Gesturing to Promote Higher-Order Thinking: Cross-Cultural Differences

Lindsey E. Richland (l.e.richland@uci.edu)
Department of Education, University of California, Irvine
2001 Berkeley Place, Irvine, CA 92697-5500 USA

Abstract

Cross-cultural differences in teachers' gestures during instructional analogies were examined in United States, Hong Kong, and Japanese mathematics classrooms. Teachers in all three regions frequently made analogies, but patterns of gestures differed in consequential ways. All teachers used gestures in a high proportion of the analogies, but teachers in the higher achieving regions, Hong Kong and Japan, were more likely to use gestures that physically linked the entities being compared. They were also more likely to tailor their gesture use to the novelty of the comparison, using a higher quantity or more informative gestures during comparisons that involved more novel source analogs. U.S. teachers conversely showed less sensitivity to the challenges inherent in higher order reasoning by novices.

Keywords: Gesture, Analogy, Education, Cross-Cultural Research, Mathematics Education

Analogies can be powerful tools for building flexible, conceptual knowledge (e.g. Clement, 1993; Gentner, Lowenstein & Thompson, 2003; Rittle-Johnson & Star, 2007). However, unless learners have appropriate expertise or instructional support, they often fail to notice or benefit from such relational comparisons (Bransford, Brown & Cocking, 1999; Gentner & Rattermann, 1989; Gick & Holyoak, 1980, 1983; Bassok & Holyoak, 1989).

Teachers' gestures are a readily available, naturally occurring, resource for directing learners' attention to relevant relational similarity during instructional analogies. In addition, much recent work has revealed that gestures are very closely tied to learning and cognition, and can provide strong support for instruction (Singer & Goldin-Meadow, 2005). While little work has yet explored relations between gesture use and analogical thinking, Alibali and colleagues (see Alibali et al, 1997b, Alibali & Nathan, 2007, Hostetter et al, 2006) have drawn attention to the related construct of "linking gestures." These are gestures produced to draw links between instructed representations, and may play an important role in leading learners' attention to abstract structure. The current analysis draws on this construct to explore the subset of such gestures that explicitly support relational comparisons.

Gesture as Instruction

Empirical and observational treatments of teachers and students' gestures during mathematics instruction have demonstrated that these actions can be a source of powerful insights into classroom learning (see Goldin-Meadow, 2002; Nathan & Alibali, 2007). Gestures provide windows onto the gesturer's thinking (see Abrahamson, 2007; Goldin-Meadow, 2002; Kendon, 1994), and there is growing cognitive evidence that instructors' gestures impact novices' representations of taught information (e.g., Glenburg & Kaschak, 2000, Goldin-Meadows, 2001). In fact when teachers' gestures differ from their instructional explanations, students remember and use strategies only implied by their gestures at a fairly substantial rate, in addition to using the verbalized explanation (Singer & Goldin-Meadow, 2005).

Analogies as Instruction

Similarly, comparative reasoning is well established as a basic component of mathematical thinking and learning (e.g. English, 2004; Novick & Holyoak, 1991; Reed, Dempster, & Ettinger, 1985). Instructional comparisons in particular have the potential to clarify similarities and differences between concepts or problems (Gentner & Gunn, 2001; Rittle-Johnson & Star, 2007), thereby fostering conceptual understanding and transfer (Bransford, Brown & Cocking, 1999; National Research Council, 2001).

Mathematics reform efforts have emphasized teaching that involves connecting mathematics concepts (see Research Advisory Committee for NCTM, 1996; Stigler & Fernandez & Yoshida, 1995), as well as engaging students in drawing comparisons between multiple solution strategies to single problems (see Carpenter, Fennema, Franke, Levi, & Empson, 1999). Recent empirical work supports these assertions, demonstrating that encouraging students to make comparisons between solution strategies can promote student learning and understanding of early algebra topics (Rittle-Johnson & Star, 2007).

Why look at Gesture in Relational Comparisons?

