

ANALOGIES VS. CONTRASTS: A COMPARISON OF THEIR LEARNING BENEFITS

Norma C.-Ming

norming@berkeley.edu

Education in Mathematics, Science, and Technology

Graduate School of Education, 4643 Tolman Hall #1670

University of California

Berkeley CA 94720-1670

ABSTRACT

An ongoing challenge for educators is helping their students recognize connections between related information while also appreciating distinctions between only seemingly related information. Although instructing students to compare pairs of problems may help call attention to the appropriate features of the domain, the literature is mixed regarding the types of comparisons that are most beneficial, analogical comparisons or contrasting comparisons. Experiments on learning exploratory data analysis suggest a possible reconciliation of the competing claims, by distinguishing between the learning gains produced by each comparison type. By precisely assessing students' knowledge before and after the comparisons, these studies reveal that analogical comparisons promote generalization across surface features, while contrasting comparisons promote discrimination between structures. Results indicate that these skills vary across individuals and should be assessed separately to optimally match instruction to their needs.

INTRODUCTION

Novices often have difficulty identifying what is important when learning a new domain, giving undue weight to surface features rather than deep structure. This is evident both in their inappropriate use of surface features to categorize and solve new problems (Chi, Felto- vich, & Glaser, 1981; Ross, 1984), and in their failure to transfer relevant knowledge to other problems lacking familiar surface fea-

tures (Gick & Holyoak, 1980; Spencer & Weisberg, 1986). While increasing the variability in the training examples may promote better understanding, long-term retention, and transfer (Schmidt & Bjork, 1992), introducing too much variability can backfire due to cognitive overload (Paas & van Merriënboer, 1994). Directing students to compare two or more problem situations simultaneously may help to focus attention on key problem features by limiting the variability.

Previous research has demonstrated that asking students to compare problems improves their problem-solving ability and understanding (Catrambone & Holyoak, 1989; Cummins, 1992; Gentner & Gunn, 2001; Kurtz, Miao, & Gentner, 2001; Linn, 2005; Schwartz & Bransford, 1998). The comparison instructions may be as simple as identifying global similarities between problems, or as directive as requesting an analogical mapping of the structural correspondences between the problems. These differences reflect varying degrees of support rather than calling into question the power of comparisons in promoting learning.

In contrast, the design of the cases being compared reveals qualitative differences, with some studies presenting analogues and others presenting contrasts. The research on analogical transfer argues for comparisons between problems bearing a common structure, emphasizing the benefits of variability in promoting generalization and the importance of shared structural features in supporting analogical transfer (Gick & Holyoak, 1983; Holyoak & Koh, 1987; Gentner & Gunn, 2001). As Cummins explained, "by comparing problems that share deep structures but differ in surface features, novices may come to abstract the

crucial problem features that define expert category membership” (1992, p. 1103). According to the structural alignment view, shared structure should be more effective than shared surface in advancing students’ learning (Markman & Gentner, 1993).

Other research emphasizes the value of comparing different structures in helping students classify and solve new problems. Training participants on superficially similar but structurally different problems provides helpful feedback on the irrelevance of the surface features and highlights the critical attributes in determining the different categories of problems (Winston, 1975; Tennyson, 1973; Quilici & Mayer, 1996). Organizing problems by common cover stories improved problem-solving more than organizing problems by common relationships (Scheiter & Gerjets, 2003). As Schwartz and Bransford suggest, “Contrasting cases help people notice specific features and dimensions that make the cases distinctive” (1998, p. 479). Linn makes a similar claim when describing the benefits of pivotal cases: “By contrasting two situations that differ on a crucial dimension, pivotal cases invite students to test their ideas about the dimension,” (2005, p. 42). These studies highlight the advantage of designing comparisons between cases that vary in the deep features that distinguish the cases.

This apparent contradiction in the literature regarding the relative benefits of analogical and contrasting comparisons may be resolved through a closer examination and more precise assessment of the learning gains. In particular, analogical transfer measures generalization to new surface features, while problem classification measures discrimination between different structures. Consequently, the effectiveness of the different types of comparisons is likely to depend on the current abilities and needs of the individual learner. Students who fail to recognize problem structures in new contexts would benefit more from analogical comparisons that boost their generalization skills. On the other hand, students who are prone to overgeneralization or who rely exclusively on surface features to identify

problem structures would benefit more from contrasting comparisons that boost their discrimination skills. The experiment presented here tests this hypothesis by assessing students’ generalization and discrimination skills before and after each type of comparison.

