Journal of Geoscience Education

ISSN: 1089-9995 (Print) 2158-1428 (Online) Journal homepage: http://www.tandfonline.com/loi/ujge20

Training of rock classifications: The use of computer images versus physical rock samples

Brian J. Meagher, Kirstyn Cataldo, Bruce J. Douglas, Mark A. McDaniel & Robert M. Nosofsky

To cite this article: Brian J. Meagher, Kirstyn Cataldo, Bruce J. Douglas, Mark A. McDaniel & Robert M. Nosofsky (2018) Training of rock classifications: The use of computer images versus physical rock samples, Journal of Geoscience Education, 66:3, 221-230, DOI: 10.1080/10899995.2018.1465756

To link to this article: https://doi.org/10.1080/10899995.2018.1465756

Published online: 30 May 2018.

Submit your article to this journal

Article views: 69

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=ujge20

JOURNAL OF GEOSCIENCE EDUCATION
2018, VOL. 66, NO. 3, 221–230
https://doi.org/10.1080/10899995.2018.1465756

Routledge
Taylor & Francis Group

Check for updates

Training of rock classifications: The use of computer images versus physical rock samples

Brian J. Meagher[a], Kirstyn Cataldo[b], Bruce J. Douglas[b], Mark A. McDaniel[c,d], and Robert M. Nosofsky[a]

[a]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405; [b]Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, Indiana 47405; [c]Department of Psychological and Brain Sciences, Washington University St. Louis, St. Louis, Missouri 63130; [d]Center for Integrative Research on Cognition, Learning, and Education, Washington University St. Louis, St. Louis, Missouri 63130

ABSTRACT

A highly controlled laboratory experiment was conducted that suggested computer-based image training of rock classifications can provide a useful supplement to physical rock training. Two groups of participants learned to classify samples of 12 major types of rocks during a training phase. One group was trained using computer images of the rock samples, and another group was trained with physical rock samples. A third group that was familiarized with images of the samples but did not receive initial classification training served as a control. The participants' ability to generalize their training to the classification of novel, physical rock samples from the 12 types was then assessed in a test phase. All groups received trial-by-trial feedback during this test phase; still, the image-based and physical rock training groups maintained a significant performance advantage (75.2% correct) compared to the control group (37.5% correct). The group that received physical rock training performed only slightly better overall (77.8% correct) than the image-based training group (72.5% correct), although the advantage for the physical rock training group was substantial for some specific types of rocks from the complete set. The results provide documentation for the potential benefits of using image-based classification-training methods as a means of supplementing physical rock classification-training methods.

ARTICLE HISTORY
Received 10 July 2017
Revised 07 March 2018
Accepted 13 April 2018
Published online 30 May 2018

KEYWORDS
Rock classification; learning and memory

Introduction

Teaching rock classifications is one of the early goals in geoscience education. Introductory college-level geology textbooks devote multiple chapters to the classification of rocks (e.g., Marshak, 2015; Tarbuck & Lutgens, 2017). The chapters provide detailed descriptions of the major types of igneous, metamorphic, and sedimentary rocks and attempt to characterize the major dimensions that organize and compose them. Reviews of science education research stress the idea that the development of expertise in a domain often relies on pattern-recognition abilities exhibited on multiple scales (Petcovic & Libarkin, 2007). One such pattern-recognition ability involves rock classification: Expert geoscientists can often recognize diverse rock types by relying on various diagnostic perceptual features of the rocks. Indeed, the teaching of rock classification is fundamental because it is one of the key components involved in allowing geologic scientists to reason and make inferences about the geologic history of given terrains (e.g., Petcovic, Libarkin, & Baker, 2009).

Learning to classify rocks presents significant challenges and can sometimes be difficult even for expert observers. One reason is that the perceptual cues needed for discriminating between rock types are often quite subtle and difficult to discern. For example, telling apart certain samples of *diorite* from certain samples of *granite* may depend on the detection of subtle, fine-grained quartz crystals in the granite samples. A second reason is that individual samples of the same type of rock can often display remarkable variability. For example, depending on their mineralogy, granites can be predominantly white, pink, gray, or orange in color.

Thus, rather than simply providing descriptions of rock types, an essential component of teaching the classifications involves providing experience with a range of samples within each of the types. In this article, the particular samples used in initial training of the rock types will be referred to as training samples. Ideally, because the goal is to classify novel samples in the real (physical) world, the best training samples would likely be physical, real-world samples. However, in practical terms, there

CONTACT Robert M. Nosofsky nosofsky@indiana.edu Department of Psychological and Brain Sciences, Indiana University, 1101 E. Tenth Street, Bloomington, IN 47405, USA.
© 2018 National Association of Geoscience Teachers

are often limits on training with physical samples. For example, a student is unlikely to always be in close proximity to all the relevant variations of physical samples of the different rock types. Even in a college-level geology classroom or actual field situations, instructors may have available only a limited number of physical samples to illustrate the large number of to-be-learned types. Given the subtle discriminating cues and the extreme variability of the samples within the individual rock types that need to be learned, providing numerous distinct training samples is likely to be beneficial.

