Article

Policy Insights from the Behavioral and Brain Sciences
2019, Vol. 6(1) 21–28

DOI: 10.1177/2372732218814861
journals.sagepub.com/home/bbs
SAGE

# Recommendations From Cognitive Psychology for Enhancing the Teaching of Natural-Science Categories

**Robert M. Nosofsky[1] and Mark A. McDaniel[2]**


## Abstract

Because of their complex structures, many natural-science categories are difficult to learn. Yet achieving accuracy in classification is crucial to scientific inference and reasoning. Thus, an emerging theme in cognitive-psychology and cognitive-science research has been to investigate better ways to instruct about categories. This article briefly reviews major findings that will help inform policies for teaching categories in the science classroom. Many of the examples come from our specific project that examines teaching rock classifications in the geologic sciences. This project uses formal models of human category learning—developed in cognitive psychology—to search for optimal teaching procedures. The model-suggested category-teaching procedures often lead to better learning outcomes than do alternative procedures motivated by teachers' and students' intuitive judgments. In addition to reviewing these enhanced procedures for teaching natural-science categories, the article points to recent broader efforts for fostering collaborations between cognitive-science researchers and education researchers.



## Keywords

category learning, science education, instruction, formal models, discipline-based education research


## Tweet

Animal species, rocks, and trees . . . How do we best teach categories like these? Research from cognitive psychology now reveals principles for enhancing students' learning of categories in the science classroom!

## Key Points

- Scientific categories are fundamental to reasoning and inference in science, but these categories can be challenging to learn.
- An emerging theme in cognitive-psychology research has been to investigate better ways to instruct about categories.
- Formal (computer) models of human category learning have enhanced the search for optimal teaching procedures.
- Recommendations include making use of multiple, variable training items from the categories; intermingled presentation sequences of training items; and incorporating active testing into learning sessions.
- The next step is to translate the core principles and recommendations to authentic educational settings and the science classroom.

## The Nature and Importance of Category Learning in Science Education

Science learning, especially at higher levels of education, requires learning the key categories of each domain. Expert botanists have great skill in identifying different plants, entomologists are expert in identifying varieties of insects, and expert geologic scientists have great facility in identifying rock types. In recent years, our intensive project has the goal of enhancing the teaching of rock categories in introductory undergraduate geoscience classes (Nosofsky, Sanders, & McDaniel, 2018a, 2018b), so many of this article's examples come from this domain. However, as described in the following, the key principles from our research are likely generalizable across scientific domains.

For readers who believe that learning to classify rocks is easy, consider an example. The top row of Figure 1 displays

[1]Indiana University Bloomington, USA
[2]Washington University in St. Louis, USA

**Corresponding Author:**
Robert Nosofsky, Department of Psychological and Brain Sciences, Indiana University, 1101 E. Tenth Street, Bloomington, IN 47405, USA.
Email: nosofsky@indiana.edu

