DISCRIMINATIVE CONTRAST AND THE ROLE OF CATEGORY ORGANIZATION ON CATEGORY LEARNING

By

ADDISON LEIGH POAGE BABINEAU

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

Master of Science, 2020
Texas Christian University
Fort Worth, Texas

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

Doctor of Philosophy in Experimental Psychology

May 2023

DISCRIMINATIVE CONTRAST AND THE ROLE OF CATEGORY ORGANIZATION ON CATEGORY LEARNING

by

ADDISON LEIGH POAGE BABINEAU

Dissertation Approved:

Dr. Sarah "Uma" K. Tauber, Major Professor

Dr. Kenneth Leising

Dr. Anna Petursdottir

Dr. Naomi Ekas

Dr. Cathy Cox

Dr. Tim Barth, for the College of Science & Engineering

ACKNOWLEDGEMENTS

I would like to begin by thanking my mentor, Dr. Sarah “Uma” Tauber. Dr. Tauber’s expertise and guidance have been indispensable to my success here at Texas Christian University. However, it is her constant kindness and support that have led to the wonderful graduate experience I have had for the past five years. I am incredibly grateful for the mentor/mentee relationship that I have with Dr. Tauber, and I am looking forward to learning more from her in the future. I would also like to thank Dr. Kenneth Leising, Dr. Anna Petursdottir, Dr. Naomi Ekas, and Dr. Cathy Cox for serving on my dissertation committee. Their feedback and expertise have been essential to the development of my dissertation, and I am grateful for their support.

Next, I would like to thank my friends and fellow graduate students, Dr. Julie Swets and Dr. Juliana Oliveira. Their friendship has been critical to my experience here at TCU and I am grateful to have found lifelong friends during my time here. As well, I would like to thank my partner, Cleyton Grinde for his love and support as I worked on my dissertation. Finally, I would like to thank my family. It is with their never-ending support that I can accomplish my goals, and I am grateful to have them cheer on my every accomplishment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Discriminative Contrast and the Role of Category Organization on Category Learning

We encounter a vast amount of information in our day-to-day lives, and we must organize that information to successfully interact with our environment. To organize this information, we develop concepts and categories. A concept is a mental representation of a thing, person, or idea. For example, we have a concept of what a rock is and have access to this conceptual information in the absence of external stimuli. Categories are the physical representations of concepts. For instance, we can hold a piece of rock in our hands and have the conceptual knowledge that the rock is a piece of granite. Categories are composed of exemplars, or examples for the category. For instance, we may encounter multiple pieces of granite, each of those pieces of granite is an exemplar.

Learning concepts is essential, and often challenging. This is particularly true in educational settings, in which students often rely on concepts from previous courses as building blocks for future learning, particularly in STEAM courses (Science, Technology, Engineering, Art, & Math). In the present work, participants learned to classify geology concepts (cf. Babineau et al., 2022). Learning geological concepts is an essential component of many introductory geology courses, and successfully learning geological concepts can impact a student's ability to pursue further education in geology (e.g., Marshak, 2015; Tarbuck & Lutgens, 2017). Geological categories can differ from each other in both structure and organization. Researchers are beginning to understand the impact of category structure on concept learning; however, minimal research has explored the impact of *category organization* (basic categories versus subcategories) on concept learning. First, I discuss how learning concepts differs from learning other types of material. Next, I discuss category structure and category organization. Then, I discuss how category structure can impact concept learning and how category organization could play an impactful and novel role in concept learning. Finally, I discuss a pilot experiment and two theoretically-motivated experiments aimed at exploring the best way to learn geological categories.

Concept Learning

Concept learning is often more challenging than other forms of semantic learning (e.g., Clary & Wandersee, 2006, Entwistle & Entwistle, 2003; Nosofsky & McDaniel, 2019). Consider learning basic semantic information in an educational setting, students may be asked to memorize specific facts about the content or recall details from a reading. In these rote memorization tasks, the important information is often directly provided, and students must recall the facts later on. For example, students in a geology course must learn that granite is formed through the slow cooling of molten rock. This is an example of a rote memorization task for learning facts. To succeed in rote memorization tasks, students can adopt a variety of strategies, some of which require minimal effort. Using simple strategies such as rote repetition of the information, students can successfully learn this type of semantic information (Roediger & Karpicke, 2006).

In contrast with learning basic semantic information, when learning concepts, students must develop an abstract understanding of the content, and accurately use that concept to interact with later material (Clary & Wandersee, 2006; Nosofsky & McDaniel, 2019). For instance, geology students need to develop an understanding for how to identify granite and use that knowledge to classify it in the field. Concepts are formed through inductive learning. Inductive learning involves making generalizations from a few exemplars to develop an understanding of the concept (Goldstone et al., 2018; Weiskopf, 2009). When learning geological concepts, students must determine which features are most essential to the category (Mervis & Rosch, 1981; Rosch 1975; Rosch & Mervis, 1975). For example, to learn how to classify granite, after observing several exemplars of granite a geology student would determine that granite typically has large, visible rock crystals that are light in color. Then, when shown rocks in the future, they would use the abstracted features to classify the rock as granite or a different type of rock. In this way, learning concepts involves developing knowledge of variations between categories across multiple dimensions, whereas rote memorization does not (Clary & Wandersee, 2006; Nosofsky & McDaniel, 2019). The aim of the present work is to determine the best way to learn these more complicated STEAM concepts. To fully understand STEAM concept learning, a category and its relationship to other to-be-learned categories must be considered. Two dimensions that are critical for all concept learning are the structure and organization of the category. These two dimensions vary concurrently, and often interact with each other.

Category Structure

Category structure can vary in two ways: distinctiveness and family-resemblance (Nosofsky et al., 2018; Rosch & Mervis, 1975). Distinctiveness refers to the degree of similarity between exemplars in one category as compared to exemplars in another category. To illustrate, consider a geology student learning to classify various categories of rocks. When studying, the student could compare an exemplar of granite (i.e., one category of rock) and an exemplar of gneiss (i.e., a different category of rock). Nosofsky et al. (2018) conducted a normative study to provide distinctiveness ratings for numerous rock categories. To do so, participants were shown pairs of exemplars from different rock categories (e.g., an exemplar of granite and an exemplar of gneiss). For each pair, participants judged how similar the rock exemplars were to each other on a 9-point scale (1 = most dissimilar to 9 =

most similar). These similarity ratings were used to calculate a distinctiveness rating for each rock category. A low rating of similarity between categories indicated that they were distinct – they had few features in common. In contrast, a high rating of similarity between categories indicated that they were similar (i.e., not distinct) – they had multiple features in common. Categories with low distinctiveness ratings are often more challenging to learn than are categories with high distinctiveness ratings, as it is harder to distinguish between categories with many features in common (Goldstone, 1996).

The other component of category structure, family-resemblance, refers to the degree of cohesiveness within a category. For instance, a geology student could compare a pair of granite exemplars (i.e., two exemplars from the same category of rock) when studying. To establish normative ratings of family-resemblance, Nosofsky et al. (2018) presented participants with pairs of rock exemplars from the same category (e.g., two exemplars of granite), and they judged how similar the exemplars were to each other on a 9-point similarity scale (1 = most dissimilar to 9 = most similar). From these judgments, the authors calculated the family-resemblance rating for each rock category. A low rating of similarity indicated that the category was low in family-resemblance – within category exemplars had few features in common. A high rating of similarity indicated that the category was high in family-resemblance – within category exemplars had many features in common. Categories with low family-resemblance ratings lack a cohesive set of features for identification and are often more challenging to learn than are categories that are cohesive with high family-resemblance ratings (Goldstone, 1996).

Taken together, category structure has a clear impact on concept learning. However, categories exist in a hierarchy of knowledge, and much of the prior research on concept learning has primarily focused on category structure at one level of the hierarchy. Importantly, no prior research has examined the impact of category structure on concept learning for other levels of the hierarchy. Exploring this novel issue was the primary aim of the current research.

Category Organization

Category organization refers to the location of the category within a hierarchy of information. Within a hierarchy, exemplars are classified into a basic category, which is composed of subcategories. For instance, sandstone is a subcategory that belongs to the basic category of sedimentary rocks (see Figure 1). Importantly, a geology student may need to learn to classify sandstone into the correct basic category of sedimentary rocks, or they may need to learn classify the sandstone into the correct subcategory of sandstone. An important feature of a categorical hierarchy is that the categories are formed by the same set of exemplars. In this way, exemplars may belong to the same basic category (e.g., sedimentary rocks) but belong to different subcategories (e.g., sandstone versus rock salt).

Organizing exemplars into basic categories or subcategories changes the degree of similarity that exists between them. For example, the rock exemplars used in the present studies belong to one of three basic categories: igneous, metamorphic, or sedimentary. Within each basic category, exemplars belong to one of three subcategories (see Figure 1). Grouping exemplars together at the subcategory levels results in higher ratings of family-resemblance than grouping categories together at the basic category level. This is because ratings of family-resemblance for basic categories also include distinctiveness ratings between exemplars from different subcategories. The similarity ratings between subcategories, even for subcategories within the same basic category, can be relatively low (i.e., they are distinct); thus, including them lowers the ratings of family-resemblance for a basic category. Taken together, a category's degree of distinctiveness and family-resemblance can be impacted by its organization within the hierarchy of knowledge. The interaction between category structure and category organization could significantly impact concept learning.

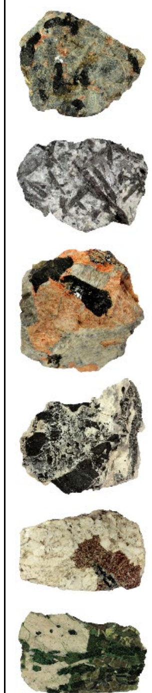
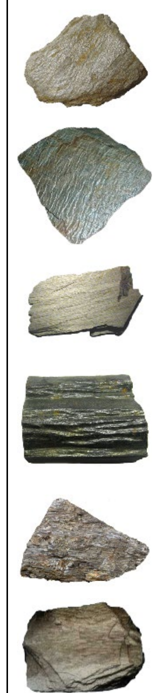
Superordinate Category	Rocks								
Basic Category	Sedimentary			Igneous			Metamorphic		
Subcategory	Sandstone	Rock Salt	Conglome-rate	Granite	Pegmatite	Peridotite	Gneiss	Marble	Phyllite
Exemplars									

Figure 1. Geological categories used in the present research. To determine category distinctiveness, exemplars from one category (e.g., sandstone) are contrasted with exemplars from another (e.g., conglomerate). To determine family-resemblance within a category, exemplars within a category (e.g., sandstone) are contrasted. Note that when viewing exemplars at the subcategory level there is a relatively high degree of family-resemblance, as compared to viewing the exemplars in a basic category.

Importantly, the distinction between basic categories and subcategories can have real implications for student learning. Introductory geology courses can emphasize learning basic geology categories over subcategories; however, research on concept learning has primarily focused on learning subcategories. Thus, there is a disparity between research on concept learning and the type of concepts that students are actually learning. This disparity is important to address, as it is unknown if students learn basic categories in the same manner as subcategories or if study strategies for learning subcategories can be extended to learning basic categories. Further, there is minimal research examining which strategies are best for learning basic level categories (e.g., Meagher et al., 2017; Miyatsu et al., 2020; Noh et al., 2014; Nosofsky et al. 2019; Palmeri, 1999; Tanaka et al., 2005). My goal was to systematically evaluate conceptual learning of categories with varying organizations to determine if study strategies that are effective for subcategory learning can be extended to basic category learning.

Study Strategies for Category Learning

According to the discriminative contrast hypothesis, study strategies that allow for contrasting exemplars can be critical for category learning (Carvalho & Goldstone, 2014). A learner can make two types of comparisons: between-category comparisons and within-category comparisons. A between-category comparison involves comparing an exemplar in one category (e.g., granite) to an exemplar in another category (e.g., gneiss). A within-category comparison involves comparing exemplars that belong to the same category (e.g., comparing an exemplar of granite to another exemplar of granite). The type of comparison that is best for category learning, is impacted by the distinctiveness and family-resemblance of the category.

Carvalho and Goldstone (2014) conducted a study in which they artificially manipulated the family-resemblance and distinctiveness of categories. One set of categories had high family-resemblance and low distinctiveness, and the other set of categories had low family-resemblance and high distinctiveness. Additionally, Carvalho and Goldstone (2014) manipulated the study order strategy for each set of categories. One set of categories was studied in a blocked order. Studying in a blocked order involves studying exemplars grouped together by category. For example, studying several exemplars of granite, then several exemplars of marble, and then several exemplars of sandstone, and so on. The other set of categories was studied in an interleaved order. Studying in an interleaved order involves mixing exemplars together from multiple categories. For example, studying one exemplar of granite, then one exemplar of marble, and then one exemplar of sandstone, and so on. The authors hypothesized that manipulating the order of the categories during study would encourage participants to look for different features of the categories during learning.

For categories with high family-resemblance and low distinctiveness, participants learned best when the exemplars were in an interleaved order. When learning categories with a low degree of distinctiveness, interleaving was more beneficial than blocking because it encouraged between-category comparisons, which helped participants discriminate between categories that had many features in common. By contrast, for categories with low family-resemblance and high distinctiveness, participants learned best when the exemplars were in a blocked order. When learning categories with low ratings of family resemblance, blocking was more beneficial than interleaving because it encouraged within-category comparisons, which helped participants identify the common feature for a category that had few commonalities across exemplars. Carvalho and Goldstone's (2014) findings support the discriminative contrast hypothesis and reveal that the most beneficial study order strategy for category learning, blocked study or interleaved study, can depend on the family-resemblance and distinctiveness of the to-be-learned categories.

Only one study has explored how study order strategy may influence learning subcategories as compared to basic categories. Miyatsu et al. (2020) conducted a study in which study order strategy was a secondary interest. In their third experiment, Miyatsu et al. (2020) had participants learn to classify basic categories or subcategories of rocks by studying in an entirely interleaved order. Then in the fourth experiment, they had additional participants complete the same task; however, the study phase was conducted in a pseudo-blocked order in which the odds of studying the same category twice were increased. No direct comparison of blocked study and interleaved study across the two experiments was provided, but there was little difference in classification performance between the two experiments. This may be due to a variety of reasons, including no random assignment and the use of the pseudo-blocking procedure, in which participants saw exemplars from different categories mixed together, similar to the interleaved order. Thus, the question of which study order strategy is more beneficial for concept learning for subcategories versus basic category remains to be answered.