These considerations suggest that the use of sample *images*—pictures of rocks—could provide the student with a significant head start on the learning of rock classifications, or perhaps provide an important supplement to training with physical rock samples. Besides being easily transportable, systematic computer algorithms might be developed that optimize the selection of training samples, their sequence of presentation, and so forth (Kirchoff, Delaney, Horton, & Dellinger-Johnston, 2014; Nosofsky, Sanders, & McDaniel, 2018). In addition, whereas the number of physical samples to which a student has ready access is almost certainly limited, an essentially unlimited number of training samples can be made available using image-training techniques.

From certain perspectives, the claim that image-based training techniques can be valuable in teaching rock classification seems noncontroversial. After all, introductory college-level texts and lab manuals make extensive use of representative pictures of rock samples for illustrating the major rock types. Furthermore, there is an extensive set of videos available on the internet as well as interactive programs that are designed to help with rock identification and classification. Nevertheless, there is little if any actual empirical evidence of the extent to which use of such techniques may allow students to effectively generalize image-based training to the classification of novel, physical rock samples. The purpose of the present research was to begin to fill this gap by testing, under highly controlled laboratory conditions, whether students could generalize their knowledge of rock classifications derived from image-based sample training techniques to the classification of novel, physical rock samples. The tests were conducted by adapting from classic paradigms used in the fields of cognitive science and the cognitive psychology of category learning.

Importantly, the hypothesis pursued here was not that image-based classification training can provide a substitute for forms of training that rely on use of physical rock samples. Indeed, there are well-documented benefits of direct field experience in geoscience education (Mogk & Goodwin, 2012), and such benefits suggest that training rock classifications with images of rocks is likely to be significantly inferior to training classification with physical rock samples. Such a view is bolstered by the fact that there are numerous attributes of rocks that cannot be communicated via visual images alone, including their density, magnified view, magnetism, mineral content, results of chemical tests, and so forth. Moreover, mere viewing of pictures does not provide for forms of embodied learning that are critical—for example, when students manipulate rocks for observation (Mogk & Goodwin, 2012).

Instead, the present research explored the weaker hypothesis that use of computer image training techniques might provide a significant head start compared to no classification training at all. In addition, a secondary purpose was to evaluate whether computer image training techniques might produce learning—at least within the current type of training paradigm—that was nearly as good for supporting generalization to physical rock samples as that produced by physical rock training. If either of these outcomes was observed, it would provide initial empirical evidence that computer image training could provide a useful supplement to methods based on physical rock training.

The experimental paradigm

Participants learned to classify multiple samples of rocks into 12 distinct types comprising a mixture of igneous, metamorphic, and sedimentary rocks. The rock types were amphibolite, basalt, breccia, conglomerate, diorite, gabbro, gneiss, micrite, obsidian, pumice, sandstone, and schist. All of these types are among the common ones that are described and listed in tables in introductory college-level geology texts (e.g., Marshak, 2015; Tarbuck & Lutgens, 2017), so use of these types seemed like a reasonable approach to initiating this investigation. Of course, students are capable of learning to classify rocks at finer levels of detail than at the level tested here. Future research would be needed to test the limits of image-based classification training techniques at these more detailed levels of classification.

In all conditions, there was an initial training phase followed by a test phase. The experimental manipulation involved the nature of the training phase, in which different groups of participants received alternative forms of training. In the test phase, all participants were evaluated on their ability to generalize their initial training to the classification of novel, physical rock samples.

There were three main conditions, distinguished solely by the type of initial training participants received. Importantly, the specific individual rock samples used in training were held constant across all three groups; only the method of classification training was varied. In a physical rock training condition, participants received classification training using actual physical samples of the rock types. In an image-training condition, participants received classification training using computer-displayed pictures of the same physical samples as were used in the physical rock training condition. In a control condition, participants were familiarized with the rock images, but did not receive classification training. The control condition was included to provide a baseline of comparison for the physical rock and image training conditions; it was not intended as a plausible type of training method that might be used in a classroom setting. Following training, all groups engaged in a test phase that involved the classification of novel, physical rock samples. The nature of the test phase was held constant across the groups.

Participants learned via a trial-by-trial test-feedback method. On each trial, a single rock sample was presented, the learner attempted to classify it into one of the 12 types, and immediate corrective feedback was provided. The trial-by-trial test-feedback method was continued during the test phase of the experiment, in which all participants classified the physical rock samples. Thus, there was the possibility of continued learning during the test phase. This continued feedback procedure during the test phase was implemented as we were interested in whether the learning that took place during the initial training phase would have long-lived benefits once teaching with the actual physical rock samples began. If participants in the control group, who received no initial classification training, immediately "caught up" with the training groups once the test phase started, such a result would suggest that little practical benefit was achieved by the initial training phase. By contrast, if the training groups maintained a significant advantage compared to the control group, even when feedback was being provided during the test phase, such a result would speak even more strongly to the utility of the initial training and emphasize the potential benefits of training with computer images of rocks.