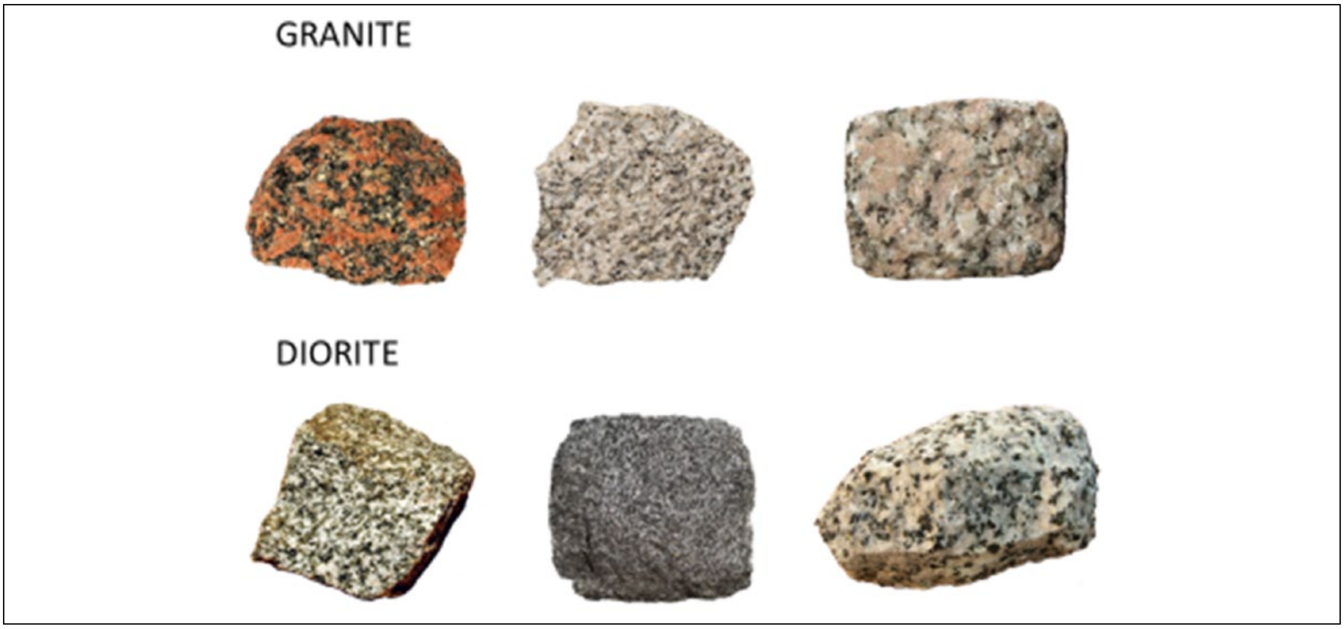


**Figure 1.** Example pictures of three samples of *granite* (top row) and three samples of *diorite* (bottom row).

pictures of three samples of *granite*, a type of rock familiar to many people. The bottom row displays pictures of three samples of a closely related rock known as *diorite*. Learning the subtle features that discriminate between these similar categories of rocks would clearly be a challenging task.

As with other natural-science categories, rock types have a graded structure, with clear prototypical instances at their centers, but with many less typical instances as well (Rosch, 1973; Smith & Medin, 1981). Thus, individual samples of the same type of rock often display remarkable within-category variability. For example, granites vary enormously in their color, ranging from black/white to pink or orange to red. Also, as with most natural categories, the boundaries dividing different rock types are often fuzzy, and contrasting categories can sometimes even overlap. A salient example is the granite/diorite samples in Figure 1. In short, rock categorization appears to be both a challenging and representative example of natural-science category learning.

As with category learning in other sciences, teaching rock classifications is an important and early goal in geoscience education. Introductory college-level geology textbooks devote multiple chapters to classifying rocks (e.g., Marshak, 2015; Tarbuck & Lutgens, 2017). The chapters describe in detail the major types of rocks and characterize their organizing attributes. Reviews of science-education research stress that developing expertise in a domain relies on pattern-recognition abilities—exhibited on multiple scales (e.g., a geographic region, a layered rock outcropping, an individual rock; Petcovic & Libarkin, 2007). One such pattern-recognition ability involves rock classification: Expert geoscientists can often recognize diverse rock types by relying on diagnostic perceptual features (e.g., color, granularity, heft).

Teaching fundamental categories, such as rock types in geology, is a core component of science curricula for good reason: Categories are the building blocks of our basic thought processes. They provide an efficient means to allow us to reason about the nature of the world and draw inferences. The role of rock classification in reasoning and inference abounds in geology. For example, as middle-school earth-science classes teach, *igneous* rocks form from the cooling and solidification of magma, the hot fluid under the earth's crust. Categories of igneous rocks can be *intrusive* versus *extrusive*. Intrusive igneous rocks, such as granite, form when magma solidifies at depth. In this case, the magma cools slowly, allowing large crystalline mineral structures to develop, resulting in a coarse grain. By contrast, extrusive igneous rocks, such as rhyolite, form when magma solidifies in a surface environment. In this case, the magma cools quickly, resulting in a fine-grained crystalline structure. A geologist examining a terrain might, therefore, obtain clues about its history by determining whether the rocks that compose the terrain are intrusive or extrusive igneous rocks, as evidenced by the grain size of the rocks. Note that although the geologist may ultimately be most interested in the causal and historical processes that led to the formation of the rocks and their appearance in the current environment, those causal processes occurred eons ago and are no longer directly accessible. Instead, the act of categorizing the rocks helps guide the reasoning processes that lead the geologist to infer the history of particular terrains.

