The Role of Feedback in the Transfer of Category Learning

By

ADDISON LEIGH POAGE BABINEAU

Bachelor of Science, 2018
Colorado State University
Fort Collins, Colorado

A Thesis Defense Document
Submitted to the Graduate Faculty of the
College of Science and Engineering
Texas Christian University
in partial fulfillment of the requirements
for the degree of

Master of Science in Experimental Psychology

May 2020

ACKNOWLEDGMENTS

I would like to begin by thanking my mentor, Dr. Sarah "Uma" Tauber. Her expertise and guidance has been paramount to my success here at TCU. Dr. Tauber has encouraged me to excel at every turn and has helped me have a positive graduate school experience. I have learned so much from her in my time here at TCU, and I am eager to learn much more from her in the years to come. I would also like to thank Dr. Mary Hargis and Dr. Kenneth Leising for serving on my thesis committee. Their feedback was essential to my thesis and I am grateful for their support.

Next, I would like to thank my fellow graduate students. Dr. Amber Witherby, Vishal Thakkar, and Paige Northern have been incredibly supportive as I completed my thesis. They brought laughter and friendship into my office each day, and provided insightful advice as I worked on my thesis. I would also like the thank Julie Swets and Juliana Oliveira for lightening my life I was worked through this project. Their friendship has been critical my experience here at TCU and I am grateful to have found lifelong friends during my time here.

Finally, I would like to thank my family. I appreciate the love and support they provide me every day and I would not be able to accomplish my goals in life without them. They always know how to cheer me up after a hard day. Although they are in Colorado, I know they are here with me in spirit, cheering on my every accomplishment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

The Role of Feedback in the Transfer of Category Learning

Through our day-to-day lives we are exposed to a vast amount of information, and a variety of important tasks. Organizing this information into categories can simplify how we interact in our environment by providing structure to the world around us. A *category* is knowledge about a set of items. Categories are made up of a multitude of *exemplars*, which are individual examples of that category. Categories act as the building blocks of a concept; a *concept* is a mental representation of something (such as objects, ideas, or even people). As an example, most adults know what a tree is (i.e., concept representation), can determine the appropriate behavior if faced with a tree in their day-to-day lives (i.e., category information), and may be able to identify which plants are trees (relative to other categories such as bushes, flowers, or weeds), or even the type of tree (i.e., exemplar knowledge).

Even though categorizing plants as trees or other vegetation may seem elementary, concepts and categories exists across all aspects of life. Detectives categorize types of fingerprints (Searston & Tangen, 2017), bird watchers categorize kinds of birds (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013), and chemists categorize chemicals (Eglington & Kang, 2017), all of which are examples of real-world categories. Learning categories can be challenging, and researchers have recently begun investigating how to improve category learning. Even so, little is known about the degree to which feedback enhances categorical learning. As well, no researchers to-date have investigated whether the form of feedback during learning is critical for applying categorical knowledge on transfer tests. The primary goal of the present research is to explore these issues. To begin, I will briefly discuss the history of the concept and category literature in comparative research with nonhuman animals as well as with humans. I will then review current research on category learning and the roll that feedback may have on improving this skill. Finally, I will discuss the present research that evaluated if feedback could benefit category learning on transfer tests (as well as studied exemplar tests), and if so, which type of feedback would support the largest improvement.

Concept Formation in Nonhuman Animals

Research with nonhuman animals has been foundational to researchers investigating how humans form and represent concepts. Comparative research has established that concept formation is not unique to humans, which is important for understanding categorical learning processes in nonhuman animals and humans alike. Work conducted with nonhuman animals has considered four different forms of concept learning, which are: perceptual concept learning, associative concept learning, relational concept learning, and analogical reasoning (for a review, see Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008).

Perceptual concept learning involves learning to sort stimuli into different classes by their perceptual similarities. An example is categorizing stimuli as cats, chairs, or cars (e.g. Aust & Huber, 2002). Humans and nonhuman animals can this do by using discrimination and generalization to identify physical difference and similarities between stimuli (Bhatt, Wasserman, Reynolds, & Knauss, 1988).

Associative concept learning involves learning to associate a category of stimuli with an equivalent. For instance, a person might learn the word (e.g., airplane) that corresponds with an object (i.e., an image of an airplane) as functional equivalents. For nonhuman animals, a matching-to-sample paradigm is often used during which animals might learn to associate a red light with a correct response of a vertical line, and a green light with a correct response of a horizontal line. Nonhuman animals have the capability to learn associative concepts, although it appears to be more limited than humans.

Relational concept learning involves learning the relationships between stimuli and determining whether they are the same or different from one and other. As an example, nonhuman animals might be given an array of icons that are all the same or different and then select the correct response that matches the relation (same or different) of the sample (i.e., a relational matching-to-sample task; e.g., Fagot, Wasserman, & Young, 2001). Nonhuman animals can achieve this type of learning under the proper circumstances.

The fourth type of concept learning is *analogical reasoning*, which involves transfer of a relationship between stimuli to another set of unrelated stimuli. An example would be knowing that the functional relationship between food and humans is the same as the relationship between fuel and a car. There is little less evidence of analogical reasoning in nonhuman animals, although some apes have demonstrated this skill.

How are Concepts Mentally Represented?

Research with nonhuman animals has increased how much we know about categorization abilities generally, and factors that can influence it. Cognitive psychologists have also focused on developing theory of how concepts are mentally represented. Three prominent kind of theories have emerged to explain how people form and represent concepts: rule-based theories, exemplar-based theories, and prototype theories (for a review, see Goldstone, Kersten, & Carvalho, 2018).

Rule-based theories suggest that people use rules to distinguish what belongs in a category from what does not (Anderson, Klein, & Beasley, 1979; Bruner, Goodnow, & Austin, 1956; Smith, Langston, & Nisbett, 1992). The leading ideology is the hypothesis-testing approach, which suggests that as people learn concepts, they form hypotheses about what distinguishes a category from others, and then they test those hypotheses as they learn (Bruner et al., 1956). As an example, if you were trying to learn how to categorize shapes (e.g., squares, circles), you might make the hypothesis that squares have four sides of equal length. Using this rule would lead you to the correctly categorize most exemplars. However, when faced with a parallelogram you would need to amend the rule to include four sides of equal length and four right angles (Bourne, 1970). One instrumental factor for the success of rule-based approaches is that each exemplar that belongs to the category must have at least one feature in common with all other exemplars (Medin & Smith, 1984; Smith & Medin, 1981). Given this necessity, using rules to form concepts is sufficient for some stimuli; however, it falls short for many of the concepts we encounter every day. Rule-based theories have been heavily criticized on the basis of three main points. First, exemplars of a category do not always share one common feature and instead only need a family resemblance (Wittgenstein, 1953). Second, not all people agree on which exemplars belong to a category (McCloskey & Glucksberg, 1978). Third, exemplars within a category can be ranked as typical or atypical category members (Rosch & Mervis, 1975). Thus, some exemplars can appear to be better examples of a category relative to others. If people always used rules to form concepts, then exemplars should not differ in typicality.

As opposed to a rule-based approach, prototype approaches to concept formation suggest that people form concepts by building a model of the most representative features of a category (Mervis & Rosch, 1981; Rosch 1975; Rosch & Mervis, 1975). To do so, key features of exemplars are abstracted and aggregated into one prototype, and that prototype is used when trying to identify new members of a category. To illustrate, consider the concept of an airplane. When first learning about airplanes, you might have seen different examples like jets, passenger planes, and biplanes. From these examples, the most characteristic features can be drawn together into a prototype made up of two main wings on either side of a cabin, smaller wings in the back, a cockpit in the front, and wheels to land. When you encounter a new type of object, you would compare it to your airplane prototype to determine if the new object is likely to be an airplane or not. This perspective (including the family resemblance theory proposed by Mervis and Rosch, 1981) can account for some of the short-comings of rule-based approaches. By representing concepts as prototypes, exemplars do not need to have one common feature, but instead only need a general resemblance with each other (Hampton, 1993). Even so, prototype approaches do have limitations. For instance, by abstracting key features together, some smaller details that could by critical for later categorization may be overlooked (e.g., Nosofsky, 1986).

Exemplar-based approaches suggest that instead of representing concepts by forming a prototype, concepts are formed by storing all the members of a category in memory, and then comparing new potential members to the stored exemplars. Consider again the concept of an airplane. When presented with a new object, you might recall every airplane you have ever seen (e.g., jet, passenger plan, biplane), and you would compare the new object to each of those plane exemplars to determine if it is an airplane or not. According to exemplar-based approaches, new exemplars are compared to all previously stored exemplars on a series of attributes and features (context model, Medin & Schaffer, 1978). A new exemplar will be categorized by the exemplar with which it has the most features in common. In the context of the previous example, if you are trying to decide if a machine is an airplane or a car, you would compare the features the machine has to the features of each airplane exemplar and each car exemplar to make your decision. Exemplar-based approaches to concept representation can produce different predictions for categorization than do prototype approaches, and they can often better account for new exceptions and outliers within a category (Posner & Keele, 1968). However, a limitation of exemplar-based theories is that it is not economical for humans to store every exemplar that they have ever encountered.

Each concept formation theory has its strengths (and weaknesses), and which theory most accurately represents concept formation and representation is still debated. As well, methodological variability can contribute to how people make categorization decisions. The type of materials used during concept formation can have a significant impact on how people form their concepts. If the exemplars shown to participants can be defined by a rule, then participants will likely learn them using a rule. If the participants can easily form a prototype, they will likely do so. Thus, there is not only one way to mentally represent concepts. Instead, there may be multiple methods, and the way in which we learn is at least partially dictated by the to-be-learned materials (Goldstone et al., 2018; Weiskopf, 2009). Researchers continue to investigate how concepts are mentally represented; however, they have only recently begun evaluating how to best support concept formation. One recent perspective aimed at improving category learning is the discriminative contrast hypothesis.

Discriminative Contrast Hypothesis

The discriminative contrast hypothesis posits that making comparisons between different exemplars is crucial for successful category learning (e.g., Carvalho & Goldstone, 2017; Kornell & Bjork, 2008). Exemplars can be compared in two ways: between categories or within categories. Between category comparisons involve comparing exemplars from one category to exemplars in a different category. As an example, if you are studying different categories of birds, comparing an exemplar of a finch to an exemplar of a chickadee would be a between category comparison. Comparing exemplars between categories emphasizes how each category is unique. Within category comparisons involve comparing an exemplar to others in the same category, this emphasizes the similarities within a category. In the previous example, comparing a finch to other exemplars of finches would be a within category comparison. The type of comparison (between categories or within category) that is best for category learning is dependent on the structure of the category that is being learned.

When the exemplars that form a category are highly similar, it is best to compare between categories so that the differences can be identified (Carvalho & Goldstone, 2017). To revisit the prior example, if you are studying different categories of birds, and all of them are the same color and size, comparing between the categories of birds would help you notice the difference between the categories and would improve performance on a later test. However, when the exemplars that form a category have few similarities, it is best to focus on the similarities within a category so that the key features of the category can be identified. In the prior example, if you are studying different categories of birds, and every category includes birds of different colors and sizes, it may be difficult to detect the common feature of the category. Thus, comparing exemplars between categories is unlikely to be beneficial (Carvalho & Goldstone, 2017; Zulkiply & Burt, 2013). In these cases, comparing exemplars within each category would assist learners in identifying key features necessary to understand the category. Most of the researchers examining the discriminative contrast hypothesis have used categories that benefit from between category comparisons as opposed to within category comparisons.

One way to facilitate comparisons between categories is to manipulate the order in which categories are studied. In seminal research evaluating the discriminative contrast hypothesis, Kornell and Bjork (2008) had participants learn to categorize paintings by the artist who painted each. Half of the paintings were randomly mixed by artist (i.e., interleaved) and the other half were grouped together by artist (i.e., blocked). The logic is straightforward. Interleaved order likely encourages participants to make comparisons between different artists' paintings, whereas blocked order encourages comparisons within an artists' paintings. Consistent with the discriminative contrast hypothesis, participants correctly categorized more paintings on a transfer test (also called a novel classification test) after learning paintings that were interleaved relative to those that were blocked (Carvalho & Goldstone, 2017; Eglington &Kang, 2017; Zulkiply, McLean, Burt, & Bath, 2012).

The discriminative contrast hypothesis has been investigated across a wide breadth of materials (e.g., Eglington & Kang, 2017; Sakamoto & Love, 2010), and across a variety of presentation formats (Carvalho & Goldstone, 2014). Whereas most researchers have considered discriminative contrast by order during study (interleave vs block), this is not the only way to encourage comparisons between category exemplars. Feedback may be another avenue to incite exemplar comparisons across different categories.

Feedback and Category Learning

Feedback has been investigated across a wide range of learning environments. In comparative psychology, the impact of a reinforcement schedule has been extensively evaluated with a variety of nonhuman animals. Similarly, in research on human learning (including associative learning, semantic tasks, episodic tasks), researchers often give learners feedback during the study trials. In this research, providing feedback typically benefits later performance (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Pashler, Cepeda, Wixted & Rohrer, 2005). Despite the breadth of evidence promoting the use of feedback, it is not clear the degree to which feedback during a category learning task is beneficial for later classification performance. Investigating this novel issue was the primary aim of the present research. Moreover, it is an open question as to which form of feedback is best. This issue was also addressed in the present work.