Method

Participants

Forty-five participants were recruited from the Indiana University Bloomington community through general advertisements. They participated in exchange for payment, which was $24 per 2-hour session plus a possible $6 bonus for good performance. Participants were randomly assigned to the three conditions, with 15 participants per condition. All participants reported little or no prior knowledge of rock classifications. Because assignment of participants to conditions was randomized, any differences in prior knowledge across conditions would likely be a minor source of experimental noise.

Materials

The experiment was programmed in MATLAB using the Psychophysics Toolbox (Brainard, 1997) and was conducted on a Windows-based computer with a 16-inch LCD monitor.

The complete set of samples used in the experiment consisted of 96 rocks. The rocks came from the 12 different types listed earlier, with eight samples of each type. The 96 rocks were divided into two sets of 48 samples each. Within each set there were four samples of each rock type. One set was used for the training phase, and one was used for the test phase. The assignment of a particular set to each phase was counterbalanced within each condition. Under the guidance of the geoscience expert of the research team (the third author), we attempted to make the sets as balanced as possible, in the sense that for each rock sample in one set, there was a corresponding, similar-looking rock sample in the second set. However, given the limited number of physical rock samples we had available for use, it was not always possible to achieve this goal for all the rock samples. The rocks varied substantially in size, from approximately an inch to nine inches long.

Under the guidance of the geoscience expert of the research team, we took photographs of each rock, which were used in the control and image-training conditions. The photographs were displayed at a fixed size at the center of the computer screen, and subtended a visual angle of approximately $7^\circ \times 7^\circ$. We attempted to take the photographs in a manner that captured the rocks' most diagnostic features—for example, the holes that appear in pumice and the embedded fragments that appear in breccia and conglomerate. We also attempted to preserve a sense of the relative size of each rock in the photographs, while still making them large enough to display their detail.

The two sets of physical rocks were located on either side of the computer table where the participant sat during the experiment. Both sets were organized into four rows of 12 rocks each, with one rock from each of the 12 types in each row. To prevent the use of location cues as a basis for classifying the rocks, the order of the rocks in each row was randomized. Each rock was assigned an alphanumeric label, consisting of a letter for its row and a number for its column position within its row. To prevent the labels of the two sets of rocks from being confused with one another, the rows of the first set were labeled A through D, and the rows of the second set were labeled E through H. This labeling system was used to improve the efficiency with which the experimenter (either one of the co-authors or a lab research assistant) could locate the rocks during the training and test phases.

Procedure

The experiment consisted of an initial training phase followed by a test phase, with a mandatory 10-minute break between the phases. Within each condition, and regardless of whether the training samples were physical rocks or images, eight participants had Set 1 assigned as the training set and Set 2 as the test set, whereas the remaining seven participants had Set 2 as the training set and Set 1 as the test set. At the start of the experiment, participants were informed that if they performed well during the test phase, they would receive a $6 pay bonus, but they were not told the percentage of correct responses they would need to earn this bonus. In the control condition, the criterion for receiving the bonus was 50% correct, whereas in the image training and physical rock training conditions, the criterion was 65% correct. All participants were tested individually in a private testing room.

During the initial training phase, participants were seated in front of a computer and—depending on condition—were told that they would be (a) learning to classify physical rocks (physical rock training), (b) learning to classify pictures of rocks (image training), or (c) making a series of ratings regarding different characteristics of pictures of rocks, such as how attractive they were (control). The structure of this initial training phase of the experiment is summarized in the top panel of Table 1. Regardless of condition, this training phase involved four main blocks of 48 trials each (for a total of 192 trials), with each main block further divided into four subblocks of 12 trials each. Each trial consisted of the presentation of one particular physical rock or rock image (depending on the condition), with an accompanying response from the participant (with the type of response also varying across the conditions). In all conditions, the rocks were presented in a random order, subject to the constraints that (a) each one of the individual 48 training samples was presented exactly once during each main block of 48 trials and (b) exactly one sample of each of the 12 rock types appeared in each subblock of 12 trials. The same 48 training samples were then presented in a new random order in each subsequent main block of the training phase, subject to the same presentation order constraints just described. In sum, during each of the four main blocks of the training phase, participants viewed 48 trials in which they saw four unique samples of each rock type, with the samples of each type spread roughly evenly across the subblocks of training. Furthermore, this procedure was repeated across the four main training blocks. Each main block lasted approximately 10 minutes, so the training phase lasted approximately 40 minutes.

Table 1. Organizational scheme for methods.

