Promoting transfer and mastery of evolution concepts with category construction

Purpose

Improving students' ability to acquire and transfer scientific knowledge is a central goal of K-12 science instruction. Studies over the past three decades have been instrumental in illustrating students' difficulties understanding and applying their knowledge of evolution (Bishop & Anderson, 1986; Nehm & Reilly, 2007; Nieswandt & Bellomo, 2009; Settlage, 1994). A key constraint on students' use of their prior knowledge is their ability to recognize when a novel situation is related to what they already know. Unfortunately, instructional practices often fail to emphasize the relationships among concepts and abstract principles that promote this kind of thinking. As such, recognition of relational similarity between examples often fails because memory retrieval relies on the surface commonalities of concepts and situations rather than the relations between them (Gentner, Ratterman, & Forbus, 1993; Gick & Holyoak, 1980, 1983; Ross & Kennedy, 1990).

The key goal of this research is to promote the development of accessible and transferrable conceptual knowledge. We hypothesize that the limitations of human, similarity-based retrieval can be overcome by focusing on teaching scientific principles as categories (Gentner & Kurtz, 2005; Bernardo, 1994; Blessing & Ross, 1996; Cummins, 1992). This paper reports on findings from two classroom experiments focused on a well-studied psychological task known as category construction (Imai & Garner, 1965, 1968; Medin, Wattenmaker, & Hampson, 1987). Historically, category construction techniques have been used as a means of assessing individuals' knowledge after a learning activity (as opposed to using category construction as a tool for promoting individuals' mastery of complex concepts and their ability to apply this knowledge in distinct domains. The following questions guide this paper:

- How does category construction compare to common instructional techniques?
- What is the effect of corrective feedback on the category construction task?
- What is the optimal temporal placement of the category construction task in relation to classroom instruction?

Theoretical Framework:

Categories hold privileged status in our minds as critical organizers of information. They compose the world into taxonomic kinds of things (capturing their basic nature or identity), offer stability as well as flexibility in reasoning and provide a basis for interpretation and prediction of the environment (Goldstone & Kersten, 2003). These properties make a categorization-based approach enticing as an untapped instructional tool because transferrable knowledge is relation-based and abstract. Learning examples of a scientific principle as members of a category will help students recognize the important relations of the principle and minimize focus on the idiosyncratic features present in any one concrete example.

Our work with category construction builds on research exploring comparison as a learning tool (Alfieri, Nokes-Malach & Schunn, 2013). This research has shown that comparison supports learning by promoting structural alignment between the targets of comparison, highlighting their shared relational structure, and promoting abstract understanding of relationally similar concepts (Gentner, 1983). Category construction sets itself apart from comparison in that it does not rely on learners to *implicitly* acquire relational knowledge. Instead,

category construction *explicitly* promotes this process because it is a constructive activity that helps students develop a tool for future reasoning – a relational category based on a scientific principle. In order to advance our understanding of the observed learning advantages of category construction in science classrooms and laboratory settings (Author, in preparation), we investigated the effect of learning with category construction under different conditions in two classroom-based experiments.

Data Source

This research was conducted in a suburban middle school in upstate New York. The school served 934 students during the 2011-2012 school year with approximately 40% qualifying for reduced-price lunch. The demographic breakdown is 88% Caucasian, 6.3% African American, 2.4% Asian/Pacific Islander, 2.2% Latino and the less than 1% Native American, with 13.9% of the district's students classified as having special needs. The experiments were conducted in 7th grade Life-Science classrooms by three certified teachers holding master's degrees in science education. The curriculum aligns with the New York State Intermediate Level Standards.

Method

Implementation of the category construction technique was similar for both of the presented experiments. The learning phase included the experimental intervention and a lesson targeting a sub-principle of evolution. The testing phase consisted of free-response and multiple choice questions designed to assess transfer and mastery of the target principles.

Students were given all materials on paper (see supplementary documents) and instructions were read aloud. Those assigned to the category construction condition were given cards containing vignettes and a mat (on which to sort the cards). Three cards were examples of the target principle and three cards were not. The mat contained instructions about the activity (sorting cards into two equal groups based on target principle) and a statement of the principle. The testing phase was conducted immediately upon sort completion. When two minutes remained in the class period, students were instructed to finish the sentence they were writing and hand in the assessment.