Overview of Pilot Experiment, Experiment 1, and Experiment 2

The main goal of the current studies was to determine the best way to learn basic categories and subcategories. Thus, the present studies are the first to directly compare learning for subcategories and basic categories while manipulating study order strategy. According to the discriminative contrast hypothesis, between-category comparisons (encouraged by studying in an interleaved order) should be most beneficial for learning categories with high family-resemblance and low distinctiveness (Carvalho & Goldstone, 2014). Consistent with prior research, when learning to classify subcategories of concepts, which typically have high ratings of family-resemblance, interleaving should be most beneficial for classification performance (Carvalho & Goldstone, 2014). In contrast, the discriminative contrast hypothesis suggests that within-category comparisons (encouraged by studying in a blocked order) should be most beneficial for learning categories with low family-resemblance and high distinctiveness (Carvalho & Goldstone, 2014). Therefore, when learning to classify basic categories, which typically have low ratings of family-resemblance, I hypothesized that blocking should be most beneficial for classification performance.

Pilot Experiment

I used geological categories in the pilot study, Experiment 1, and Experiment 2 (cf. Babineau et al., 2022) because they (a) are consistent with information geology students need to learn in authentic classroom contexts, and (b) have been previously normed to establish measures of distinctiveness and family-resemblance (Nosofsky et al., 2018). Despite prior research that has found interleaving to be beneficial for subcategory learning (Kornell & Bjork, 2008), researchers have not previously examined study order strategy when learning *geological* subcategories. Thus, the aim of the pilot experiment was to determine if the interleaving benefit commonly observed for learning subcategories replicates with geological stimuli. All experiments were preregistered with the Open Science Framework (OSF) prior to data collection (https://osf.io/ypksg/), and all data and materials are freely available.

Method

Participants

A between-participants design was implemented with two study order strategies (blocked & interleaved). A power analysis was conducted using G*Power to estimate the number of participants necessary for each group (Faul et al., 2007). Using an effect size of d = .94, (Kornell & Bjork, 2008, Experiment 1b) with alpha error at .05, and power at .95, I estimated that 31 participants would be sufficient for each group for a well-powered Pilot experiment.

A total of 70 undergraduate Texas Christian University students were randomly assigned to study order strategy group and participated for partial credit in psychology courses. A technological error resulted in incomplete data for eight participants, and their data were excluded from analyses. Thus, a total of 62 participants (n = 31 per group) were included in the final data set.

The sample was college-aged (blocked group, M = 19.65 years, SE = .34; interleaved group, M = 18.87 years, SE = .43), and most participants identified as women (blocked group, n = 23, 74.2%, 8 men; interleaved group, n = 30, 96.8%, 1 man) and White (blocked group, n = 24, 77.4%, 3 Asian, 2 White and Latino, 1 Black, & 1 Asian and other; interleaved group, n = 24, 77.4%, 3 Latino, 1 White and Latino, & 3 preferred not to respond). The groups did not significantly differ on most demographic characteristics (age, $t(60) = 1.41$, $p = .16$; ethnicity, χ^2 (6) = 11.33, $p = .08$); although significantly more participants in the interleaved group identified as women than did those in the blocked group, χ^2 (1) = 6.37, $p = .012$. Few participants reported experience in majors or hobbies related to geology (see Table 2). There was no significant difference in self-reported prior experience in geology between groups, χ^2s ≤ 2.07, ps ≥ .151, and the groups did not significantly differ in participants' self-rated knowledge of geology, $t(60) = 1.69$, $p = .10$.

Materials

Table 1. *Normative Ratings of Family-resemblance and Distinctiveness for Subcategories and Basic Categories that were Used in the Pilot Experiment, Experiment 1 & Experiment 2*

Basic Category	Basic Category Family-resemblance	Basic Category Distinctiveness	Subcategory	Subcategory Family-resemblance	Subcategory Distinctiveness
Igneous	4.48	4.04	Granite	6.59	4.28
			Pegmatite	5.41	3.92
			Peridotite	6.01	4.25
Metamorphic	4.27	3.95	Gneiss	6.58	4.11
			Marble	5.65	4.01
			Phyllite	5.43	3.97
Sedimentary	3.86	3.88	Conglomerate	5.88	4.03
			Rock salt	6.05	3.77
			Sandstone	5.46	3.52

Note. Ratings were made on a scale of 1 to 9 (1 = most dissimilar to 9 = most similar). Thus, high ratings of family-resemblance indicate that exemplars within the category are highly similar. In contrast, low ratings of distinctiveness indicate that the exemplars in the category are highly distinct from other categories. Values indicate the mean ratings for each category, calculated from Nosofsky et al. (2018).

Participants learned to classify nine subcategories of rocks. The subcategories were selected such that three subcategories belonged to each basic category (see Figure 1). The subcategories can be found in Table 1. For each subcategory, nine exemplars were randomly assigned to the study phase and the remaining three exemplars were assigned to the novel classification test for each participant. Thus, participants viewed 81 exemplars during the study phase and studied classification test, and 27 exemplars during the novel classification test.

Each rock subcategory was previously normed by Nosofsky et al. (2018). A paired sample t-test revealed that subcategories had a significantly higher degree of family-resemblance (M = 5.90, SE = .15) than distinctiveness (M = 3.99, SE = .08), $t(8)$ = 15.49, p <

.001, $d = 5.24$. In other words, exemplars in a subcategory had more features in common with other exemplars in the same subcategory than with exemplars from different subcategories. Further, a repeated measures analysis of variance (ANOVA) was used to explore the degree of distinctiveness between subcategories. The results indicated that the degree of distinctiveness between subcategories did not significantly differ, $F(8, 56) = 1.89$, $p = .079$; thus, no subcategory was significantly more distinct than any other subcategory.

Procedure

The Pilot experiment was conducted remotely due to the COVID-19 pandemic. Participants completed the study individually with a trained research assistant using the videoconferencing software Zoom (Zoom Video Communications Inc., 2016). Participants entered all their own responses, and the research assistant was present throughout the duration of the study to ensure that participants followed directions and to answer participants' questions. To begin the experiment, participants completed an informed consent and demographic survey. Next, participants answered five prior experience questions. The prior experience questions (see Table 2) were included to determine participants' previous experience with and interest in geology. Of the prior experience questions, four questions were answered with a fixed yes/no response, and one question was answered on a fixed Likert scale (i.e., 1 to 7). Participants were allowed to take as long as they liked to answer each question, and the prior experience questions were shown one-at-a-time, in a fixed order.

Table 2. *Participants' Reponses to each Prior Experience Question for the Pilot Experiment*

Question	**Pilot Experiment**	
	Blocked Group	**Interleaved Group**
1. Are you currently or have you ever been a Geology Major or Minor?	0%	0%
2. Are you currently or have you ever been an Applied Geoscience Major?	0%	0%
3. Do you collect rocks?	0%	*n* = 2, 6.5%
4. Do you own a rock and mineral field guide?	0%	0%
5. What is your own rated level of expertise at identifying rocks?	1.2 (.07)	1.4 (.14)

Note. Values for questions 1- 4 indicate the percentage of participants that responded with "yes" for the prior geology experience questions. Values for question 5 indicate the *M* (*SE*) on a scale from 1 to 7 (i.e., 1 = novice to 7 = expert).

Following the prior experience questions, participants answered six prior knowledge questions. During the prior knowledge questions, participants were asked to classify exemplars into their correct subcategory, as well as their correct basic category. Prior knowledge items were structured to include subcategory knowledge and basic category knowledge in preparation for the following experiments, in which category organization (i.e., basic category vs. subcategory) was manipulated. To begin the prior knowledge questions, participants were randomly shown one of three exemplars (basalt, breccia, migmatite), one-at-a-time, in the center of the screen. Below the exemplar, nine subcategory names (basalt, breccia, bituminous coal, chert, diorite, gabbro, hornfels, migmatite, schist) were shown in a fixed alphabetical order for all participants. There was also an "*I don't know*" button. Participants were instructed to select "*I don't know*" if they were unsure how to classify the exemplar. After responding to the subcategory classification question, participants responded to the basic category classification question. To do so, the same exemplar remained on the screen and participants were instructed to classify the exemplar by selecting the corresponding button for the basic category type (igneous, metamorphic, or sedimentary), or to select the *"I don't know"* option. Participants proceeded through the prior knowledge questions in this order until each of the three prior knowledge exemplars had been classified into both a subcategory and a basic category. The order of the exemplars was randomized for each participant, and participants took as long as they needed on each question. No feedback was given during any portion of the prior knowledge questions. None of the prior knowledge exemplars or the subcategories (either the correct subcategories or the lure subcategories) were included in the study phase or test phase of the experiment. In this way, the prior knowledge questions revealed participants' existing knowledge of rock classifications without interacting with the rock subcategories used in the experimental task.

After completing the prior knowledge questions, the participants began the study phase. Participants were informed that their goal was to learn the categories well enough so they could classify new exemplars into the categories on an upcoming test. In the blocked group, participants studied the exemplars in groups by subcategory. Specifically, the study phase exemplars for each subcategory were shown consecutively such that participants studied all of the exemplars in one subcategory before studying all of the exemplars in another subcategory. The order of the exemplars within each subcategory was randomized for each participant. Further, the order of the nine subcategories within the study phase was randomized for each participant. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding subcategory label shown beneath it. Each exemplar was presented for 3s before the next exemplar was automatically presented. There was a brief .5s inter-stimulus-interval between each exemplar that consisted of a blank white screen.

The study phase was nearly identical for participants in the interleaved group; however, the order of the study phase exemplars differed. Those in the interleaved group studied the subcategories in a pseudo-random order such that the exemplars were distributed into 9 groups with one exemplar from each subcategory randomly allocated to each group for each participant. The order of the exemplars within each group was randomized per participant; however, two exemplars from the same subcategory were never shown consecutively. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding subcategory label shown beneath it. Each exemplar was presented for 3s before the next exemplar was automatically presented. There was a brief .5s inter-stimulus-interval between each exemplar that consisted of a blank white screen. Thus, the duration of the study phase was held constant between the interleaved and blocked groups.

After completing the study phase, participants in both groups completed a brief 15s distractor task. During the distractor task, participants counted out loud, backwards by threes from 479. Then, participants in both groups began the test phase.

The test phase consisted of two tests, the novel classification test and the studied classification test. The order of the two tests was counterbalanced between-participants. During the novel classification test, participants were shown a never-before-seen exemplar belonging to one of the subcategories that was learned during the study phase. The names of each subcategory were displayed on buttons in alphabetical order beneath the novel exemplar. Participants were instructed to classify each exemplar into its correct subcategory by selecting the button with the corresponding subcategory label. Each novel test exemplar was presented one-at-a-time, in the center of the screen, and participants took as much time as needed to make their classification. Participants did not receive any feedback on their answers, and the order of the exemplars on the novel test was random for each participant. The "*I don't know*" option was not present during either the novel classification test or the studied classification test; thus, participants were required to respond to each exemplar on the test. The studied classification test was nearly identical to the novel classification test; however, the exemplars presented were the same as the exemplars shown during the study phase.

After completing both classification tests, participants responded to the follow-up question probing for their beliefs about study order strategy. Participants were informed of the two study order strategies (i.e., interleaved or blocked) and were provided with a brief example of each. Participants in the blocked condition were shown the following:

> "*In this study you learned how to classify various rocks into their correct categories. During the study phase, you were shown the* ***examples of each rock type grouped together by category****. For instance, you may have studied several examples of Marble, followed by several examples of Peridotite, and then several examples of Gneiss, etc.*
>
> *Other students in the experiment studied the examples of each rock type in a different order. They studied the* ***examples of each rock type intermixed between categories****. For instance, they may have studied one example of Marble, followed by one example of Peridotite, and then one example of Gneiss, etc.*
>
> *Importantly, these students studied the same number of rocks as you, the only difference was the order of their study*."

Participant in the interleaved condition were shown the following:

> "*In this study you learned how to classify various rocks into their correct categories. During the study phase, you were shown* ***the examples of each rock type intermixed***

between categories. *For instance, you may have studied one example of Marble, followed by one example of Peridotite, and then one example of Gneiss, etc. Other students in the experiment studied the examples of each rock type in a different order. They studied the* ***examples of each rock type grouped together by category***. *For instance, they may have studied several examples of Marble, followed by several examples of Peridotite, and then several examples of Gneiss, etc. Importantly, these students studied the same number of rock as you, the only difference was the order of their study.*"

Participants were asked which order they thought would help them learn best. Participants selected one of three options: "*studying examples of each rock type grouped together by category*" (i.e., blocked), *"studying examples of each rock type mixed together"* (i.e., interleaved), or *"my learning would have been about the same with either study order*". The three options were shown in a random order on the screen for each participant. After completing the follow-up question, participants were debriefed, thanked for their time, and granted partial credit in psychology courses.

Pilot Experiment Results

Analyses are reported in the order in which participants completed each task. First, participants' performance on the prior knowledge questions is reported. Then participants' performance on the novel and studied classification tests is examined between groups, as well as between test counterbalance orders. Participants' responses to the belief follow-up question are reported last.

Prior Knowledge

Participants' performance on the prior knowledge questions can be found in Figure 2. Subcategory performance and basic category performance were calculated separately for each participant ("*I don't know*" responses were counted as incorrect). Averages were calculated separately for subcategory performance and basic category performance and aggregated for each group. Most important, participants' mean prior knowledge performance ($M = .09$, $SE = .02$) indicated low levels of geology classification knowledge. Participants in the interleaved group responded with "*I don't know*" on a majority ($M = .61$, $SE = .06$) of the prior knowledge items (subcategory items, $M = .79$, $SE = .07$; basic category items, $M = .44$, $SE = .07$). Participants in the blocked group responded with "*I don't know*" on a majority ($M = .68$, $SE = .05$) of the prior knowledge items (subcategory items, $M = .84$, $SE = .05$; basic category items, $M = .53$, $SE = .07$). Further, one-sample *t*-tests revealed that both groups performed significantly below chance for the subcategory items (chance levels of performance = .11; $t\text{s} \geq 4.38$, $p\text{s} < .001$) and significantly below chance for the basic category items (chance levels of performance = .33; $t\text{s} \geq 3.72$, $p\text{s} < .001$).