Block	Subblock	Trials per subblock	Total trials per block	Total trials per phase
Training Phase				
1	1, 2, 3, 4	12	48	
2	5, 6, 7, 8	12	48	
3	9, 10, 11, 12	12	48	
4	13, 14, 15, 16	12	48	192
Testing Phase				
1	1, 2, 3, 4	12	48	
2	5, 6, 7, 8	12	48	96

Physical rock training condition

At the beginning of each trial in the physical rock training condition, the computer displayed on screen the alphanumeric label (e.g., *B4*) corresponding to the training rock selected for that trial. The participant reported the label to the experimenter, who then located the selected rock in the 4 × 12 array of physical rocks that lay on the table. The experimenter handed the rock to the participant. The participant was free to examine the rock in any manner she or he desired, and then classified it into one of the 12 types by pressing one of 12 labeled keys at the top row of the computer keyboard. Following the response, the participants were provided with feedback: The word "CORRECT" appeared onscreen after a correct response, and the word "INCORRECT" after an incorrect response. Regardless of whether the response was correct or incorrect, the name of the rock type was also displayed, centered below the correct/incorrect message. The feedback remained on screen for three seconds, followed by a blank screen intertrial interval of one second. The participant then handed the rock back to the experimenter, who placed it back into the training set. At the end of each block, the computer reported to the participants their percentage of correct responses during that block.

Image-training condition

On each trial in the image training condition, a picture of a rock would be displayed on the computer screen, and participants classified it into one of the 12 types in the same manner as in the physical rock training condition. The feedback procedure was the same as in the physical rock condition, except rather than there being a physical rock present, the picture of the rock continued to be displayed on the screen along with the feedback. The feedback and the picture remained on screen for three seconds, followed by a blank screen intertrial interval of one second. At the end of each block, the computer reported to participants their percentage of correct responses during that block.

Control condition

The purpose of the control condition was to force participants to look at the rocks in order to provide them with general familiarity about the visual characteristics of the rocks but without providing any form of classification training. Again, this condition was included to provide a baseline for comparison with the physical rock training and image training groups and not as a plausible training method that might be used in a classroom setting. Participants were asked to make ratings, using a 9-point rating scale (1 = lowest, 9 = highest), of each of the rock pictures on four dimensions: attractiveness, interestingness, strangeness, and value. These dimensions were chosen to be uninformative and to avoid directing participants' attention to any features of the rocks that could be used for purposes of aiding their scientific classifications. As in the other conditions, there were again four blocks of 48 trials each, with the presentation schedule of the rock samples the same as in the other conditions. Participants rated one of the four dimensions per block and made their responses by pressing the number keys on the computer keyboard. Due to the subjective nature of the judgments, no corrective feedback was provided. For each participant, the dimensions were assigned to the blocks in a random order. On each trial, a single rock picture was displayed on the computer screen, and the participant entered the relevant rating.

Test phase

After the 10-minute break, participants were again seated in front of the computer. Regardless of their original training condition, participants were asked to classify physical rocks from the set of samples they did *not* use for their training; for example, if Set 1 was used for training, Set 2 was used for testing. As shown in the bottom panel of Table 1, the test phase consisted of two blocks of trials, with the procedure in each block being identical to that used in the training phase of the physical rock training condition. In particular, each test block was again divided into four subblocks of 12 trials each, with a single physical sample from each of the 12 rock types tested within each subblock. Note that feedback was again provided on each trial during the test phase; accordingly, learning could conceivably continue during that phase. As explained in the introduction to this experiment, this procedure was adopted in order to ascertain how quickly the control group participants might catch up with the image training and physical rock training participants once they started receiving feedback training. Note also that because the same specific rock samples used in the first test block were used in the trials in the second test block, the second test block is of less theoretical interest, because much of the performance may involve rote memorization of recently classified samples rather than category-based generalization.

At the end of the test phase, participants were debriefed and received their payment.

Results

Three participants' data were removed due to unusual behavior, such as rushing through the experiment or not following instructions. All three of these participants were from the control condition, leaving 12 participants in the control. The mean proportion of correct responses for the control condition during the first test block was 0.375 (SD = .124). The participants whose data were removed all scored below this mean (0.271, 0.292, and 0.333). As reported below, performance in the control condition was clearly worse than in the two training conditions; thus, the removal of these participants' data is a conservative procedure with respect to finding differences in performance across the groups. In particular, had these three participants' data been included, mean performance in the control condition would have been even worse. The analyses reported below are based on the data from the remaining 42 participants.

Training phase

Figure 1 displays the overall mean proportion of correct responses during the training phase as a function of subblocks in the image training and physical rock training conditions. Because there were four subblocks in each of the four training blocks, the subblocks are labeled consecutively from 1 to 16 (see Table 1 for reference). Also, because the participants in the control condition did not engage in classification training, there are no training data to report for the control group. As can be seen from Figure 1, performance improved dramatically during the course of training, with the image and physical rock training conditions showing similar trajectories. This observation is confirmed by a two (condition: image vs. physical rock) × 16 (subblock: 1 through 16) mixed-model ANOVA on the training data, with condition as a between-subjects factor and subblock as a within-subjects factor. The main effect of the subblock factor was significant—$F(15, 420) = 72.97$, MSE (mean-squared error) = .013, $p < .0001$—confirming that substantial learning occurred as training progressed. There was no main effect of the condition factor, $F(1, 28) < 1$; and no interaction between the condition and subblock factors, $F(15, 420) = 1.19$, $MSE = 0.013$, $p = 0.34$. Thus, by the end of training, participants in the image and physical rock conditions had learned to classify the rocks at