METHOD

This study evaluated the relative benefits of analogical comparisons and contrasting comparisons as a function of students’ generalization and discrimination abilities when learning exploratory data analysis. Pre- and post-tests diagnosed the skills needing the most improvement, while also characterizing students’ overall knowledge patterns. One condition (*treatment condition*) presented participants with the type of comparison hypothesized to best serve their learning needs, while the other condition (*control condition*) presented the opposite comparison type from that predicted to match their diagnosis.

Study participants received instruction in when and why to use side-by-side boxplots, scatterplots, or contingency tables to visualize the relationships in a set of data in order to answer a specific question about the data. For each of the three display types, instruction began with a videotaped lecture of a university statistics professor explaining the display type, followed by an example problem demonstrating its use. Participants’ knowledge of when to use that display type was then assessed through their answers to a series of three diagnostic problems that varied from the example problem in their structure (structure-varying), surface features (surface-varying), or both structure and surface (both-varying). For the purposes of this study, problem structure was determined by the number and type(s) of variables relevant to answering the problem: A problem that asks about two quantitative variables should be answered using a scatterplot; two categorical variables, a contingency table; and one categorical and one quantitative variable, side-by-side boxplots. Surface features

refer to the cover story and question wording of the problem.

Each problem asked whether the display type of interest should be used to answer some question about a hypothetical dataset described in the problem. Having three “yes/no” questions resulted in eight different response patterns (knowledge diagnoses) characterizing students’ knowledge about that display type. Correctly rejecting the structure-varying diagnostic problem as an example of the display type would constitute successful discrimination between exemplars and nonexemplars of that problem structure; correctly classifying the surface-varying diagnostic problem as an example of the display type would constitute successful generalization of the problem structure to new surface features.

According to the hypothesis outlined above, analogical comparisons promote generalization, while contrasting comparisons promote discrimination. The knowledge diagnoses were used to determine the type of comparison that would be most helpful in meeting each student’s learning needs, as summarized in Table 1. Of the eight diagnoses, four (middle four rows) have unambiguous predictions regarding the more useful comparison type: two show successful discrimination on the structure-varying problem but failure to generalize on the surface-varying problem (i.e., candidates for analogical comparisons), while another two show the reverse pattern (i.e., candidates for contrasting comparisons). Of the remaining four, two (top two rows) show success at both generalizing and discriminating, while two (bottom two rows) show failure at both skills. Failing to discriminate is arguably the greater error than failing to generalize for these last two diagnoses, which indicate a student who may be focusing inappropriately on the problem’s surface features or who may be completely confused (as evidenced by answering all three problems incorrectly). These diagnoses were therefore deemed candidates for contrasting comparisons. Consequently, the

diagnoses showing successful generalization and discrimination were considered candidates for analogical comparisons, in that a student who answered all three problems correctly might still benefit from learning to generalize to a broader range of surface features. This also balances the number of diagnoses assigned to each comparison type.

Subsequent to answering the pre-test diagnostic problems, participants were asked to compare the example problem to a new problem that varied either in structure (contrasting comparison) or surface (analogical comparison), in keeping with their knowledge diagnosis and experimental condition (treatment vs. control). After completing the comparison, participants repeated the diagnostic problems to assess whether and how their knowledge diagnosis may have changed, and solved a new set of transfer problems to assess the robustness and generality of their knowledge. If analogical comparisons promote generalization while contrasting comparisons promote discrimination, then participants in the treatment condition should outperform those in the control condition.

Design

The experiment used a between-group design with random assignment to condition, manipulating whether participants received the comparison type predicted to match their knowledge diagnosis (treatment: comparison matches diagnosis; control: comparison does not match diagnosis). Knowledge diagnosis at pre-test may be considered as an eight-category individual-difference factor with which the experimental condition may interact. The dependent variables were participants’ answers to the same diagnostic problems at post-test (measured by score or by knowledge diagnosis), and their answers to a set of transfer problems.