A between-participants *t*-test comparing the two study order strategies (blocked group, interleaved group) revealed that prior knowledge classification performance for the subcategory items did not significantly differ between the interleaved group ($M = .03$, $SE = .02$) relative to the blocked group ($M = .03$, $SE = .02$), $t(60) < .001$, $p = 1.00$. In addition, prior knowledge classification performance for the basic category items did not significantly differ between the interleaved group ($M = .17$, $SE = .04$) relative to the blocked group ($M = .13$, $SE = .04$), $t(60) = .73$, $p = .47$.

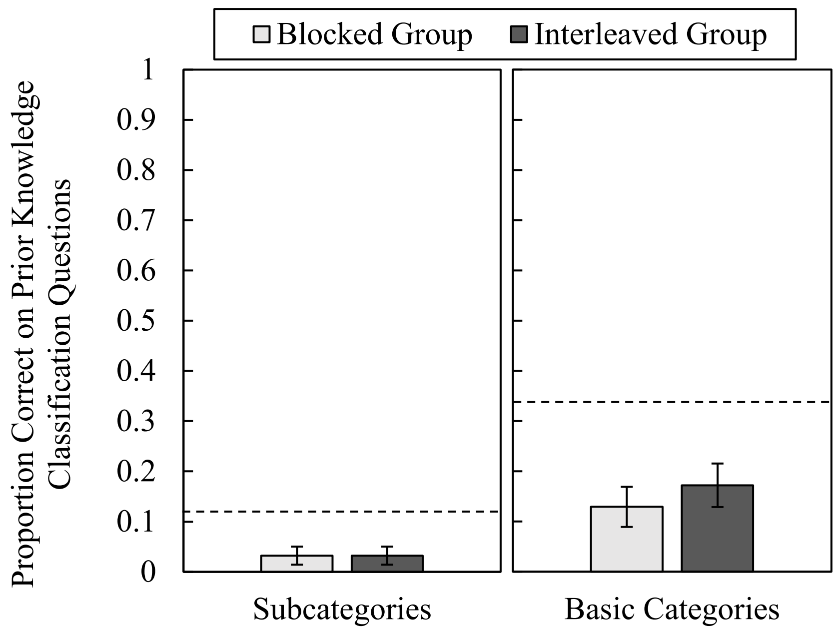


Figure 2. Mean proportion correct for prior knowledge classification questions for participants in the Pilot experiment. Errors bars represent the standard error of the mean. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance is .11, whereas chance levels of performance for basic categories is .33.

Novel Classification Test Performance

Participants' performance on the novel classification test can be found in Figure 3 (left side). The proportion correct was calculated separately for each participant and then averaged per group. Classification performance for all classification tests (Pilot Experiment, Experiment 1, & Experiment 2) was calculated in the same manner as used for the pilot experiment. Participants' performance on the novel classification test was examined between groups, as well as between counterbalanced test orders.

A 2 (study order strategy: blocked group, interleaved group) x 2 (counterbalance: novel first, novel second) between-participants ANOVA revealed that the interleaved group performed numerically better on the novel classification test than did the blocked group, although the results were not statistically significant, $F(1, 58) = 4.02$, $p = .050$, $\eta^2 = .06$. There was not a significant main effect of the counterbalanced test order, $F(1, 58) = .069$, $p = .794$; however, this was qualified by a significant interaction between study order strategy and counterbalance, $F(1, 58) = 8.71$, $p = .005$, $\eta^2 = .12$. Follow-up analyses revealed that test order did not significantly impact novel classification performance for participants in the interleaved group (Novel first, $M = .64$, $SE = .04$; Novel second, $M = .53$, $SE = .03$), $t(29) = 1.95$, $p = .061$. However, test order did matter for participants in the blocked group such that those who took the novel test second ($M = .57$, $SE = .04$) performed significantly better on it than did those who took the novel test first ($M = .44$ $SE = .04$), $t(29) = 2.22$, $p = .035$, $d = .83$. Thus, test order had a significant impact on participants' novel classification performance.

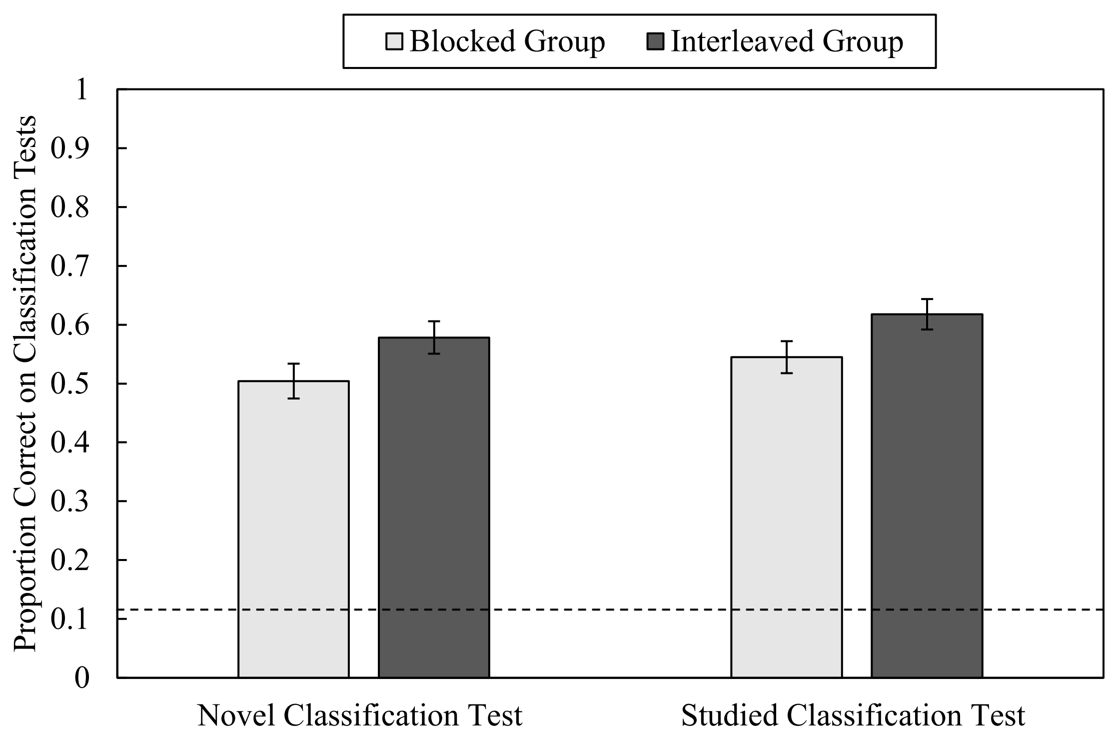


Figure 3. Mean proportion correct on the novel classification test and studied classification test for participants in the Pilot experiment. Error bars represent standard error of the mean. The dashed line represents chance levels of performance, which was .11.

Studied Classification Test Performance

Participants' performance on the studied classification test can be found in Figure 3 (right side). Studied classification test performance was examined between groups as well as the counterbalanced test order. A 2 (study order strategy: blocked group, interleaved group) x 2 (counterbalance: studied first, studied second) between-participants ANOVA revealed that the interleaved group performed significantly better on the studied classification test than did the blocked group, $F(1, 58) = 4.36$, $p = .041$, $\eta^2 = .06$. There was not a significant difference between counterbalanced test order, $F(1, 58) = .004$, $p = .947$. However, this main effect was qualified by a significant interaction between study order strategy and counterbalance, $F(1, 58) = 6.63$, $p = .013$, $\eta^2 = .096$. Follow-up analyses revealed that for participants in the blocked group, those who took the studied test first ($M = .59$, $SE = .04$) did somewhat better on it than did those who took the studied test second ($M = .50$, $SE = .03$), though the difference was not statistically supported, $t(29) = 1.83$, $p = .078$. By contrast, for participants in the interleaved group, those who took the studied test first ($M = .58$, $SE = .04$) did somewhat worse on it than did those who took the studied test second ($M = .67$, $SE = .04$) though the difference was not statistically supported, $t(29) = 1.82$, $p = .080$. Thus, the interaction was driven by a numerical difference in performance for the blocked group dependent upon test order that was apparent in the other direction for the interleaved group.

Beliefs about Study Order Strategy

Participants' responses to the follow-up question about their beliefs can be found in Table 3. Consistent with prior research (Kornell & Bjork, 2008; Tauber et al., 2013; Yan et al., 2016; for an individual difference perspective, see Babineau et al., 2022), most participants (71%) indicated that they thought studying in a blocked order would help them learn geological categories best. Participants' responses to the belief follow-up question did not significantly differ between the blocked and interleaved groups, χ^2 (2) = 1.20, p = .548.

Table 3. *Participants' Reponses to the Beliefs about Study order Strategies Question for the Pilot Experiment*

	Pilot Experiment	
"Which study order do you think would help you learn best?"	**Blocked Group**	**Interleaved Group**
Studying examples of each rock type grouped together by category.	67.74%	74.19%
Studying examples of each rock type mixed together.	12.90%	16.13%
My learning would have been about the same with either study order.	19.36%	9.68%

Note. Values indicate the percentage of participants who selected each response for the follow-up question probing for participants' beliefs about study order strategies.

Pilot Experiment Discussion

When learning to classify geological subcategories, participants who learned in an interleaved order performed better on the studied classification test than did participants who learned in a blocked order. Further, participants who learned in an interleaved order performed numerically better on the novel classification test than did participants who learned in a blocked order. These findings lend support to the discriminative contrast hypothesis (Carvalho & Goldstone, 2014). The subcategories used in the Pilot experiment had a higher degree of family-resemblance than degree of distinctiveness. Learning such categories, according to the discriminative contrast hypothesis, will typically be benefitted by interleaving as it encourages participants to make between-category comparisons. Taken together, the findings from the Pilot experiment are consistent with prior research on subcategory learning (Kornell & Bjork, 2008).

Whereas studied classification performance was significantly better when participants studied in an interleaved order than in a blocked order, there was not a significant difference between study order strategies for the novel classification test. This was surprising because prior research examining study order strategies has typically used novel classification performance to investigate performance differences between study order strategies. Although an a priori power analysis was conducted to ensure adequate power for the pilot experiment, the power analysis was conducted using effects from prior research that employed different stimuli (i.e., paintings; Kornell & Bjork, 2008). Thus, when learning geological categories, the effect of study order strategy may be smaller and require more power to detect. In Experiment 1, the sample size was determined using the effect size from the pilot experiment, which resulted in a well-powered experiment to detect study order effects for geological materials.

Though not predicted, the order of the classification tests had a significant impact on novel classification performance for participants in the blocked group, but not for participants in the interleaved group. Participants in the blocked group who completed the novel test second performed better on it than did those who completed the novel test first. Test order also numerically impacted performance on the studied classification test. These outcomes suggest that when using both a novel classification test and a studied classification test, researchers should consider the influence of test order on classification performance. This is particularly important given that previous work has designed classification tests by mixing novel and studied exemplars on the same test (Miyatsu et al., 2020), fixing test order for all participants (Tauber et al., 2013), or only using one type of classification test (Kornell & Bjork, 2008). Thus, in Experiments 1 and 2, test order was examined though analysis of the test order counterbalance.

Experiment 1

One aim of Experiment 1 was to replicate the Pilot experiment with a new sample and larger sample size. Specifically, I investigated the impact of study order strategies on subcategory learning. Replicating the Pilot experiment, I hypothesized that participants learning subcategories would perform better on the novel classification test when learning in an interleaved order relative to a blocked order. Further, I anticipated that participants learning subcategories would perform better on the studied classification test when learning in an interleaved order relative to a blocked order.

In addition to replicating the Pilot experiment, I extended the design to include basic category learning. Far less is known about which study order strategy is best when participants learn to classify basic categories. As suggested by the discriminative contrast hypothesis, I hypothesized that participants learning basic categories would perform better on the novel classification test when learning in a blocked order relative to an interleaved order. I also hypothesized that participants learning basic categories would perform better on the studied classification test when learning in a blocked order relative to an interleaved order.

Importantly, the classification tests differed between the subcategory groups and the basic category groups. The subcategory groups learned to classify nine rock categories, therefore on the novel and studied classification tests participants made classification attempts out of nine potential categories (chance levels of performance = .11). The basic category groups learned to classify three rock categories, therefore on the novel and studied classification tests participants made classification attempts out of three potential categories (chance levels of performance = .33). In this way, comparing performance classifying subcategories to performance classifying basic categories would not be statistically valid.

Thus, I did not statistically compared performance for the subcategory groups relative to performance for the basic category groups for the novel classification test and studied classification test. Instead, study order was investigated independently for the subcategory groups and basic category groups.

Method

Participants

Study order strategy (blocked, interleaved), and category organization (subcategories, basic categories) were manipulated in a between-subjects design. Using an effect size of d = .47, (Pilot experiment, novel classification performance main effect) with alpha error at .05, and power at .85, a power analysis was conducted with G*Power, which estimated that 83 participants randomly assigned to each group would be sufficient for a well-powered experiment (Faul et al., 2007). A total of 332 undergraduate Texas Christian University students participated for partial credit in psychology courses, with 83 participants randomly assigned to each group.

Demographic characteristics for participants in Experiment 1 are displayed in Table 4. The groups did not significantly differ on demographic characteristics (age, $F(3, 327) = 1.33$, $p = .264$; gender, χ^2 (12) = 11.27, $p = .506$; ethnicity, χ^2 (36) = 35.81, $p = .478$). Further, few participants reported experience in majors or hobbies related to geology (see Table 5). There was no significant difference in self-reported prior experience in geology between groups, χ^2s ≤ 3.01, ps $\geq .390$, and the groups did not significantly differ in participants' self-rated knowledge of geology, $F(3, 328) = .21$, $p = .892$.