Response scoring proceeded in two stages for each free-response item and was blind to condition. First, responses were grouped based on similarity (i.e. responses that suggested that a species will be more resistant to disease with variation were grouped together), resulting in up to 21 categories per question. Second, response categories were compared and discussed within the research team and assigned scores based on adherence to the target principle. This process produced ordered categorical scores for each free-response question (see example provided in supplementary materials). The multiple choice questions were scored as correct/incorrect.

Experiment I targeted the evolutionary sub-principle: *variation can be an advantage* when the environment changes, and examined the effect of providing feedback during category construction. Experiment I also tested the effectiveness of category construction as compared to standard practice – a fill-in-the-blank, text-based worksheet. After the introductory lesson, students (N=191) were randomly assigned to one of four conditions: worksheet (n=48), category construction with no feedback (n=46), category construction with accuracy feedback (i.e., correct/incorrect) (n=54) and category construction with accuracy feedback and guidance (n=43). Students in the category construction with accuracy feedback condition were informed if their

work was incorrect and asked to try again. In addition to that, students in the category construction condition with accuracy feedback and guidance were directed to change one of the incorrect cards on their mat. WS students completed a worksheet illustrating the principle followed by 10 questions.

Experiment I assessment included two free-response transfer items and three freeresponse mastery questions centered on the principle of variation. Two of the mastery questions and all of the transfer questions had an associated multiple-choice component. Each question tested students' understanding and application of the variation principle. Designing targeted and valid items has proven to be difficult because existing validated materials focus on the entirety of the process of evolution by natural selection as opposed to the sub-components we identified in the New York State standards. Accordingly, the test questions we used were developed by the research team, though questions from NAEP and TIMSS science assessments were adapted when possible. All items were examined by domain experts from the project's advisory board and the participating middle school teachers.

Experiment II focused on the principle *living things compete for limited resources such as food and shelter.* Students in the category construction condition were asked to identify the examples that have *a group of living things that might evolve or go extinct because of competition.* Experiment II also examined the most effective positioning of category construction within a class period: before the related lesson, after the lesson, or the next day following the lesson. Unlike experiment I, classes (rather than students) were randomly assigned to condition. The category construction procedure was identical to that of experiment I except for the timing of the introductory lesson, the target principle and the feedback provided (all students received accuracy feedback with guidance consistent with that condition from experiment I).

Experiment II assessment contained two free-response transfer items, two free-response mastery items and two multiple-choice mastery items. Each question tested students' understanding and application of the competition principle. Assessment development paralleled that of experiment I.

Results

Response data were nested within teacher and classroom, and a non-negligible proportion was missing due to non-response (approximately 5% for experiment I and 7% for experiment II). The data were missing completely-at-random according to Little's MCAR test (Little, 1998). Accordingly, multiple imputation was conducted with the MI software package to compute missing values (Su, Gelman, Hill, & Yajima, 2011). Each experiment's assessment produced a set of ordered categorical and binomial variables. These variables were initially collapsed into two scores, one consisting of an average of all transfer items in a given assessment and one for all mastery items. After regression analyses uncovered no statistically-reliable differences between groups, an item-level analysis was conducted. The question of assessment validity will be addressed in the discussion. Cumulative link mixed model regressions were used (Christensen, 2012) to analyze each assessment item with ordinal scores. Three factors were included in each regression: condition was included as a nunordered categorical variable, classroom nested within teacher was included to control for general student aptitude.

Experiment I Results. Recall that the goals of experiment I were twofold: first, to investigate the effectiveness of category construction as compared to standard practice; and second, to explore the power of feedback during category construction. Our analyses uncovered reliable differences between two of the five assessment items. Students who received accuracy feedback (and no guidance about which card(s) were incorrect) were more likely to provide a high-quality response to our mastery question based on *variation among mammals* when compared to students in the worksheet condition (coefficient=0.781, *SE*=0.40, *z*=1.966, *p*<.05). Students with accuracy feedback also performed better than students in the no feedback condition (coefficient=0.904, *SE*=0.44, *z*=2.049, *p*<.05) and the accuracy feedback with guidance condition (coefficient=0.805, *SE*=0.44, *z*=1.847, *p*<.07) on the *team Ping-Pong* transfer question. No other items had reliable differences between conditions.