Analogies vs. Contrasts: A comparison of their learning benefits

Response given			Score	Diagnosis description & Abbreviation	Generalizes	Discriminates	Comparison Prescribed
Structure varies	Surface varies	Both vary					
N	Y	N	3	Correct (Corr)	Y	Y	Analogy
N	Y	Y	2	Slip? (Slip)	Y	Y	Analogy
Y	Y	N	2	Misconception (MC-TF)	Y	N	Contrast
Y	Y	Y	1	Overly general (OG)	Y	N	Contrast
N	N	N	2	Overly specific (OS)	N	Y	Analogy
N	N	Y	1	False alarm? (FA)	N	Y	Analogy
Y	N	N	1	Misconception (MC-F)	N	N	Contrast
Y	N	Y	0	Incorrect (Inc)	N	N	Contrast

Table 1. Summary of response patterns to diagnostic problems, associated scores and knowledge diagnoses, success at generalizing and discriminating, and the type of comparison hypothesized to best support those learning needs.

Materials

Instructional and test materials for each display type incorporated six problems: one example problem, three diagnostic problems, and two comparison problems. The three diagnostic problems varied from the example in structure, surface, or both, while the comparison problems varied from the example in either structure or surface. Abbreviated examples from one section follow. For each problem, participants were asked to decide whether side-by-side boxplots would be an appropriate display type to use for answering the given question.

Example problem

To investigate concerns about grade inflation, a university is examining data from graduating seniors in 1990 and in 2000. Is there a difference between the two groups' grades?

Variable Name	Description
Year	Student's graduation year (1 = 1990, 2 = 2000)
GPA	Student's grade-point average (e.g., 3.67)
Courseload	Total number of courses taken (e.g., 18)

Structure-varying diagnostic problem

To determine how well its classes serve students of different math backgrounds, a statis-

tics department is analyzing its course evaluations. Is there a difference in students' satisfaction ratings based on their SAT math scores?

Variable Name	Description
SATM	Student's SAT math score (e.g., 620).
Grade	Student's course grade (e.g., B+).
Rating	Student's overall evaluation of the course (e.g., 7.5).

Surface-varying diagnostic problem

The human resources department is compiling data on employees' salaries and their history with the company. Is there an association between employees' department and length of employment?

Variable Name	Description
Department	Place of work (1 = marketing, 2 = public-relations).
Income	Annual salary (e.g., \$82,250).
History	Number of months at this company (e.g., 87).

Structure- and surface-varying diagnostic problem

The 2000 U.S. Census asked people to classify themselves as belonging to one race or having

a mixed-race heritage. Does the percentage of multiracial people vary across the different regions of California?

Variable Name	Description
Age	Resident's age, in years (e.g., 24).
Race	Racial background. (1 = single race, 2 = mixed-race)
Region	Area of residence. (0 = Northern CA, 1 = Southern CA)

Contrasting comparison problem (structure-varying)

To investigate concerns about grade inflation, a university is examining data from graduating seniors in 1990 and in 2000. Is there a difference between the two groups' grades?

Variable Name	Description
Year	Student's graduation year (1 = 1990, 2 = 2000).
LtrGrade	Student's avg grade (1 = A, 2 = B, 3 = C, 4 = D, 5 = F).
Courseload	Total number of courses taken (e.g., 18).

Analogical comparison problem (surface-varying)

A city is holding a referendum to determine how its residents feel about ending welfare. Is the percentage of residents voting to end welfare related to their salary?

Variable Name	Description
Precinct	Voting precinct (e.g., 1 = Park Place, etc.).
Income	Annual salary (e.g., \$76,063).
Vote	Opinion (1 = end welfare, 2 = keep welfare program).

Participants received feedback showing the expected answers and the appropriate display type after studying the example problem and after completing the comparison between

the example and the new problem, but no feedback upon completing the diagnostic problems.

Procedure

The experiment consisted of five phases: (1) initial instruction (videotaped lecture plus example problem), (2) pre-test knowledge diagnosis, (3) comparison, (4) post-test knowledge diagnosis, (5) transfer problems. Phases 1-4 were repeated in sequence in three separate blocks, once for each of the three display types (side-by-side boxplots, scatterplots, contingency tables). The order of the display types was randomized for each participant. For the comparison, participants studied either the analogical problem or the contrasting problem, depending on their diagnosis and condition, then compared that problem to the original study example. The comparison entailed listing similarities and differences between the problems.