## Psychological-Science Research and Guidance From Formal Models

In sum, because of their complex graded structures and fuzzy boundaries, natural categories are often challenging to learn. Yet achieving accurate classification is crucial to scientific reasoning and inference. Thus, an emerging theme in cognitive-science research communities has been to investigate factors that facilitate the *instruction* of categories (e.g., Jacoby, Wahlheim, & Coane, 2010; Kang & Pashler, 2012; Kornell & Bjork, 2008; Nosofsky, Sanders, Zhu, & McDaniel, 2018; Pashler & Mozer, 2013). As these investigations have underscored, determining the optimal instruction procedures can be a daunting task.

Consider just a few of the fundamental decisions that arise in prescribing how to best foster category learning:

i. Which training instances should be used? For example, some might favor presenting students with a few prototypical samples of a category, so that students can get a sense of the central features of the category. Others might favor presenting a range of samples of the category so that students can understand the variability of category features. In our experience, science teachers disagree about which procedure produces better outcomes.
ii. In what order should the instances be presented? For example, when several categories must be learned (e.g., granite, diorite, rhyolite), should instances from the same category be blocked together during instruction or should instances from the different categories be intermingled? Students generally prefer the blocked regimen—but does that produce the best learning?—we look at the evidence in the next section.
iii. What mix of study versus testing should be applied? In numerous educational settings, tests are viewed primarily as instruments of assessment. However, influential work in cognitive psychology demonstrates that testing can also enhance later retention, beyond what can be achieved by equivalent amounts of restudy alone.
iv. Should the focus be on teaching general rules? For many science categories, general rules will not completely capture the category—there are often many exceptions to the general rules. Or should students learn by studying many instances and inducing the category on their own?

Conducting empirical studies to systematically navigate through the vast set of combinations of teaching possibilities would be an extraordinarily time-consuming and expensive process. Our work on rock category learning has implemented a research approach to surmount this challenge. The approach involves making use of sophisticated, formal (computer) models—developed in the field of cognitive psychology—that predict human categorization performance in quantitative detail (for a comprehensive review, see Pothos & Wills, 2011). For example, suppose students are tasked with learning to classify a collection of rocks into a set of 10 different igneous-rock categories. The formal models can do a reasonably good job of predicting the precise probability that students will classify each individual rock sample into the competing categories and how this classification behavior will vary across different teaching conditions (e.g., Nosofsky, Sanders, Zhu, & McDaniel, 2018). Although this article cannot describe these formals models, it does document teaching strategies validated by these models and their empirical tests.

Thus, the formal models can help guide the search for effective procedures for teaching the categories. The idea is to use the successful models of human category learning to simulate the outcome of different teaching techniques, including a focus on the fundamental instructional questions described earlier (e.g., Khajah, Lindsey, & Mozer, 2014; Mathy & Feldman, 2016; Nosofsky, Sanders, Zhu, & McDaniel, 2018; Patil, Zhu, Kopec, & Love, 2014). The significant advantage is that cognitive psychologists can then focus empirical studies on those techniques that the models predict would be most successful.

The following section illustrates such an approach. In doing so, we highlight findings to guide policy recommendations for translation to undergraduate science education. Many of the examples come from our own research on rock-category learning because our project applies formal models of category learning to identify promising instructional procedures that have thus far been vetted with empirical studies using rock categories.