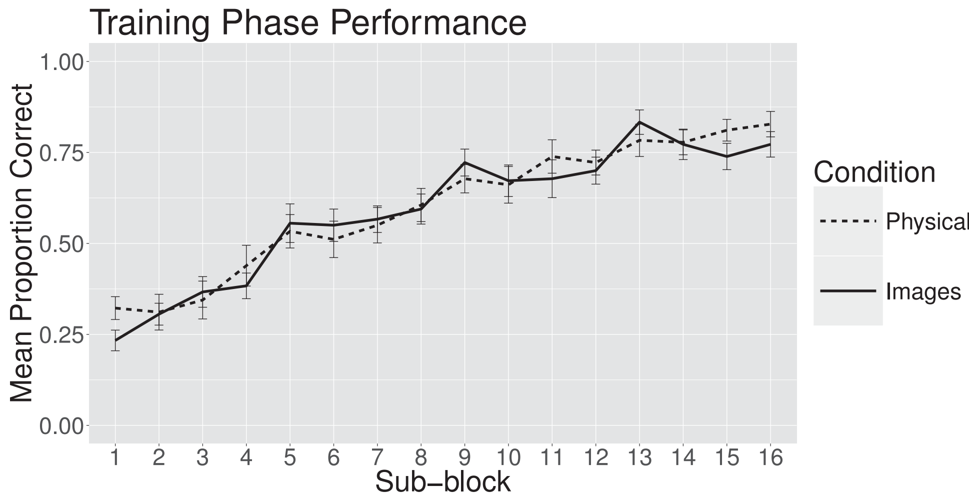

Figure 1. Mean proportion of correct responses during the training phase as a function of conditions (physical rock training versus image-based training) and subblocks, with error bars representing one standard-error of the mean. Performance in both conditions largely overlapped during this phase of the experiment, suggesting that participants in both conditions had learned to classify the rocks at roughly the same level of accuracy.

roughly the same level of accuracy, providing equal baselines with which to assess performance in the test phase.

Test phase

Because the central interest is in the participants' ability to generalize what they had learned during training—that is, classify novel physical samples of the 12 rock types—the major analyses focus on only the first test block. Because the same specific rocks were simply repeated during the second test block, performance during the second block may entail a good deal of rote memorization. Thus, the second block test results are reported in secondary analyses.

Figure 2 displays the overall mean proportion of correct classification responses during the first test block as a function of conditions and the four subblocks. As can be seen, performance in the image and physical rock training conditions was similar, whereas performance in the control condition was much worse than in the other two conditions. To support this observation, a 3 (condition: control, image, physical rock) × 4 (subblock: 1–4) mixed-model ANOVA was conducted, with condition as a between-subjects factor and subblock as a within-subjects factor. The analysis revealed a significant main effect of the condition factor, $F(2, 39) = 61.41$, $MSE = .040$, $p < .0001$. The main effect of the subblock factor was not significant—$F(3, 117) = 1.16$, $MSE = .013$, $p =$

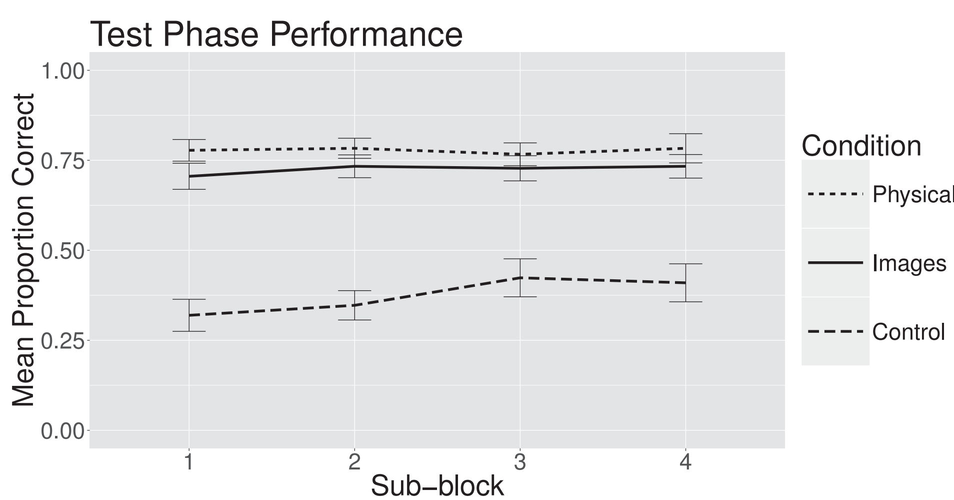

Figure 2. Mean proportion of correct responses during the first test block as a function of conditions (physical rock training, image-based training, control) and subblocks, with error bars representing one standard error of the mean. Performance in both the image and physical rock training conditions was better than performance in the control condition, with performance in the physical rock condition trending somewhat higher than in the image-based condition. These test block results suggest that classification training with images of rocks can serve as a valuable supplement to physical rock based training.

