%	10.94%
End-of-Chapter Problems	4.79%	0.68%	3.25%	37.78%	53.16%
Animations	17.78%	1.20%	13.68%	48.03%	18.97%
Definitions	23.25%	1.54%	9.57%	42.39%	23.08%
WebAssign Homework	0.34%	2.74%	8.21%	49.40%	39.15%
WebAssign Tutorials	2.05%	0.51%	2.91%	26.50%	67.69%
WebAssign Quizzes	0.34%	4.79%	11.45%	42.05%	40.85%
Chapter Goals	28.21%	1.54%	17.61%	41.37%	10.77%
Real World Examples	20.85%	3.76%	24.62%	32.82%	17.78%
Solutions Manual	29.91%	0.68%	6.15%	32.65%	30.26%
Strategy Maps	28.55%	0.85%	10.60%	39.49%	20.00%
Problem Solving Tips	17.09%	0.51%	7.86%	46.32%	27.18%
Review & Check Questions	7.86%	0.51%	2.74%	42.05%	46.32%
Rated on a 5-point Likert scale					

Table A3: Spring 2016 Participants' Rating of Helpfulness of Textbook Features

Features	*I don't use this/NA*	*Very unhelpful*	*Unhelpful*	*Helpful*	*Very helpful*
Weekly Reading Assignment Outline	5.38%	5.73%	12.19%	51.97%	24.73%
Written text	10.79%	5.40%	18.35%	52.52%	12.95%
Images/Visualizations Photographs/Figures	4.32%	2.88%	10.43%	57.19%	25.18%
Text Solution to Example Problems	6.79%	2.50%	4.64%	48.57%	37.50%
Mastery Questions	22.14%	2.86%	12.50%	45.00%	17.50%
Animations or simulations	9.71%	1.80%	14.03%	50.72%	23.74%
Definitions/Glossary	27.86%	3.21%	12.86%	42.50%	13.57%
WebAssign homework problems	0.36%	1.43%	4.29%	48.57%	45.36%

WebAssign tutorials	7.53%	2.87%	8.60%	39.43%	41.58%
WebAssign progress checks	0.36%	10.00%	17.50%	48.21%	23.93%
Video Solution to Example Problems	13.93%	2.86%	7.14%	43.21%	32.86%
Interactive Figures & Review Questions	5.38%	2.15%	7.89%	58.78%	25.81%
End of Chapter Questions (Review or Challenge Problems)	23.93%	2.50%	10.00%	46.43%	17.14%
Tables	13.97%	2.94%	10.66%	55.15%	17.28%
Create Notecards	34.64%	3.57%	16.43%	33.21%	12.14%
Supplemental Information	23.93%	1.79%	11.43%	50.71%	12.14%
Highlighting Function	56.43%	3.57%	15.36%	17.50%	7.14%
Search Function	46.43%	2.50%	11.43%	26.43%	13.21%
Study Center	54.71%	2.54%	11.23%	23.91%	7.61%
QuizMe (Custom Quiz Function)	66.31%	3.58%	8.96%	16.49%	4.66%
Study Guide (Create Custom Study Guide)	53.41%	2.87%	8.24%	25.81%	9.68%
Tutor Solutions to Mastery Questions	34.29%	1.43%	6.43%	36.43%	21.43%
Activities	40.79%	3.61%	11.19%	35.38%	9.03%

Table A4: Fall 2016 Participants' Rating of Helpfulness of Textbook Features (N =769)

Features	*I don't use this/NA*	*Very unhelpful*	*Unhelpful*	*Helpful*	*Very helpful*
Weekly Reading Assignment Outline	9.36%	1.30%	8.84%	49.67%	30.56%
Written text	7.02%	1.82%	10.14%	56.96%	23.28%
Images/Visualizations Photographs/Figures	2.73%	1.95%	6.24%	50.85%	37.84%
Text Solution to Example Problems	2.34%	1.30%	3.77%	38.36%	54.10%
Video Solution to Example Problems	16.12%	1.56%	6.24%	34.33%	41.87%
Animations or simulations	6.37%	2.60%	15.34%	48.11%	27.05%
Definitions/Glossary	23.67%	1.69%	14.56%	41.09%	18.86%
Mastery Questions	11.83%	1.17%	6.11%	37.19%	43.56%
WebAssign tutorials	3.38%	11.18%	39.01%	28.61%	17.56%
WebAssign progress checks	0.13%	8.45%	15.47%	48.11%	27.83%
Video Solution to Example Problems	16.12%	1.56%	6.24%	34.33%	41.87%