In spite of the key role of comparative reasoning in mathematical proficiency, ensuring that students notice and learn from comparisons is not a trivial task. Novices are known to overlook, over-extend, or erroneously transfer structure based on irrelevant appearance of similarity (e.g. Bassok 1997; Chi, Feltovich, & Glaser, 1981; Chi, Glaser & Clement & Gentner, 1991; Gick & Holyoak, 1980, 1983; Reed 1987; Richland, Morrison & Holyoak, 2006; Ross, 1987, 1989). Thus, one cannot expect that simple use of an instructional comparison will inevitably lead to learning. Rather, the manner in which comparative structure is processed can be crucial to whether reasoners are swayed by irrelevant similarity or benefit from deep structure (e.g., Gick & Holyoak, 1983; Clement & Genter, 1991; Pedone, Holyoak & Hummel, 2001).

International Differences in Classroom Mathematics Comparisons

Early interest in the role of international differences during comparisons within mathematics instruction emerged from cross-cultural studies that demonstrated the regular usage of highly connected lessons in Japanese classroom teaching (Hiebert & Stigler, 1999; Stevenson & Stigler, 1992). These researchers documented that Japanese teachers regularly designed instruction so that students used and compared multiple solution strategies to the same problem. Also, lessons were coherent, when compared to U.S. lessons, such that each instructional activity was clearly aligned with the prior lesson content. This had direct implications for student learning (Stigler & Fernandez, 1995). Paired with the recurrent finding that Japanese children scored higher on international mathematics tests than U.S. children, these studies suggested that comparing solution strategies, or comparative prior to new knowledge, might be an important component of teaching mathematics with conceptual understanding.

In a more specific treatment of teachers' use of analogies internationally, Richland, Zur and Holyoak (2007) used observational frequency coding to identify differences in mathematical comparisons within U.S. Japanese, and Hong Kong classrooms. A subset of videotapes collected as part of the Third International Mathematics and Science Study (TIMSS-R) were analyzed by identifying every instance of comparisons and coding them for their alignment with strategies predicted by empirical work to promote higher-level thinking and learning. These data revealed striking patterns that suggested both Hong Kong and Japanese teachers spontaneously used strategies that were more aligned with basic cognitive principles of reasoning and learning from analogy than U.S. teachers. Most important for the current paper, the Hong Kong and Japanese teachers were more likely to use hand gestures that explicitly moved between the focal objects. The current project explores this finding in more depth by re-analyzing these data with additional codes corresponding to teachers' gestures.

This paper's aims are twofold. First, analyses were conducted to reveal existing cross-cultural commonalities and differences in production of gestures during instructional analogies. Because of the large-scale nature of this aim, codes focused on broad-stroke differences between gesture patterns. This is an alternate method from much of the important work in gesture that has more closely analyzed the details of hand and body movements within smaller samples of interactions. Second, analyses were conducted to better understand context and use of teachers' use of comparative gestures that support students' likelihood of noticing and learning from higher-order instructional comparisons.

METHOD

Video Data

As described in previous analyses of these data (Richland, Zur, Holyoak, 2007), videotaped classroom lessons were randomly sampled from the corpus of data collected as part of the Third International Mathematics and Science Study – Repeat (TIMSS-R; Hiebert et al., 2003). The TIMSS-R dataset was collected as a randomized probability sample of all lessons taught in public, private, and parochial schools throughout the country in one year. Each selected classroom was videotaped on one occasion during a normal class period. Lessons were typically 50 min long, yielding approximately twenty-five hours of videotaped data that were examined in this project.

Ten lessons were randomly selected from classes videotaped in each of the United States, Hong Kong, and Japan (thirty total). Hong Kong and Japan were selected for comparison to the United States because their students outperform U.S. students on international tests, and because their classroom teaching styles are quite different from each other as identified in the primary TIMSS-R findings.

Observational Coding

Lessons were analyzed using a temporally linked transcript and video. Qualitative codes were used to generate quantitative frequency data. Lessons were first divided into units of analysis. All identifiable instances of relational comparisons were marked. In a series of passes, these relational comparisons were then categorized according to previously designed codes. At least two coders divided the data for each pass, and reliability was calculated for all coders on approximately 20% of the data. Reliability was calculated as the number of agreements divided by number of total comparison units coded. All disagreements were resolved through discussion.