Table 4. *Demographic information for participants in Experiment 1 & Experiment 2*

Groups	Age	Gender Identification	Ethnic Identification
Experiment 1			
Blocked Subcategories group	19.31 (.12)	*n* = 60 Women, 23 Men	*n* = 63 White, 9 Latino, 4 Latino & White, 3 Black, 3 Asian, 1 Asian & White
Interleaved Subcategories group	19.29 (.14)	*n* = 60 Women, 22 Men, 1 Woman & Gender Diverse	*n* = 66 White, 6 Latino, 4 Latino & White, 2 Black, 2 Asian & White, 1 Asian, 1 Native American, 1 Asian, Native American, & Latino
Blocked Basic categories group	19.50 (.18)	*n* = 64 Women, 16 Men, 2 Woman & Gender Diverse, & 1 no response	*n* = 57 White, 8 Latino, 6 Black, 6 Asian, 2 Asian & White, 1 Latino & White, 1 Native American & Latino, 1 White, Asian, & Pacific Islander, 1 no response
Interleaved Basic categories group	19.72 (.23)	*n* = 62 Women, 20 Men, 1 Woman & Gender Diverse	*n* = 60 White, 9 Latino, 5 Latino & White, 5 Asian, 1 Asian & White, 1 White & Black, 1 White, Asian, & Latino, 1 Hispanic
Experiment 2			
Blocked Basic categories group	19.06 (.12)	*n* = 47 Women, 29 Men, 5 Woman & Gender Diverse, 1 Woman & Man, 1 no response	*n* = 61 White, 4 Hispanic, 4 Asian, 3 Hispanic & White, 3 Black, 2 Pacific Islander & White, 1 Hispanic, Native American & White, 1 Latino, 1 Native American & White, 1 unsure, 1 no response, 1 listed a specific country
Interleaved Basic categories group	19.10 (.18)	*n* = 61 Women, 18 Men, 2 Woman & Gender Diverse, 1 gender Diverse, 1 no response	*n* = 62 White, 4 Latino, 4 Hispanic, 2 Latino & White, 2 Hispanic & White, 2 Black, 1 Hispanic & Native American, 1 Black & White, 1 Black & Middle Eastern, 1 Asian, 1 Asian & White, 2 no response
Blocked Equated Subcategories group	18.96 (.29)	*n* = 62 Women, 16 Men, 2 Gender Diverse, 2 Woman & Man, 1 Woman & Gender Diverse	*n* = 55 White, 6 Hispanic, 3 Black, 3 Asian, 2 Pacific Islander, 2 Latino, 2 Hispanic & White, 2 Asian & White, 1 Native American & White,1 Latino & White, 1 Black & White, 1 Black & Hispanic, 1 Asia, Pacific Islander & White, 2 listed a specific country, 1 no response
Interleaved Equated Subcategories group	18.95 (.18)	*n* = 62 Women, 18 Men, 2 Woman & Man, 1 Woman & Gender Diverse	*n* = 57 White, 5 Hispanic, 5 Black, 5 Latino, 4 Asian, 3 Hispanic & White, 1 Middle Eastern, 1 Asian & White, 1 Latino & White
Blocked Equated Basic categories group	18.70 (.11)	*n* = 57 Women, 19 Men, 3 Woman & Gender Diverse, 2 Gender Diverse, 1 Woman & Man, 1 no response	*n* = 60 White, 5 Black, 3 Hispanic & White, 2 Hispanic, 2 Latin, 2 Asian & White, 1 Pacific Islander & White, 1 Asian & Latino, 1 Native American & White, 1 Black & White, 1 Black & Hispanic, 1 Asian, 1 Latino & White, 1 Hispanic, Middle Eastern & White, 1 listed a specific country, 1 no response
Interleaved Equated Basic categories group	18.91 (.17)	*n* = 64 Women, 13 Men, 4 Woman & Gender Diverse, 2 Woman & Man	*n* = 62 White, 8 Black,4 Hispanic, 3 Asian, 1 Black & Native American,1 Middle Eastern, 1 Asian & White, 1 Latino & White, 1 Asian, Pacific Islander & White, 1 listed a specific country

Note. Values for Age indicate the *M* (*SE*) years of age for each group. Values for Gender Identification and Ethnic Identification indicate the number of participants from each group who reported each identity.

Table 5. *Participants' Reponses to each Prior Experience Question for the Experiment 1*

Question	Subcategory Groups		Basic Category Groups	
	Blocked Group	**Interleaved Group**	**Blocked Group**	**Interleaved Group**
1. Are you currently or have you ever been a Geology Major or Minor?	0%	*n* = 1, 1.2%	0%	0%
2. Are you currently or have you ever been an Applied Geoscience Major?	0%	0%	0%	0%
3. Do you collect rocks?	*n* = 3, 3.6%	*n* = 3, 3.6%	*n* = 2, 2.4%	*n* = 5, 6.0%
4. Do you own a rock and mineral field guide?	*n* = 1, 1.2%	0%	*n* = 1, 1.2%	*n* = 1, 1.2%
5. What is your own rated level of expertise at identifying rocks?	1.42 (.07)	1.48 (.08)	1.39 (.07)	1.45 (.09)

Note. Values for questions 1- 4 indicate the percentage of participants that responded with "yes" for the prior geology experience questions. Values for question 5 indicate the *M* (*SE*) on a scale from 1 to 7 (i.e., 1 = novice to 7 = expert).

Materials

Participants learned to classify the categories of rocks used in the Pilot experiment (see Table 1). The subcategories were identical to those in the Pilot experiment. The basic categories (i.e., igneous, metamorphic, sedimentary) were each composed of three subcategories (see Figure 1). For each subcategory, nine exemplars were randomly assigned to the study phase, and the remaining three exemplars were assigned to the novel classification test for each participant. Thus, participants viewed 81 exemplars during the study phase and studied classification test, and 27 exemplars during the novel classification test, regardless of their category organization (i.e., subcategory or basic category). Importantly, the level at which the exemplars were classified – subcategory or basic category – was the only difference between groups.

Stimuli were selected from Nosofsky et al. (2018) so that subcategories had a significantly higher degree of family-resemblance than distinctiveness, and basic categories did not have a significant difference between ratings of family-resemblance and ratings of distinctness, $t(8) = 1.72$, $p = .123$. In other words, I selected exemplars in a basic category that were as similar to each other as they were to exemplars in other basic categories. Further, I selected stimuli for which the degree of family-resemblance did not significantly differ between the basic categories, $F(2, 6) = 1.44$, $p = .308$. Likewise, the degree of distinctiveness did not differ between basic categories, $F(2, 6) = .47$, $p = .648$. Thus, a given basic category was not significantly more cohesive or more distinct than any other basic category. Additionally, subcategories and basic categories did not significantly differ in degree of distinctiveness, $t(8) = .78$, $p = .457$. In other words, subcategories were not more distinct than were basic categories. However, I selected subcategories that had a significantly higher degree of family-resemblance relative to basic categories, $t(8) = 10.83$, $p < .001$. In this way, subcategories were comprised of exemplars that had more similarities between them than did basic categories.

Procedure

Experiment 1 was conducted during the COVID-19 pandemic. Thus, similar to the Pilot experiment, data collection for Experiment 1 occurred virtually, during which participants completed the study individually with a trained research assistant using Zoom (Zoom Video Communications Inc., 2016). Virtual participants entered all their own responses, and the research assistant was present throughout the duration of the study to ensure that participants follow directions and to answer participants' questions. These virtual data collection procedures have been adapted for category learning research over the 2.5 years and have been successful in maintaining participant engagement and research integrity (e.g., Ariel, et al., under review; Babineau & Tauber, 2022; Witherby et al., 2022).

For participants randomly assigned to learn subcategories, the procedure was identical to the Pilot experiment. For participants assigned to learn basic categories, the procedure was nearly identical to the Pilot experiment; however, participants learned to classify exemplars into their basic categories rather than into their subcategories.

During the study phase, participants assigned to learn basic categories in a blocked order studied the exemplars in groups by basic category. Specifically, the study phase exemplars for each basic category were shown consecutively such that participants studied all of the exemplars in one basic category before studying all of the exemplars in another basic category. The order of the exemplars within each basic category was randomized for each participant. Further, the order of the 3 basic categories within the study phase was also randomized for each participant. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding basic category label displayed beneath it. Each exemplar was presented for 3s before the next exemplar was automatically presented. There was a brief .5s inter-stimulus-interval between each exemplar that consisted of a blank white screen.

During the study phase, participants assigned to learn basic categories in an interleaved order studied the exemplars in a pseudo-random order. Specifically, participants studied the basic categories mixed together, such that the order of the exemplars within each group was randomized per participant; however, two exemplars from the same basic category were never shown consecutively. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding basic category label shown beneath it. Each exemplar was presented for 3 seconds before the next exemplar was automatically presented. There was a brief .5 second inter-stimulus-interval between each exemplar that consisted of a blank white screen.

During the novel classification test, participants assigned to learn basic categories were shown a never-before-seen exemplar belonging to one of the basic categories. The names of each basic category were displayed on buttons in alphabetical order beneath the novel exemplar. Participants were instructed to classify each exemplar into its correct basic category by selecting the button with the corresponding basic category label. Each novel test exemplar was presented one-at-a-time, in the center of the screen, and participants were allowed to take as much time as they would like to make their classification. Participants did not receive any feedback on their answers, and the order of the exemplars on the novel test were random for each participant. The "*I don't know*" option was not present during either the novel classification test or the studied classification test. Thus, participants were required to respond to each exemplar in the test phase. The studied classification test for basic categories was nearly identical to the novel classification test; however, the exemplars presented were the same as the exemplars shown during the study phase. After completing both classification tests, participants assigned to learn basic categories responded to the follow-up question probing for their beliefs about study order strategy. Nearly all aspects of the follow-up question were identical to those used in the pilot experiment; however, the examples demonstrating an interleaved study order as compared to a blocked study order involved basic categories.

Experiment 1 Results

Analyses are reported in the order in which participants completed each task. First, participants' performance on the prior knowledge questions are reported. Then participants' performance on the novel and studied classification tests are reported between groups, as well as between test counterbalance orders. Participants' responses to the belief follow-up question are reported last.

Prior Knowledge Performance

Participants' performance on the prior knowledge questions can be found in Figure 4. Subcategory performance and basic category performance were calculated in the same manner as the Pilot experiment. Participants' mean prior knowledge performance ($M = .13$, $SE = .01$) indicated low levels of geology classification knowledge. Further, participants responded with "*I don't know*" on a majority of the prior knowledge items (blocked subcategories group, overall, $M = .64$, $SE = .04$, subcategory items, $M = .82$, $SE = .03$; basic category items, $M = .47$, $SE = .05$; interleaved subcategories group, overall, $M = .53$, $SE = .04$, subcategory items, $M = .69$, $SE = .04$; basic category items, $M = .37$, $SE = .04$; blocked basic categories group, overall, $M = .62$, $SE = .04$, subcategory items, $M = .74$, $SE = .04$; basic category items, $M = .49$, $SE = .05$; interleaved basic categories group, overall, $M = .64$, $SE = .04$, subcategory items, $M = .78$, $SE = .04$; basic category items, $M = .49$, $SE = .05$). One-sample *t*-tests revealed that for the interleaved subcategories group, performance on the subcategory prior knowledge items was not significantly different from chance (chance levels of performance .11; $t(82) = .68$, $p = .501$). For all other groups, performance on the subcategory prior knowledge items was significantly lower than chance levels of performance ($ts \geq 2.26$, $ps \leq .027$). For the basic category prior knowledge items, all groups performed significantly below chance (chance levels of performance = .33; $ts \geq 4.10$, $ps < .001$).

For the subcategory groups, a between-participants *t*-test indicated that prior knowledge classification performance for the subcategory items did not significantly differ between the interleaved subcategories group (*M* = .10, *SE* = .02) and the blocked subcategories group (*M* = .06, *SE* = .02), $t(164) = 1.40$, $p = .176$. Further, prior knowledge classification performance for the basic category items did not differ between the interleaved subcategories group (*M* = .20, *SE* = .03) and the blocked subcategories group (*M* = .17, *SE* = .03), $t(164) = .86$, $p = .392$.

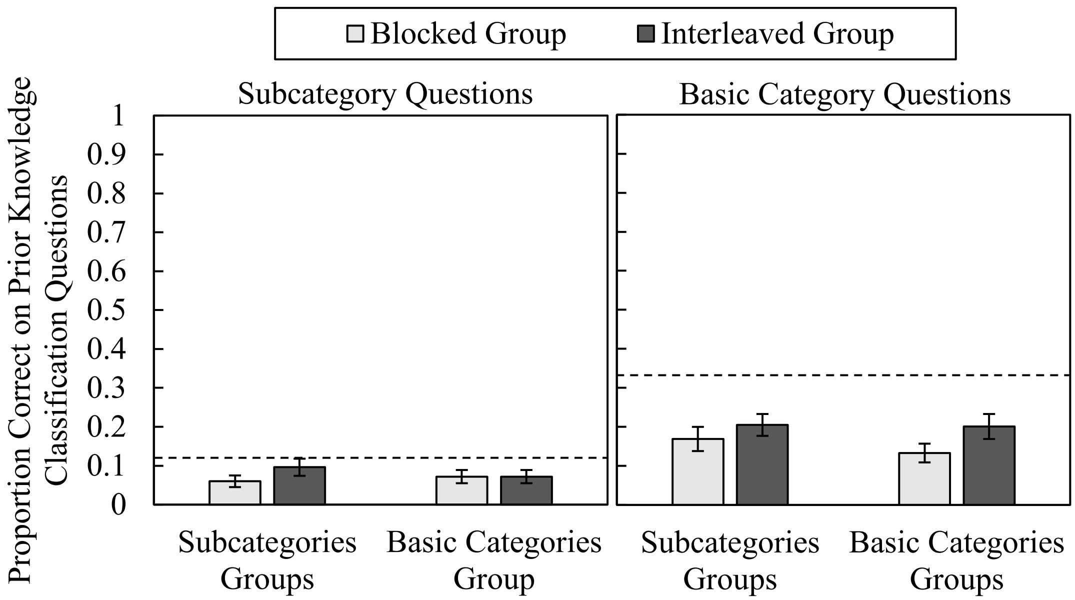


Figure 4. Mean proportion correct for prior knowledge classification questions for Experiment 1. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for basic categories was .33.

For the basic category groups, a between-participants *t*-test indicated that prior knowledge classification performance for the subcategory items did not significantly differ between the interleaved basic categories group (*M* = .07, *SE* = .02) and the blocked basic categories group (*M* = .07, *SE* = .02), $t(164) < .001$, $p = 1.00$. As well, prior knowledge classification performance for the basic category items did not differ between the interleaved basic categories group (*M* = .20, *SE* = .03) and the blocked basic categories group (*M* = .13, *SE* = .02), $t(164) = 1.69$, $p = .093$. These outcomes are consistent with the pilot experiment and indicate that participants did not have substantial prior knowledge for geological classifications and that prior knowledge did not significantly differ between groups.

Novel Classification Test Performance

Participants' performance on the novel classification test can be found in Figure 5. For the subcategory groups, novel classification test performance was explored with a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: novel first or novel second) between-participants ANOVA. Novel classification performance did not significantly differ between the interleaved subcategories group (M = .53, SE = .02) and the blocked subcategories group (M = .54, SE = .02), $F(1, 162)$ = .006, p = .938. Further, novel classification performance did not significantly differ between those who completed the novel test first (M = .52, SE = .02) and those who completed the novel test second (M = .55, SE = .02), $F(1, 162)$ = 2.47, p = .118,

nor was there a significant interaction between study order and test counterbalance on novel classification performance, $F(1, 162)$ = .023, p = .879.