During Phase 5, participants studied a short explanation, example, and solution on when to use two new statistical display types, namely pie charts and histograms. They were then presented with 25 new problems in a similar format to the problems previously shown, and asked which (if any) of the five display types they had studied would be most appropriate for answering the question given. Since there were no significant differences in participants' performance on these problems, they will not be discussed further here.

Participants

Participants were college students or recent graduates (bachelor's degree within the last two years), all of whom had no statistics training beyond high school. Due to random assignment, 44 participants were in the treatment group and 39 were in the control group.

RESULTS

Distribution of Knowledge Diagnoses at Pre-test and Over Time

At pretest, 82% of participants in both conditions (68 out of 83) successfully generalized on the surface-varying diagnostic problem, whereas only 54% (45 out of 83) successfully discriminated on the structure-varying diagnostic problem. Across the duration of the experiment, both groups showed improvement, but more in the treatment condition, as shown in Figures 1 and 2.

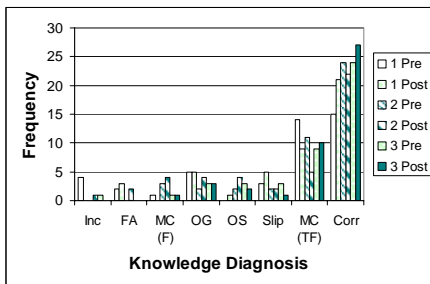


Figure 1. Frequency of knowledge diagnoses from pre- to post-test across all three blocks of the experiment, for the treatment condition.

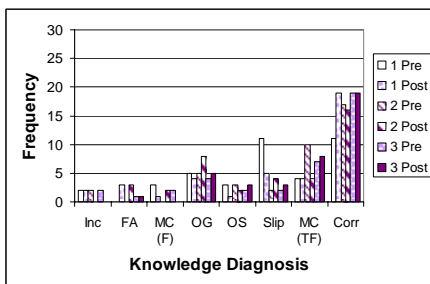


Figure 2. Frequency of knowledge diagnoses from pre- to post-test across all three blocks of the experiment, for the control condition.

Differences between Condition in Problem-Categorization Accuracy

As predicted, the treatment condition outperformed the control condition in categorizing the diagnostic problems after performing the comparisons prescribed. Since

the conditions did not differ until after the first set of three diagnostic problems, participants' answers to these questions may serve as a pretest measure, with their answers to the subsequent five sets of diagnostic problems serving as a dependent measure. Analyzing these scores by a 2 (condition) × 4 (pretest score) ANOVA reveals a significant condition-by-pretest interaction ($F_{1,3} = 3.709$, $SS = 67.800$, $MSE = 22.600$, $p = .015$), with the treatment being particularly beneficial for students with low pretest scores (0 or 1) and essentially leveling the playing field. Figure 3 depicts these trends.

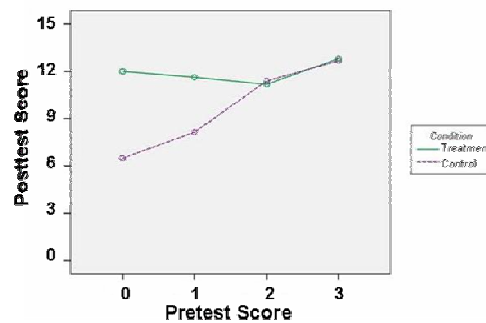


Figure 3. Posttest accuracy as a function of condition and pretest score.

Further examination of these differences in terms of the individual knowledge diagnoses shows a significant condition-by-pretest-diagnosis interaction ($F_{1,7} = 4.136$, $SS = 112.017$, $MSE = 22.403$, $p = .002$) from a 2 (condition) × 8 (pretest diagnosis) ANOVA. This interaction derives from participants diagnosed with overly general knowledge, whose scores show a large difference between conditions ($M_T = 12.80$, $SD_T = 1.924$, $N_T = 5$; $M_C = 8.00$, $SD_C = 2.236$, $N_C = 5$; $p = .007$). As shown in Figure 4, the difference between conditions varies with knowledge diagnosis, even for equivalent pretest scores. This highlights the value of intervening based on students' specific knowledge, rather than just their overall accuracy. Although receiving both comparison types may have helped some participants, this occurred in both conditions and was a poorer predictor of posttest scores.