Experiment II Results. Experiment II sought to uncover the optimal timing of the category construction activity in relation to instruction on the same concept. In the experiment, only one of our four free response assessment items produced a reliable difference between the groups. The students who completed the category construction task after the related lesson were more likely to produce high-quality responses on the herbicide mastery question than students who completed the sort activity before the related lesson (coefficient=0.907, *SE*=0.45, *z*=2.011, p<.05). No other items had reliable differences between conditions.

Scholarly Significance

This initial investigation of category construction embedded in science classrooms has uncovered promising results which will inform the continued development of the category construction technique. In support of prior laboratory data (Authors, in preparation), the results suggest that the task promotes better learning outcomes than textbook-based worksheets. On the other hand, experiment I analyses uncovered some unanticipated results. Although we hypothesized that informing students about which card(s) were inaccurate in their sort would result in the best outcomes, students who were only told whether they were right or wrong performed better on our assessment. We attribute this result to the specific implementation of feedback in the condition that received extra guidance – due to time constraints, students were only given feedback if the result of their sort was incorrect; they were never given affirmative feedback. Allowing students to engage in category construction until correct may have promoted better outcomes.

The results of experiment II suggest that the sequence of category construction relative to the related instruction is an important implementation variable. Engaging in category construction after instruction served students better than completing the task before (or one day after) instruction, suggesting that category construction will be most effective as reinforcement for recently acquired knowledge. These findings are contrary to Bransford and Schwartz (1999) who report that direct instruction after training with contrasting cases was a potent combination. In our case, category construction may have facilitated students' ability to relate example scenarios to their background knowledge; resulting in greater support and consolidation of the target material.

A limitation of these results is the validity of our assessment questions. While it was preferred to use existing, validated assessment items, this was not possible due to the nature of our target principles. This point has consequences for this research and the community at large. It suggests that more work is needed for assessment development, validation and norming in follow-up investigations. More importantly, the lack of pre-existing questions that address subconcepts of evolution highlights the fact that evolution by natural selection is frequently taught as a singular process, rather than the combination of many processes related to variation, competition, and selection. We propose that this approach may hinder students' conceptual grasp of evolution and that decomposing evolution into its sub-concepts may be a more effective way to promote acquisition of the phenomenon as a whole (Gerjets, Scheiter and Catrambone, 2004).

As part of our multi-year investigation, this work will play a critical role in future implementations of the category construction task in the classroom, culminating in an empirically-supported, lightweight intervention for promoting the mastery and transfer of complex concepts.

References

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. Educational Psychologist, 48(2), 87-113.
- Bernardo, A. B. I. (1994). Problem-specific information and the development of problem-type schemata. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 379-395.
- Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching*, 27, 415-427.
- Blessing, S.B., & Ross, B.H. (1996). Content Effects in Problem Categorization and Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 792-810.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61-100.
- Christensen, R. H. B. (2012). Ordinal—regression models for ordinal data R package version 2012.01-19. Vienna: R Foundation for Statistical Computing. Available: www.cran.r-project.org/package=ordinal.
- Cummins, D. D. (1992). The role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1103-1124.
- Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), Categorization inside and outside the lab. (pp. 151-175). Washington, DC: APA.
- Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, 25, 524-575.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science*, 32(1-2), 33-58.
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306-355.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1-38.
- Goldstone, R. L., & Kersten, A. (2003). Concepts and Categories. In A. F. Healy & R. W. Proctor (Eds.) Comprehensive handbook of psychology, Volume 4: Experimental psychology. (pp. 591-621). New York: Wiley.

- Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, 69, 596-608.
- Imai, S., & Garner, W. R. (1968). Structure in perceptual classification. Psychonomic Monograph Supplements, 2(9, Whole No. 2).
- Little, R. J. A. (1998). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198 1202.
- Medin. D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*, 242-27.
- Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *Bioscience*, 57, 263-272.
- Nieswandt, M., & Bellomo, K. (2009). Written extended-response questions as classroom assessment tools for meaningful understanding of evolutionary theory. *Journal of Research in Science Teaching*, 46, 333-356.
- Settlage, J. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching*, 31, 449-457.
- Yu-Sung Su, Andrew Gelman, Jennifer Hill, Masanao Yajima (2011). Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *Journal of Statistical Software*, 45(2), 1-31.

E1 category construction mat

You are about to read some short passages about animals. Be sure to read them all carefully and think about the main idea of each passage.