## Promising Strategies for Enhancing the Teaching of Science Categories

### *Use Variable Samples*

One key finding is that, if the goal is to enhance students' ability to generalize to category items that the students have not yet seen, then training with multiple, variable samples of the categories is better than training on a few central samples of each category (e.g., Homa & Vosburgh, 1976; Nosofsky, Sanders, Zhu, & McDaniel, 2018; Posner & Keele, 1968; Wahlheim, Finn, & Jacoby, 2012). For example, as noted earlier, the well-known rock type granite can be extraordinarily variable in its appearance. Training with multiple samples that capture this variability is better than training with only a few central, clear-cut samples (Nosofsky, Sanders, Zhu, & McDaniel, 2018).

Well-established formal models of human category learning correctly predict this result for several reasons. First, when training with multiple, variable samples, the training samples will more completely cover each category's multidimensional features. Thus, if the student is presented with a newly seen item from the category, that item is more likely to have a highly similar training sample as a neighbor, given multiple, variable training samples. The presence of the high-similarity neighbor boosts the student's ability to correctly classify the new item. Second, participants are far less likely to attend to idiosyncratic (often misleading) features of the training samples when training-set size is large rather than small. For example, if trained on some very small set of granite samples, the learner might resort to strategies such as remembering that one of the particular granites had (by circumstance) a triangular shape, but this idiosyncratic property would be useless in allowing the learner to generalize to as yet unseen items of the category. By learning instead to focus attention on category-level features (e.g., granite's interlocking crystals and coarse grain size) rather than on idiosyncratic ones, students are in better position to generalize correctly to novel items of the trained categories.

## Use Idealized Training When Categories Overlap

A second key finding places limits on the principle of training with multiple, variable samples. In particular, when categories include items from overlapping categories (e.g., in terms of their perceptual features), "idealized" training, rather than training with the actual overlapping items, is better (e.g., Giguère & Love, 2013; Hornsby & Love, 2014; Nosofsky & Stanton, 2005). Idealized training excludes samples from the less likely category in the overlapping regions. For example, a key attribute that generally discriminates between the igneous rocks rhyolite and andesite is lightness/darkness of color. In idealized training, students would not train with a very dark-colored sample of rhyolite that overlaps into the dark region that is more typical of andesite. As anticipated by successful formal models of human category learning (e.g., Nosofsky & Palmeri, 1997), this form of idealized training greatly reduces the retrieval of inconsistent forms of evidence that are harmful to making optimal classification decisions (Giguère & Love, 2013; Nosofsky & Stanton, 2005). (Supplementary forms of instruction could, however, make students aware of the possibility of the overlapping distributions.)

## Use Mingled, Not Blocked, Samples

A third main finding involves optimal presentation sequences of training samples when teaching multiple categories. For many natural categories, including those taught in science classes, presenting samples from the multiple categories in an intermingled fashion is better than presenting samples from each of the single categories in a massed, blocked fashion (e.g., Eglington & Kang, 2017; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2012; but see Carvalho & Goldstone, 2014). For example, suppose that students are training to classify rocks into 10 different igneous-rock categories, and that performance assessment later tests the students on novel items from the trained categories. Generalization performance during the test will be enhanced if the original training sequences provided intermingled presentations across the multiple categories rather than massed, blocked presentations of each single category (e.g., Eglington & Kang, 2017; Kornell & Bjork, 2008).

Presenting items in intermingled fashion improves comparison processes by allowing students to focus attention on features that provide effective contrasts between competing categories. So, for example, students would learn to focus attention on the attribute of lightness/darkness of color to discriminate between the rock types rhyolite and andesite but would learn to focus attention on the attribute of rounded versus angular fragments to discriminate between the rock types conglomerate and breccia. The importance of focusing attention on relevant dimensions in optimizing learners' classification performance has been central to formal modeling of human categorization for decades (e.g., Nosofsky, 1986; Shepard, Hovland, & Jenkins, 1961).