.33—nor was the interaction between the condition and subblock factors, $F(6, 117) = 0.81$, $MSE = .013$, $p = .57$.

In addition, statistical tests were conducted that examined the difference between each individual pair of conditions during this first test block. The performance difference between the control and image training conditions was statistically significant—$F(1, 39) = 81.20$, $MSE = .010$, $p < .0001$—as was the performance difference between the control and physical rock training conditions, $F(1, 39) = 107.53$, $MSE = .010$, $p < .0001$. By contrast, the overall performance difference between the image and physical rock training conditions did not reach significance—$F(1, 39) = 2.08$, $MSE = .010$, $p = .16$—although the trend was for the physical rock condition to have a slight advantage.

Performance on rock type by condition

To achieve a more detailed comparison of performance across the image and physical rock training conditions, performance on each of the individual rock types was measured across the conditions. The results are displayed in Figure 3. Although the focus here is on the image versus physical rock conditions, for completeness the figure displays the results for the control condition as well. Viewed from this perspective, overall performance is better in the physical rock condition than in the image condition for a couple of specific rock types. In addition, the rock types vary considerably in their overall level of difficulty. To confirm these observations, a 2 (condition: image vs. physical rock) × 12 (rock type) mixed-model ANOVA was conducted on the proportion-correct data. Although the trend was for overall performance to be better in the physical rock training condition, this main effect did not reach statistical significance, $F(1, 28) = 2.62$, $MSE = .096$, $p = .12$. There was a significant effect of the rock type factor—$F(11, 308) = 28.44$, $MSE = 0.044$, $p < .0001$—reflecting that the types varied considerably in their overall difficulty. Most importantly, the interaction between the condition and rock type factors was also statistically significant—$F(11, 308) = 2.87$, $MSE = 0.044$, $p = .001$—reflecting that the effectiveness of physical rock training versus image-based training differed across the individual rock types.

Prior to conducting the experiment, a key hypothesis was that physical rock training would be particularly beneficial, compared to image-based training, for the rock types *schist, sandstone*, and *pumice*. For schist, a person's ability to observe its schistocity is enhanced if the rock is manually rotated; for sandstone, a person's perception of the sandy texture of the rock is enhanced by tactile manipulation; and for pumice, a person's perception of its extreme low density is a cue that becomes available only through hefting the rock, not through visual inspection of an image. As can be seen in Figure 3, there was indeed marked improvement in classifying schist and sandstone in the physical rock condition compared to the image training condition; for pumice, however, meaningful comparisons cannot be made because the level of performance reached ceiling in both conditions. Pairwise t-tests revealed that schist was indeed classified significantly more accurately in the physical rock training condition than in the image training condition—$t(28) = 5.03$, $p < .001$—and that the same pattern of significant results was observed for sandstone, $t(28) = 2.12$, $p < .05$. An additional pairwise t-test was conducted between the image and physical rock conditions for the average of all rock types except for schist,

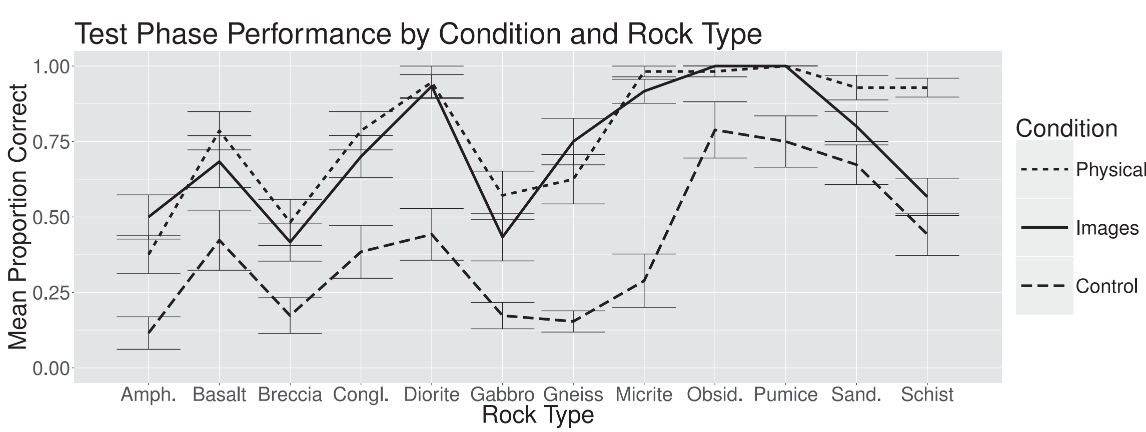

Figure 3. Mean proportion of correct responses during the first test block as a function of conditions (physical rock training, image-based training, control) and rock type, with error bars representing one standard error of the mean. Performance across all conditions differed significantly by rock type, indicating that some rock types are more difficult to learn than others. Importantly, performance on a couple of particular rock types—schist and sandstone—was significantly better in the physical rock condition than in the image-based condition. The difference in performance across the physical rock and image-based conditions was not statistically significant for any of the other rock types. Note: Amph. = amphibolite, Congl.= conglomerate, Obsid. = Obsidian, Sand. = sandstone.