Three of the cards you will read are examples of how variation (or differences) can be an advantage when the environment changes. In addition to those cards, there are three cards that do not show variation as an advantage.

Your job is to find the three examples that demonstrate an advantage of variation.

The most important part is making sure you can tell the difference between the cards with advantages because of variation and the others. When you are confident you have picked the correct examples, make sure to place the variation examples on the green locations and the other examples below.



E1 category construction cards

Principle Examples

Sardines

Sardines are fish that live in the ocean. While some sardines swim fast, others swim more slowly. If predators come, faster sardines will be better able to escape.

Frogs

Frogs use their tongues to catch insects to eat. Some frogs have longer tongues, while others have shorter tongues. When there are not enough insects for frogs to eat, the frogs with longer tongues will tend to catch the food they need.

Wolves

Wolves grow more fur in the winter. Some wolves grow very thick coats and some only grow a little more fur. If there is a harsh winter, the wolves with thicker fur will be better able to fight off the cold.

Non-examples

Whales

Whales get food by swimming through the water with their mouths open. Doing this allows them to scoop small fish into their mouth. If there are no small fish in the water, the whales cannot get enough food to survive because they have no other way to get their food.

Lizards

There is a species of lizard that has green skin. Their skin color helps them hide near green plants. If their habitat becomes very dry for too long a time, then the surrounding plants lose their green color. When this happens, predators can easily find and eat the lizards.

Cats

Cats clean their bodies by licking their fur. If a disease comes that spreads by licking fur, it is possible that many cats would die. This is because every cat uses the same method for cleaning.

E1 worksheet condition

Variation and Evolution

People understand that variation is a good thing. It's better to have many choices than only one or two choices. Variation is common in nature.

Where do we see variation in nature?

Variation is seen all around us. For example, there are many different types of organisms (Bacteria, Protozoa, Fungi, Plants, and Animals). Within these kingdoms, there are many variations. For example, the Kingdom Animalia has dogs, tigers, elephants, and goldfish. There is also variation within a single species. That is, all members of a given species have naturally occurring, random differences. For example, cats can have many colors of eyes. Some of these differences are the result of tiny, random genetic changes from one generation to the next. Even siblings (brothers and sisters) who share much of their genetic material have huge differences in how they look. Many of these differences don't really matter. But some differences will affect how well the organism is able to survive and reproduce.

How variation helps a species:

Colds are caused by viruses. The cold virus is a species that has lots of **variation**. The cold virus is present in humans all over the world. As long as there are humans, there will be cold viruses that infect us. Imagine that you have a cold. When you have a cold, a cold virus is using your cells to make new cold viruses. At the same time, you body kicks into high gear, fighting off the cold virus. Eventually your body becomes very good at finding the original virus. It kills all the original virus and then you no longer have a cold. Your body remembers how to fight the original virus so you do not get sick from the original virus again. However, not all the new cold viruses look like the original virus, and you body doesn't see a new virus as a problem.

So, if all cold viruses were identical to the original cold virus, you would get sick once, your body would fight off the cold, and you would be forever immune to the cold virus. Sadly, there is enough **variation** in the virus population, that about every year a few viruses change enough to no longer be recognized by your body. To fight off this new form of a cold virus, you have to get sick with a cold again, while your body learns to recognize the new cold virus.

A few questions:

6.

1	is seen all around us.	
	ingdoms,,	
common within each	,	, and it also is
	within a species	
4. Cold viruses use your	to make	
5. You have likely had a viral col good at remembering?	ld (fever, stuffy nose, sore throat) more than once. Ho	ow is that possible if your body is

_____ in the virus population results in you getting a cold each year.

E1 assessment



Questions that uncovered reliable differences denoted with red asterisks.

Response scoring example for Mammals question in E1

When mammals reproduce, the offspring are not just like their parents. But since the parents were able to survive and reproduce, it seems like a good idea to have perfect copies of them for the next generation.

What is the advantage of NOT having perfect copies?