Interactive Figures & Review Questions	3.90%	1.04%	8.32%	49.67%	36.28%
End of Chapter Questions (Review or Challenge Problems)	11.44%	0.65%	5.59%	35.11%	47.07%
Tables	10.53%	1.43%	12.74%	57.35%	16.91%
Create Notecards	39.79%	2.86%	14.69%	26.79%	15.73%
Supplemental Information	24.58%	1.95%	12.22%	46.16%	15.08%
Highlighting Function	51.11%	3.51%	16.38%	20.29%	8.58%
Search Function	43.17%	1.30%	8.97%	32.64%	14.04%
Study Center	49.93%	1.82%	8.84%	26.01%	13.39%
QuizMe (Custom Quiz Function)	57.74%	1.30%	9.10%	9.10%	19.90%
Study Guide (Create Custom Study Guide)	47.72%	1.69%	7.02%	27.96%	15.60%
Tutor Solutions to Mastery Questions	23.28%	0.52%	6.11%	32.90%	37.06%
Activities	31.08%	0.91%	8.84%	44.60%	14.56%

APPENDIX B: ANIMATION EYE-TRACKING

Which component of the G-protein complex interacts with phospholipase C?

a. Alpha subunit

b. Beta subunit

c. GDP

d. Gamma subunit

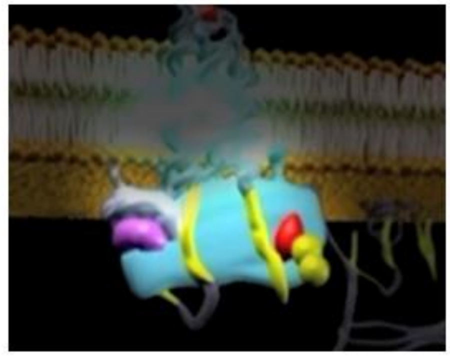


Figure B 1: an example of the multiple choice (recall) items on the post-test exam for the alpha-receptor pathway; the correct answer choice is highlighted.

Your doctor prescribes a drug that blocks the binding sites for IP_3 molecules. Which of the following steps in the pathway would **NOT** occur in response to this treatment? *(for this question, there may be more than one correct answer, select all that apply)*

a. Activation of phospholipase C

b. Release of calcium from the endoplasmic reticulum.

c. Activation of calmodulin molecules.

d. Separation of the beta-gamma subunits from the receptor.

Figure B 2: an example of the multiple select (conceptual reasoning) items on the post-test exam for the alpha-receptor pathway; the correct answer choices are highlighted, both were required for full credit.

Table B1: summary statistics of post-test scores for both experimental conditions

Visualization Type	*Mode*	*Median*	*Mean*	*St. Dev.*
Dynamic	100	78.95	72.81	21.43
Static	78.95	78.95	71.60	16.73

APPENDIX C: WORKED EXAMPLES EYE-TRACKING

Table C1: Pre-test items administered as clicker questions; adapted from the Kotz textbook.[105]

Conceptual Reaction Order:

Which of the following will confirm that the decomposition of $N_2O_5(g)$ to $NO_2(g)$ and $O_2(g)$ is a first-order process?

(a) A graph of $[N_2O_5]$ vs. time gives a positive slope.
(b) A graph of $\ln[N_2O_5]$ vs. time gives a positive slope.
(c) A graph of $1/[N_2O_5]$ vs. time gives a positive slope.
(d) A graph of $\ln[N_2O_5]$ vs. time gives a negative slope.
(e) A graph of $[N_2O_5]$ vs. time gives a negative slope.
(f) A graph of $1/[N_2O_5]$ vs. time gives a negative slope.

Algorithmic Reaction Order:

Given that:
$2NO + 2H_2 \rightarrow N_2 + 2\,H_2O$

Experiment	[NO]	$[H_2]$	Initial Reaction Rate at 300K
1	0.352	0.329	0.0554
2	0.704	0.329	0.222
3	0.352	0.658	0.111

What is the reaction order with respect to H_2?
(a) First-order
(b) Second-order
(c) Zero-order
(d) Pseudo-first order

Conceptual Half-Life:

The decomposition of sucrose to fructose and glucose in acid solution is a first-order process. How many half-lives would be required to decompose 85% of the sample?