Pass 1: Identification of Relational Comparisons. Two expert coders (researchers with relevant doctorates) separately identified all units of relational comparisons within every lesson. All disagreements were resolved by discussion. The definition of a relational comparison derived from Gentner's (1983) structure-mapping theory. Relational comparison was defined as a higher-order relationship between a source (or base) object and a target object. Based upon this definition, there were several criteria used to mark an interaction as a relational comparison. First, a source and a target had to have been clearly identifiable. For example, an instance in which the teacher stated "lets solve this problem like you used to do them" was not marked because the precise source was not readily apparent to us or to learners. Second, for a positive identification there had to be some clear indication of the intention to compare the source and target items. These could be an explicit verbalization (e.g., "X is just like Y") or a less explicit verbalization signifying a same or difference comparison (e.g., "We just finished with X, now lets do another one (Y)"). Notification could also include serial or parallel spatial positioning on the board or on a worksheet or textbook page. Finally, gesture that motioned between two or more representations could signify a comparison.

Teachers in all countries produced a high number of relational comparisons during the ten 8^{th}-grade mathematics lessons. As reported previously (Richland, Zur & Holyoak, 2007), a total of 195 units were identified in the U.S. lessons, with a mean of 20 and a range of 9-30 per lesson. A total of 185 units were identified in Hong Kong lessons, with a mean of 18 and a range of 7-27 per lesson. A total of 139 comparisons were identified in Japanese lessons, with a mean of 14 and a range of 9 to 25 per lesson. Thus relational comparisons were common in all countries.

Source and Target Simultaneously Visible. Every unit of relational comparison was then categorized in a binary code for whether or not the source was visible to students while the target was taught. A positive score was given if the source was easily visible to students while they were expected to be reasoning about the target. Generally, this meant that the source was written or drawn on a classroom chalkboard, on an overhead projector, or on a class worksheet, and was left in that location while the teacher moved their attention to the target. This code was considered important as a strategy for supporting students' higher order reasoning. Retrieval and working memory demands increase when the source is not available (Cho, Holyoak & Cannon, 2007). The reasoner must successfully identify and retrieve the source object, as well as manipulate the representation in working memory to determine alignment and structural relationship to the target object. Processing demands may be further compounded if the source object is novel and the reasoner cannot take advantage of expertise to chunk relations (Chi, Glaser & Rees, 1982; Kimball & Holyoak, 2000).

Eight units were excluded from analysis due to an inability to code the materials. Reliability between coders was calculated as 87%.

As reported previously (Richland, Zur & Holyoak, 2007), U.S. teachers made the source visible during discussion of the target at the lowest rate across all three countries (X^2 (2) = 28.4%, $p < .001$). U.S. teachers were least likely to provide the students with visual access to the source during comparisons to the target. Visual representations can highlight relational structure and reduce both retrieval and working memory demands for representing complex relations, so this difference may indicate that U.S. students face higher processing loads than students in Hong Kong and Japanese classrooms.

Source Familiarity. Each source analog from every relational comparison was coded for the level of students' presumed familiarity with the object. Coding was based upon information explicitly provided about students' level of knowledge about the source. Generally, this was coded using cues verbalized during the relational comparison (e.g., "yesterday we learned concept X. Today we will see how that concept is similar to Y"). Source familiarity was coded in a binary code as either 'likely to be well-known' (standard cultural knowledge and mathematics learned the prior year or before) or 'new information' (information taught in the current or immediately prior lesson). Japanese coders provided supplemental information about whether a source was a part of standard cultural knowledge for Japanese students. Because data were not available for the students in the specific classrooms for which lessons were analyzed, this code measured a presumption of familiarity based on information from the lesson rather than an independent verification of students' familiarity. The code served to assess the teacher's judgments about students' level of familiarity with the sources the teacher initiated.

If sufficient information about source familiarity was not provided, the comparison was excluded from analysis. Eight units were excluded from this code. Reliability was 90% between coders. Data analysis revealed that there was a significant difference in the use of familiar sources by country, X^2 (2) = 6.6, $p < .05$. As shown in Figure 1B, teachers in the U.S. were least likely to use familiar sources while those in Hong Kong and Japan were more likely to use sources familiar to the students.

As reported previously (Richland, Zur & Holyoak, 2007), students' presumed familiarity with the source objects differed by country (X^2 (2) = 8.8, $p < .05$). While most sources in all countries were more likely to be new information, Hong Kong and Japanese teachers were more likely than U.S. teachers to use sources that invoked well-known information.