Two common types of feedback are corrective feedback and elaborative feedback. *Corrective feedback* includes information about the correctness of the answer and the correct answer (e.g., "*Correct! This is an Organochloride.*"). *Elaborative feedback* includes information about the correctness of an answer, the correct answer, and it also includes additional information. Elaborative feedback is a broad term that includes a variety of different types of feedback because the elaborative information often depends on the learning task. For instance, elaborative feedback can include details about the correct answer, an explanation about why an answer is correct, or a passage about the correct answer (Butler, Godbole, & Marsh, 2013).

To examine performance differences between feedback types, Butler et al. (2013) had participants read short passages and take a practice test with corrective feedback, elaborative feedback, or no feedback. Following the study phase, participants took two different tests. One of the tests had the same questions as the practice tests and did not require transfer of knowledge. Participants who received feedback (corrective or elaborative) performed better than did participants who received no feedback; however, there were no performance differences between corrective feedback and elaborative feedback on this test. Butler and colleagues speculated that when participants where familiar with the questions, it was important to know the correct answer, but not any additional information to respond correctly. The second test had questions that the participants had never answered, requiring them to transfer their knowledge from the practice test to a novel question. For this test, participants who received feedback performed better than did participants who did not receive feedback. Of most interest, participants who received elaborative feedback performed better than did participants who received corrective feedback. Elaborative feedback may provide the participants with the additional information necessary to effectively transfer their knowledge to new test questions. In this way, the results of feedback are conditional on the type of final test.

In the context of category learning, a form of elaborative feedback that encourages comparisons between exemplars may best support learning, which is consistent with the discriminative contrast hypothesis. In my research, participants learned to categorize different types of organic chemistry compounds. The chemical categories I used have been found to benefit from between category comparisons (Eglington &Kang, 2017). Thus, elaborative feedback that directs the learner to the differences between categories could aid in later performance on a transfer test with novel exemplars. Similar to the findings of Butler et al. (2013), I expected that performance differences between participants who receive corrective feedback and elaborative feedback would be conditional on the type of exemplars shown at test. When the test consists of studied exemplars, there would be no difference between the two feedback types; but when the test has novel exemplars (and therefore requires transfer to new items) the participants who received elaborative feedback would perform better relative to those who received corrective feedback. I also predicted that participants who received corrective feedback or elaborative feedback would perform better on transfer and non-transfer tests than would participants who did not receive any feedback.

Experiment 1

A key feature of Experiment 1 (and Experiment 2) was to investigate students' categorical learning with educationally-relevant materials. Within our education system, the ability to successfully form concepts is critical. Student achievement hinges on their understanding of a course's concepts and categories, particularly within STEM (Science, Technology, Engineering, & Math) education. The transition to investigating real-world categories from the artificial categories used in most concept representation studies is relatively recent. Using real-world categories from STEM fields improves the ecological validity of the category learning literature. Thus, it seems appropriate that category learning research should be conducted with the materials that students may actually need to learn.

In Experiment 1, participants learned to classify organic chemistry compounds into their respective chemical category. Of most interest, participants were given no feedback, corrective feedback, or elaborative feedback after classifying each exemplar during study. For corrective feedback, participants were informed if their answer was correct and were reminded of the correct chemical category. For elaborative feedback, participants were similarly informed if their answer was correct and were provided with the exemplar's correct category. In addition, elaborative feedback included information about the features of the correct category and the features of the category they incorrectly selected, and participants were asked to determine how the two categories were different from each other. This encouraged participants to compare and contrast the two categories, so I refer to this new form of feedback as *contrast feedback*. Participants took final tests classifying chemical compounds. One test required participants to transfer their knowledge to novel chemical compounds (novel exemplar classification test) and the other test did not require participants to transfer their knowledge of organic chemistry compounds (studied exemplar classification test).

Method

Participants and Design

A power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) using an effect size, d = .85, (from Butler et al., 2013, Experiment 1) with alpha error at .05 and power at .95, estimated that 37 participants would be needed for each group. As such, 112 Texas Christian University undergraduate students participated for partial course credit in psychology courses and were randomly assigned to group. Of these participants, 8 who were randomly assigned to the contrast feedback group (21% of group sample) did not follow directions during feedback. Thus, these 8 participants were excluded from the analyses. To reach the target sample size, additional data were collected by randomly assigning participants to the three feedback groups (no feedback, corrective feedback, contrast feedback), and 8 participants assigned to the contrast feedback group were added to the sample giving a total sample of $N = 112$ ($n = 37$ for the no feedback and contrast feedback groups, $n = 38$ for the corrective feedback group). The sample was college-aged ($M = 18.63$ years, $SD = .91$), most participants identified as female (75%; 85 females, 24 males, and 3 gender non-binary) and Caucasian (75%). The three feedback groups did not differ in age, $F(2,109) = .49$, $p = .612$, $\eta^2_p = .01$, gender identity, $\chi^2(4, N = 112) = 2.56$, $p = .653$, or ethnicity, $\chi^2(14, N = 112) = 12.78$, $p = .544$.

Materials

Stimuli consisted of six chemical categories (Epoxide, Nitrile, Organochloride, Organophosphate, Pyrethroid, and Sulfone). Stimuli were obtained from Eglington and Kang (2017). Each category included 20 exemplars, and each exemplar was a two-dimensional chemical compound in black and white (See Figure 1, top panel). For each category, 12 of the exemplars were randomly assigned to the study phase and the remaining 8 exemplars were assigned to the novel exemplar classification test. There were 120 total exemplars, with 72 exemplars in the study phase and 48 exemplars in the novel test phase.

For each chemical category there was one piece of additional information for the contrast feedback group (See Figure 1, bottom panel). The additional information was the diagnostic feature of the category, meaning that it was necessary for category membership. The additional information was developed from the materials provided by Eglington and Kang (2017).

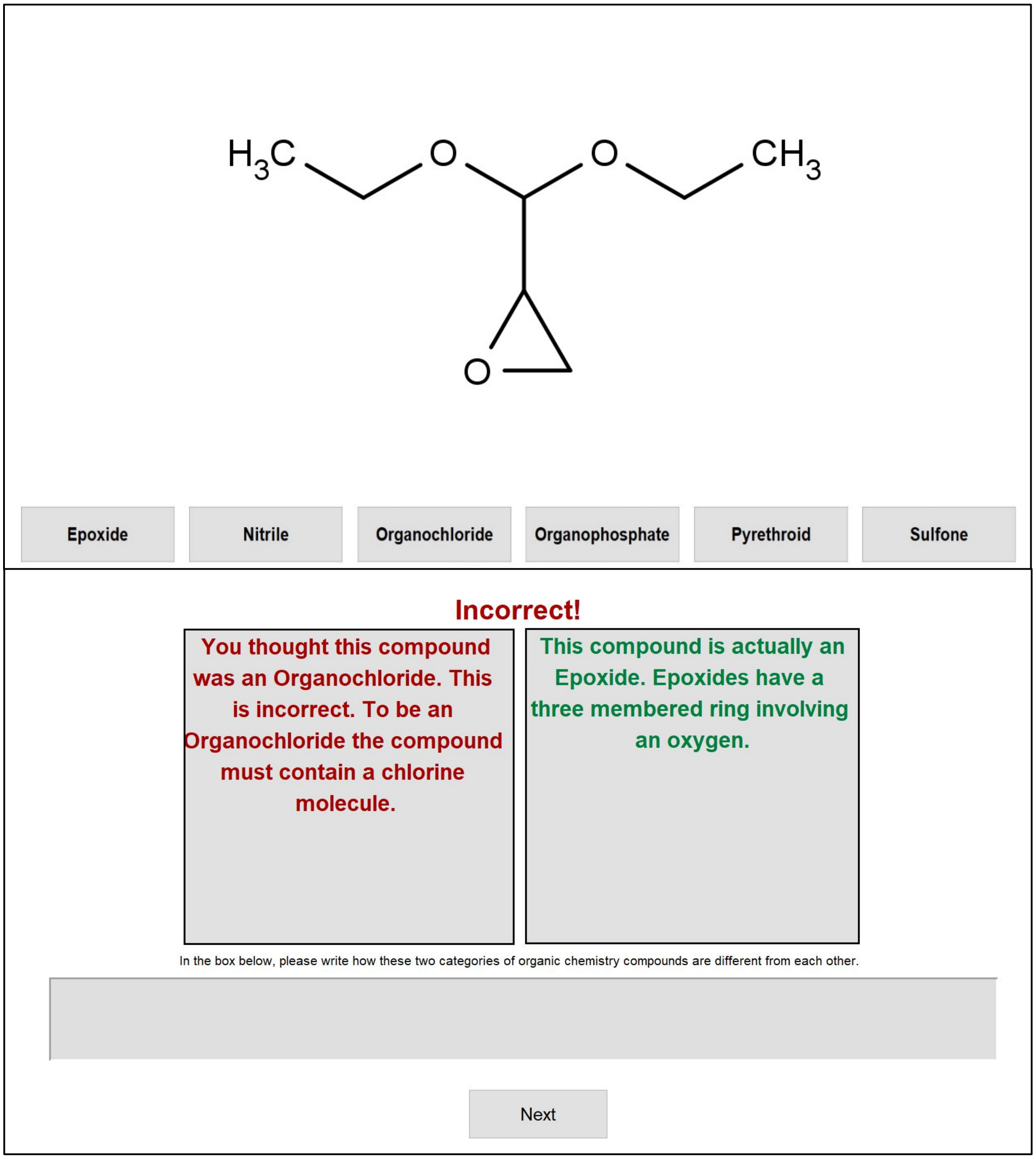

Figure 1. Top panel: Example of the practice phase. Bottom panel: Example of contrast feedback, with the feature for the incorrectly categorized chemical category on the left, and the feature for the correct chemical category on the right.

Procedure

To begin the experiment, participants were asked a brief series of questions probing for their previous knowledge and experience with organic chemistry (See Table 1). Participants reported their major(s) and self-rated their knowledge of organic chemistry. Participants also answered four general chemistry questions. The four general chemistry questions were developed from introductory chemistry principles to determine participants' actual prior knowledge of chemistry.

Next, all participants were given a brief tutorial on how to view and interpret an organic chemistry compound (see Figure 2). The tutorial presented each participant with information on how different molecules are represented in chemical structures, as well as what bonds look like. To ensure that each participant could interpret the stimuli, the tutorial ended with a test that required participants to identify key elements of six trial exemplars. The six trial exemplars were taken from chemical categories that were not included in the actual experiment. During the tutorial, participants were given the option to return to the beginning of the tutorial at any time if they wanted to review the material again.

After completing the tutorial, participants in all groups were instructed to study the exemplars in preparation for a novel classification test. During the initial study phase, all 72 exemplars assigned to the study phase were presented one-at-a-time for 4 s each. The correct chemical category was presented beneath each exemplar. The exemplars were shown in a random order for each participant.

Following the initial study phase, participants began the practice phase during which the 72 study phase exemplars were shown one-at-a-time for study. Underneath each exemplar, there were six buttons each labeled with the name of a chemical category. The buttons were in a single horizontal line and in alphabetical order. Participants were instructed to select the chemical category that correctly corresponded with the exemplar shown on the screen and had as much time as necessary to do so.

After practice classification, participants in the contrast feedback group were directed to a feedback screen. Following an incorrect answer, participants in the contrast feedback group received the statement, “Incorrect!” in red font, as well as two pieces of additional information (see Figure 1, bottom panel). One piece of additional information contained the diagnostic feature for the chemical category the participants thought the compound belonged to; this was shown as the statement in red font, “You thought this compound was a [incorrectly identified chemical category.] This is incorrect. To be a [incorrectly identified chemical category] to compound must have a [diagnostic feature for the incorrectly identified chemical category]”. The other piece of additional information contained the diagnostic feature for the correct chemical category and was shown in green font as the statement, “This compound is actually a [correct chemical category]. To be a [correct chemical category] the compound must have a [diagnostic feature for the correct chemical category]”. To proceed to the next exemplar, participants typed how the two chemical categories were different from each other into the provided text box. This encouraged participants to contrast the two features to better understand how to differentiate the correct chemical category from the incorrect chemical category. Following a correct answer, they were shown the statement, “Correct! This is a [correct chemical category].” in green font. Participants typed the name of the correct chemical category into a text box on the screen to ensure that the participants were interacting with the feedback regardless of correctness or feedback group assignment.

After practice classification, participants in the corrective feedback group were directed to a feedback screen. Following a correct answer, they were shown the statement, "Correct! This is a [correct chemical category]." in green font. If the participant incorrectly identified the exemplar, the statement, "Incorrect! This is a [correct chemical category]." was shown in red font. Regardless of correctness on the practice trial, participants typed the name of the correct chemical category into a text box on the screen before they were allowed to proceed to the next exemplar.