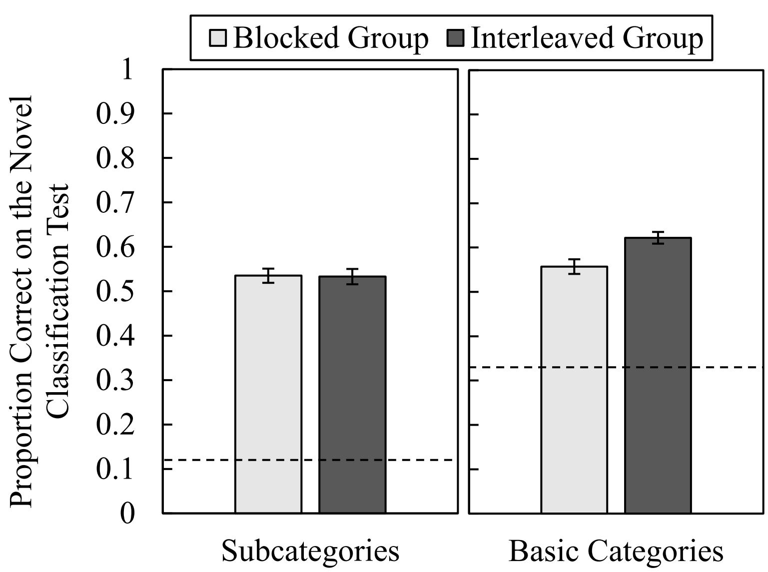


Figure 5. Mean proportion correct on the novel classification test in Experiment 1. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for basic categories was .33.

For the basic category groups, a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: novel first or novel second) between-participants ANOVA revealed that the interleaved basic categories group (*M* = .62, *SE* = .01) performed significantly better on the novel classification test as compared to the blocked basic categories group (*M* = .56, *SE* = .02), $F(1, 162) = 9.51$, $p = .002$, $\eta^2 = .055$. Novel classification performance did not significantly differ between those who completed the novel test first (*M* = .59, *SE* = .02) and those who completed the novel test second (*M* = .59, *SE* = .02), $F(1, 162) = .022$, $p = .880$. The interaction between study order and test counterbalance on novel classification performance was not significant, $F(1, 162) = 1.46$, $p = .229$.

These outcomes are inconsistent with the pilot experiment and hypothesized results. Specifically, study order did not significantly impact novel classification performance for subcategories whereas the pilot experiment suggested a positive effect of interleaving on novel classification performance. Further, I hypothesized that basic category learning would benefit most from learning in a blocked study order as compared to an interleaved study order, however, the results indicate that interleaving was most beneficial for novel classification performance with basic categories.

Studied Classification Test Performance

Participants' performance on the studied classification test can be found in Figure 6. For the subcategory groups, studied classification test performance was explored with a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: studied first or studied second) between-participants ANOVA. Studied classification performance did not significantly differ between the interleaved subcategories group (*M* = .58, *SE* = .02) and the blocked subcategories group (*M* = .56, *SE* = .02), $F(1, 162) = 1.07$, $p = .303$. Studied classification performance did not significantly differ between counterbalanced test orders, $F(1, 162) = 3.68$, $p = .057$, (studied test first, $M = .59$, $SE = .01$; studied test second, $M = .55$, $SE = .02$). Further, there was not a significant interaction between study order and test counterbalance on studied classification performance, $F(1, 162) < .001$, $p = .990$.

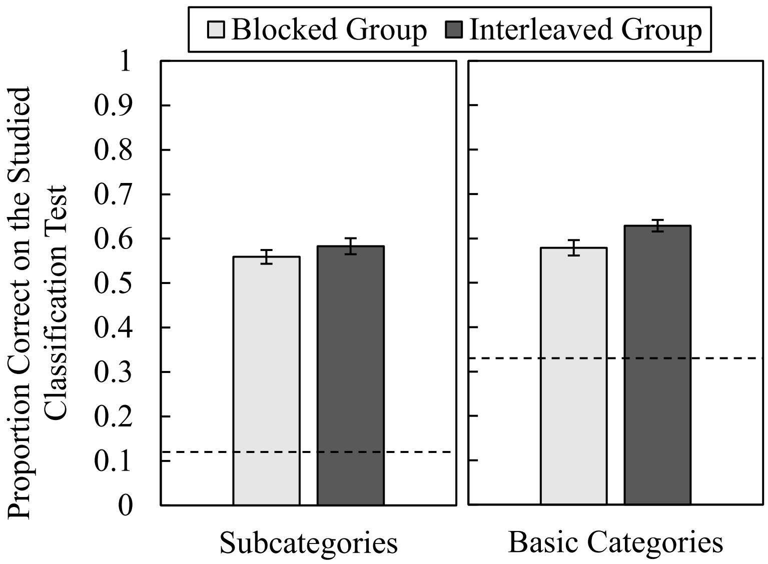


Figure 6. Mean proportion correct on the studied classification test in Experiment 1. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for basic categories was .33.

For the basic category groups, a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: studied first or studied second) between-participants ANOVA revealed that the interleaved basic categories group ($M = .58$, $SE = .02$) performed significantly better on the studied classification test as compared to the blocked basic categories group ($M = .63$, $SE = .01$), $F(1, 162) = 5.23$, $p = .024$, $\eta^2 = .031$. Studied classification performance did not significantly differ between those who completed the studied test first ($M = .61$, $SE = .02$) and those who completed the studied test second ($M = .60$, $SE = .01$), $F(1, 162) = .64$, $p = .426$. There was not a significant interaction between study order and test counterbalance on studied classification performance, $F(1, 162) = .071$, $p = .790$.

Similar to the results of the novel classification test, outcomes for the studied classification test are inconsistent with the pilot experiment and hypothesized results. When learning subcategories, study order did not significantly impact studied classification performance. Also unexpected, when learning basic categories, studying in an interleaved order was more beneficial to studied classification performance than was studying in a blocked order.

Beliefs about Study Order Strategy

Participants' responses to the follow-up question about their beliefs can be found in Table 6. Most participants (70%) indicated that they thought studying in a blocked order would help them learn geological categories best. This outcome is consistent with the Pilot experiment and prior research (e.g., Kornell & Bjork, 2008; Tauber et al., 2013; Yan et al., 2016: for an individual difference perspective, see Babineau et al., 2022). Participants' responses were evaluated between groups with 6 chi-square analyses with a Bonferroni correction to adjust the significance threshold to .008 for each analysis. Participants' responses to the belief follow-up question significantly differed between interleaved subcategories group and the interleaved basic categories group, χ^2 (2) = 13.84, $p < .001$, and between the interleaved subcategories group and the blocked basic categories group, χ^2 (2) = 9.85, $p = .007$. Follow-up analyses revealed that more participants in the interleaved subcategories group reported that blocking would be best for their learning ($n = 70$, 84.3%) as compared to the interleaved basic categories group ($n = 54$, 65.1%), χ^2 (1) = 8.16, $p = .004$; and compared to the blocked basic categories group ($n = 49$, 59.0%), χ^2 (1) = 13.1, $p < .001$. Further, fewer participants in the interleaved subcategories group reported that interleaving would be best for their learning ($n = 6$, 7.23%) as compared to the interleaved basic categories group ($n = 20$, 24.1%), χ^2 (1) = 8.94, p = .003; and compared to the blocked basic categories group (n = 21, 59.0%), χ^2 (1) = 9.95, p < .001. All other comparisons were nonsignificant, χ^2s ≤ 5.51, ps ≥ .064.

Table 6. *Reponses to the Beliefs about Study order Strategies Question for Experiment 1*

	Subcategory Groups		**Basic Category Groups**	
"Which study order do you think would help you learn best?"	**Blocked Group**	**Interleaved Group**	**Blocked Group**	**Interleaved Group**
Studying examples of each rock type grouped together by category.	69.88%	84.34%	59.04%	65.06%
Studying examples of each rock type mixed together.	18.07%	7.23%	25.30%	24.10%
My learning would have been about the same with either study order.	12.05%	8.43%	15.66%	10.84%

Note. Values indicate the percentage of participants who selected each response for the follow-up question probing for participants' beliefs about study order strategies.

Experiment 1 Discussion

The results of Experiment 1 revealed that when learning geological subcategories, study order did not significantly impact classification performance. However, when learning geological basic categories, studying in an interleaved order was beneficial to classification performance as compared to studying in a blocked order. These outcomes were unexpected and inconsistent with the Pilot experiment and the discriminative contrast hypothesis (Carvalho & Goldstone, 2014).

When learning subcategories, participants in Experiment 1 performed similarly when studying examples in an interleaved order as compared to studying in a blocked order. However, in the Pilot experiment, participants' classification performance indicated that there may be a benefit of interleaving on classification performance as compared to blocking. To more conclusively determine the effect of study order on category learning, I conducted two *continuously cumulating meta-analyses* (CCMAs) to aggregate outcomes from the Pilot experiment and Experiment 1 as suggested by Braver and colleagues (2014). There was not a significant effect of interleaving on novel classification performance relative to blocking (Pilot Experiment, Mdiff = .07, Spooled = .16; Experiment 1, Mdiff = .002, Spooled = .15; pooled d = 0.12, 95% CI [-0.38, 0.15]). Similarly, there was not a significant effect of interleaving on studied classification performance relative to blocking (Pilot Experiment, Mdiff = .07, Spooled = .15; Experiment 1, Mdiff = .02, Spooled = .15; pooled d = 0.25, 95% CI [-0.51, 0.01]). Thus, when examined with all available data in the reported experiments, studying geological *subcategories* in an interleaved order did not significantly impact classification performance as compared to studying in a blocked order.

The subcategory outcomes do not support the discriminative contrast hypothesis, which suggests that when learning subcategories, performance is best after studying in an interleaved order because it encourages participants to make between-category comparisons that should be most beneficial for learning categories that are relatively high family-resemblance (Carvalho & Goldstone, 2014). Even so, these results are consistent with recent meta-analyses of the interleaving effect, which have shown that STEM materials may demonstrate only a small interleaving benefit (Brunmair & Richter, 2019; Firth et al., 2021). As important, these results could have meaningful implications for interleaving research by identifying materials for which study order does not significantly impact performance.

When learning basic categories, participants in Experiment 1 performed better on the classification tests when studying examples in an interleaved order as compared to studying in a blocked order. This outcome is the opposite of my hypothesis and the discriminate contrast hypothesis, which suggests that when learning basic categories, performance would be best after studying in a blocked order because this order encourages participants to make within-category comparisons that should be most beneficial for learning categories that are relatively low in family-resemblance (Carvalho & Goldstone, 2014). One aim for Experiment 2 was to replicate this effect from Experiment 1 to provide further insight into it.

Another aim of Experiment 2 was to directly examine how category organization and study order impacted classification performance. In Experiment 1, participants who learned subcategories classified exemplars into nine categories, whereas participants who learned basic categories classified exemplars into three categories. This was done intentionally, as hierarchal information naturally results in a fewer number of basic categories than subcategories. This also mimics both a classroom setting, in which geology students would likely be required to learn fewer basic categories than subcategories and experimental contexts used in prior research. However, by learning a different number of categories, participants' performance on the classification tests could not be compared between the subcategory groups and the basic category groups. Thus, in Experiment 2, groups that are equated on number of categories and number of exemplars per category were added to directly compare performance between subcategories and basic categories.

Experiment 2

In Experiment 2, participants learned to classify basic categories in a blocked order or in an interleaved order. Additionally, equated subcategory groups and equated basic category groups were included in Experiment 2. In the equated subcategory groups, one equated subcategory group learned to classify three subcategories in a blocked order, and the other equated subcategory group learned to classify three subcategories in an interleaved order. In the equated basic category groups, one equated basic category group learned to classify the three basic categories in a blocked order, and the other equated basic category group learned to classify the three basic category in an interleaved order; however, each basic category had fewer exemplars to match the equated subcategory groups. The inclusion of the equated subcategory groups and equated basic category groups allowed for a direct comparison of classification performance between participants learning subcategories and basic categories while the number of categories and number of exemplars was kept constant.

The impact of the number of categories and number of exemplars during learning has not been directly evaluated from the perspective of the discriminative contrast hypothesis. Even so, I expected that participants' performance in the equated subcategory groups and equated basic category groups would be consistent with the discriminative contrast hypothesis (Carvalho & Goldstone, 2014). Specifically, I predicted that participants in the equated subcategory groups would perform better on the novel classification test and studied classification test when learning in an interleaved order as compared to learning in a blocked order. Whereas I predicted that participants in the equated basic category groups would perform better on the novel classification test and studied classification test when learning in a blocked order as compared to learning in an interleaved order. Manipulating the number of to-be-learned categories should not impact category structure because the degree of family-resemblance of a category is independent of the number of categories during learning (Nosofsky et al., 2018). As well, subcategories selected did not significantly differ in ratings of distinctness (Nosofsky et al., 2018). Thus, studying three subcategories should not significantly impact the degree of distinctiveness as compared to learning nine subcategories.

In addition, manipulating the number of exemplars for the equated basic category groups should not impact ratings of distinctiveness and family-resemblance, as the exemplars were selected at random for each participant. I also predicted that subcategories may be easier to learn as they have a higher family-resemblance rating. Thus, participants in the equated subcategory groups may perform better than participants in the equated basic category groups.

Method

Participants

Study order strategy (blocked or interleaved), and category organization (basic categories, equated subcategories, and equated basic categories) were manipulated in a between-participants design. Using an effect size of d = .47, (Experiment 1, novel classification performance interleaving effect for the basic categories groups) with alpha error at .05, and power at .85, a power analysis was conducted with G*Power estimated that 83 participants randomly assigned to each group would be sufficient for a well-powered experiment (Faul et al., 2007). Thus, a total of 498 undergraduate Texas Christian University students participated for partial credit in psychology courses, with 83 participants randomly assigned to each group.

Demographic characteristics for participants in Experiment 2 are displayed in Table 4. The groups did not significantly differ on demographic characteristics (age, $F(5, 487) = 541$, $p = .745$; gender, $\chi^2\ (25) = 27.41$, $p = .336$; ethnicity, $\chi^2\ (120) = 103.51$, $p = .859$). Further, few participants reported experience in majors or hobbies related to geology (see Table 7). There was no significant difference in self-reported prior experience in geology between groups, χ^2s ≤ 8.10, ps $\geq .151$, and the groups did not significantly differ in participants' self-rated knowledge of geology, $F(5, 491) = .79$, $p = .557$.