(a) 5
(b) Between 4 and 5
(c) 3
(d) Between 3 and 4

Algorithmic Half-Life:

Radioactive iodine-125, which has a first-order rate constant for decay of $0.011d^{-1}$, is used for studies of thyroid function. If you are given a dose containing 2.76×10^{15} atoms, how many atoms remain after 72 hours?

(a) 1.25×10^{15} atoms
(b) 2.67×10^{15} atoms
(c) 2.71×10^{15} atoms
(d) 3×10^{15} atoms

Table C2: average viewing time for each example in each of the possible viewing orders

Worked Example Type	*AOI Name*	*AOI Type*	*Percent of Screen (%)*
Algorithmic	Problem_Text	Text	2.34
	Alg_Step1	Text	2.06
	Alg_Equil	Math/Image	3.17
	Alg_Step2	Text	0.89
	Alg_ICE	Math/Image	3.12
	Alg_Step3	Text	1.50
	Alg_Quad	Math/Image	3.79
	Alg_Step4	Text	0.73
	Alg_pH	Math/Image	1.07
Conceptual	Problem_Text	Text	5.89
	Conceptual_Image	Math/Image	18.41
	Problem_Text2	Text	2.32
	Solution_Text	Text	8.42
	Solution_Image	Math/Image	0.69
	Solution_Text2	Text	5.69
Integrated	Problem_Text	Text	2.62
	Int_Text1	Text	8.85
	Int_Step1	Text	1.20
	Int_Equil	Math/Image	3.00
	Int_Text2	Text	5.83
	Int_Step2	Text	0.53
	Int_ICE	Math/Image	3.77
	Int_Text3	Text	6.86
	Int_Step3	Text	2.11
	Int_Quad	Math/Image	4.29
	Int_Text4	Text	2.67
	Int_Step4	Text	0.52
	Int_pH	Math/Image	1.58

Problem:
What is the pH of a 0.030M solution of acetic acid? The K_a of acetic acid is 1.8×10^{-5}.

Solution:

Step 1: Write the equilibrium and acid dissociation expression.

$$CH_3CO_2H(aq) \leftrightarrow CH_3CO_2^-(aq) + H^+(aq)$$

$$K_a = \frac{[H^+][CH_3CO_2^-]}{[CH_3CO_2H]}$$

Step 2: Organize an ICE table.

	CH_3CO_2H	H_2O	$CH_3CO_2^-$	H^+
Initial	0.030		0	0
Change	-x		+x	+x
Equilibrium	0.030-x		x	X

Step 3: Plug the equilibrium unknowns into the acid dissociation expression and solve for x.

$$K_a = \frac{(x)(x)}{0.030 - x} \cong \frac{x^2}{0.030}$$

$$x = \sqrt{K_a \times [CH_3CO_2H]} = \sqrt{1.8 \times 10^{-5}(0.030)}$$

$$x = [H^+] = 5.4 \times 10^{-7}\text{ M}$$

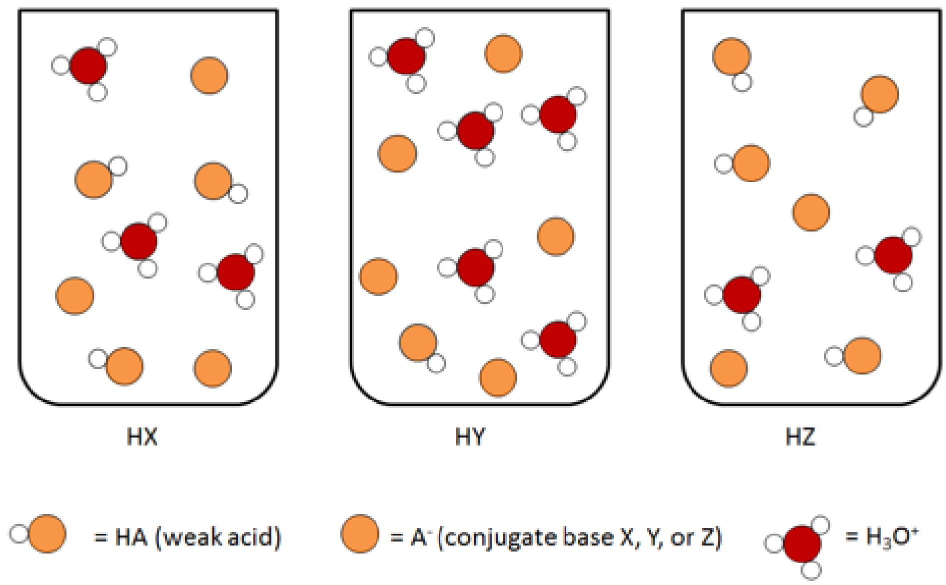
Step 4: Convert $[H_3O^+]$ to pH.