Although the formal categorization models correctly predict the benefits of presenting high-variability training instances and of using intermingled presentation schedules, these outcomes sometimes do not match up with the students' own intuitions about what types of category training aid learning. For example, one study compared students' learning of different artistic styles by using intermingled versus blocked training sequences (Kornell & Bjork, 2008). In self-reports recorded after training, the students thought that blocking was more effective than intermingling; in fact, their objective test performance showed the opposite pattern. Likewise, students learned different bird categories in a study that manipulated the variability of training instances (Wahlheim et al., 2012). The students' ability to generalize correctly to novel samples of the trained categories was better in high-variability conditions than in low-variability ones, yet the students provided self-report judgments that their category knowledge was better in the low-variability conditions. These two instances of "metacognitive illusions" illustrate that intuitions about the best instructional situations can be dramatically off base. The pervasiveness of these metacognitive illusions (Bjork, 1994) strongly reinforces the potential benefits of using formal models and careful empirical research to help guide the search for effective teaching techniques, rather than relying on intuitive judgments.

## Use Active Learning

A fourth well-established finding from the cognitive science of category learning is that active learning also promotes category instruction. For example, quizzes (rather than repeated study) aid memory and learning (e.g., Karpicke & Blunt, 2011), and this result holds for learning natural concepts (Jacoby et al., 2010) and for learning in the science classroom (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011). Active testing promotes learning because it provides students with practice in the kinds of memory retrieval needed when generalizing to novel items of the learned categories.

The research on active testing is closely related to another well-researched issue in the psychology of category learning, namely, comparisons between observational versus feedback learning. In observational learning, the student is presented with sequences of items along with their category labels and studies each item-category pair on each trial. By comparison, in feedback learning, after being presented with each item, the student first attempts to generate or recall its associated category label and only then receives feedback concerning the correct answer. Note that the feedback-learning procedure basically involves the use of testing on each individual trial, so it presumably also provides the learner with extensive practice in memory retrieval. With the exception of certain category structures defined by exceedingly simple single-feature rules, feedback learning generally leads to better outcomes than does observational learning in the domain of perceptual categorization (e.g., Ashby, Maddox, & Bohil, 2002). This finding suggests that instructors could assist students in science category learning by incorporating a lot of feedback-learning exercises into the class.

## *Other Promising Procedures: Fading, Coaching, Varying Levels*

Additional research from the cognitive-science community has pointed to still other procedures that show promise for enhancing the teaching of natural categories. First, *fading* procedures are likely to enhance natural-science category learning. In fading, a highly salient cue is introduced at the outset of training, making correct responding relatively easy. The strength of the cue is then gradually removed ("faded") to more typical or challenging levels. For example, early in training, one might present very light-colored samples of rhyolite and very dark-colored samples of andesite; as training proceeds, the lightness/darkness contrast gradually reduces, but this relevant stimulus property still controls the classification response. Fading procedures have been particularly effective for enhancing generalization to novel test samples of categories when the samples of those categories comprise large numbers of features, many irrelevant to the goal of telling the categories apart (Pashler & Mozer, 2013). For example, lightness/darkness is a key feature for discriminating between rhyolite and andesite; numerous other attributes such as the hue, size, and shape of the samples might also vary, but they are irrelevant to achieving the correct classification.

Explicit coaching can also enhance instance-based training of natural categories. For example, in a study that taught a series of challenging rock classifications, *feature highlighting* was incorporated into the training samples (Miyatsu, Gouravajhala, Nosofsky, & McDaniel, in press; see also Eglington & Kang, 2017). In this procedure, highly diagnostic but subtle features for discriminating between different rock types were explicitly listed on the computer screen, with surrounding ovals on the rock images to highlight the presence and location of those features. For example, to enhance students' ability to classify items as breccia versus conglomerate, training explicitly identified the angular versus rounded fragments that compose these rock types. Overall, the feature-highlighting technique significantly improved category generalization, compared to control conditions that did not provide this explicit coaching.