Table 7. *Participants' Reponses to each Prior Experience Question for the Experiment 2*

Question	Basic Category Groups		Equated Subcategory Groups		Equated Basic Category Groups	
	Blocked Group	Interleaved Group	Blocked Group	Interleaved Group	Blocked Group	Interleaved Group
1. Are you currently or have you ever been a Geology Major or Minor?	0%	0%	0%	*n* = 1, 1.2%	*n* = 2, 2.4%	0%
2. Are you currently or have you ever been an Applied Geoscience Major?	0%	0%	0%	0%	0%	0%
3. Do you collect rocks?	*n* = 1, 1.2%	*n* = 3, 3.6%	*n* = 4, 4.8%	*n* = 4, 4.8%	*n* = 5, 6.0%	*n* = 2, 2.4%
4. Do you own a rock and mineral field guide?	*n* = 3, 3.6%	*n* = 1, 1.2%	0%	*n* = 2, 2.4%	0%	0%
5. What is your own rated level of expertise at identifying rocks?	1.51 (.09)	1.57 (.09)	1.37 (.07)	1.41 (.08)	1.51 (.09)	1.42 (.08)

Note. Values for questions 1- 4 indicate the percentage of participants that responded with "yes" for the prior geology experience questions. Values for question 5 indicate the *M* (*SE*) on a scale from 1 to 7 (i.e., 1 = novice to 7 = expert).

Materials

Participants learned to classify the categories of rocks used in the Pilot experiment and Experiment 1 (see Table 1). The subcategories and basic categories were identical to those used in Experiment 1. Participants in the equated subcategory groups learned three subcategories that were selected randomly for each participant. In the equated subcategory groups, participants viewed 27 exemplars during the study phase and studied classification test, and nine exemplars during the novel classification test. Participants in the equated basic category groups learned the three basic categories; however, each basic category had a reduced number of exemplars that were selected randomly for each participant. In the equated basic category groups, participants viewed 27 exemplars during the study phase and studied classification test, and nine exemplars during the novel classification test. Thus, the equated subcategory groups and equated basic category groups studied the same number of categories and the same number of exemplars, they only differed in category organization.

Procedure

Data collection for Experiment 2 occurred in-person. Participants completed the study individually using computers made available in the Metacognition, Memory, and Aging lab at TCU. A trained research assistant was present to ensure that participants followed directions and to answer participants' questions. For participants assigned to the basic category groups, the procedure for Experiment 2 was identical to the procedure for Experiment 1. For participants in the equated basic category groups, the procedure was nearly identical to the procedure for the basic category groups in Experiment 1; however, they had a fewer number of studied exemplars and test exemplars. For participants in the equated subcategory groups, the procedure was nearly identical to the procedure for the subcategory groups in Pilot experiment and Experiment 1; however, they learned to classify a fewer number of subcategories.

In the study phase, participants assigned to learn the equated subcategories in a blocked order studied the exemplars in groups by subcategory. Specifically, the study phase exemplars for each subcategory were shown consecutively such that participants studied all of the exemplars in one subcategory before they studied all of the exemplars in another category. The order of the exemplars within each subcategory was randomized for each participant. Further, the order of the three subcategories within the study phase was randomized for each participant. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding subcategory label shown beneath it. Each exemplar was presented for 3s before the next exemplar was automatically presented. There was a brief .5s inter-stimulus-interval between each exemplar that consisted of a blank white screen.

The study phase was nearly identical for participants assigned to learn the equated subcategories in an interleaved order; however, the order of the study phase exemplars differed. The exemplars were studied in a pseudo-random order such that the exemplars were distributed into nine groups, with one exemplar from each subcategory randomly allocated to each group for each participant. The order of the exemplars within each group was randomized per participant; however, two exemplars from the same subcategory were never shown consecutively. Each exemplar was shown one-at-a-time in the center of the screen, with the corresponding subcategory label shown beneath it. Each exemplar was presented for 3s before the next exemplar was automatically presented. There was a brief .5s inter-stimulus-interval between each exemplar that will consist of a blank white screen.

During the novel classification test, participants assigned to the equated subcategory groups were shown a never-before-seen exemplar belonging to one of the three subcategories. The names of the subcategory were displayed on buttons in alphabetical order beneath the novel exemplar. Participants were instructed to classify each exemplar into its correct subcategory by selecting the button with the corresponding subcategory label. Each novel test exemplar was presented one-at-a-time, in the center of the screen, and participants were allowed to take as much time as they would like to make their classification. Participants did not receive any feedback on their answers, and the order of the exemplars on the novel test was random for each participant. The "*I don't know*" option was not present during either the novel classification test or the studied classification test. Thus, participants were required to make a response for each exemplar in the test phase. The studied classification test was nearly identical to the novel classification test; however, the exemplars presented were the same as the exemplars shown during the study phase.

After completing both classification tests, participants in the equated subcategory groups responded to the follow-up question probing for their beliefs about study order strategy. Nearly all aspects of the follow-up question were identical to those used in Experiment 1; however, the examples demonstrating an interleaved study order as compared to a blocked study order involved the subcategories studied by the participant. All other aspects of the procedure for Experiment 2 were identical to those used in Experiment 1.

Experiment 2 Results

As with prior experiments, analyses are reported in the order in which participants completed each task. Participants' performance on the prior knowledge questions is reported first, followed by participants' performance on the novel and studied classification tests. Last, participants' responses to the belief follow-up question are reported.

Prior Knowledge Question Performance

Participants' performance on the prior knowledge questions can be found in Figure 7. Subcategory performance and basic category performance were calculated in the same manner as in previous experiments. Overall, participants' mean prior knowledge performance (M = .13, SE = .01) indicated low levels of geology classification knowledge. Participants in the blocked basic categories group responded with "*I don't know*" on a majority (M = .56, SE = .04) of the prior knowledge items (subcategory items, M = .70, SE = .04; basic category items,

$M = .42$, $SE = .05$). Participants in the interleaved basic categories group responded with "*I don't know*" on a majority ($M = .52$, $SE = .04$) of the prior knowledge items (subcategory items, $M = .68$, $SE = .04$; basic category items, $M = .36$, $SE = .05$). Participants in the blocked equated subcategories group responded with "*I don't know*" on a majority ($M = .58$, $SE = .03$) of the prior knowledge items (subcategory items, $M = .76$, $SE = .04$; basic category items, $M = .41$, $SE = .05$). Participants in the interleaved equated subcategories group responded with "*I don't know*" on a majority ($M = .61$, $SE = .04$) of the prior knowledge items (subcategory items, $M = .77$, $SE = .04$; basic category items, $M = .45$, $SE = .05$). Participants in the blocked equated basic categories group responded with "*I don't know*" on a majority ($M = .52$, $SE = .04$) of the prior knowledge items (subcategory items, $M = .66$, $SE = .05$; basic category items, $M = .37$, $SE = .05$). Participants in the interleaved equated basic categories group responded with "*I don't know*" on a majority ($M = .57$, $SE = .04$) of the prior knowledge items (subcategory items, $M = .73$, $SE = .04$; basic category items, $M = .41$, $SE = .05$).

One-sample *t*-tests revealed that performance for the subcategory prior knowledge items was not significantly different from chance for the blocked equated subcategories group (chance levels of performance .11; $t(82) = 2.00$, $p = .05$), or for the interleaved equated basic categories group, $t(82) = 1.15$, $p = .255$. For all other groups, performance for the subcategory prior knowledge items was significantly below chance levels of performance, $ts \geq 2.29$, $ps < .001$. For the blocked equated basic categories group, performance on the basic category prior knowledge items was not significantly different from chance (chance levels of performance .33; $t(82) = 1.92$, $p = .059$). All other groups performed significantly below chance for the basic category prior knowledge items, $ts \geq 4.72$, $ps < .001$.

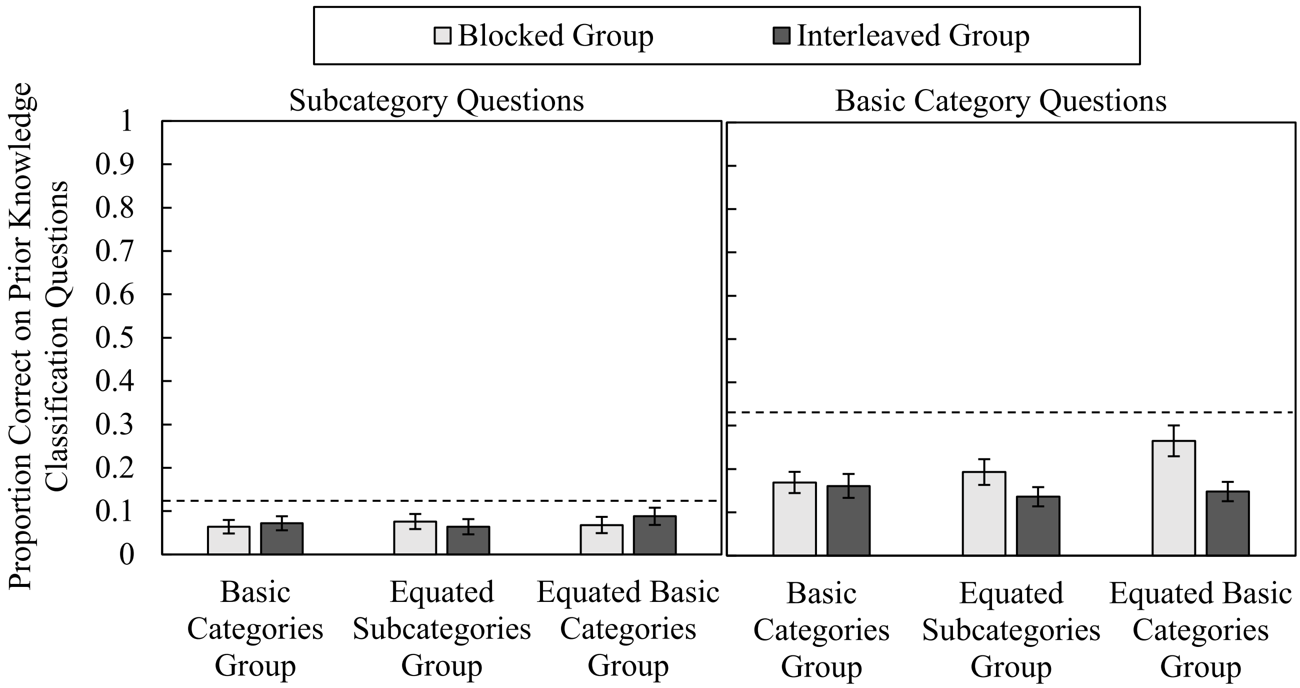


Figure 7. Mean proportion correct on the prior knowledge classification questions for Experiment 2. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for basic categories was .33.

For the basic category groups, a between-participants *t*-test indicated that prior knowledge classification performance for the subcategory items did not significantly differ between the interleaved basic categories group (M = .07, SE = .02) and the blocked basic categories group (M = .06, SE = .02), $t(164)$ = .36, p = .722. Further, prior knowledge classification performance for the basic category items did not differ between the interleaved basic categories group (M = .16, SE = .03) and the blocked basic categories group (M = .17, SE = .02), $t(164)$ = .22, p = .828. These results are consistent with prior experiments.

For the equated groups, a 2 (study order strategy: blocked group, interleaved group; between-participants) x 2 (category organization: equated subcategories, equated basic categories; between-participants) between-participants ANOVA indicated that participants' performance for the subcategory prior knowledge items did not significantly differ between the blocked groups (M = .07, SE = .01) and the interleaved groups (M = .08, SE = .01), $F(1, 328)$ =

.05, $p = .828$. Subcategory prior knowledge classification performance did not significantly differ between the basic categories groups ($M = .07$, $SE = .01$) and the subcategories groups ($M = .08$, $SE = .01$), $F(1, 328) = .19$, $p = .663$. Nor was the interaction between study order and category organization significant for the subcategory prior knowledge items, $F(1, 328) = .76$, $p = .384$.

To examine performance on the basic category prior knowledge items for the equated groups, a 2 (study order strategy: blocked group, interleaved group; between-participants) x 2 (category organization: equated subcategories, equated basic categories; between-participants) between-participants ANOVA was conducted. There was a significant main effect of study order such that participants in the blocked groups ($M = .23$, $SE = .02$) performed better on the basic category prior knowledge items than did participants in the interleaved groups ($M = .14$, $SE = .02$), $F(1, 328) = 9.50$, $p = .002$, $\eta^2 = .028$. Follow-up analyses with a Bonferroni correction to adjust the significance threshold to .008 for each analysis revealed that participants in the blocked equated basic category group ($M = .27$, $SE = .04$) performed significantly better on basic category prior knowledge items as compared to the interleaved equated subcategories group ($M = .14$, $SE = .02$; $t(164) = 3.09$, $p = .003$, $d = .45$), and as compared to the interleaved equated basic category group ($M = .15$, $SE = .02$; $t(164) = 2.77$, $p = .006$, $d = .43$). Basic category prior knowledge classification performance did not significantly differ between the basic categories groups ($M = .17$, $SE = .02$) and the subcategories groups ($M = .21$, $SE = .02$), $F(1, 328) = 2.27$, $p = .133$. The interaction between study order and category organization was not significant, $F(1, 328) = 1.16$, $p = .283$. Significant differences in prior knowledge classification performance between groups were unexpected. As such, for novel classification performance and studied classification performance analyses for the equated groups, I explored prior knowledge classification performance as a control variable.

Novel Classification Test Performance

Participants' performance on the novel classification test can be found in Figure 8. For the basic category groups, novel classification test performance was explored with a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: novel first or novel second) between-participants ANOVA. Novel classification performance did not significantly differ between the interleaved basic category group (M = .59, SE = .02) and the blocked basic category group (M = .58, SE = .02), $F(1, 162)$ = .288, p = .593. Novel classification performance did not significantly differ between those who completed the novel test first (M = .60, SE = .02) and those who completed the novel test second (M = .56, SE = .02), $F(1, 162)$ = 2.79, p = .097, nor was there a significant interaction between study order and test counterbalance on novel classification performance, $F(1, 162)$ = 2.51, p = .115. For the basic category groups, study order did not impact novel classification performance. This outcome was unexpected and is inconsistent with the findings of Experiment 1.