$$pH = -\log[H_3O^+] = -\log(5.4 \times 10^{-7}) = 6.27$$

Figure C 1: the algorithmic worked-out example shown on the Tobii eye tracker.

Problem:

The following pictures represent aqueous solutions of three acids HA (in which the conjugate base is X^-, Y^-, or Z^-); water molecules have been omitted for clarity:



Assuming [HX] = [HY] = [HZ], arrange the three acids in order of increasing acid strength. Which acid, if any, is a strong acid? What is the percent dissociation in solution HX?

Solution:

A strong acid dissociates (ionizes) completely in water. Most acids only dissociate weakly, with the resultant ions in equilibrium with the weak acid. This equilibrium can be expressed as an acid dissociation constant K_a. The acid dissociation constant K_a quantifies the amount of ions formed from the dissociation of a weak acid in water.

$$K_a = \frac{[H_3O^+][A^-]}{[HA]}$$

As the K_a increases, the amount of ions produced increases. A strong acid would have a $K_a > 1$ because it ionizes completely. Therefore, ranking the acids in order of increasing acid strength is a matter of ranking them according to increasing K_a value: HZ < HX < HY. Since none of the acids ionize completely, none of these acids would be considered strong acids.

The extent of this dissociation can be quantified as percent dissociation.

$$\%\ dissociation = \frac{[H_3O^+]}{[HA]} \times 100\%$$

The percent dissociation of HX can be calculated by counting the ratio of hydronium ions to acid molecules: $\frac{3}{6} \times 100\% = 50\%$

Figure C 2: the conceptual worked-out example that participants viewed on the eye-tracker. This example's figure, problem text, and problem's solution were adapted from McMurray and Fay.[146]

Problem:
Calculate the pH of a 0.055M solution of benzoic acid ($C_6H_5CO_2H$), given that the K_a is 6.3×10^{-5}.

Solution:
Benzoic acid does not ionize completely in water, and an equilibrium between the weak acid and its resultant ions exists. An appropriate acid dissociation expression can be used to quantify the acid dissociation.

Step 1: Write the equilibrium and acid dissociation expression.

$$C_6H_5CO_2H(aq) \leftrightarrow C_6H_5CO_2^-(aq) + H^+(aq)$$

$$K_a = \frac{[H^+][C_6H_5CO_2^-]}{[C_6H_5CO_2H]}$$

An ICE table shows the initial and equilibrium concentrations and the change based on reaction stoichiometry between them. The molarity of the ions at equilibrium is represented by x, the molarity that dissociated.

Step 2: Organize an ICE table.

	$C_6H_5CO_2H$	H_2O	$C_6H_5CO_2^-$	H^+
Initial	0.055		0	0
Change	-x		+x	+x
Equilibrium	0.055-x		x	x

The molarity of each ion is represented by x because of the stoichiometric ratio between the ions. Because the extent of dissociation x is small compared to the initial molarity of the acid, an approximation can be used.

Step 3: Plug the equilibrium unknowns into the acid dissociation expression and solve for x.