After practice classification, participants in the no feedback group typed the word "next" into a text box to proceed to the next exemplar. This was done to ensure that participants in each feedback group were required to type a response between each practice classification trial.

Upon completing the practice phase, participants proceeded to the test phase. During the novel exemplar portion of the test phase, participants in all groups were presented with the 48 novel exemplars. These exemplars had not been previously studied (thus, they were new), but belonged to the same six chemical categories that participants studied earlier. Each exemplar was presented one-at-a-time with the six chemical categories displayed beneath it. The six chemical category buttons were in the same order as the practice phase. Participants were instructed to select the correct chemical category, and they were given as much time as needed to make their selection. The studied exemplar test followed the exact same structure as the novel exemplar test; however, the exemplars tested were the 72 exemplars that had been previously studied. No feedback was given during any portion of the test phase. Finally, participants were thanked for their time, granted credit, and debriefed.

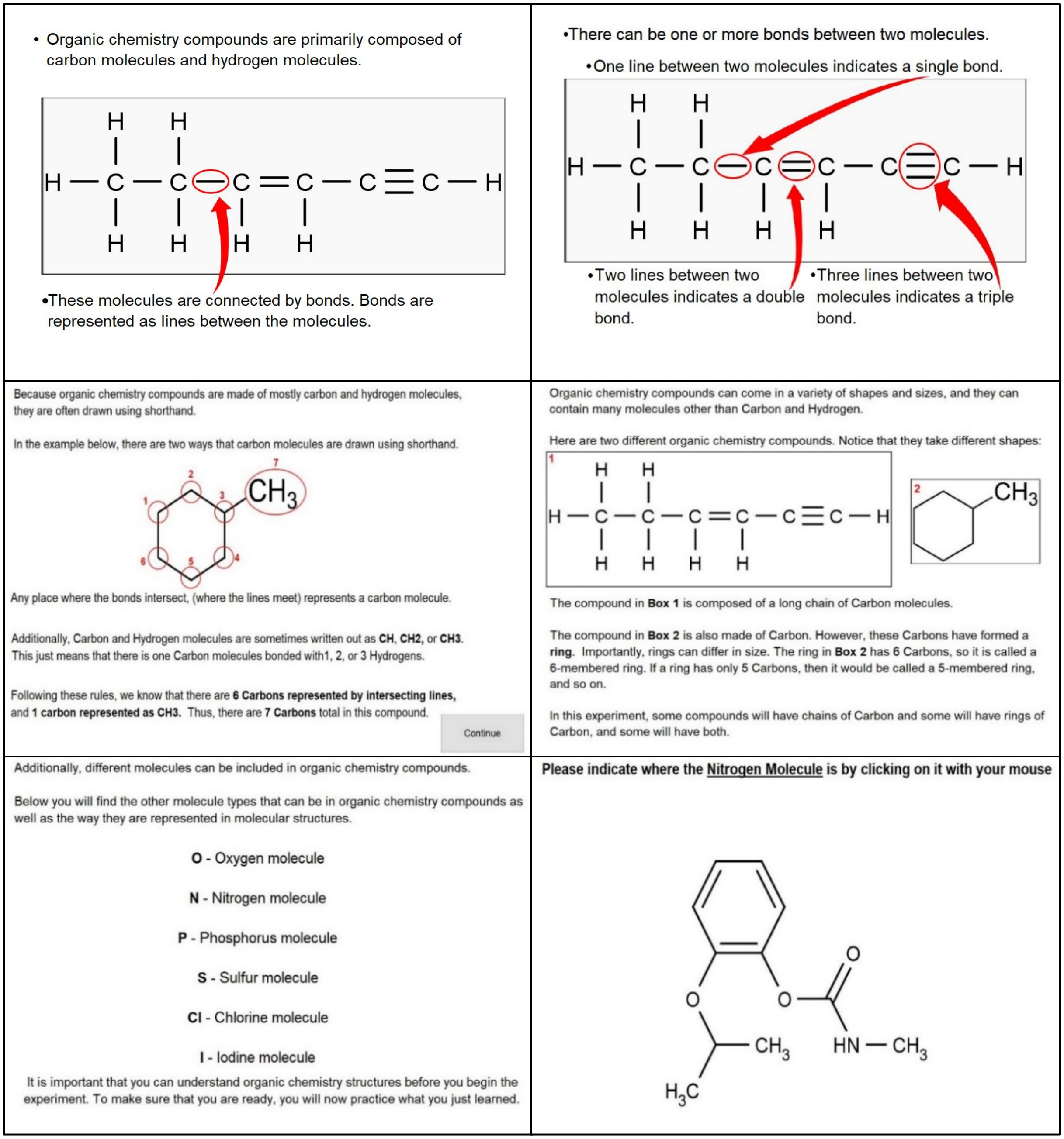

Figure 2. Each screen of the organic chemistry tutorial, which was shown in a fixed order (left to right, and top to bottom). On the final screen, participants clicked on the appropriate chemical component to proceed.

Results and Discussion

Although performance on the novel exemplar test was of most interest, participants' performance on the prior knowledge questions is reported first to establish participants' knowledge of chemistry prior to learning the chemical categories in this experiment. This was followed by analyses of participants' performance on the novel exemplar test. Then, participants' performance on the studied exemplar test is reported. Participants' performance during the practice phase and feedback preference is reported last.

Prior Knowledge Questions

Outcomes for the prior knowledge questions are located in Table 1. The correct answer to each of the four open-ended questions was a single word, so responses were scored on a scale from 0 (omission or commission error) to 1 (correct), and scores for the 4 questions were summed and a proportion of correct answers was calculated for each participants' prior knowledge score. Overall, participants' prior knowledge of organic chemistry was low ($M = .11$, $SE = .02$), and did not differ between the three feedback groups, $F(2,109) = 1.16$, $p = .317$, $\eta^2_p = .02$. Further, participants' self-reported knowledge of chemistry (1 = novice to 7 = expert), was low ($M = 1.54$, $SE = .02$), and did not differ between the three feedback groups, $F(2,98) = 1.01$, $p = .438$, $\eta^2_p = .02$.

Few participants identified as chemistry majors or minors (6 students; 5.4% of sample), or as biochemistry majors or minors (5 students; 4.5% of sample). The distribution of chemistry and biochemistry majors and minors did not differ between feedback groups, χ^2s ≤ 2.94, ps $\geq .230$ (See Table 1). In sum, participants had low prior knowledge of organic chemistry, and few students identified as chemistry and biochemistry majors.

Table 1. Experiment 1 Prior Knowledge Questions	No Feedback	Corrective Feedback	Contrast Feedback
1. Are you currently or have you ever been a Chemistry Major or Minor?*	8.1%	2.6%	5.4%
2. Are you currently or have you ever been a Biochemistry Major or Minor?*	0%	5.3%	8.1%
3. What is your own rated level of expertise at chemistry?	1.69 (.18)	1.59 (.26)	1.35 (0.12)
Prior Knowledge Composite Score**	.09 (.04)	.15 (.04)	.08 (.02)
4. What is the name of the compound with the formula C_6H_{14}?	.05 (.04)	.10 (.05)	.03 (.03)
5. What is the name of the hydrocarbon series that has a double bond between carbon atoms?	.08 (.05)	.08 (.04)	.03 (.03)
6. How many valence electrons does an oxygen molecule have?	.14 (.06)	.37 (.08)	.30 (.08)
7. Benzene rings mainly undergo which type of chemical reaction?	.08 (.05)	.05 (.04)	.00 (.00)

M(*SE*) are reported for questions 3 through 7.
*Binary (yes/no) prior knowledge questions. Percentage of "yes" responses per group sample is reported.
** Composite score comprised of questions 4 through 7.

Classification performance on the novel exemplar test

Performance on the novel exemplar test was quantified per participant by summing the total correctly classified and dividing by the total possible (48) to obtain the proportion correct (see Figure 3). Feedback – contrast feedback or corrective feedback – enhanced classification performance for novel exemplars. Further, contrast feedback enhanced novel exemplar classification performance more so than did corrective feedback. In support of these conclusions,a 3-level (feedback type: no feedback, corrective feedback, contrast feedback) one-way analysis of variance (ANOVA) on the proportion correct on the novel exemplar test revealed that there was a significant difference between the three feedback groups, $F(2,109) = 13.03$, $p \leq .001$, $\eta^2_p = .19$. Participants who received contrast feedback performed significantly better than did those who received corrective feedback, $t(73) = 2.33$, $p = .022$, $d = .54$, or no feedback, $t(72) = 4.88$, $p \leq .001$, $d = 1.13$. Further, participants who received corrective feedback performed significantly better than did those who did not received feedback, $t(73) = 2.85$, $p = .006$, $d = .66$.

Outcomes from performance on the novel exemplar test in Experiment 1 supported my hypotheses. Relative to no feedback and corrective feedback, providing students with elaborative feedback lead to superior classification performance on the transfer test with novel stimuli. This outcome is consistent with the discriminative contrast hypothesis and suggests that chemistry students would benefit most from contrast feedback when learning to classify chemical categories. This outcome is also consistent with Butler et al. (2013), in that providing additional information was beneficial to students as they learned to classify chemical categories. As well, relative to no feedback, providing students with corrective feedback enhanced performance on the transfer test.

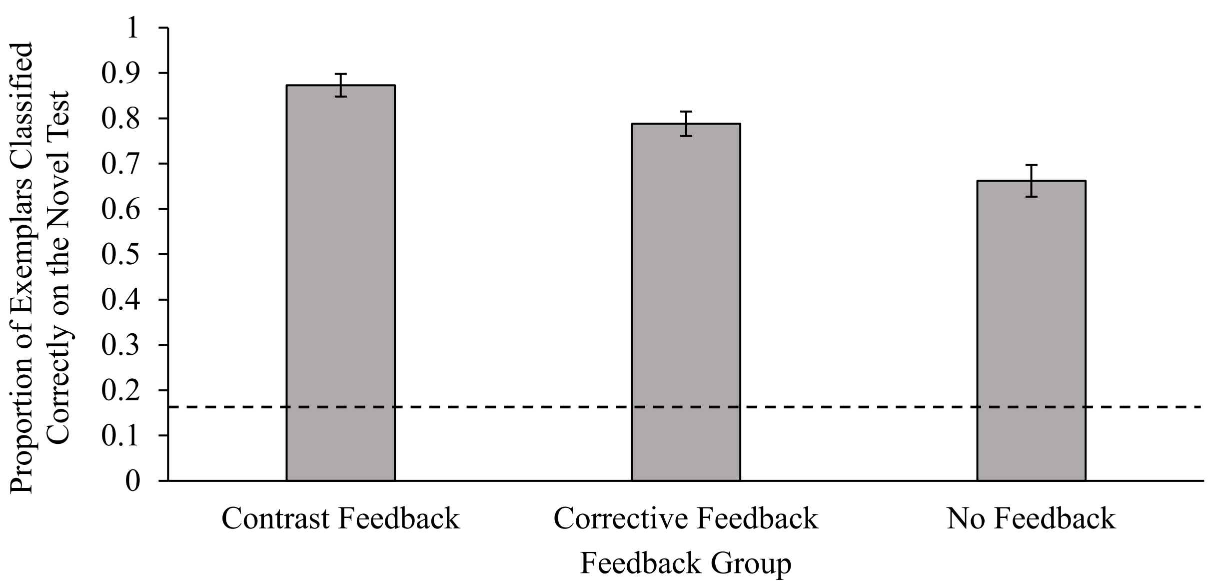

Figure 3. Classification performance on the novel exemplar test for Experiment 1.Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

The relationship between time spent on feedback and classification performance on the novel test

It is possible that classification performance on the novel exemplar test was influenced by the amount of time students spent on feedback during the practice phase rather than by the type of feedback per se. Indeed, the amount of time students spent on feedback differed by feedback group, $F(2,109) = 94.45$, $p \leq .001$, $\eta^2_p = .63$ (see Figure 4). Participants in the contrast feedback group and in the corrective feedback group spent significantly longer on feedback than did participants in the no feedback group (who were not given feedback, but instead typed “next” to progress to the next exemplar), *t*s ≥ 11.86, *p*s ≤ .001, *d*s ≥ 2.76. Participants in the contrast feedback group and corrective feedback group did not differ in the amount of time spend on feedback, $t(73) = 1.92$, $p = .058$, $d = .44$. Given the differences in median reaction time during feedback between the groups, performance differences between feedback groups may have resulted from the contrast and corrective feedback groups spending more time with the material than did the no feedback group. As such, I investigated the degree to which time spent on feedback contributed to the effect of feedback on classification performance on the novel exemplar test.

First, a search for outliers in time spent on feedback revealed one outlier ($Mdn_{RT} = 13.68$, $M = 20.94$, $SD = 19.12$) that was 2 standard deviations greater than the average median time spent on feedback. A Pearson product-moment correlation was calculated to determine if – across all feedback groups – the amount of time spent on feedback was related to performance on the novel exemplar test. When the outlier was included in the analysis, there was no significant relationship between time spent on feedback and novel exemplar classification performance, $r = .14$, $N = 112$, $p = .129$, $R^2 = .02$. When the outlier was excluded from the analysis, however, there was a significant positive relationship between time spent on feedback and novel exemplar classification performance, $r = .25$, $N = 111$, $p = .009$, $R^2 = .06$. These results suggest that the outlier had a significant impact on the outcomes of the analyses. Thus, even though all other outcomes maintain with the outlier included in the data set, it was removed from all subsequent analyses of time on feedback.