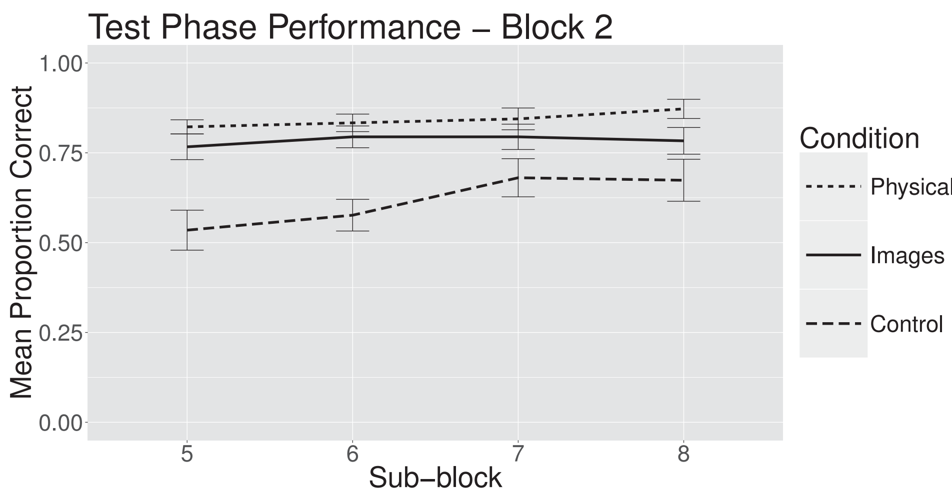

Figure 4. Mean proportion of correct responses during the second test block as a function of conditions (physical rock training, image-based training, control) and subblock, with error bars representing one standard error of the mean. Performance in the image-based and physical rock training conditions was again higher than in the control condition, suggesting that the benefits of both of these types of training were long-lived. Performance in the control condition began to improve during this second test block. However, because the samples in this block were identical to those of the first test block, the improvement likely reflects some rote memory for some of the repeated test samples, rather than an improved ability to generalize to novel samples of the rock types.

sandstone, and pumice. The difference in the grand means for these other rock types did not approach statistical significance, $t(28) = .42$, $p = .68$.

Second test block

The mean proportion of correct responses during the second test block, in which the samples from the first block were simply repeated, is shown as a function of conditions and subblocks in Figure 4. Although performance continues to improve in the control condition, it remains significantly below performance in both the physical rock training condition and the image-based training condition. An ANOVA revealed a main effect of the condition factor, $F(2, 39) = 15.21$, $MSE = .047$, $p < .001$; and a main effect of the subblock factor, $F(3, 117) = 4.35$, $MSE = .010$, $p < .01$. The interaction between the condition and subblock factors did not reach statistical significance, $F(6, 117) = 1.67$, $MSE = .010$, $p = .14$. A direct pairwise statistical comparison revealed that performance in the image training condition remained significantly higher than in the control condition, $F(1, 39) = 15.97$, $MSE = .012$, $p < .001$. The overall performance difference between the image and physical rock conditions again did not reach significance—$F(1, 39) = 2.16$, $MSE = .012$, $p = .15$—although performance in the physical rock condition again trended slightly higher. Hence, the performance advantages yielded by the initial training in the image-based and physical rock conditions, compared to the control condition, were long-lived: Across the two test blocks, there were 96 study-test-feedback trials with the physical rocks. Nevertheless, it is important to note that performance in both the image-based and physical rock training conditions appears to be reaching an imperfect asymptote, despite the continued feedback received on the repeated rock samples. This issue is discussed at greater length in the final section of this article.

Discussion

Summary

Our research has provided evidence that, compared to a control condition, classification training with images of rocks can provide long-lived significant enhancement of observers' ability to classify novel samples of physical rocks into their scientific categories. Moreover, with a couple of exceptions, the image-based classification training led to generalization performance on samples of physical rocks that was nearly as good as that yielded through physical rock training.

Given that the distinctions between different rock types are often subtle, and that there is often extensive variability of appearance within rock classification types, it seems important that students be exposed to a wide variety of examples of the different rock types in learning to classify them. However, presenting wide varieties of physical samples is not always feasible. Thus, training with computer-generated rock images may provide an important supplement to physical rock training, a possibility that is bolstered by the present research results.