Another key issue concerning the teaching of natural-science categories is that most are arranged in hierarchies, and the instructor will want students to be able to classify at *multiple levels of the hierarchy.* Rocks, for example, have three main high-level divisions: igneous (discussed previously), metamorphic, and sedimentary. (Metamorphic rocks undergo profound physical or chemical changes in their form after being subjected to intense heat or pressure; sedimentary rocks form when mineral and organic particles are deposited on the floor of bodies of water and are eventually cemented together.) Each of the three high-level divisions includes numerous lower-level types. Igneous-rock types include granite, obsidian, and pumice; metamorphic-rock types include marble, quartzite, and slate; and sedimentary-rock types include conglomerate, shale, and sandstone.

Now, suppose the geology instructor's primary goal is to teach students to classify correctly at the high-level division of igneous versus metamorphic versus sedimentary. To meet this primary goal, an intuitively sensible approach would be to focus training at this high level, without requiring the student to also learn distinctions at the lower level of the different types of igneous, metamorphic, and sedimentary rocks. Logically, however, training could occur simultaneously at both the high division level and at the lower level of types of igneous, metamorphic, and sedimentary rocks. For example, rather than learning to classify an item as simply an igneous rock, the student might be required to learn simultaneously that the rock's high-level division is igneous *and* that its specific type is granite.

The preferred method of training turns out to depend on the details of the category structures being learned (Nosofsky, Sanders, Gerdom, Douglas, & McDaniel, 2017). In particular, consider the case in which the high-level category divisions form *compact* clusters, with all rock types within each division being similar to one another, and with rock types that belong to contrasting divisions being dissimilar. In this case, if the goal is to learn to classify correctly at the high-division level, it is indeed best to focus training at this high level, without the requirement that the lower-level rock-type names also be learned. By contrast, suppose instead that the high-level divisions are structured in a disorganized and *dispersed* manner, with numerous rock types *within* each division being *dissimilar* to one another, and numerous rock types belonging to *different* divisions being highly *similar*. In this case, it is better to require the student to simultaneously learn the high-level division *and* the specific rock-type name for each rock. Granted, the student is being required to learn additional forms of information (specific rock-type names) that are not logically needed to achieve the primary goal (learning the high-level division to which each rock sample belongs). Nevertheless, the result was predicted a priori by a well-known formal model of how humans learn to classify at multiple levels of category hierarchies (for details, see Nosofsky et al., 2017).

In the study, students learned only small subsets of the high-level categories of igneous, metamorphic, and sedimentary rocks (Nosofsky et al., 2017). The small subsets were purposely selected to produce either highly compact or highly dispersed categories that needed to be learned. It turns out that if one considers the full battery of igneous, metamorphic, and sedimentary rocks that exist in nature, the category structures are likely closer to the dispersed end of the continuum than toward the compact end (for preliminary evidence, see Nosofsky et al., 2017; Nosofsky, Sanders, Meagher, & Douglas, 2018). For example, numerous igneous rock types are highly dissimilar from one another yet highly similar to various metamorphic and sedimentary rocks. Ongoing research examines whether the pattern of results for the dispersed categories (Nosofsky et al., 2017) may also be observed when students are taught full batteries of the types of igneous, metamorphic, and sedimentary rocks that are encountered in introductory, undergraduate geoscience classes.

## Individual Differences in Students' Category-Learning Strategies

Formal models of category learning like those that underpin the research featured in this article have established two distinct ways that learners can construct their categories. One is to learn the particular instances of each category that are presented during training (*exemplar-based* approach; for example, Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). New instances from the categories can then be categorized based on how similar they are to the collection of trained instances. Specifically, if the new instance is more similar to the trained instances from Category A than to trained instances from the other categories, then the new instance is classified as a member of Category A. A different approach is to try to abstract an underlying rule or scheme that ties all of the trained instances together (e.g., Bourne, 1974; Posner & Keele, 1968). New instances from the categories are then classified according to the rule or scheme abstracted from the training items.