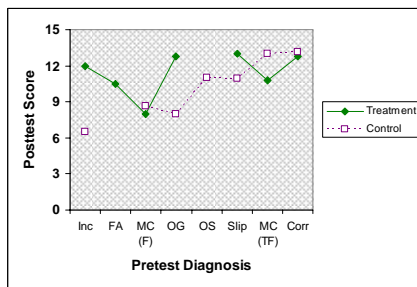


Figure 4. Posttest accuracy as a function of experimental condition and pretest diagnosis.

Discrimination and Generalization Ability

Examining participants' learning in terms of discrimination and generalization ability confirms that the two comparison types produced distinct benefits. Accuracy on problems with different deep structures provides a measure of discrimination, while accuracy on problems bearing the same structure but different surface features provides a measure of generalization. For all participants, regardless of condition, contrasting comparisons produced better discrimination scores, as shown by a main effect of comparison type ($F_{1,7} = 12.156$, $SS = .739$, $MSE = .739$, $p = .001$) according to a 2 (comparison) \times 8 (pretest diagnosis) ANOVA. A similar analysis on generalization scores reveals a marginally significant main effect of comparison type ($F_{1,7} = 3.444$, $SS = .266$, $MSE = .266$, $p = .065$), with higher generalization scores resulting from analogical comparisons.

Performance on Comparison Problems

Participants' performance on the two comparison problems sheds further insight on the reasons underlying their different effects. On average, treatment-condition participants show improvement in solving the contrasting comparison problem across the three blocks, with below-chance performance (16.7%, 4 out of 24) in Block 1 ($t = 4.543$, $df = 24$, $p < .001$), at-chance performance (64.3%, 9 out of 14) in Block 2 ($t = .488$, $df = 15$, $p = .633$), and better-than-chance performance (87.5%, 14 out of

16) in Block 3 ($t = 5.102$, $df = 17$, $p < .001$). Although these are not entirely the same participants receiving this intervention across the three blocks, they do represent the same pretest diagnoses and the same failure to reject the structure-varying diagnostic problem at pretest for each block.

Participants in the control condition would be expected to do better than those in the treatment condition on the same problem based on their superior pretest scores and their initial accuracy in rejecting the surface-similar, structure-varying diagnostic problem at pretest. While their performance on this question in Block 1 was significantly better than that for the treatment condition ($t = 2.908$, $df = 47$, $p = .006$), it is no different from chance (52.0%, or 13 out of 25; $t = .189$, $df = 26$, $p = .852$), suggesting that perhaps the knowledge that produced their initial success in answering the structure-varying diagnostic problem at pretest may not be so robust. On subsequent blocks, their performance on the contrasting comparison problem increased modestly to 57.1% in Block 2 (16 out of 28; no different from chance: $t = 1.099$, $df = 29$, $p = .281$) and 68.2% in Block 3 (15 out of 22; marginally above chance: $t = 1.994$, $df = 22$, $p = .0586$). In spite of their initial head start on the pretest, control participants did not maintain their advantage over treatment participants on the contrasting comparison problem, as Figure 5 depicts.

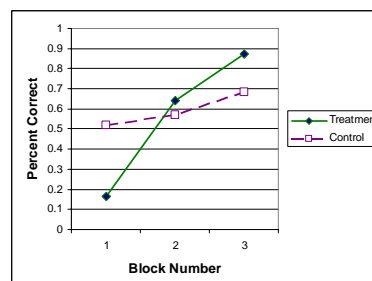


Figure 5. Accuracy on contrasting comparison problem across blocks.