Instructional implications

As clarified at the outset, we make no claim that image-based training will suffice on its own to allow adequate generalization for properly classifying rocks in the real world. For example, various nonvisual cues are critical to rock classification, including characteristics such as density, mineral content, outcomes of chemical tests, and so forth. In addition, various visual characteristics become stronger when the observer can manipulate physical rocks through rotation and view the rocks at different angles and scales. Indeed, in the present experiment, clear evidence was obtained that observers learned to classify *schist* and *sandstone* with significantly higher accuracy in the physical rock training condition than in the image-based condition. The most likely reasons are that a person's ability to observe schist's mica grains is enhanced through manual rotation of the rock, and the perception of sandstone's texture is enhanced through tactile manipulation. Thus, rather than replacing physical rock classification training with image-based training, the suggestion here is a weaker one: namely, that image-based classification training can provide a useful supplement to classification training with physical rocks. Possibly, image-based training might provide students with an initial foothold on the rock types that need to be learned, and could also provide students with an appreciation of the great variability across individual samples that exists within many of the rock types. Such initial training could provide a valuable head start and be followed with experience in the real-world field situations needed to develop true geologic expertise.

Limitations and future research directions

As noted earlier, and as can be seen in Figures 2 and 4, even in the physical rock training condition of the present experiment, learners appear to have reached a far-from-perfect asymptote in their classification performance of the present rock types. Another limitation of the experiment was that testing for generalization occurred only 10 minutes after training; consequently, generalization performance during the test phase might be even less accurate after a longer delay. Thus, various techniques would be needed to enhance the present forms of sample-based training, regardless of whether the training is through images or physical rocks. It seems likely that the performance limit is reflecting a difficulty in discriminating between specific pairs of rock types with rather subtle diagnostic characteristics. For example, the participants in our experiment had a great deal of difficulty discriminating between *breccia* and *conglomerate* and between *amphibolite* and *gabbro*. As discussed in the introduction, the boundary lines between diagnostic characteristics that separate between different rock types are often subtle. This seems to be the case, for example, with respect to the rounded fragments that compose conglomerate and the angular fragments that compose breccia. A teacher may need to draw special attention to this subtle characteristic to improve learners' classification performance. Regarding amphibolite and gabbro, both are dark-colored, coarse-textured rocks; however, amphibolite often has a weakly foliated texture reflecting a planar alignment of minerals; this type of structure is not present in gabbro. In the samples of rocks used in this experiment, this characteristic was again extremely subtle; thus, performance would likely be enhanced if steps were taken to draw explicit attention to this subtle characteristic.

There are various methods that might be used to improve these subtle discriminations. One issue concerns the sequencing of the training samples. In the present paradigm, the rock types were presented in a random order, subject to the constraint that each type was viewed once per subblock of 12 trials. Various studies have shown that explicit interspersing of difficult-to-discriminate real-world categories can enhance classification learning (e.g., Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim, Dunlosky, & Jacoby, 2011), so this direction seems like a promising one to pursue. Other recent studies have demonstrated that presentations of category members in organized, simultaneous visual displays can enhance the learning of dispersed categories—that is, ones in which there is a good deal of variability within categories and subtle discriminating characteristics between categories (Meagher, Carvalho, Goldstone, & Nosofsky, 2017). Although in the Meagher et al. (2017) study such displays were used to enhance the teaching of the broad categories of igneous, metamorphic, and sedimentary rocks, similar principles ought to apply in teaching classifications at the present, more specific level of types of rocks. Still other approaches place emphasis on active forms of learning, in which observers decide for themselves which samples they should view on any given trial (e.g., Markant & Gureckis, 2014). Another domain of current exploration involves techniques in which the teacher provides explicit verbal instruction along with simultaneous visual highlighting of subtle discriminating features to enhance sample-based training of rock classifications (Miyatsu, Gouravajhala, Nosofsky, & McDaniel,). Such enhanced techniques are likely to be valuable regardless of whether the sample-based training is purely image-based or involves the use of physical rocks.

Funding

This work was supported by National Science Foundation Grant 1534014 (EHR Core Research) to Robert Nosofsky, Bruce Douglas, and Mark McDaniel.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436. doi:10.1163/156856897X00357

Carpenter, S.K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*, 369–378. doi:10.1007/s10648-012-9205-z

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103. doi:10.1002/acp.1801

Kirchoff, B. K., Delaney, P. F., Horton, M., & Dellinger-Johnston, R. (2014). Optimizing learning of scientific category knowledge in the classroom: The case of plant identification. *CBE-Life Sciences Education*, *13*, 425–436. doi:10.1187/cbe.13-11-0224

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

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*, 94–122. doi:10.1037/a0032108

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

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. doi:10.3758/s13423-017-1251-6

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*.

Mogk, D. W., & Goodwin, C. (2012). Learning in the field: Synthesis of research on thinking and learning in the geosciences. *Geological Society of America Special Papers*, *486*, 131–163. doi:10.1130/2012.2486(24)

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018). 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. doi:10.1037/xge0000369

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

Petcovic, H. L., Libarkin, J. C., & Baker, K. M. (2009). An empirical methodology for investigating geocognition in the field. *Journal of Geoscience Education*, *57*, 316–328. doi:10.5408/1.3544284

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

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. doi:10.3758/s13421-010-0063-y