For the equated groups, a 2 (study order strategy: interleaved or blocked) x 2 (category organization: equated subcategories, equated basic categories) x 2 (counterbalanced test order: novel first or novel second) between-participants ANOVA revealed that performance on the novel classification test did not significantly differ between participants who studied in an interleaved order (M = .71, SE = .01) as compared to those who studied in a blocked order (M = .71, SE = .01), $F(1, 324)$ = .006, p = .940. The interaction between study order and category organization on novel classification performance was not significant, $F(1, 324)$ = 2.40, p = .122; nor was the interaction between study order and test counterbalance, $F(1, 324)$ = .041, p

= .840; or the three-way interaction, $F(1, 324) = .025$, $p = .874$. Of most interest, study order did not significantly impact novel classification performance for the equated groups, nor did study order interact with category organization or test counterbalance. Further, novel classification performance did not significantly differ between those who completed the novel test first ($M = .71$, $SE = .01$) and those who completed the novel test second ($M = .71$, $SE = .01$), $F(1, 324) = .126$, $p = .723$. However, those who learned equated subcategories ($M = .87$, $SE = .01$) performed significantly better on the novel classification test than did those who *learned* equated basic categories ($M = .56$, $SE = .01$), $F(1, 324) = 263.30$, $p < .001$, $\eta^2 = .45$. The interaction between category organization and test counterbalance on novel classification performance was not significant, $F(1, 324) = .692$, $p = .406$.

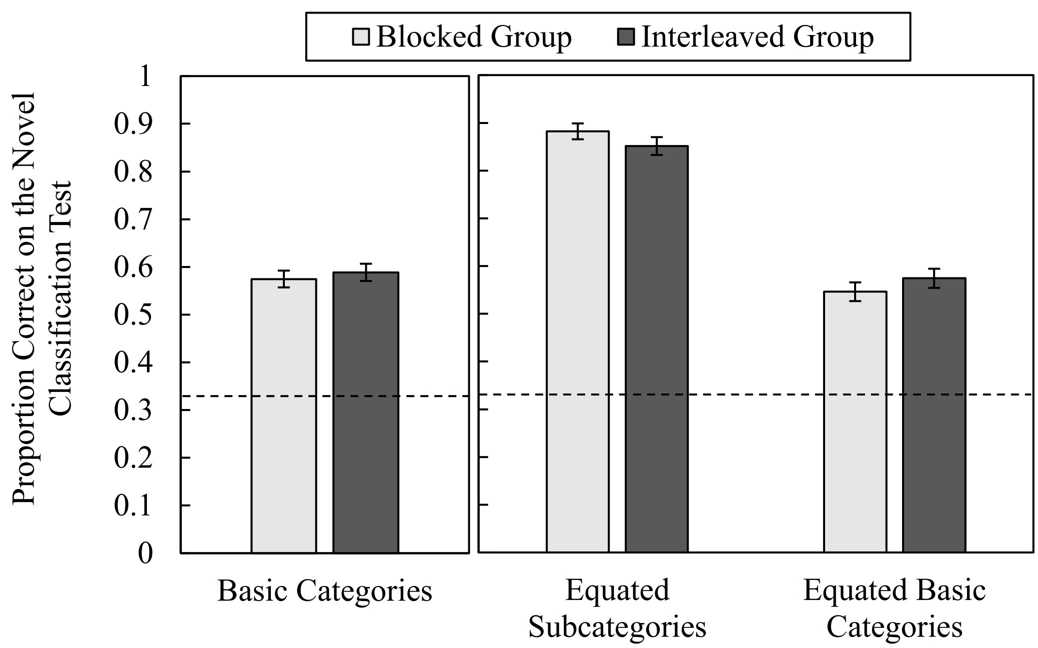


Figure 8. Mean proportion correct on the novel classification test in Experiment 2. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for abbreviated subcategories and basic categories was .33.

To explore prior knowledge classification performance as a control variable, a multiple linear regression was conducted on novel classification performance for the equated groups. Category organization condition (dummy coded; 0 = basic categories, reference, 1 = subcategories), study order condition (dummy coded; 0 = blocked, reference, 1 = interleaved), test counterbalance, (dummy coded, 0 = novel test first, reference, 1 = novel test second), and overall prior knowledge classification performance (uncentered) were entered as predictors. The three, 2-way interactions and one 3-way interaction were also entered in the model. Most important, outcomes from the multiple regression were consistent with the outcomes from the ANOVA, indicating that category organization was the only significant predictor of novel classification performance, $b = .32$ ($SE = .04$), $t = 8.34$, $p < .001$, $R^2 = .46$. Prior knowledge classification performance did not significantly predict novel classification performance for the equated groups, $b = .02$ ($SE = .06$), $t = .33$, $p = .745$, $R^2 = .001$. In sum, the equated subcategory groups performed significantly better on the novel classification test than did the equated basic category groups.

Studied Classification Test Performance

Participants' performance on the studied classification test can be found in Figure 9. For the basic category groups, studied classification test performance was explored with a 2 (study order strategy: interleaved or blocked) x 2 (counterbalanced test order: studied first or studied second) between-participants ANOVA. Studied classification performance did not significantly differ between the interleaved basic category group ($M = .61$, $SE = .02$) and the blocked basic category group ($M = .61$, $SE = .02$), $F(1, 162) = .04$, $p = .851$. Studied classification performance did not significantly differ between those who completed the studied test first ($M = .61$, $SE = .02$) and those who completed the studied test second ($M = .61$, *SE* = .02), *F*(1, 162) = .09, *p* = .763. Nor was there a significant interaction between study order and test counterbalance on studied classification performance, *F*(1, 162) = 2.92, *p* = .09. Thus, for the basic category groups, study order did not significantly impact studied classification performance.

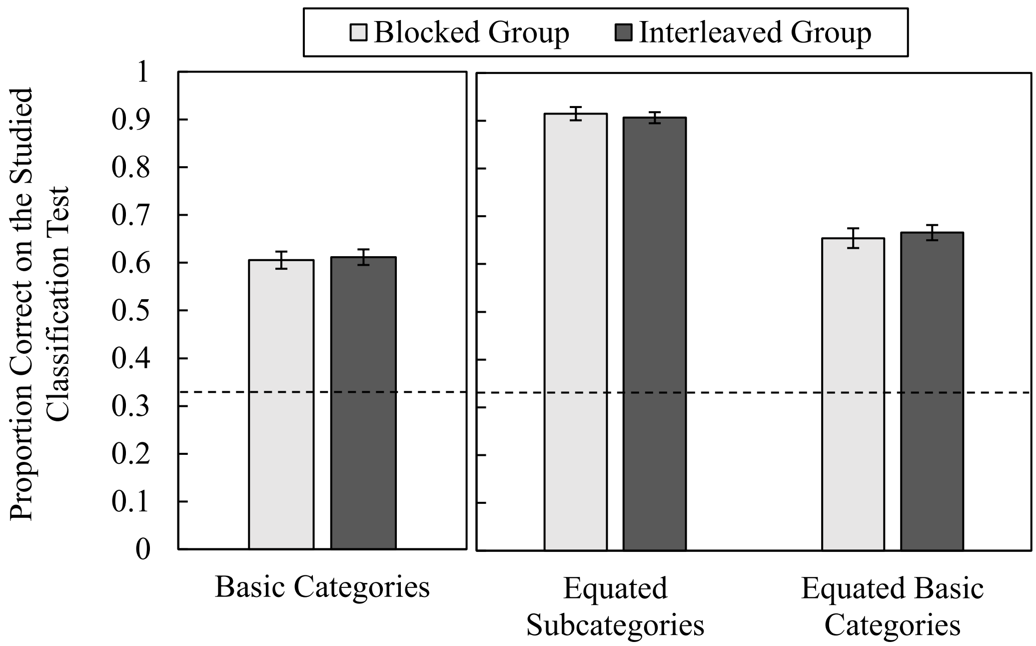


Figure 9. Mean proportion correct on the studied classification test in Experiment 2. Dashed lines present chance levels of performance. When classifying subcategories, chance levels of performance was .11, whereas chance levels of performance for abbreviated subcategories and basic categories was .33.

For the equated groups, a 2 (study order strategy: interleaved or blocked) x 2 (category organization: equated subcategories, equated basic categories) x 2 (counterbalanced test order: studied first or studied second) between-participants ANOVA revealed that performance on the studied classification test did not significantly differ between participants who studied in an interleaved order (*M* = .79, *SE* = .01) as compared to those who studied in a blocked order (*M* = .78, *SE* = .01), *F*(1, 324) = .01, *p* = .925. Further, the interaction between study order and category organization on studied classification performance was not significant, *F*(1, 324) = .42, *p* = .520. Nor was the interaction between study order and test counterbalance, $F(1, 324) = .61$, $p = .437$, or the three-way interaction, $F(1, 324) = .11$, $p = .739$. Thus, study order did not significantly impact studied classification performance for the equated groups, and study order did not significantly interact with the other variables. Further, studied classification performance did not significantly differ between those who completed the studied test first ($M = .79$, $SE = .01$) and those who completed the studied test second ($M = .79$, $SE = .01$), $F(1, 324) = .001$, $p = .974$. However, category organization was significant, $F(1, 324) = 249.01$, $p < .001$, $\eta^2 = .43$, such that those who learned equated subcategories ($M = .91$, $SE = .01$) performed significantly better on the studied classification test than did those who learned equated basic categories ($M = .66$, $SE = .01$). The interaction between category organization and test counterbalance on studied classification performance was not significant, $F(1, 324) = 2.01$, $p = .158$.

Similar to the novel classification test, prior knowledge classification performance was explored as a control variable by conducting a multiple linear regression on studied classification performance for the equated groups. Category organization condition (dummy coded; 0 = basic categories, reference, 1 = subcategories), study order condition (dummy coded; 0 = blocked, reference, 1 = interleaved), test counterbalance, (dummy coded, 0 = studied test second, reference, 1 = studied test first), and overall prior knowledge classification performance (uncentered) were entered as predictors. The three, 2-way interactions and one 3-way interaction were also entered in the model. Outcomes from the multiple regression were consistent with the outcomes from the ANOVA, indicating that category organization was the only significant predictor of studied classification performance, $b = .29$ ($SE = .03$), $t = 9.19$, $p < .001$, $R^2 = .51$. Prior knowledge classification performance did not significantly predict studied classification performance for the equated groups, *b* = .02 (*SE* = .05), *t* = .43, *p* = .669, R^2 = .001.

Beliefs about Study Order Strategy

Participants' responses to the follow-up question about their beliefs can be found in Table 8. Most participants (54%) indicated that they thought studying in a blocked order would help them learn geological categories best. This outcome is consistent with the Pilot experiment and Experiment 1. To evaluate whether participants' responses differed between groups, 15 chi-square analyses with a Bonferroni correction to adjust the significance threshold to .003 for each analysis were conducted. The analyses revealed that the interleaved equated subcategories group differed significantly as compared to the interleaved basic categories group, χ^2 (2) = 18.09, $p < .001$; and that the interleaved equated subcategories group differed significantly as compared to the interleaved equated basic categories group, χ^2 (2) = 13.31, $p < .001$. Follow-up analyses indicated that fewer participants in the interleaved equated subcategories group reported that blocking would be best for their learning (*n* = 29, 34.9%) as compared to the interleaved basic categories group (*n* = 55, 66.3%), χ^2 (1) = 16.29, $p < .001$; and compared to the interleaved equated basic categories group (*n* = 52, 62.7%), χ^2 (1) = 12.75, $p < .001$. Further, more participants in the interleaved equated subcategories group reported that interleaving would be best for their learning (*n* = 32, 38.6%) as compared to the interleaved basic categories group (*n* = 12, 14.5%), χ^2 (1) = 12.37, $p < .001$. All other comparisons were nonsignificant, χ^2s ≤ 8.23, *p*s $\geq .016$.

Table 8. *Reponses to the Beliefs about Study order Strategies Question for Experiment 2*

"Which study order do you think would help you learn best?"	**Basic Category Groups**		**Equated Subcategory Groups**		**Equated Basic Category Groups**	
	Blocked Group	**Interleaved Group**	**Blocked Group**	**Interleaved Group**	**Blocked Group**	**Interleaved Group**
Studying examples of each rock type grouped together by category.	50.60%	66.26%	53.01%	34.94%	56.63%	62.65%
Studying examples of each rock type mixed together.	27.71%	14.46%	24.10%	38.55%	22.89%	25.30%
My learning would have been about the same with either study order.	21.69%	19.28%	22.89%	26.51%	20.48%	12.05%

Note. Values indicate the percentage of participants who selected each response for the follow-up question probing for participants' beliefs about study order strategies.

Experiment 2 Discussion

Overall, the results of Experiment 2 revealed that when learning basic geological categories, study order did not significantly impact classification performance. Further, study order did not significantly impact classification performance when directly comparing category organization; however, participants who learned to classify equated subcategories performed better on the classification tests than did participants who learned to classify equated basic categories.

When learning basic categories, participants in Experiment 1 performed better when studying exemplars in an interleaved order as compared to studying in a blocked order. However, in Experiment 2, participants' classification performance indicated that study order did not significantly impact classification performance. Thus, I conducted two CCMAs to aggregate outcomes across Experiment 1 and Experiment 2 as suggested by Braver et al. (2014). Basic category groups were not included in the Pilot experiment, thus, no data from the Pilot experiment was included in these analyses. The analysis revealed a significant beneficial effect of interleaving on novel classification performance relative to blocking when learning basic categories (Experiment 1, *M*diff = .07, *S*pooled = .14; Experiment 2, *M*diff = .01, *S*pooled = .16; pooled *d* = 0.28, 95% CI [0.06, 0.50]). However, there was not a significant effect of interleaving on studied classification performance relative to blocking (Experiment 1, *M*diff = .05, *S*pooled = .14; Experiment 2, *M*diff = .006, *S*pooled = .16; pooled *d* = 0.20, 95% CI [-0.41, 0.02]). Thus, when examined with all available data from the reported experiments, studying *basic* geological categories in an interleaved order improved novel classification performance as compared to studying in a blocked order; however, study order did not significantly impact studied classification performance. Further, when the number of basic category exemplars was reduced for the equated basic category groups, study order did not impact classification performance. This was unexpected and highlights the nuanced context in which interleaving is beneficial to basic category learning.