Final Score	Group number	Response summary	
x	х	Blank scores were calculated with multiple imputation	
0	0	Nonsensical	
0	1	Life would be boring/confusing without variation	
0	2	So that offspring will be different from their parents	
0	3	So that they can/cannot reproduce	
1	4	Stating variation with no further explanation	
2	5	Variation helps the next generation to be better than their parents	
3	6	Variation allows the mammals to eat different kinds of food	
3	7	Variation makes the mammals more likely to survive	
3	8	Variation makes the mammals more resistant to disease or disability	
4	9	Variation supports the mammals' adaptation to their environment	
4	10	Variation supports the radiation of species	
4	11	Variation is needed so that the mammals can evolve	
4	12	Variation allows the mammals to adapt to the environment when it changes	

E2 category construction mat

You are about to read some short passages about animals. Be sure to read them all carefully and think about the main idea of each passage.

In this task you are going to read cards that describe how living things interact in their environment. Living things compete for limited resources such as food and shelter.

Your job is to find the cards that have a group of living things that might evolve or go extinct because of competition. You need to separate those cards from the cards that do not show competition.

The most important part is making sure you understand when competition can change a population of animals. When you are confident you have picked the examples that show competition, make sure to place the competition examples on the green locations and the other examples below.



E2 category construction cards

Principle Examples

Non-examples

Red Squirrels and Grey Squirrels

Red squirrels and grey squirrels need to store nuts for the winter, but there are never enough nuts for all the squirrels. Grey squirrels are quicker and store more nuts than the red squirrels. Many red squirrels will not survive the winter.

Bluebirds and Sparrows

Bluebirds and Sparrows like to build their nests in small holes in trees. There are never enough holes for all the birds. Sparrows can get the best nest locations because they are more aggressive. Without a nest, the bluebirds cannot raise their young.

Alligators and Pythons

Alligators and pythons hunt other animals for food, but the food supply has recently shrunk. Pythons are better hunters and eat more of the food supply. Soon there will be fewer alligators in the habitat. Honey Bees and Butterflies Honey bees and butterflies spend a lot of their time looking for the nectar of flowers. Bees can fly faster than butterflies and are able to get nectar more quickly. Even though bees are faster, butterflies can fly much higher and reach more flowers.

Squirrel Monkeys and Spider Monkeys

Squirrel monkeys and spider monkeys eat fruit from trees. Squirrel monkeys are much smaller, so they can eat fruit that grows in places that spider monkeys can't go. Spider monkeys move fast to get fruit that is closer to the ground.

Dolphins and Sharks

Dolphins and sharks hunt groups of fish. Dolphins have a way of swimming around to collect the fish in one spot. The sharks and dolphins both benefit from this and do not have to fight to get the fish they need.

E2 assessment

	-	
You take a trip to a town you've never visited before and soon discover that it has two bakeries. One baker specializes in bread and the other specializes in desserts. They are both quite good. You read in the newspaper that there used to be many more bakeries in town. What do you think happened to all the other bakeries?	Carl's, a local-owned grocery store, has been around for 50 years. Recently, a new store called Biggie's opened up. Biggie's has all of the same products, but they are cheaper than at Carl's. Unfortunately, Carl's can't afford to make their prices any lower. What do you think will happen to Carl's and why?	
Why do you think these are the last two in town?	What might Carl's be able to do?	
Sometimes members of a species find their way to a new and different habitat. They become invasive species which means animals or plants living in an environment where they do not belong.	* When farmers want to plant wheat, they first treat their fields with a chemical called an herbicide. Herbicides kill plants. After they treat the fields, farmers will plant the wheat. Herbicides are very expensive.	
	Why do you think farmers spend money on herbicides?	
If an invasive species starts to succeed in a new environment, what effects can this have on other living things in the habitat?	Animals compete for limited resources (for example, food and water). Which of the following BEST demonstrates the idea that animals compete for limited	
 Months after a bad fire, some plants that never grew in the forest started to grow in large numbers. Why do you think that these plants were able to grow better after the fire? A. The plants were able to learn survive without water. B. The plants learned to use the ash as food. C. The plants were able to thrive because many of the other plants of the forest died. D. The plants tried to change to adjust to the new forest floor. 	 A. Plants growing in the ocean have adapted special mechanisms for dealing with salt water. Some plants deal with salt water better than other plants. B. Some wolves run faster than others. These wolves hunt better and catch more animals. This means that the pack is better fed, and the pack grows bigger. C. Bats and Birds eat insects. Between the two of them, there are more insects than both can eat. D. If you build a bird house, a strong, aggressive type of bird will move in and less aggressive birds will be chased off. 	

Questions that uncovered reliable differences denoted with red asterisks.