Performance on the analogical comparison problem does not show a meaningful difference between conditions for any of the three blocks ($t_1 = .286$, $df_1 = 30$, $p_1 = .777$; $t_2 =$

1.310, $df_2 = 30$, $p_2 = .200$; $t_3 = .915$, $df_3 = 31$, $p_3 = .367$). In both conditions, accuracy on this problem followed a generally increasing trend, as shown in Figure 6. That this is higher than the average performance on the contrasting comparison problem suggests again that students were more successful at generalizing than at discriminating. Treatment-condition accuracy on the analogical comparison problem increased from chance ($t = 1.096$, $df = 20$, $p = .286$) on Block 1 (61.9%, 13 out of 21) to above-chance ($t_2 = 3.395$, $df_2 = 29$, $p_2 = .002$; $t_3 = 6.60$, $df_3 = 27$, $p_3 < .001$) on Block 2 (76.7%, 23 out of 30) and Block 3 (89.3%, 25 out of 28). Control-condition accuracy showed a similar pattern, starting at chance on Block 1 (66.7%, 9 out of 15; $t = 1.323$, $df = 14$, $p = .207$) and rising above chance on Block 2 (91.7%, 11 out of 12; $t = 5.000$, $df = 11$, $p < .001$) and Block 3 (78.9%, 15 out of 19; $t = 3.012$, $df = 18$, $p = .007$).

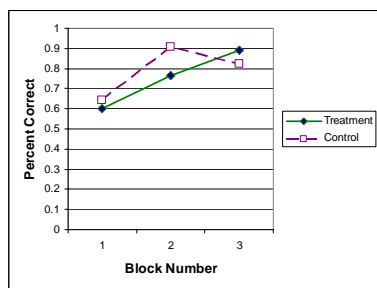


Figure 6. Accuracy on analogical comparison problem across blocks.

Comparison Processes

Follow-up studies on the processes by which participants performed the two types of comparisons reveal meaningful differences in attention to structural features. In a related study in which participants learned the same material but performed the same type of comparison throughout, participants in the contrasting-comparison condition ($N = 35$) were significantly more likely than those in the analogical-comparison condition ($N = 46$) to mention the most critical structural feature, variable type, during the comparison ($t = 7.985$, $df = 79$, $p < .001$), as shown in Figure 7.

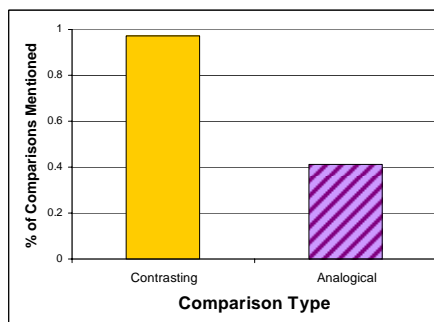


Figure 7. Frequency of mentioning variable type on contrasting vs. analogical comparisons.

This difference was replicated and explored more fully in a subsequent protocol study ($N = 18$) examining students' problem-comparison processes, with all participants performing both types of comparisons. A 2 (contrasting vs. analogical comparison) \times 2 (similarities vs. differences) ANOVA on the number of features mentioned during the comparisons revealed a significant interaction ($p = .005$), with the number of similarities identified in contrasting comparisons exceeding all other similarities and differences, as shown in Figure 8. Note that the same features (all surface features) could also have been listed as differences in analogical comparisons yet were not, suggesting that these features were more salient as similarities between the contrasting cases than as differences between analogues.

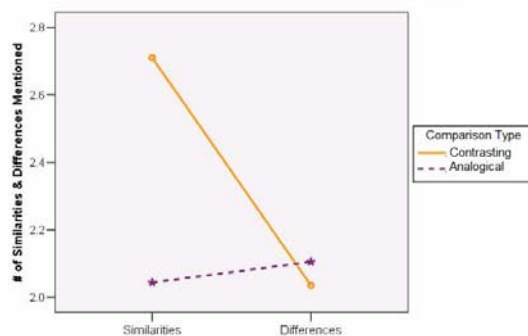


Figure 8. Similarities and differences identified during comparisons, by type.

These findings reveal that contrasting comparisons are particularly effective in calling students' attention to key structural features in these problems, especially identifying relevant variable types, a crucial step which students do not always perform (Lovett & Chang, 2007). Their effectiveness may stem from the large number of similar surface features present in the comparisons, which are easy to identify and may facilitate aligning the problems being compared, as has been found in previous research on the influence of surface similarities in retrieving and mapping across analogues (Holyoak & Koh, 1987; Kotovsky & Gentner, 1996). Such alignment may then increase the likelihood of noticing the different structural features, which are fewer in number and more salient against the backdrop of multiple surface similarities. What is particularly noteworthy here is that these surface similarities may promote alignment in spite of concomitant structure-violating differences, and support alignment more effectively than structural similarities.