To further explore the role of time spent on feedback on performance, analyses were conducted for each of the feedback groups. Specifically, Pearson product-moment correlations were calculated between the amount of time spent on feedback and novel exemplar classification performance for each individual feedback group. In the no feedback group, there was no significant relationship between the amount of time spent on feedback and novel test performance, $r = -.14$, $N = 37$, $p = .425$, $R^2 = .02$. For the other two groups, the amount of time spent on feedback was significantly related to novel test performance (corrective feedback group: $r = -.33$, $N = 38$, $p = .042$, $R^2 = .10$; contrast feedback group: $r = -.46$, $N = 36$, $p = .005$, $R^2 = .21$).

To determine if the relationship between type of feedback and novel test performance was accounted for by time spent of feedback, a multiple linear regression was conducted on the proportion correct on the novel exemplar test. Time spent on feedback was entered into block one, and feedback group (dummy coded; corrective feedback = 0, and contrast feedback = 1) was entered into block two. The results revealed that after controlling for time spent on feedback, participants who received contrast feedback performed significantly better on the novel exemplar classification test than did participants who received corrective feedback, $b = .12$ ($SE = .03$), $t = 3.60$, $p = .001$, $R^2 = .14$.

Even though the amount of time spent on feedback significantly differed between the groups, outcomes maintained when controlling for this variable. Specifically, participants who received contrast feedback performed significantly better on the novel exemplar test relative to those who received correct feedback or no feedback. As well, participants who received corrective feedback performed significantly better on the novel exemplar test relative to those who received no feedback.

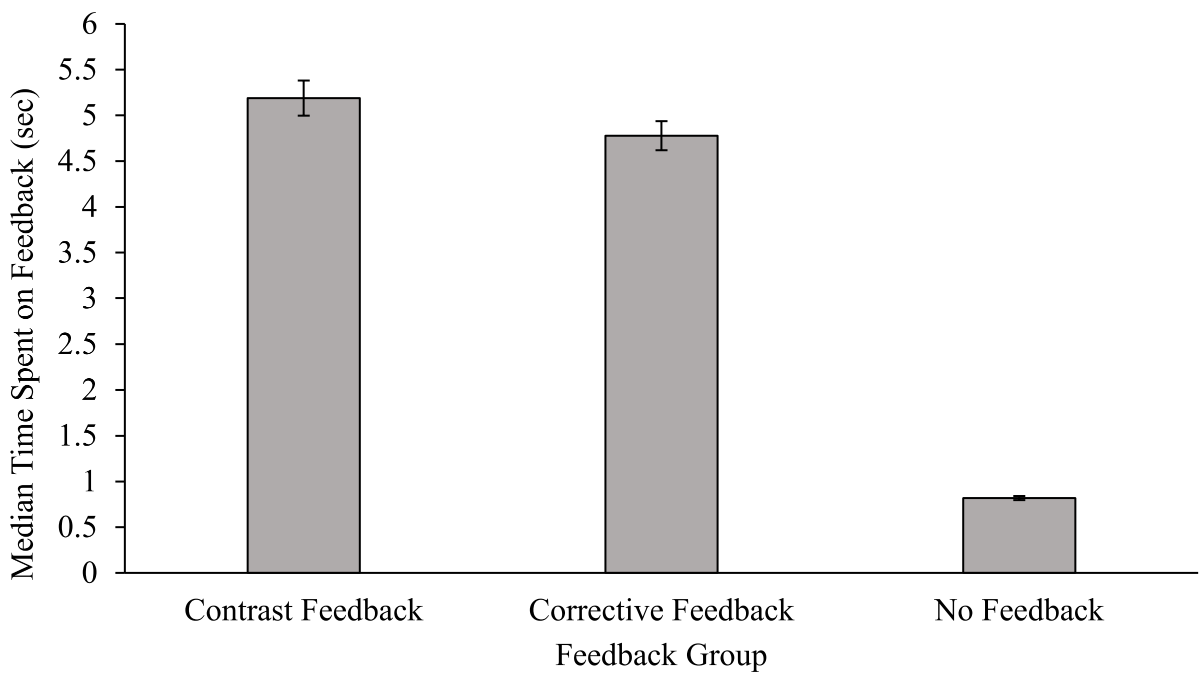

Figure 4. Median time spent on feedback for Experiment 1. Error bars represent one standard error of the mean.

Classification performance on the studied exemplar test

Performance on the studied exemplar test was quantified per participant by summing the total correctly classified and dividing by the total possible (72) to obtain the proportion correct (see Figure 5). Classification performance for studied exemplars was enhanced by feedback, although the type of feedback did not matter. In support of these results, a 3-level (feedback type: no feedback, corrective feedback, contrast feedback) one-way ANOVA revealed that there was a significant difference in classification performance between the three feedback groups, $F(2,109)$

= 12.38, $p \leq .001$, $\eta^2_p = .19$. Specifically, participants who received contrast feedback or corrective feedback performed significantly better than did participants who did not receive feedback, *t*s ≥ 2.93, *p*s ≤ .004, *d*s ≥ .68. However, participants who received contrast feedback did not perform significantly different on the studied exemplar classification test than did participants who received corrective feedback, $t(73) = 1.815$, $p = .074$, $d = .42$.

Participants who received feedback – corrective or contrast feedback – performed better on the studied exemplar classification test than did participants who did not receive feedback. Thus, providing participants with feedback was beneficial to learning chemical categories. Additionally, the contrast feedback and corrective feedback groups did not differ in performance on the studied exemplar test. This is consistent with prior research examining elaborative feedback (Butler et al., 2013).

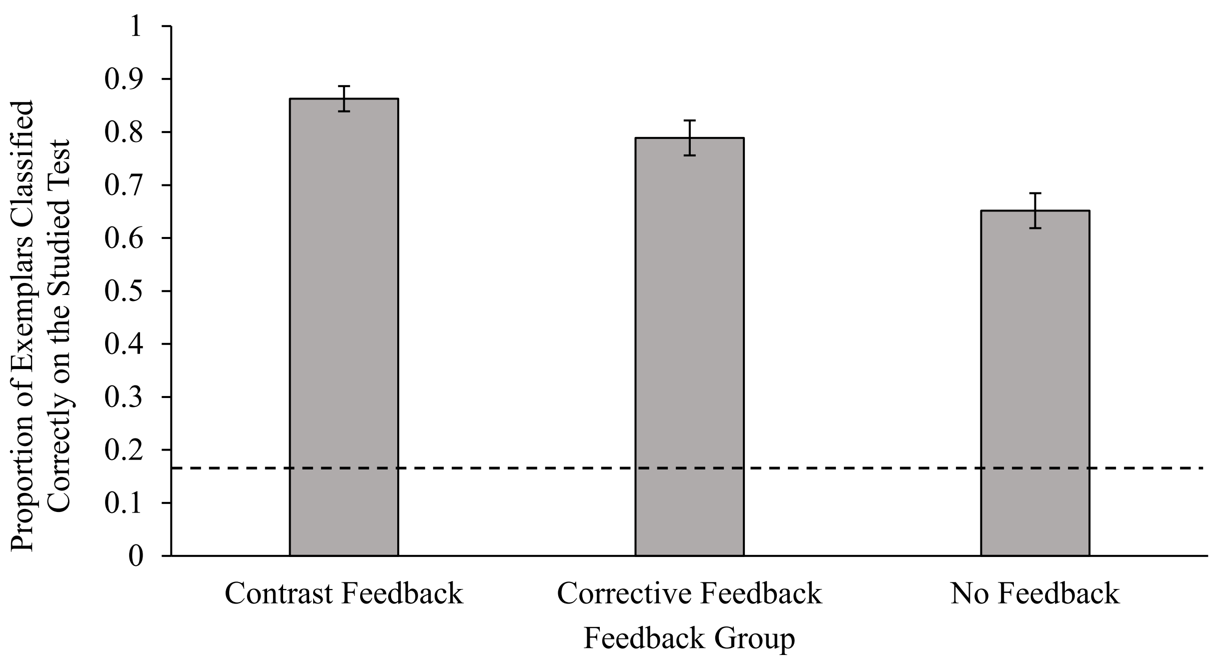

Figure 5. Classification performance on the studied exemplar test for Experiment 1. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

The relationship between time spent on feedback and classification performance on the studied test

Similar to analyses of novel exemplar test performance, I explored the degree to which time on feedback contributed to studied exemplar test performance. There was a significant relationship between the amount of time spent on feedback and studied exemplar classification performance, $r = .21$, $N = 111$, $p = .025$, $R^2 = .04$. Pearson product-moment correlations were conducted between the time spent on feedback and performance on the studied exemplar test for each individual feedback group. For the no feedback group, there was no significant relationship between the amount of time spent on feedback and novel test performance, $r = -.23$, $N = 37$, $p = .176$, $R^2 = .05$. For the other two feedback groups, the amount of time spent on feedback significantly related to novel test performance (corrective feedback group: $r = -.33$, $N = 38$, $p = .046$, $R^2 = .11$; contrast feedback group: $r = -.60$, $N = 36$, $p \leq .001$, $R^2 = .36$).

To examine the relationship between time spent on feedback and proportion correct on the studied exemplar test, a multiple linear regression was conducted on studied test performance. Time spent on feedback was entered into block one, and feedback group (dummy coded; corrective feedback = 0, contrast feedback = 1) was entered into block two. The results revealed that while controlling for time spent on feedback, participants who received contrast feedback performed significantly better on the studied exemplar test than did those who received corrective feedback, $b = .11$ ($SE = .04$), $t = 3.03$, $p = .003$, $R^2 = .10$.

The results of the multiple regression indicate that, although time spent on feedback differed between feedback groups, students who received contrast feedback did better on the studied exemplar classification test than did students who received corrective feedback. In contrast to the results of the ANOVA, this finding suggest that contrast feedback was better for performance on the studied exemplar test than was corrective feedback.

Classification Performance during the Practice Phase

Classification performance during the practice phase was quantified per participant by summing the total correctly classified and dividing by the total possible (72) to obtain the proportion correct (see Figure 6). Contrast feedback enhanced performance during the practice phase relative to no feedback. Further, corrective feedback was not beneficial as compared to the no feedback group or the contrast feedback group. A 3-level (feedback type: no feedback, corrective feedback, contrast feedback) one-way ANOVA was conducted on the proportion of items correctly classified during the practice phase. The results revealed that there was a significant difference between the three feedback groups, $F(2,109) = 7.281$, $p \leq .001$, $\eta^2_p = .12$. Follow-up *t*-tests revealed that the contrast feedback group performed significantly better during the practice phase than did the no feedback group, $t(72) = 4.06$, $p \leq .001$, $d = .94$. The corrective feedback group did not differ from the contrast feedback group nor from the no feedback group, *t*s ≤ 1.91, *p*s ≥ .060, *d*s ≤ .44.

Serial position was also examined across the practice phase. Performance was quantified as the proportion of correct responses for each item position in the practice phase, then responses were collapsed into 6 bins (e.g. items 1-12, 13-24, etc.). As evident from inspection of Figure 7, the contrast feedback group performed the best across the practice phase, as compared to the corrective feedback and no feedback groups. The corrective feedback group performed better than did the no feedback group across the practice phase. Further, while performance improved across the practice phase for the contrast feedback and corrective feedback groups, the no feedback group did not appear to improve with additional trials.