Comparative Gestures. Gesture has the potential to carry information beyond verbalized content (see Goldin-Meadow, 2001; Kendon, 1994), so teachers' gestures produced during construction of a relational comparison were considered potential cues to draw learners' attention to the comparison, to aid alignment of source and target relational structure, and to support mapping and inferences between the analogs. Coders first marked presence or absence of any gesture. Gestures were defined as hand and arm movements that conveyed information that was relevant to instruction (e.g. pointing at the board would be included, while pointing at a student would be excluded). Gestures that were unclear as to their intention were not included (e.g., "beat" gestures that moved with speech).

Once gesture was identified within a unit of comparison, all gestures during the comparison were analyzed to determine whether any were used to compare the objects. This generally meant one or more hand motions moving directly between one object and the other object. If such a gesture was identified, the comparison was marked as containing a comparative gesture. If the gesture (s) were deemed to only reference one or the other of the objects being compared, the comparison was coded as NOT containing a comparative gesture.

Twelve comparisons were excluded for an inability to see teachers' gestures in video. Intercoder reliability was 91%.

As previously reported, the likelihood of using gestures that explicitly linked source and target objects differed by country X^2 (2) =36.0, $p < .001$. Teachers in the United States produced comparative gestures in a far smaller percentage of comparisons than Hong Kong or Japanese teachers (U.S. = 13%, H.K. = 35%, Japan = 44%). Thus Japanese teachers are more than three times as likely to use comparative gestures than U.S. teachers.

Current Analyses

The current analyses build upon these cross-cultural established differences. Analyses examined the total quantity of gestures produced, along with a better measure of the frequency of comparative gestures. Specifically, comparative gestures were examined in the subset of analogies in which both source and target analogs were visible simultaneously. Finally, relationships between gesture usage and learners' familiarity with the source analog were assessed. These allowed for determining whether teachers tailored their gestures to the presumed difficulty of the comparison. Familiarity was used as a proxy of difficulty since expertise is known to facilitate attention to relational structure (Chi et al, 1981).

RESULTS

Gesture

The two levels of gesture codes were analyzed separately to examine any international differences.

Presence or Absence of Gesture. Teachers in all three countries used gesture in the majority of the relational comparisons they produced, (mean percentage of all comparisons containing gesture: U.S. = 83%, Hong Kong = 90%, Japan = 90%). Thus, gesture was a very frequent part of instruction during the segments of the lesson we analyzed. Accordingly, there was no statistical difference in the rates of using gesture within instructional comparisons across countries, X^2 (2) = 4.9, p = .09. Very few comparisons were produced without gesture (U.S. = 17%, H.K. = 10%, Japan = 10%). Note as well that these data under-represent the total number of gestures during instructional comparisons, since gesture was only coded once per comparison as present or absent. If more than one gesture was produced per comparison (and anecdotally, this was common), this was not recorded. This finding supports and extends other predictions about the ubiquity of gesture across cultures and within instruction. Specifically, it shows that gestures are clearly a regular part of mathematics instruction around the world, and further, are common when teachers are conveying abstract, comparative information.

Comparative Gestures versus Gesture to Source or Target only. There were important differences in the types of gestures across countries. In all countries, the majority of gestures produced were directed toward either the source or target objects individually X^2 (1) =75.3, $p < .001$. However, as noted above, the likelihood of using gestures that explicitly moved between source and target objects differed by country. These comparative gestures are cues to signal the relevance of alignment and structure mapping between the objects being compared, and have they have the potential to be strong cognitive supports for enabling students to notice and benefit from relational comparisons. This result therefore could have important consequences for students' higher-order thinking and learning.

Comparative Gestures and Source Visibility. The above analysis of comparative gestures could; however, be reflecting differences in teachers' use of visual representations. International variations in board use are well documented (Stevenson & Stigler, 1992). Thus in a stronger test of the relationship between country and use of comparative gestures, a second analysis examined only the comparisons in which the source and target objects were visibly available simultaneously. This reduced confounds between teachers' uses of visual supports and use of comparative gestures. If the two objects were not visible simultaneously, teachers could not use gestures to physically link them. In this subset of all comparisons, there remained a significant difference in rates of using comparative gestures across countries, X^2 (2) =23.5, $p <$ .001. As shown in Figure 1, U.S. teachers used overall fewer analogies in which the source and target were visible simultaneously. Then when those proportions were further separated into comparative versus non-comparative gestures, it is clear that U.S. teachers were least likely to use comparative gestures, even when both source and target analogs were clearly visible.