For artificial laboratory categories, some learners prefer an exemplar approach to learning categories, whereas other learners prefer to acquire an underlying rule or abstraction that summarizes the general characteristics of the category (e.g., rocks classified as conglomerate have rounded fragments cemented together; Little & McDaniel, 2015; McDaniel, Cahill, Robbins, & Wiener, 2014; Wahlheim, McDaniel, & Little, 2016). In terms of educational applications, college students appear to adopt these different strategies when confronted with learning complex concepts in their science courses (in particular chemistry; Frey, Cahill, & McDaniel, 2017; McDaniel et al., 2018). A potential implication of this finding is that category-learning instruction might be most successful if the nature of the instruction is tied to the category-learning strategy preferred by individual students. For example, instruction that emphasizes underlying rules that tie together training instances might fare well with learners who prefer a rule-learning strategy; by contrast, instruction that encourages students to learn a variety of training instances may fare well with learners who prefer an exemplar strategy. At the present juncture, these possibilities stand only as hypotheses, and new studies need to test them.

## Forming Collaborations Between Cognitive Science and Education Researchers

To develop policies for improving and transforming undergraduate Science, Technology, Engineering and Mathematics (STEM) education, a clear requisite is to build an evidence base that informs such policies. The research described in this article is a part of that effort. On a broader scale, the need to develop a comprehensive evidence base has spurred the development of a constellation of new disciplines that are known as discipline-based education research (DBER), such as physics education research, biology education research, chemistry education research, and geoscience education research (Singer, Nielsen, & Schweingruber, 2012). DBER has primarily been driven by grassroots efforts of faculty in the respective disciplines (e.g., individuals trained as biologists "switching" to studying teaching and learning of biology). Accordingly, much DBER work has been conducted without clear connections or integration with fields that have a long tradition of conducting research on teaching, learning, and thinking, such as cognitive science and educational psychology. In parallel to DBER efforts, many cognitive scientists have started focusing their efforts on research and theory that is inspired by educational issues and learning in authentic educational contexts. Yet, much of this work has not been directly connected to particular discipline-based content or learning challenges addressed by DBERs. This lack of crosstalk between DBER scholars and cognitive psychologists has meant that potentially useful teaching and learning strategies remain unknown or unused by most STEM faculty (see, for example, Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013).

We (among many other DBERs and cognitive scientists) believe that policy efforts to foster and stimulate cognitive science–DBER collaborations would have a high payoff in providing a strong evidence base for improvement of STEM education. One of the challenges, however, is that cognitive science and DBER are independent fields, each with diverse subfields, and no natural mechanism exists for enhancing collaborations. A recent positive trend is that cognitive scientists and DBERs are starting to hold meetings to discuss ways to surmount the obstacles for cognitive science–DBER collaborations and to share research findings (e.g., National Science Foundation sponsored meetings in Washington DC, September 2016, and at Washington University in St. Louis, September 2018). To promote this cross-fertilization in deeper ways, federal funding agencies—such as the National Science Foundation and the Institute for Education Sciences—and private foundations could sustain funding programs for promoting cognitive psychology–DBER collaborations. As some examples, the NSF CORE program (Education and Human Resources [EHR] Core Research [ECR], Fundamental Research in STEM Education) developed with such collaborations in mind and currently funds such work (indeed, much of the work outlined in this article was funded by this program). The NSF EHR IUSE program (Improving Undergraduate STEM Education) also encourages collaborative research and development to improve undergraduate teaching and learning. With these forward-looking funding efforts, compelling evidence to inform science education can help leverage effective science instruction for all students.

## Conclusion and Policy Insights

The evidence on methods of enhancing the teaching of categories is well established in the cognitive-psychology laboratory. Future research is needed to more fully document that the methods will translate well to authentic educational settings and the science classroom. Nevertheless, although continued refinement will likely take place, it is almost certain that many of the core principles—including the use of multiple variable training samples, intermingled presentation orders, and active testing—would lead to significantly enhanced category-learning outcomes in the authentic classroom settings. The recommendations could be easily implemented by science teachers in educational and professional settings with little or no cost and high potential benefits. Many of the techniques could likely also be implemented in the form of easily transportable computer-based training programs that could provide useful supplements to forms of instruction delivered in the classroom and the field. The recommendations contribute to the goal of enhancing STEM education for the diverse set of students in this country.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Science Foundation Grant 1534014 (EHR Core Research) to Robert Nosofsky, Bruce Douglas, and Mark McDaniel.