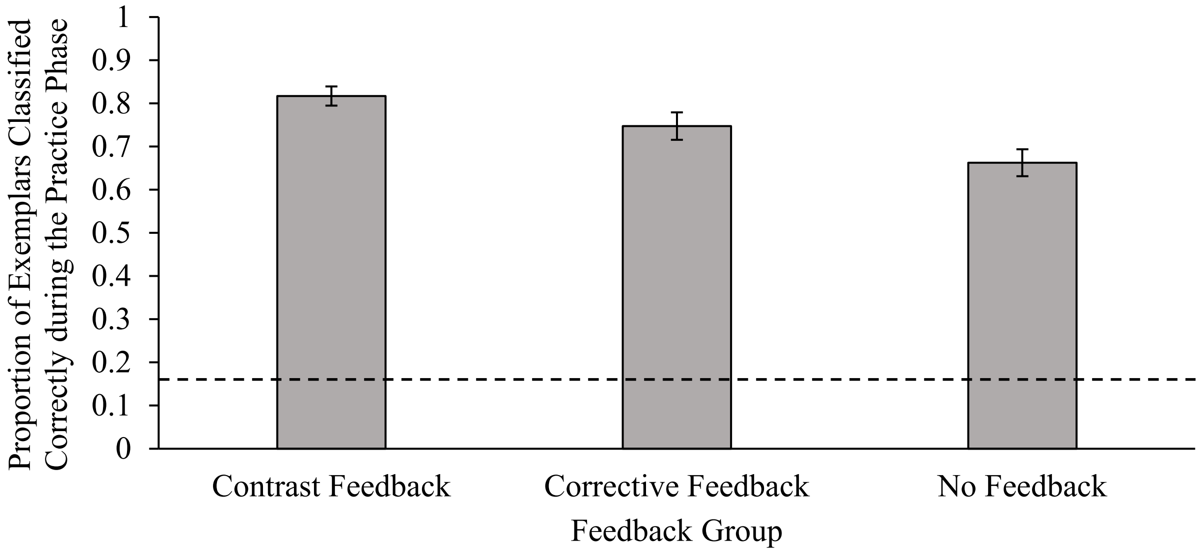

Figure 6. Classification performance during the practice phase for Experiment 1. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

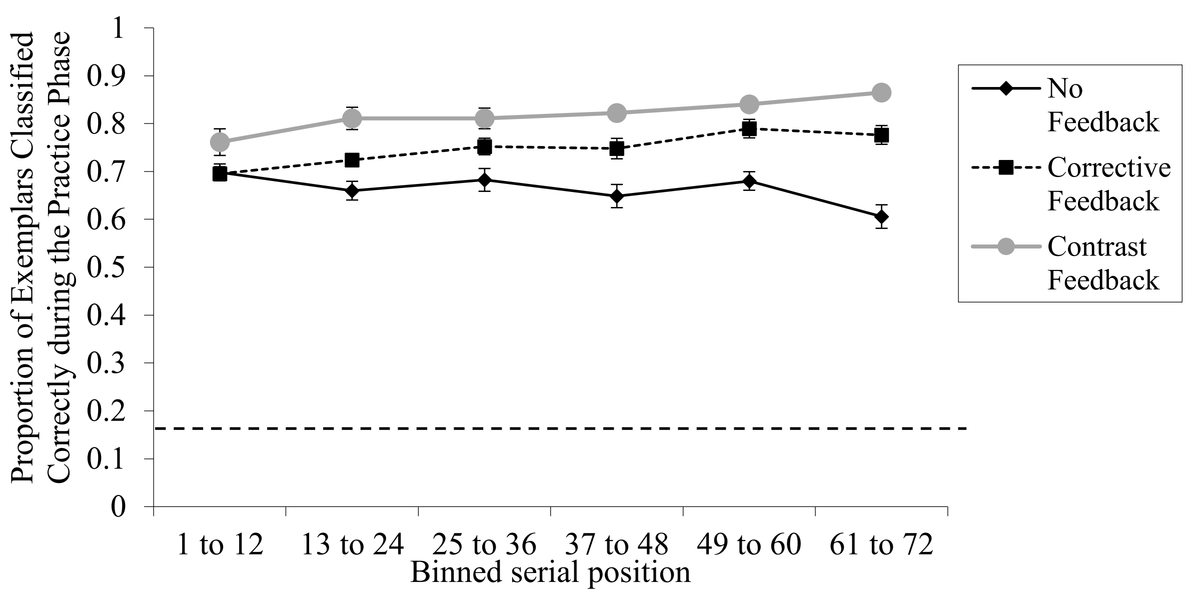

Figure 7. Classification performance by serial position during the practice phase for Experiment 1. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

Feedback Preferences

Most participants (54.5%) reported that receiving contrast feedback would be the most beneficial for their learning. Fewer participants reported a preference for corrective feedback (32.1%) or for no feedback (13.4%). Participants' preferences did not differ as a function of feedback group assignment, $\chi^2(6, N = 112) = .91, p = .923$.

Participants preferred contrast feedback, regardless of the type of feedback they received during category learning. This suggests that participants preferred the type of feedback – contrast feedback – that was most beneficial to classification performance.

Experiment 2

The primary goal of Experiment 2 was to replicate and extend the findings of Experiment 1. In Experiment 1, I established that contrast feedback was beneficial for category learning on both the novel exemplar test and the studied exemplar test. Contrast feedback may enhance category learning for two different reasons.

Researchers have previously suggested that the benefits of other kinds of elaborative feedback may be attributed to the additional information included in the elaborative feedback (Butler et al., 2013). Thus, the additional information about each chemical category provided during contrast feedback in Experiment 1 may have produced participants' superior classification performance rather than contrast processing per se. In other words, based on the outcomes from Experiment 1 alone, I cannot separate the impact of having the diagnostic feature from engaging in comparative processing on participants' learning in the contrast feedback group.

Alternatively, the performance differences observed in Experiment 1 may be attributed to comparisons between categories. The discriminative contrast hypothesis suggests that making comparisons between categories is beneficial for category learning. Thus, the performance benefit of contrast feedback may have occurred because participants compared and contrasted the chemical categories. The discriminative contrast hypothesis would indicate that making these comparisons resulted in the performance benefit observed in the contrast feedback group, above and beyond the benefit provided by additional information.

The primary goal of Experiment 2 was to replicate Experiment 1 and to further explore why contrast feedback resulted in the best performance on the novel and studied exemplar tests. To do so, the three feedback groups from Experiment 1 were used in Experiment 2 (to serve as a direct replication) and a fourth feedback group was added. The main purpose of the fourth feedback group was to explore if contrast feedback enhanced categorization performance as much as a different type of elaborative feedback. The second type of elaborative feedback was called *feature feedback*. Participants who received feature feedback saw one piece of information about the correct chemical category (see Figure 8). This piece of information described the diagnostic feature of the correct category. Both contrast feedback and feature feedback were provided with the diagnostic feature for the chemical category, thus, they have the same amount of information for each chemical category. Most important, however, only the contrast feedback group directly compared the correct category and the incorrect category that was selected. Accordingly, whereas contrast feedback and feature feedback were matched in the amount of information provided to learners, they encouraged different types of processing during learning.

I anticipated that results from Experiment 2 would support the discriminative contrast hypothesis. Specifically, the contrast feedback group would perform better than would the feature feedback group, corrective feedback group, and no feedback group on the novel exemplar test. Additionally, I predicted that the feature feedback group would perform better than would the no feedback and corrective feedback groups on the novel exemplar test, and that the corrective feedback group would perform better than would the no feedback group. For the studied exemplar test, I predict the same pattern of results as the novel exemplar test, which would replicate and extend the results from Experiment 1.

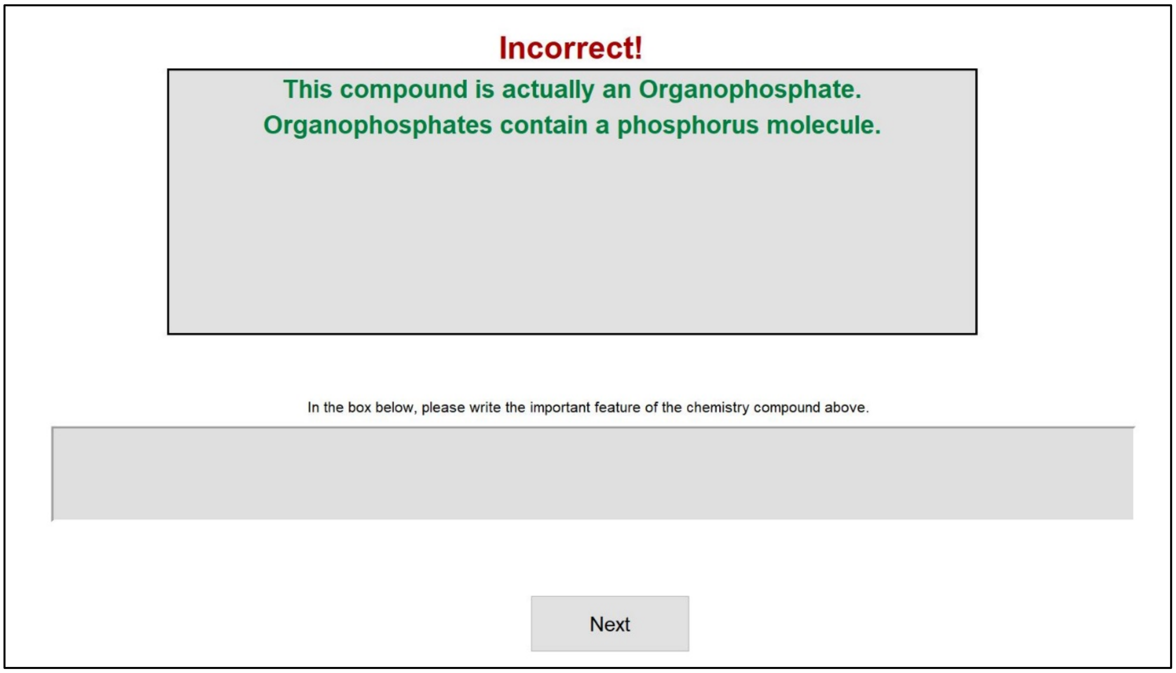

Figure 8. Feature feedback.

Method

Participants and Design

A power analysis was conducted with G*Power to estimate the number of participants necessary for each group (Faul et al., 2007). Using an effect size, d = .54, (classification performance of the contrast feedback versus corrective feedback groups on the novel test from Experiment 1) with alpha error at .05 and power at .95, it was estimated that 91 participants would be needed for each group. Therefore, 364 undergraduate Texas Christian University students participated for partial credit in psychology courses. Of these participants, 28 who were randomly assigned to the contrast feedback group (31% of group sample) and 27 who were randomly assigned to the feature feedback groups (30% of group sample) did not follow directions during feedback. Thus, these participants were excluded from the analyses. To reach the target sample size, additional data were collected by randomly assigning participants to the contrast feedback group or the feature feedback group. From these data, 28 participants were randomly assigned to the contrast feedback and 27 participants were randomly assigned to the feature feedback group, and their data were added to the sample giving a total sample of $N = 364$ ($n = 91$ per group). The sample was college-aged ($M = 19.12$ years, $SD = 2.07$), most identified as female (65%; 235 females, 110 males, and 18 gender non-binary), and Caucasian (76%). The four feedback groups did not differ in age, $F(3,358) = .27$, $p = .849$, $\eta^2_p = .002$, gender identity, $\chi^2(6, N = 363) = 9.64$, $p = .141$, or in ethnicity, $\chi^2(30, N = 363) = 23.70$, $p = .786$.

Materials

The materials were the same as those used in Experiment 1. Further, the additional information developed for the contrast feedback group was also used in the feature feedback group (see Figure 8). Feature feedback consisted of the diagnostic feature of correct chemical category.

Procedure

The procedure for Experiment 2 was the same as the procedure for Experiment 1. That is, all participants began by answering several prior knowledge questions about their organic chemistry experience and viewed a brief organic chemistry tutorial before the initial study phase with 72 exemplars. Following the initial study phase, participants proceeded to the practice phase. During the practice phase, participants in the no feedback, corrective feedback, and contrast feedback groups received feedback in the same manner as Experiment 1.

The feature feedback group was very similar to the contrast feedback group. Following a correct answer, participants in the feature feedback group were shown the statement, “Correct! This is a [correct chemical category].” in green font and were asked to type the name of the correct chemical category to progress to the next trial. Following an incorrect answer, participants in the feature feedback group received the statement, “Incorrect!” in red font, as well as one piece of additional information (see Figure 8). The piece of feedback contained the diagnostic feature for the correct chemical category and was shown in green font as the statement, “This compound is actually a [correct chemical category]. To be a [correct chemical category] the compound must have a [diagnostic feature for the correct chemical category]”. To proceed to the next exemplar, participants typed the diagnostic feature of the chemical category into the provided text box. This encouraged participants to engage with the diagnostic feature and maintained consistency with the typed response required for the other feedback groups. As in Experiment 1, after completing all 72 exemplars in the practice phase, the participants took both the novel exemplar test (48 novel exemplars) and the studied exemplar test (72 exemplars) before they were thanked for their time, granted credit, and debriefed.

Results and Discussion

Responses to prior knowledge questions are reported first to explore participants’ pre-existing chemistry knowledge. Participants’ performance on the novel exemplar test is reported next. Then, participants’ performance on the studied exemplar test is reported. Participants’ performance during the practice phase and feedback preferences is reported last.

Prior Knowledge Questions

Results for the prior knowledge questions are located in Table 2. The four open-ended questions were scored in the same manner as in Experiment 1, and the proportion of correct prior knowledge questions was computed for each participant. Participants' prior knowledge of organic chemistry was low (*M* = .11, *SE* = .01) and did not differ between the four feedback groups, $F(3, 360) = 1.51$, $p = .211$, $\eta^2_p = .01$. Additionally, participants' self-reported knowledge of organic chemistry (1 = novice to 7 = expert), was low (*M* = 1.58, *SE* = .05) and did not differ between the four feedback groups, $F(3, 303) = 1.51$, $p = .142$, $\eta^2_p = .02$.