Gestures and Source Familiarity. The next analysis examined the relationship between teachers' gesture use and learners' familiarity with the source analog. Alibali & Nathan (2007) have argued that teachers may use more gestures when topics are less familiar to learners. In order to investigate this hypothesis within this sample, rates and types of gestures were analyzed in relation to students' expected familiarity with the source analogs.

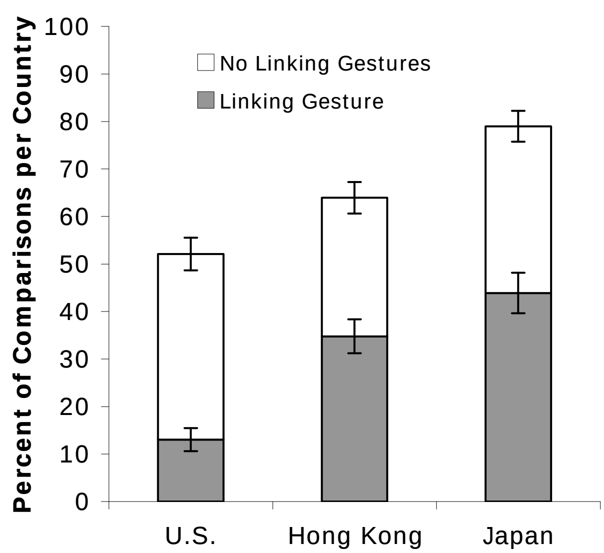

Figure 1. Rates of producing comparative versus non-comparative gestures when source and target analogs were simultaneously available

First, data for all countries were examined together, to determine whether there were common patterns in the relationships between gestures and source familiarity. Overall gesture rates for comparisons in which the source was coded "familiar" were compared to gesture rates in comparisons in which the source was coded "not familiar."

There was an overall difference X^2 (1) = 6.89, $p < .01$, such that teachers in all countries used more gestures of any type when the source was not highly familiar. However, there were no reliable relationships between the use of comparative gestures and level of source familiarity, X^2 (1) = 2.2, $p = .14$. At least, there were no differences when all countries were examined together.

In order to understand whether there were variations between countries, each country was also analyzed separately. When examined alone, U.S. teachers showed no reliable patterns of either gesturing more during less familiar comparisons X^2 (1) = 1.1, $p = .30$, nor of using comparative gestures according to source familiarity, X^2 (1) = 0, $p = .99$. In contrast, Japanese teachers' use of gesture frequency was inversely related to the familiarity of the comparison. This meant that they used higher rates of gesturing during less familiar comparisons, X^2 (1) = 7.4, $p < .01$. Their rates of using comparative gestures did not differ between familiar and less familiar comparisons, X^2 (1) = .46, $p = .55$, perhaps because Japanese teachers used fairly high rates of comparative gestures in all comparisons. Hong Kong teachers, in a different pattern of tailoring gestures to learners' familiarity, did not show a difference in the overall number of gestures used according to source familiarity, X^2 (1) = 1.8, $p = .18$, but they did use comparative gestures at a higher rate for novel source analogs than for more familiar comparisons X^2 (1) = 5.5, $p = .01$. These data suggest that both Japanese and Hong Kong teachers manipulated their gesture use in a reflection of learners' familiarity with source analogs, while U.S. teachers did not.

Discussion

Several important results emerged from these data. Least surprising, but useful as documentation, was the finding that all teachers commonly produced gestures during instructional analogies. This is not unexpected, since many studies have documented that teachers frequently use gestures during instruction (e.g. Flevares & Perry, 2001). However, the international commonality of using gestures during instructional analogies is novel, and suggests that gestures may play an important role in promoting students' learning from classroom analogies.

Most exciting are the results that indicate clear international differences in *how* and *when* gestures were produced during these instructional analogies. First, a stringent test showed that the teachers in the United States were least likely to use gestures that physically compared the two or more representations in an analogy, even when those two representations were visible simultaneously. In contrast, both Hong Kong and Japanese teachers were proportionally more likely to use comparative gestures.