### References

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*(5), 666-677.

Bjork, R. A. (1994). Memory and meta-memory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bourne, L. E., Jr. (1974). An inference model of conceptual rule learning. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 231-256). Potomac, MD: Lawrence Erlbaum.

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.

Dunlosky, J., Rawson, K., Marsh, E. J., Nathan, M., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58.

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

Frey, R. F., Cahill, M. J., & McDaniel, M. A. (2017). Students' concept-building approaches: A novel predictor of success in chemistry courses. *Journal of Chemical Education*, *94*, 1185-1194.

Giguère, G., & Love, B. C. (2013). Limits in decision making arise from limits in memory retrieval. *Proceedings of the National Academy of Sciences*, *110*, 7613-7618.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 322-330.

Hornsby, A. N., & Love, B. C. (2014). Improved classification of mammograms following idealized training. *Journal of Applied Research in Memory and Cognition*, *3*, 72-76.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1441-1451.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97-103.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775.

Khajah, M. M., Lindsey, R. V., & Mozer, M. C. (2014). Maximizing students' retention via spaced review: Practical guidance from computational models of memory. *Topics in Cognitive Science*, *6*, 157-169.

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

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Little, J., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule-abstraction. *Memory & Cognition*, *43*, 85-98.

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

Mathy, F., & Feldman, J. (2016). The influence of presentation order on category transfer. *Experimental Psychology*, *63*, 59-69.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399-414.

McDaniel, M. A., Cahill, M. J., Frey, R. F., Rauch, M., Doele, J., Ruvolo, D., & Daschbach, M. M. (2018). Individual differences in learning exemplars versus abstracting rules: Associations with exam performance in college science. *Journal of Applied Research in Memory and Cognition*, *7*, 241-251.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, *143*, 668-693.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Miyatsu, T., Gouravajhala, R., Nosofsky, R. M., & McDaniel, M. A. (in press). Feature highlighting enhances learning of complex natural-science categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-61.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266-300.

Nosofsky, R. M., Sanders, C. A., Gerdom, A., Douglas, B. J., & McDaniel, M. A. (2017). On learning natural-science categories that violate the family-resemblance principle. *Psychological Science*, *28*, 104-114.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018a). A formal psychological model of classification applied to natural-science category learning. *Current Directions in Psychological Science*, *27*, 129-135.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018b). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General*, *147*, 328-353.

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.

Nosofsky, R. M., Sanders, C. A., Zhu, X., & McDaniel, M. A. (2018). Model-guided search for optimal training exemplars in a natural-science category domain: A work in progress. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-018-1508-8

Nosofsky, R. M., & Stanton, R. D. (2005). Speeded classification in a probabilistic category structure: Contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 608-629.

Pashler, H., & Mozer, M. C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1162-1173.

Patil, K., Zhu, X., Kopec, L., & Love, B. (2014). Optimal teaching for limited-capacity human learners. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 2465-2473). NIPS 2014.

Petcovic, H. L., & Libarkin, J. C. (2007). Research in science education: The expert-novice continuum. *Journal of Geoscienceeducation*, *55*, 333-339.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353-363.

Pothos, E. M., & Wills, A. J. (Eds.). (2011). *Formal approaches in categorization*. Cambridge, UK: Cambridge University Press.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York, NY: Academic Press.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1-42.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

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

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

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory & Cognition*, *40*, 703-716.

Wahlheim, C. N., McDaniel, M. A., & Little, J. L. (2016). Category learning strategies in younger and older adults: Rule abstraction and memorization. *Psychology and Aging*, *31*, 346-357