Further, few participants identified as chemistry majors or minors (20 students, 5.5% of sample) or as biochemistry majors or minors (12 students, 3.3% of sample). The distribution of chemistry majors and minors significantly differed between feedback groups, $\chi^2(3, N = 364) \leq 11.43$, $p = .010$ (See Table 2). By contrast, the distribution of biochemistry majors and minors did not differ between feedback groups, $\chi^2(3, N = 364) \leq 4.86$, $p = .185$. In sum, although there was an uneven distribution of chemistry majors and minors between the four feedback groups, prior knowledge of organic chemistry was low.

Table 2. Experiment 2 Prior Knowledge Questions	No Feedback	Corrective Feedback	Contrast Feedback	Feature Feedback
1. Are you currently or have you ever been a Chemistry Major or Minor?*	9.6%	10.9%	0%	3.4%
2. Are you currently or have you ever been a Biochemistry Major or Minor?*	5.8%	4.6%	0%	3.4%
3. What is your own rated level of expertise at chemistry?	1.66 (.15)	1.74 (.13)	1.37 (0.07)	1.58 (.12)
Prior Knowledge Composite Score**	.10 (.03)	.08 (.02)	.05 (.01)	.09 (.02)
4. What is the name of the compound with the formula C_6H_{14}?	.07 (.03)	.06 (.03)	.00 (.00)	.06 (.03)
5. What is the name of the hydrocarbon series that has a double bond between carbon atoms?	.06 (.03)	.03 (.02)	.00 (.00)	.02 (.01)
6. How many valence electrons does an oxygen molecule have?	.20 (.05)	.22 (.05)	.20 (.04)	.24 (.05)
7. Benzene rings mainly undergo which type of chemical reaction?	.06 (.03)	.01 (.01)	.00 (.00)	.03 (.02)

M(*SE*) are reported for questions 3 through 7.
*Binary (yes/no) prior knowledge questions. Percentage of "yes" responses is reported.
** Composite score comprised of questions 4 through 7.

Classification performance on the novel exemplar test

Performance on the novel exemplar test was quantified in the same manner as in Experiment 1 (see Figure 9). Feedback enhanced classification performance of novel exemplars. Further, classification performance was best following contrast feedback and feature feedback, though performance in the two groups did not differ. In support of these findings, a 4-level (feedback type: no feedback, corrective feedback, contrast feedback, feature feedback) one-way ANOVA conducted on the proportion correct on the novel exemplar test revealed that the four groups differed significantly, $F(3, 363) = 12.40$, $p \leq .001$, $\eta^2_p = .09$.

Follow-up *t*-test established that the contrast feedback group performed significantly better on the novel exemplar test than did the corrective feedback group and the no feedback group, *t*s ≥ 3.07, *p*s ≤ .002, *d*s ≥ .46. However, classification performance on the novel exemplar test did not significantly differ between the contrast feedback group and feature feedback group, $t(180) = .32$, $p = .748$, $d = .05$. The feature feedback group performed significantly better on the novel exemplar test than did the corrective feedback group and the no feedback group, *t*s ≥ 3.39, *p*s ≤ .001, *d*s ≥ .50. Classification performance on the novel exemplar test did not significantly differ between the corrective feedback group and the no feedback group, $t(180) = 1.56$, $p = .121$, $d = .23$.

Providing participants with elaborative feedback – contrast feedback or feature feedback – was beneficial for their performance on the novel exemplar test. Further, providing participants with corrective feedback was numerically but not statistically beneficial for participants' performance in contrast with providing with participants with no feedback.

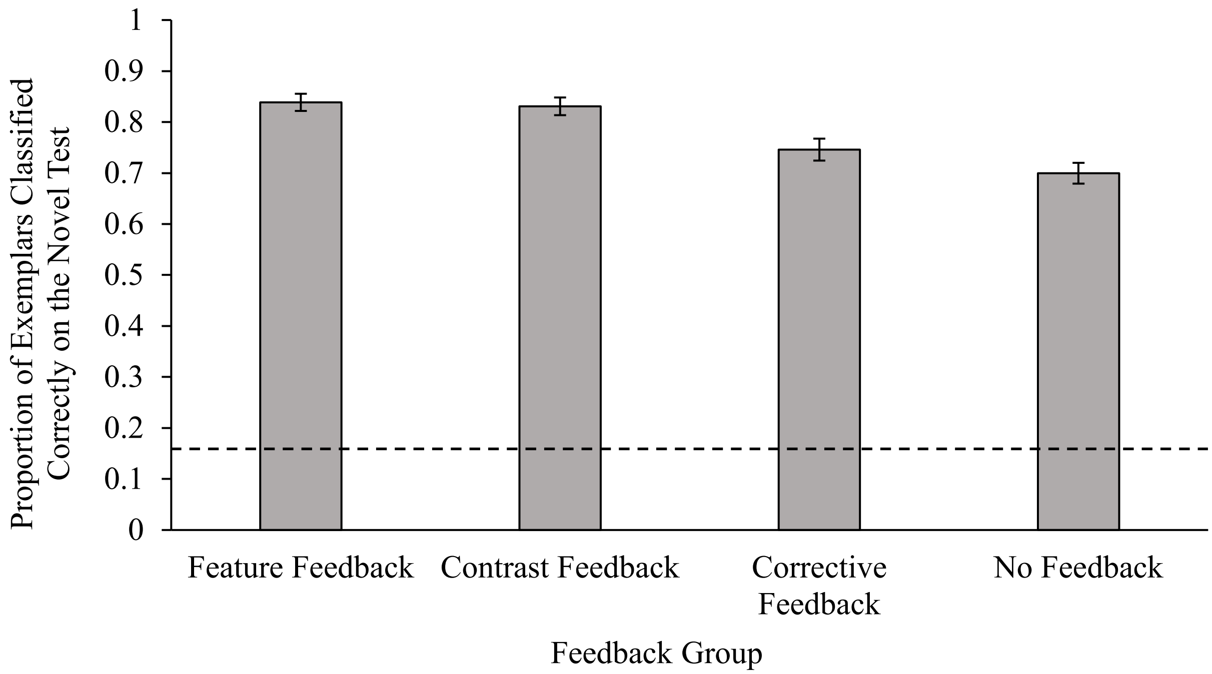

Figure 9. Classification performance on the novel exemplar test for Experiment 2. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

The relationship between time spent on feedback and classification performance on the novel test

Similar to Experiment 1, the amount of time that participants spent on feedback may have influenced their performance on the novel exemplar test instead of the type of feedback they received. Groups spent significantly different amounts of time on feedback, $F(3, 360) = 133.68$, $p \leq .001$, $\eta^2_p = .53$ (see Figure 10). Specifically, participants in the no feedback group spent less time on feedback than did the other three feedback groups, ts ≥ 17.84, ps ≤ .001, ds ≥ 2.65. Participants in the corrective feedback group also sent less time on feedback than did participants in the contrast feedback group or the feature feedback group, ts ≥ 2.39, ps ≤ .018, ds ≥ .35. The amount of time spent on feedback did not significantly differ between the contrast feedback group and the feature feedback group, *t*(180) = .86, *p* = .393, *d* = .13. Given the differences in median reaction time during feedback between the four feedback groups, I investigated the degree to which time spent on feedback contributed to the effect of feedback on classification performance on the novel exemplar test.

A search for outliers in time spent on feedback returned 10 outliers (Mdn_{RT} = 11.61, *SD* = 3.17) that were 2 standard deviations greater than the average median time spent on feedback. A Pearson product-moment correlation was calculated to determine if – across all four feedback groups – the amount of time spent on feedback was related to performance on the novel exemplar test. When the outliers are included in the analysis, participants' performance on the novel exemplar test was significantly correlated with the time spent on feedback, *r* = -.10, *N* = 364, *p* = .047, R^2 = .01. When the outliers were excluded from the analysis, however, there was not a significant relationship between the time spent on feedback and performance on the novel exemplar test, *r* = .06, *N* = 354, *p* = .283, R^2 = .004. Thus, the outliers influenced the outcomes of the analyses. Although all other outcomes maintain with the outliers included in the data set, they were removed from all of the following analyses that include time on feedback.

Pearson product-moment correlations were calculated exploring the relationship between time spent on feedback and novel exemplar classification performance for each individual feedback group. For the no feedback group, there was no significant relationship between the time spent on feedback and novel exemplar test performance, *r* = -.20, *N* = 91, *p* = .054, R^2 = .04. For the other three groups, time on feedback was significantly related to novel exemplar test performance (corrective feedback group: *r* = -.39, *N* = 90, $p \leq .001$, R^2 = .15; contrast feedback group: *r* = -.41, *N* = 85, $p \leq .001$, R^2 = .17; feature feedback group: *r* = -.53, *N* = 88, $p \leq .001$, R^2 = .28).

A multiple linear regression was conducted on the proportion of correct responses on the novel exemplar test and time spent on feedback. Given that contrast feedback and feature feedback did not differ in performance on the novel exemplar test or in time spent on feedback, these groups were collapsed into an elaborative feedback group (although it can be noted that results maintain when conducted for each group individually). Time spent on feedback was entered into block one, and feedback group (dummy coded; corrective feedback = 0, and elaborative feedback = 1) was entered into block two. The results revealed that after controlling for time spent on feedback, participants who received elaborative feedback performed significantly better on the novel exemplar test than did participants who received corrective feedback, $b = .13$ ($SE = .02$), $t = 6.29$, $p \leq .001$, $R^2 = .11$.

Providing elaborative feedback benefitted performance on the novel exemplar classification test compared to providing corrective feedback. As important, these outcomes are not due to the time spent on feedback.

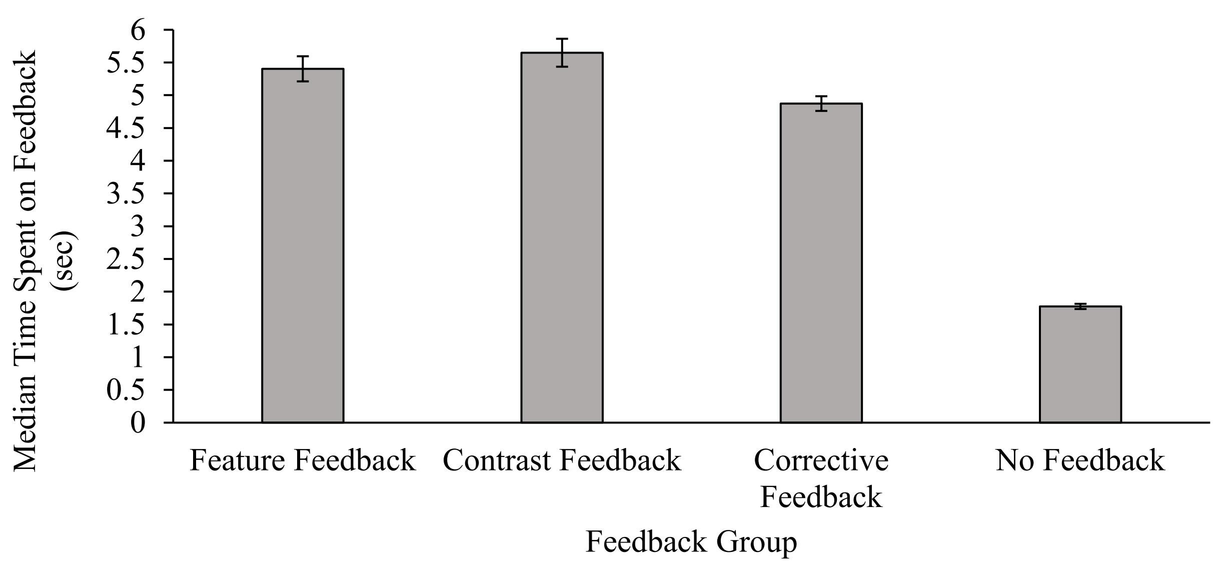

Figure 10. Median time spent on feedback for Experiment 2. Error bars represent one standard error of the mean.

Classification performance on the studied exemplar test

Performance on the studied exemplar test was quantified in the way as in Experiment 1 (see Figure 11). Feedback enhanced classification performance for studied exemplars. Specifically, contrast feedback and feature feedback were best for classification performance as compared to the corrective feedback group and the no feedback group. In support of these conclusions, a 4-level (feedback type: no feedback, corrective feedback, contrast feedback, feature feedback) one-way ANOVA revealed a significant difference in performance on the studied exemplar test between the four feedback groups, $F(3,360) = 11.78$, $p \leq .001$, $\eta^2_p = .09$. Follow-up *t*-test revealed that the contrast feedback group did not differ from the feature feedback group on studied exemplar test performance, $t(180) = .02$, $p = .981$, $d = .003$. However, both elaborative feedback groups – contrast feedback and feature feedback – performed better on the studied exemplar test than did the corrective feedback group or than did the no feedback group, ts ≥ 3.10, ps $\leq .002$, ds $\geq .46$. Additionally, the corrective feedback group and the no feedback group did not differ on performance on the studied exemplar test, $t(180) = 1.54$, $p = .127$, $d = .23$.

In sum, providing participants with elaborative feedback – contrast feedback or feature feedback – was beneficial for performance on the studied exemplar test relative to providing participants with corrective feedback. Further, providing participants with corrective feedback did not result in statistically better performance on the studied exemplar test. However, numerically, the corrective feedback group performed better on the studied exemplar test than did the no feedback group.