DISCUSSION

In summary, contrasting comparisons improved discrimination between problem types, while analogical comparisons improved generalization to new contexts and surface features. Further, contrasting comparisons produced the greatest benefit for participants who began the training with overly general knowledge, an overreliance on surface features instead of deep structure, or with no success in identifying the correct analysis method at pretest. Analogical comparisons were most helpful for the smaller number of participants who only recognized problems with both deep and surface similarities to the initial study example. The error patterns underscore the value of assessing students' abilities prior to selecting an instructional intervention, insofar as what helps one student may hurt another. In particular, students demonstrating overly general knowledge need to improve their discrimination skills and rectify their tendency toward overgeneralization, which contrasting com-

parisons improve but which analogical comparisons exacerbate.

These results underscore the importance of precise assessment, not just to go beyond raw scores to examine the underlying knowledge revealed by error patterns, but also to better integrate and reconcile the extensive literature on learning from comparisons. Improvements in performance depend on what is being measured—generalization, as captured by analogical transfer tasks, or discrimination, as captured by problem-categorization tasks. Rather than evaluating the effectiveness of analogical comparisons against contrasting comparisons *in toto*, carefully distinguishing between generalization and discrimination as separate skills enables recognition that each comparison type serves a different learning goal. These differences between generalization and discrimination emerge both in terms of the skills demanded by a particular task, and in terms of the skills promoted by a particular instructional technique.

Another important theoretical implication of this work is the demonstration that deeper common structures are not always better when designing comparisons. Were that the case, the analogical (surface-varying) comparisons should fare better in this study than the contrasting (structure-varying) comparisons, in light of their superior structural overlap (same number and type of variables, not just same number of variables). This suggests that “good-enough” structural alignment that enables the learner to perform the comparison may be sufficient for the task, and that surface features can play a meaningful role in facilitating alignment, perhaps even more effectively than deeper common structure in certain cases.

Finally, this research highlights the importance of noticing differences as well as similarities. The differences in participants' comparison protocols showed more alignment than the similarities did, perhaps since those differences emerged from an alignable structure whereas the similarities were just common attributes. Differences may demand more effort to perceive and thus yield greater learning.

REFERENCES

- Carnine, D. (1980). Three procedures for presenting minimally different positive and negative instances. *Journal of Educational Psychology, 72*, 452-456.
- Catrambone, R., & Holyoak, K.J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1147-1156.
- Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.
- Cummins, D.D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1103-1124.
- Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition, 29*, 565-577.
- 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.
- Holyoak, K.J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332-340.
- Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797-2822.
- Kurtz, K.J., Miao, C.H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*, 417-446.
- Linn, M.C. (2005). WISE design for lifelong learning—pivotal cases. In P. Gärdenfors & Petter Johansson (Eds.), *Cognition, Education, and Communication Technology* (pp. 223-255). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lovett, M.C., & Chang, N.M. (2007). Data analysis skills: What and how are students learning? In M. Lovett & P. Shah (Eds.), *Thinking with Data: The Proceedings of the 33rd Carnegie Symposium on Cognition* (pp. 293-318). Mahwah NJ: Lawrence Erlbaum Associates, Inc.
- Markman, A.B., & Gentner, D. (1993). Structural alignment in similarity comparisons. *Cognitive Psychology, 25*, 431-467.
- Paas, F.G.W.C., & Van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.
- Quilici, J.L., & Mayer, R.E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.
- Ross, B.H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371-416.
- Scheiter, K., & Gerjets, P. (2003). Sequence effects in solving knowledge-rich problems: The ambiguous role of surface similarities. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Boston MA: Cognitive Science Society.
- Schmidt, R.A., & Bjork, R.A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
- Schwartz, D.L., & Bransford, J.D. (1998). A time for telling. *Cognition and Instruction, 16*, 475-522.
- Spencer, R.M., & Weisberg, R.W. (1986). Context-dependent effects on analogical transfer. *Memory & Cognition, 14*, 442-449.
- Tennyson, R.D. (1973). Effect of negative instances in concept acquisition using a verbal learning task. *Journal of Educational Psychology, 64*, 247-260.
- Winston, P.H. (1975). Learning structural descriptions from examples. In P.H. Winston (Ed.), *The psychology of computer vision* (pp. 157-209). New York: McGraw-Hill.