$$K_a = \frac{(x)(x)}{0.055 - x} \cong \frac{x^2}{0.055}$$

$$x = \sqrt{K_a \times [C_6H_5CO_2H]} = \sqrt{6.3 \times 10^{-5}(0.055)}$$

$$x = [H^+] = 1.9 \times 10^{-3}\ \mathrm{M}$$

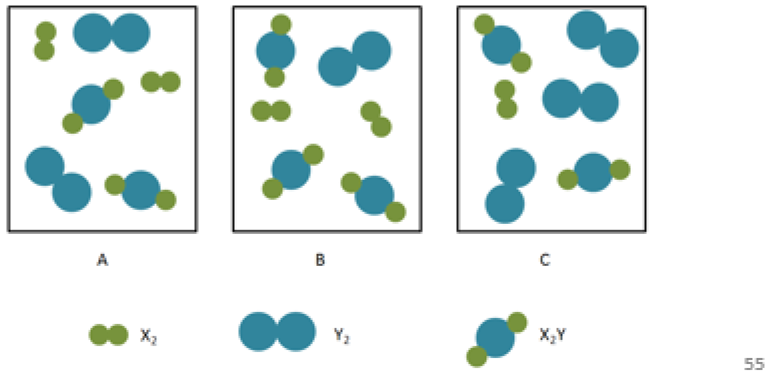
pH is a measure of the hydronium ion concentration in a solution.

Step 4: Convert $[H_3O^+]$ to pH.

$$pH = -\log[H_3O^+] = -\log(1.9 \times 10^{-3}) = 2.72$$

Figure C 3: the integrated worked-out example viewed on the Tobii eye tracker.

1. In the reaction $A(g) + 2B(g) \rightleftharpoons C(g)$, 0.041 moles of A, 0.052 moles of B, and 0.170 moles of C are introduced into a 5.00L reaction container at 600K. At this temperature, the K_c for the reaction is 3.7×10^3. Is the reaction mixture at equilibrium? If not, what is the direction of the net reaction?
2. The reaction $2X_2 + Y_2 \rightleftharpoons 2X_2Y$ has an equilibrium constant of 1. Which reaction mixture is at equilibrium, if any? For reaction mixture C, will the net reaction proceed in the forward or reverse direction?



3. Calculate the equilibrium pressures of N_2O_4 and NO_2 in the reaction $2NO_2(g) \rightleftharpoons N_2O_4(g)$ at 298K ($K_c = 215.5$) if 2.52 g of NO_2 gas is placed in the 6.00L flask initially.
4. For the reaction $2A(g) \rightleftharpoons A_2(g)$, the equilibrium state is shown on the left. In the box on the right, draw the result of increasing the volume of the container.

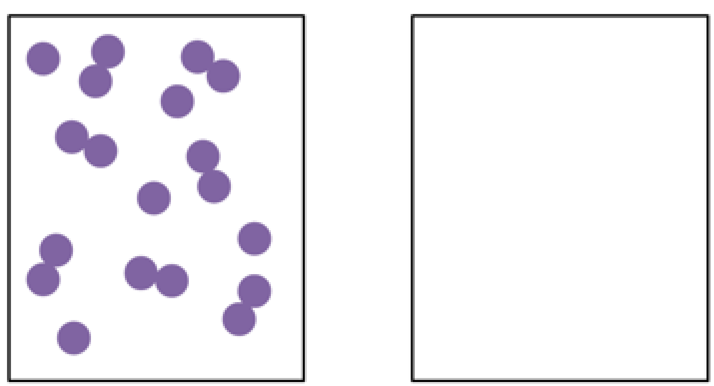


Figure C 4: The post-test items that participants solved during a semi-structured concurrent think-aloud interview. The second item was adapted from McMurry and Fay and the fourth item was inspired by a problem in McMurray and Fay.[146]

Sample code for correlation between eye-tracking metrics on each example:

```
setwd("C:/Users/elday/Documents/UGA/Chem.
Ed/Projects/Conceptual-Algorithmic Eye-
Tracking/Stats")
f.df <- read.table("Eye Tracking Metrics
integrated2.txt", header=TRUE, stringsAsFactors=FALSE)
dim(f.df)
[1] 26 81

duration_count_click.df = f.df[,10:29]
head(duration_count_click.df, 3)
duration_count_click.mx =
as.matrix(duration_count_click.df)
#converted to matrix
install.packages("Hmisc", repos='http://cran.us.r-
project.org')
library(Hmisc)
# can also use corr.test function in psych package
duration_count_click.cor =
rcorr(duration_count_click.mx)
class(duration_count_click.cor)
[1] "rcorr"
# [1] "rcorr"
# output cor, size, and p value, respectively
write.csv(duration_count_click.cor[1],
"integrated2_FL_FC.cor.csv")
write.csv(duration_count_click.cor$n,
"integrated2_FL_FC.cor_n.csv")
write.csv(duration_count_click.cor$P,
"integrated2_FL_FC.cor_p.csv")
```