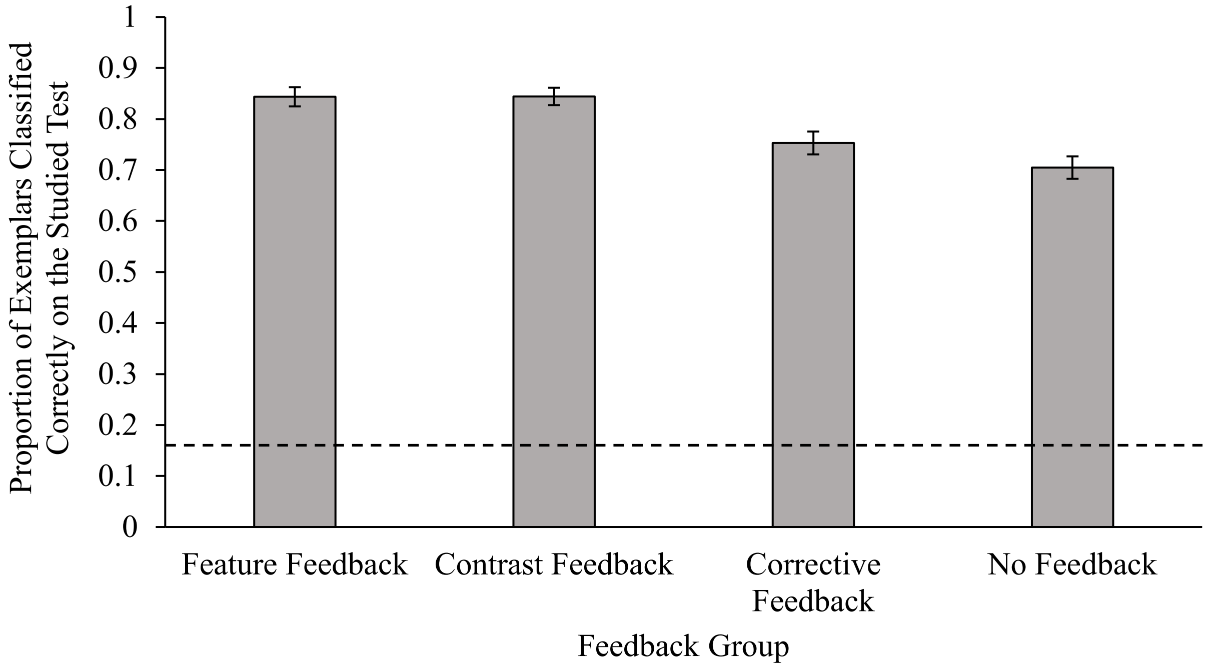

Figure 11. Classification performance on the studied exemplar test for Experiment 2. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

The relationship between time spent on feedback and classification performance on the studied test

Similar to the novel exemplar test, performance on the studied exemplar test may have been influenced by the amount of time spent on feedback. A Pearson product-moment correlation conducted across all four feedback groups revealed there was a significant relationship between the amount of time spent on feedback and studied exemplar test performance, $r = .036$, $N = 354$, $p = .494$, $R^2 = .001$.

Pearson product-moment correlations were also conducted for each individual feedback group to investigate the relationship between the time spent on feedback and performance on the studied exemplar test. For all four groups, time on feedback was significantly related to classification performance on the studied test (no feedback group: $r = -.23$, $N = 91$, $p = .030$, $R^2 = .05$; corrective feedback group: $r = -.44$, $N = 90$, $p \leq .001$, $R^2 = .19$; contrast feedback group: $r = -.40$, $N = 85$, $p \leq .001$, $R^2 = .16$; feature feedback group: $r = -.52$, $N = 88$, $p \leq .001$, $R^2 = .27$).

Contrast feedback and feature feedback did not differ in performance on the studied exemplar test or in time spent on feedback, as such, these groups were collapsed into an elaborative feedback group (although the results did maintain when analyses were conducted for each group individually). Three multiple linear regression were conducted to examine the relationship between the time spent on feedback and performance on the studied exemplar test for the no feedback group, corrective feedback group, and elaborative feedback group. To explore the comparison between the no feedback group and the elaborative feedback group, time spent on feedback was entered into block one and feedback group (dummy coded; no feedback = 0, elaborative feedback = 1) was entered into block two. The results revealed that while controlling for time spent on feedback, participants who received elaborative feedback performed significantly better on the studied exemplar test than did participants who received no feedback, $b = .42$ ($SE = .04$), $t = 10.34$, $p \leq .001$, $R^2 = .23$. To explore the comparison between the corrective feedback group and the elaborative feedback group, time spent on feedback was entered into block one and feedback group (dummy coded; corrective feedback = 0, elaborative feedback = 1) was entered into block two. The results revealed that while controlling for time spent on feedback, participants who received elaborative feedback performed significantly better on the studied exemplar test than did participants who received corrective feedback, $b = .13$ ($SE = .02$), $t = 5.65$, $p \leq .001$, $R^2 = .07$. To examine the relationship between the no feedback group and the corrective feedback group, time spent on feedback was entered into block one and feedback group (dummy coded; no feedback = 0, corrective feedback = 1) was entered into block two. The results revealed that while controlling for time spent on feedback, participants who received corrective feedback performed significantly better on the studied exemplar test than did participants who received no feedback, $b = .35$ ($SE = .07$), $t = 5.24$, $p \leq .001$, $R^2 = .13$.

These results indicate that receiving elaborative feedback was better for performance on the studied exemplar test than was corrective feedback. As well, feedback – corrective feedback, contrast feedback, and feature feedback – was beneficial for performance on the studied exemplar test as compared to no feedback.

Classification Performance during the Practice Phase

Practice phase performance was quantified in the same manner as in Experiment 1 (see Figure 12). Results followed the same pattern as the novel exemplar test and studied exemplar test, in that contrast feedback and feature feedback enhanced performance during the practice phase as compared to corrective feedback and no feedback. In support of these findings, the results of a 4-level (feedback type: no feedback, corrective feedback, contrast feedback, feature feedback) one-way ANOVA on the proportion of items correctly classified during the practice phase revealed that the groups significantly differed, $F(3,360) = 7.54$, $p \leq .001$, $\eta^2_p = .06$. The contrast feedback group performed significantly better during the practice phase than did the corrective feedback group and the no feedback group, ts ≥ 3.13, ps $\leq .002$, ds $\geq .46$. However, the contrast feedback group did not differ in practice phase performance from the feature feedback group, $t(180) = .17$, $p = .862$, $d = .03$. The feature feedback group also performed significantly better during the practice phase than did the corrective feedback group and the no feedback group, ts ≥ 3.36, ps $\leq .001$, ds $\geq .50$. Further, the corrective feedback group did not differ in practice phase performance from the no feedback group, $t(180) = .25$, $p = .806$, $d = .04$.

Serial position effect was also explored across the practice phase. Similar to Experiment 1, performance was calculated as the proportion of correct response for each item position in the practice phase, then these proportions were collapsed into 6 bins (e.g. items 1-12, 13-24, etc.). As shown in Figure 13, the contrast feedback group and feature feedback group performed the best throughout the practice phase relative to the corrective feedback group and no feedback group. The corrective feedback group also performed better than did the no feedback group across the practice phase. As the practice phase progressed, the contrast feedback group and the feature feedback group displayed the largest improvement across trials as compared to the corrective feedback group and the no feedback group. The corrective feedback group and the no feedback group, however, did not show substantial improvement across the practice phase.

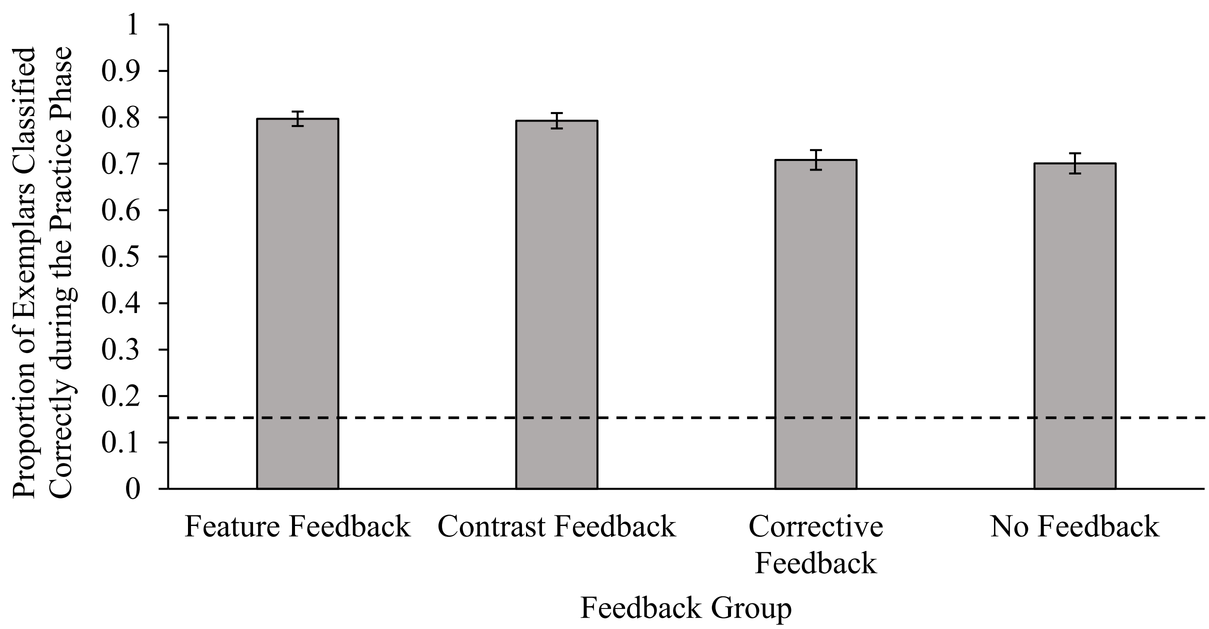

Figure 12. Classification performance during the practice phase for Experiment 2. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

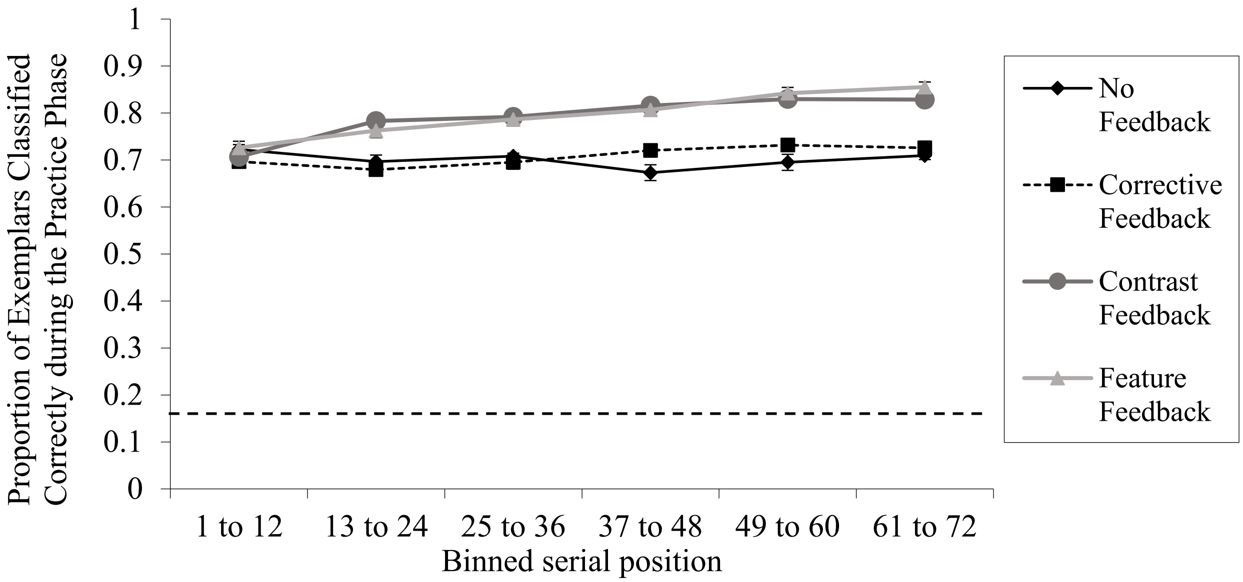

Figure 13. Classification performance by serial position during the practice phase for Experiment 2. Error bars represent one standard error of the mean. Dashed line represents chance levels of performance (chance proportion of correct exemplars = .166).

Feedback Preferences

Most participants reported that receiving contrast feedback (39.3%) would be most beneficial for their learning. Fewer participants reported a preference for feature feedback (29.9%), corrective feedback (19.5%), or for no feedback (11.3%). Participants' preferences differed as a function of their feedback group assignment, $\chi^2(9, N = 364) = 19.33$, $p = .022$ (See Table 3 for percentages).

Overall, these findings suggest that participants have a preference for elaborative feedback – contrast feedback or feature feedback – compared to corrective feedback or no feedback. This indicates that participants preferred the type of feedback that was most beneficial for learning chemical categories.