For the equated groups, study order did not impact classification performance; however, category organization did impact classification performance. Specifically, participants who learned equated subcategories performed significantly better on the classification tests than did participants who learned equated basic categories. The geological subcategories have a higher degree of family-resemblance that likely contributed to participant ability to learn to classify them.

General Discussion

The present experiments were the first to systematically explore the effects of study order and category organization on geological category learning. Studying in an interleaved order led to better novel classification performance for basic geological categories relative to studying in a blocked order (CCMA including data from Experiment 1 & Experiment 2). However, study order did not impact *studied* classification performance when learning basic categories (CCMA including data from Experiment 1 & Experiment 2). Further, study order did not impact classification performance when learning nine subcategories (CCMA including data from the Pilot experiment & Experiment 1), or when learning three subcategories (equated subcategory groups; Experiment 2), nor did study order impact classification performance when learning basic categories with fewer exemplars (equated basic category groups; Experiment 2). Taken together, these outcomes reveal that in most instances, study order did not impact classification performance with the only exception being novel classification of basic-level categories.

The present outcomes do not support the discriminative contrast hypothesis. The discriminative contrast hypothesis suggests that the study order most beneficial for category learning is related to category structure (Carvalho & Goldstone, 2014). Specifically, when learning categories with high family-resemblance, such as geological subcategories, studying in an interleaved order should be better for classification performance as compared to studying in a blocked order. However, when learning categories with low family-resemblance, such as basic geological categories, studying in a blocked order should be better for classification performance as compared to studying in an interleaved order. In the present experiments, neither studying in an interleaved order or in a blocked order resulted in significant classification improvements for most of the groups. Thus, perhaps the current experiments demonstrate a boundary condition for which the discriminative contrast hypothesis is not relevant.

The present results contribute a well-powered analysis of how study order influences classification performance for geological categories and add to the growing body of knowledge for these geological stimuli (Babineau et al., 2022; Babineau & Tauber, 2022; Lu et al., 2021; Meagher et al., 2017; Miyatsu et al., 2020). Importantly, the results suggest that interleaving may not be the most beneficial study order strategy for all materials or for all students (Brunmair & Richter, 2019; Firth et al., 2021). Identifying such boundary conditions is valuable because in prior research, conclusions about the value of interleaved study have been too broad. These results are also relevant for geology instructors and students, as the best study order strategy for learning depends upon the priorities for the course. For example, if the course goal is to classify novel exemplars into their correct basic category, instructors should recommend students study in an interleaved order; whereas, if the course goal is to classify exemplars into their correct subcategory, instructors may choose to make recommendations about other study strategies because study order is unlikely to substantially impact learning of these geological categories.

One factor that may be relevant for interpreting the obtained outcomes is participants' prior knowledge in geology (Babineau et al., 2022; Firth et al., 2021). In the present experiments, students demonstrated relatively low levels of geological prior knowledge in all experiments indicating that they were novices in rock classification. Recent research has found that for novices in rock classification, study order may not predict classification performance (Babineau et al., 2022). Thus, when students have low levels of prior knowledge, study order may not significantly impact classification performance; however, students with high levels of prior knowledge (e.g., graduate students in geology), may benefit from different study order strategies. Specifically, prior knowledge may impact how students make comparisons during category learning. The discriminative contrast hypothesis suggests that study order impacts the types of comparisons students make during learning (Carvalho & Goldstone, 2014; Firth et al., 2021). Thus, students with higher levels of geological prior knowledge may be more successful in making comparisons during learning as compared to a novice in geology (Firth et al., 2021). Future work should continue to examine how expertise may moderate the interleaving effect.

Interestingly, when learning basic geological categories with 27 exemplars in each category, interleaving was best for novel classification performance; whereas when learning basic geological categories with 9 exemplars each category, there was not a benefit of interleaving on classification performance. These outcomes are inconsistent with the discriminative contrast hypothesis, which suggests that blocking should be most beneficial for these types of categories. Current theory in interleaving suggests that an attentional component may also contribute to the interleaving effect (Firth et al., 2021). The attention attenuation hypothesis suggests that when categories are blocked, students' attention to the exemplars decreases as trials increase throughout each block (Wahlheim et al., 2011). In this way, students' attention in the blocked basic category group may have decreased as they studied each basic category, whereas students' attention in the interleaved basic category group may have been more consistent through the study phase. These differences in attention may have been more pronounced for students learning basic categories with 27 exemplars as compared to learning the equated basic categories with 9 exemplars. Future work should aim to explore how the number of exemplars in a category impacts students' attention during different study orders.

Whereas study order did not significantly impact classification performance for most groups, category structure did significantly impact classification performance. Comparisons between the equated subcategory groups and equated basic category groups (Experiment 2) revealed that classification performance was better for subcategories as compared to basic categories. This is consistent with prior research on category learning that suggests that learning categories with low-family resemblance is more challenging than learning categories with high family-resemblance (Goldstone, 1996). Subcategories have relatively high family-resemblance ratings, whereas basic categories have relatively low family-resemblance ratings. In this way, basic categories have fewer cohesive features for classification, making them more challenging to identify. This is reflected in the novel and studied classification performance outcomes for the equated groups in Experiment 2.

Although students' beliefs about study order were not my main interest, in most groups, students reported that they thought blocking would be best for their learning. This is consistent with prior research examining students' beliefs about learning (Kornell & Bjork, 2008; Morehead et al., 2016; McCabe, 2011; for beliefs about individual differences, see Babineau et al., 2022). In the current experiments, studying in a blocked order was not beneficial for classification performance. Thus, these outcomes indicate a discrepancy between students' beliefs about study order and their actual learning. Future research should explore how this discrepancy impacts students' study decisions, particularly when students self-regulate their study. Interestingly, in the interleaved equated subcategory group (Experiment 2), fewer students reported that blocking would be best for their learning and more students indicated that interleaving would be best for their learning. Students' beliefs about learning are influenced by multiple factors such as the ease or fluency of a learning task, as well as their pre-existing theories about how to learn best (Yan et al., 2016). Perhaps students' experience in the interleaved equated subcategory group influenced their beliefs about learning. Specifically, students in the interleaved equated subcategory group learned to classify 3 subcategories with 9 exemplars each; thus, the task was relatively fast, and participants performed well on the classification tests. In this way, students may have had a relatively easy and fluent learning experience while learning the equated subcategories in an interleaving order. Together, these outcomes may have contributed to students' greater endorsement of interleaving as compared to the other groups.

In conclusion, study order did not significantly impact classification performance when learning geological subcategories, equated geological subcategories, or equated basic geological categories such that studying in an interleaved order resulted in similar classification performance to studying in a blocked order. The one exception was that when learning basic geological categories with 27 exemplars each, interleaving was better than blocking for novel classification performance. These findings contribute to STEAM category learning literature, as well as to the theoretical explanations of study order effects.

References

Ariel, R., Babineau, A.L. & Tauber, S.K., (under review) A Brief Retrieval Practice Intervention Improves Memory for Medication Side Effects. *Journal of Experimental Psychology: Applied.*

Babineau, A.L. & Tauber, S.K. (2022). Students' decisions to switch between categories or stay within them are related to practice classification performance. *Memory & Cognition,* Advanced Online Publication. 1-17. https://doi.org/10.3758/s13421-022-01375-2

Babineau, A.L., Witherby, A.E., Ariel, R., Pelch, M.A., & Tauber, S.K. (2022). Do Domain Knowledge and Retrieval Practice Predict Students' Study Order Decisions? [Special issue]. *Journal of Intelligence, 10:122,* 1-20. https://doi.org/10.3390/jintelligence10040122

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9*, 333-342. DOI: 10.1177/1745691614529796

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*, 1029–1052. https://doi.org/10.1037/bul0000209

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. https://doi.org/10.3758/s13421-013-0371-0

Clary, R. M., & Wandersee, J. H. (2007). A mixed methods analysis of the effects of an integrative geobiological study of petrified wood in introductory college geology classrooms. *Journal of Research in Science Teaching: The Official Journal of the*

National Association for Research in Science Teaching, *44*, 1011-1035. https://doi.org/10.1002/tea.20178

Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: The interplay of memorising and understanding, and the development of knowledge objects. *Higher Education Research & Development*, *22*, 19-41. https://doi.org/10.1080/0729436032000056562

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. https://doi.org/10.3758/BF03193146

Firth, J., Rivers, I., & Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Review of Education, 9*, 642-684. https://doi.org/10.1002/rev3.3266

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*, 608-628. https://doi.org/10.3758/BF03201087

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

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”?. *Psychological Science*, *19*, 585-592. https://doi.org/10.1111/j.1467-9280.2008.02127.

Lu, X., Penney, T. B., & Kang, S. H. (2020). Category similarity affects study choices in self-regulated learning. *Memory & Cognition*, *49*, 1-16. https://doi.org/10.3758/s13421-020-01074-w

Marshak, S. (2015). *Essentials of geology* (5th ed.). New York, NY: W.W. Norton.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*, 462-476. https://doi.org/10.3758/s13421-010-0035-2

Meagher, B. J., Carvalho, P. F., Goldstone, R. L., & Nosofsky, R. M. (2017). Organized simultaneous displays facilitate learning of complex natural science categories. *Psychonomic Bulletin & Review*, *24*, 1987-1994. https://doi.org/10.3758/s13423-017-1251-6

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89-115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Miyatsu, T., Nosofsky, R. M., & McDaniel, M. A. (2020). Effects of specific-level versus broad-level training for broad-level category learning in a complex natural science domain. *Journal of Experimental Psychology: Applied*, *26*, 40-60. https://doi.org/10.1037/xap0000240

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*, 257-271. https://doi.org/10.1080/09658211.2014.1001992

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multilevel induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science*, *25*, 1592-1599. https://doi.org/10.1177/0956797614535938

Nosofsky, R. M., & McDaniel, M. A. (2019). Recommendations from cognitive psychology for enhancing the teaching of natural-science categories. *Policy Insights from the Behavioral and Brain Sciences*, *6*, 21-28. https://doi.org/10.1177/2372732218814861

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*, 530-556. https://doi.org/10.3758/s13428-017-0884-8

Nosofsky, R. M., Slaughter, C., & McDaniel, M. A. (2019). Learning hierarchically organized science categories: simultaneous instruction at the high and subtype levels. *Cognitive Research: Principles and Implications*, *4*, 1-17. https://doi.org/10.1186/s41235-019-0200-5

Palmeri, T. J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. *Psychonomic Bulletin & Review*, *6*, 495-503. https://doi.org/10.3758/BF03210840

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192-233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605. https://doi.org/10.1016/0010-0285(75)90024-9

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*, 145-151. https://doi.org/10.1111/j.0956-7976.2005.00795.x

Tarbuck, E. J., & Lutgens, F. K. (2017). *Earth science* (15th ed.). Boston, MA: Pearson.

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. https://doi.org/10.3758/s13423-012-0319-6

Wahlheim, C.N., Dunlosky, J. & Jacoby, L.L. (2011) Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*, 750–763. https://doi.org/10.3758/s13421-010-0063-y

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, *169*, 145-173. https://doi.org/10.1007/s11229-008-9340-8

Witherby, A. E., Tauber, S. K., & Goodrich, M. (2022). People hold mood-congruent beliefs about memory but do not use these beliefs when monitoring their learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48,* 499-519. https://doi.org/10.1037/xlm0001096

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*, 918-933. https://doi.org/10.1037/xge0000177

Zoom Video Communications Inc. (2016). Security guide. Zoom Video Communications Inc. Retrieved from https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf

VITA

Personal Information

Addison Leigh Poage Babineau

a.babineau@tcu.edu

Education

Master of Science, Experimental Psychology
Emphasis: Cognition
Texas Christian University, 2020
Mentor: Dr. Sarah "Uma" Tauber

Bachelor of Science Behavioral and Cognitive Neuroscience
Colorado State University, 2018
Undergraduate Thesis Mentor: Dr. Matthew Rhodes

Publications

Babineau, A.L. & Tauber, S.K. (2022). Decisions to Block Study are related to Practice Classification Performance: Support for the Performance Monitoring Hypothesis for Category Learning. *Memory & Cognition.* Advanced Online Publication. 1-17

Babineau, A.L., Witherby, A.E., Ariel, R., Pelch, M.A., & Tauber, S.K. (2022). Do Domain Knowledge and Retrieval Practice Predict Students' Study Order Decisions? [Special issue]. *Journal of Intelligence, 10:122*, 1-20

Sitzman, D.M., Rheams, J., Babineau, A.L., & Tauber, S.K. (2022). Older and Younger Adults' Revision of Health Misconceptions. *Memory*, 30(2), 172-188.

Experience

Research Assistantship, Texas Christian University
Supported by the James S. McDonnel Foundation 21st Century Science Initiative in Understanding Human Cognition, Collaborative Grant
2018-2019

Teaching Assistantship, Texas Christian University
2019-2021

Research Assistantship, Texas Christian University
2021-2023

ABSTRACT

DISCRIMINATIVE CONTRAST AND THE ROLE OF CATEGORY ORGANIZATION ON CATEGORY LEARNING

By Addison Leigh Poage Babineau, Ph.D., 2023
Department of Psychology
Texas Christian University

Dissertation Advisor: Dr. Sarah “Uma” Tauber, Associate Professor of Psychology

Learning to classify information into concepts and categories is an essential component to educational success. However, learning to correctly classify examples into a given category can be challenging, particularly when learning STEAM (Science, Technology, Engineering, Art, & Math) categories. Two factors that may impact STEAM category learning are category organization and study order. For example, when learning geological categories, the information may be organized into basic categories such as igneous and metamorphic rocks; or the information may be organized into subcategories such as granite and peridotite (which are types of igneous rocks). When studying either subcategories or basic categories, students may decide to study in blocks, such that a student would study several examples from the same category in a row; or they could study the categories in an interleaved order, such that the categories are mixed together. The discriminative contrast hypothesis suggests that the study order most beneficial to learning may be contingent to category organization. Thus, the goal of the present dissertation was to systematically evaluate the impact of study order and category organization from the lens of the discriminative contrast hypothesis. Across a Pilot experiment and two high-powered experiments, study order did not significantly influence classification performance when

learning most geological categories. However, interleaving categories during learning was beneficial to novel classification performance for students who learned basic categories with many exemplars. Whereas study order did not impact classification performance for most groups, category structure did impact classification performance. Specifically, students who learned to classify geological subcategories performed significantly better on the classification tests than did students who learned basic geological categories. The present results are inconsistent with the discriminative contrast hypothesis and indicate a nuanced context in which study order influences STEAM category learning.