Table 3. Experiment 3 Feedback Preference Responses

	Preferred Type of Feedback			
Feedback Group*	No Feedback	Corrective Feedback	Contrast Feedback	Feature Feedback
No Feedback	12.1%	20.9%	46.2%	20.9%
Corrective Feedback	11.0%	25.3%	41.8%	22.0%
Contrast Feedback	15.4%	13.2%	37.4%	34.1%
Feature Feedback	6.6%	18.7%	31.9%	42.9%

* Percentage of each group sample is reported.

General Discussion

The reported experiments were the first to investigate the effect of feedback on category learning. In Experiments 1 and 2, participants' classification performance on the novel test was superior following contrast feedback compared to corrective feedback or no feedback. Contrast feedback encouraged comparisons between the diagnostic features of two chemical categories. According to the discriminative contrast hypothesis, searching for differences will benefit category learning, and contrast feedback should encourage this processing. Thus, Experiments 1 and 2 support the discriminative contrast hypothesis because classification performance was superior following contrast feedback relative to corrective feedback or no feedback. However, results regarding classification performance following feature feedback may pose a challenge for the discriminative contrast hypothesis.

The feature feedback introduced in Experiment 2 included the diagnostic feature of the chemical category and contrast processing was not required. Contrast feedback also included the diagnostic feature of the chemical category as well as the diagnostic feature of another chemical category, and participants were encouraged to make comparisons between the two. Classification performance on the novel exemplar test did not differ between the contrast feedback and feature feedback groups, which is counter to predictions and does not support the discriminative contrast hypothesis. Whereas making comparisons between categories may not have been necessary, providing the diagnostic feature of the category – a component of both contrast feedback and feature feedback – may have been sufficient to promote correct classification on the transfer test. If this is correct, it is important for educators who could save students' time and effort by eliminating contrasts but note that diagnostic cues combined with corrective feedback are critical.

It is also possible that the elaborative feedback groups did not differ because making a comparison between categories was challenging. Contrast feedback required participants to consider both the previous exemplar and the exemplars from the chemical category they incorrectly classified, and then to compare those categories. This may have strained participants' working memory and eliminated the benefit of discriminative contrast (e.g., Lavie, 2010). By comparison, feature feedback only provided the diagnostic feature for the correct chemical category, reducing the strain on working memory, while still containing the additional information that was useful for classification. To explore this, participants' typed responses to feedback were examined for the contrast and feature feedback groups. On practice trials for which the participants answered correctly, they typed the name of the correct category on the feedback screen. As such, participants in all feedback groups always followed directions on correct practice trials because the program would not progress otherwise. On incorrect practice trials, however, participants in the contrast feedback group typed how the diagnostic features of the correct chemical category and incorrect chemical category were different. Thus, contrast responses following incorrect trials were scored using a scale of 0 to 1. A score of 0 indicated that the participant did not type any diagnostic feature. A score of 0.5 indicated that the participant typed one diagnostic feature of the chemical category but did not compare it with the feature of the other category. A score of 1 indicated that the participant typed both diagnostic features and compared them. The mean proportion of feedback responses correctly answered was high for the contrast group (*M* = .75, *SD* = .23), but performance was not at ceiling and there was room for improvement. Indeed, 51.1% of trials were only partially correct (i.e., contrast responses with only one of the diagnostic features – a score of 0.5) and an additional 7.2% of trials were entirely incorrect (i.e., contrast responses with neither diagnostic feature – a score of 0). Only 41.7% of trials were scored as fully correct (i.e., contrast responses with a comparison of the categories diagnostic features – a score of 1). These findings indicate that making the category comparisons in contrast feedback may have been challenging for participants.

Responses to feature feedback were scored on a scale of 0 to 1. Participants in the feature feedback group only typed the diagnostic feature of the correct chemical category. As such, a score of 0 indicated that participants did not type the diagnostic feature in their feedback response. A score of 1 indicated that the participant typed the diagnostic feature of the correct chemical category. The mean proportion of feedback responses correctly answered approached ceiling for the feature group (*M* = .94, *SD* = .12). Likewise, 91.7% of trials were entirely correct (i.e., feature responses with the diagnostic feature – a score of 1), whereas, 8.3% of trials were entirely incorrect (i.e., feature responses without the diagnostic feature – a score of 0). Thus, the feature feedback group correctly responded to feedback on most trials indicating that it was presumably not challenging for participants to complete feature feedback.

Most important, participants in the contrast feedback and feature feedback groups responded to the feedback in the same way on many trials by including only one diagnostic feature during feedback responses, despite the feedback types encouraging different kinds of processing. As such, the two elaborative feedback groups may not have differed on novel exemplar test performance because they were using the two forms of elaborative feedback in the same manner. Given this unexpected response to feedback, outcomes from Experiments 1 and 2 alone do not elucidate the degree to which contrast feedback benefits category learning above and beyond feature feedback. To more thoroughly address this issue, researchers should consider adding design features (e.g., two feedback response boxes on the feedback screen) that would encourage participants to make comparisons between the categories when given contrast feedback. Regardless of the mechanism driving the elaborative feedback effect, what is clear is that elaborative feedback enhanced participants' learning and transfer of educationally-relevant information beyond the benefit of corrective feedback. This has direct implications for educators in STEM fields. Incorporating elaborative feedback into the classroom may be a simple strategy to improve student performance on core STEM concepts and categories. Future work should investigate how elaborative feedback improves category learning in the classroom, and the extent to which it can enhance transfer to related STEM concepts.

The results of the present research also suggest that performance on the studied exemplar test (i.e., the non-transfer test) was better when participants were provided with elaborative feedback as compared to corrective feedback. This outcome is inconsistent with prior research (Butler et al., 2013). Determining the explanation for the inconsistent outcomes is a challenge because the present research examined different participants, and used different materials and procedures than did Butler et al. (2013). One key difference may have been the type of information that was provided during elaborative feedback (fact versus diagnostic feature) and the level at which it was relevant (trial versus conceptual). Future work should aim to investigate what type of elaborative feedback is most beneficial for non-transfer test performance, and through which mechanisms it enhances performance.

As a secondary goal, I found in Experiments 1 and 2 that providing participants with feedback during category learning was more beneficial to performance on transfer tests and non-transfer tests as compared to no feedback during category learning. These effects were statistically significant for both tests in Experiment 1 and for classification performance on the novel test in Experiment 2. For the studied exemplar test in Experiment 2, however, the corrective feedback group performed numerically, but not statistically, better than did the no feedback group. To more conclusively determine the effect of corrective feedback on category learning, I conducted two *continuously cumulating meta-analyses* (CCMAs) to aggregate outcomes across experiments as suggested by Braver, Thoemmes, and Rosenthal (2014). Corrective feedback resulted in better classification performance on the novel exemplar test relative to no feedback (Experiment 1, $M_{\text{diff}} = .13$, $S_{\text{pooled}} = .19$; Experiment 2, $M_{\text{diff}} = .06$, $S_{\text{pooled}} = .20$; pooled $d = 0.35$, 95% CI [0.10, 0.60]). Similarly, corrective feedback resulted in better classification performance on the studied exemplar test relative to no feedback (Experiment 1, $M_{\text{diff}} = .14$, $S_{\text{pooled}} = .20$; Experiment 2, $M_{\text{diff}} = .05$, $S_{\text{pooled}} = .21$; pooled $d = 0.35$, 95% CI [0.11, 0.60]). Thus, corrective feedback enhanced classification performance on transfer and non-transfer tests with the effects being moderate in magnitude.

Regarding performance during the practice phase, participants who received elaborative feedback improved across practice phase trials. Further, participants in both experiments reported a preference for elaborative feedback as compared to corrective feedback or no feedback. Elaborative feedback was the most demanding on participants relative to the other types of feedback, so participants' preference for it was unexpected. For participants who received elaborative feedback, perhaps enhanced performance throughout each stage of learning influenced their feedback preference. More generally, the results from the practice phase and participants' feedback preferences suggest that elaborative feedback for category learning is not a desirable difficulty. A desirable difficulty is a learning strategy that is challenging during learning, but aids later performance (Bjork & Bjork, 2009). In contrast, elaborative feedback enhanced learning at every stage of category learning. As such, elaborative feedback may be a desirable learning aid for conceptual information.

In sum, feedback was beneficial to category learning of foundational information in chemistry. Specifically, corrective feedback resulted in better novel and studied exemplar test performance than did no feedback, and elaborative feedback was beneficial to novel and studied exemplar test performance above and beyond the effect of corrective feedback. These findings suggest that educators in STEM fields should provide their students with corrective feedback in all instances. Further, elaborative feedback should be provided to students when possible. The mechanism driving the elaborative feedback effect may be the diagnostic feature of each category, the comparisons between categories promoting discriminative contrast, or from both. Future research should determine what contributes to the elaborative feedback effect, and in which contexts it can improve student learning of complex conceptual information.

References

Anderson, J. R., Kline, P. J., & Beasley Jr, C. M. (1979). A General Learning Theory and its Application to Schema Abstraction. *Psychology of learning and motivation, 13*, 277-318.

Aust, U., & Huber, L. (2002). Target-defining features in a “people-present/people-absent” discrimination task by pigeons. *Animal Learning & Behavior*, *30*, 165-176.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of educational research*, *61*, 213-238.

Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 219.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society*. (pp. 56–64). Worth Publishers.

Bourne, L. E. (1970). Knowing and using concepts. *Psychological Review*, *77*, 546.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, *9*, 333-342.

Bruner, J. S., Goodnow, J. J., & George, A. (1956). Austin. A study of thinking. *New York: John Wiley & Sons, Inc*, *14*, 330.

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*, 290.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & cognition*, *42*, 481-495.

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1699.

Eglington, L. G., & Kang, S. H. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*, 475-485.

Fagot, J., Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons (Papio papio) and humans (Homo sapiens). *Journal of Experimental Psychology: Animal Behavior Processes*, *27*, 316.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*, 175-191.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2018). Categorization and concepts. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *3*, 1-43.

Hampton, J. (1993). Prototype models of concept representation. In I. V. Mecheln, J. Hampton, R. S. Michalski, & P. Thenus (Eds.), *Categories and concepts: Theoretical Views and inductive data analysis* (pp. 67-95). London, United Kingdom: Academic Press.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”?. *Psychological science*, *19*, 585-592.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current directions in psychological science*, *19*, 143-148.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets?. *Memory & Cognition*, *6*, 462-472.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*, 207.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual review of psychology*, *35*, 113-138.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, *32*, 89-115.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*, 39.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology*, *77*, 353.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, *104*, 192.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*, 573-605.

Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *Journal of Experimental Psychology: Applied*, *16*, 361.

Searston, R. A., & Tangen, J. M. (2017). Training perceptual experts: Feedback, labels, and contrasts. *Canadian Journal of Experimental Psychology*, *71*, 32.

Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive science*, *16*, 1-40.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts* (Vol. 9). Cambridge, MA: Harvard University Press.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study?. *Psychonomic bulletin & review*, *20*, 356-363.

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, *169*, 145.

Wittgenstein, L. (1953). *Philosophical Investigations* (G. E. M. Anscombe, Trans.) New York, NY: Macmillan

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews*.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16-27.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215-221.

VITA

Addison Leigh Poage Babineau was born on May 3, 1996, in Parker Colorado. She grew up with her family in Parker, Colorado and graduated from Legend High School in 2014. She went on to pursue her Bachelor of Science in Neuroscience at Colorado State University, graduating in 2018.

In August, 2018, she began working towards her Doctorate in Experimental Psychology at Texas Christian University. During her time at Texas Christian University she acted as a Research Assistantship supported by the James S. McDonnel Foundation 21st Century Science Initiative in Understanding Human Cognition, Collaborative Grant from 2018-2019, and received a Teaching Assistantship from 2019–2020. Under the mentorship of Dr. Sarah "Uma" K. Tauber, Addison primarily investigates student learning and instruction, as well as metacognition, memory and healthy cognitive aging.

ABSTRACT

THE ROLE OF FEEDBACK IN THE TRANSFER OF CATEGORY LEARNING

By: Addison Leigh Poage Babineau, B.S., 2018
Department of Psychology
Texas Christian University

Thesis Advisor: Dr. Sarah "Uma" K. Tauber, Associate Professor of Psychology

The present research investigated what type of feedback was most beneficial to category learning. The discriminative contrast hypothesis suggests that comparisons between categories enhances learning (e.g., Kornell & Bjork, 2008). As such, feedback that gave learners additional information and encouraged between-category comparisons with additional information (i.e., contrast feedback) was predicted to improve category learning more than feedback that only gave additional information (i.e., feature feedback), feedback that gave the correct answer (i.e., corrective feedback) or no feedback. To explore this, participants learned categories of organic chemistry compounds and received different types of feedback. The results revealed that providing feedback during category learning was beneficial for learning, and that providing additional information during feedback was beneficial to later performance above and beyond corrective feedback. The contrast feedback group did not perform better than did the feature feedback group; more research should investigate the degree to which contrast feedback improves later performance.