A second set of analyses showed that there were international differences in the relationships between gesture use and the difficulty of the instructional analogy for classroom students. U.S. teachers did not show any patterns of differentiating their gesture use depending upon the novelty of the source analog. In contrast, both Hong Kong and Japanese teachers did show varying patterns of gesture depending on the circumstances. Japanese teachers used overall more gestures when the source was more novel, and Hong Kong teachers used more comparative gestures. Since expertise and familiarity with a source analog is well known to improve reasoners' likelihood of attending to relational structure (Chi et al, 1983; Holyoak, Junn & Billman, 1986), novices would be expected to require more assistance in analogies involving more novel source analogs. Thus, these cultural differences in gesture use may have real impact on the classroom students' ability to attend to and learn from relational structure.

Theoretically, these data raise intriguing questions about whether comparative gestures, or the more broad construct of linking gestures, are tied to the speaker's knowledge of the representations and relations they describe, or tied to pedagogical and/or cultural norms of communication. Speakers are known to gesture differentially when they are more versus less familiar with the content they describe, though importantly this does not seem to be the explanation in this case. Analogies were classified as more difficult based upon classroom students' likely familiarity with the source analog, which is quite likely to be different from the teachers'. A teacher who designed a lesson plan for a specific lesson might arguably even be more familiar with that source analog than with one designed for a prior day's class. Thus, there is no reason to expect that the designation of difficulty coded here from the learners' perspective would be explanatory for the teachers. Thus, the use of comparative gestures is more likely attributable to the culturally or pedagogically derived gesture routines.

Importantly, there is some evidence that gestures used to link representations are under teachers' conscious control and are trainable (Hostteter et al, 2006). That suggests that while the observed patterns in gesture use may derive from cultural norms of discourse and mathematics instructional practice, these are not beyond teachers' control and could be an available tool for improving U.S. teachers efficacy in teaching by analogy.

Acknowledgments

This work was supported in part by grants from the Institute of Education Sciences #R305H030141; and the Office of Naval Research, N000140810186. I am also grateful to the many coders who participated in this work.

References

Alibali, M. W. & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In Goldman, R., Pea, R., Barron, B. J., and Derry, S. (Eds.) Video

Research in the Learning Sciences (pp. 349-365). Mah Wah, NJ: Erlbaum.

Alibali, M.W., Flevares, L., & Goldin-Meadow, S. (1997a). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*, 183-193.

Alibali, M. W., Sylvan, E. A., Fujimori, Y., & Kawanaka, T. (1997b). The functions of teachers' gestures: What's the point? Paper presented at the Annual Meeting of the Midwestern Psychological Association, Chicago, Illinois.

Bassok, M. (1997). Two types of reliance on correlations between content and structure in reasoning about word problems. In L. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 221- 246) Hillsdale, NJ: Erlbaum.

Bransford, J. D., Brown, A. L., & Cocking, R.R. (1999). *How people learn: Brain, mind, experience, and school*. National Research Council, Commission on Behavioral & Social Sciences & Education. Committee on Developments in the Science of Learning. Washington, DC: National Academy Press.

Cho, S., Holyoak, K. J., and Cannon, T.D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition, 35* (6), 1445-1455.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30,*1241-1257.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

English, L. (2004). (Ed.), *Mathematical reasoning: Analogies, metaphors, and images.* Hillsdale, NJ: Erlbaum.

Flevares, L. M., & Perry, M. (2001). How many do you see? The use of nonspoken representations in first-grade mathematics lessons. *Journal of Educational Psychology, 93,* 330-345.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155-170.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition, 29*, 565-577.

Gentner, D., Loewenstein, J., Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95,* 393-408

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1-38.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9,* pp.558-565.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K.B., Hollingsworth, H., Jacobs, J., Chui, A.M., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study.* NCES 2003-013. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hostetter, A. B., Bieda, K., Alibali, A. W., Nathan, M. J., & Knuth, E. J. (2006). Don't just tell them, show them! Teachers can intentionally alter their instructional gestures. In R. Sun & N. Miyake (Eds.) Proceedings of The 28th Annual Conference of the Cognitive Science Society (pp. 1523-1528). Mahwah, NJ: Erlbaum.

National Research Council (2001). *Adding it up: Helping children learn mathematics.* J. Kilpatrick, J. Swafford & Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 398-415.

Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory and Cognition, 29*, 214-221.

Reed, S.K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 124-139.

Reed, S.K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 106-125.

Research Advisory Committee of the National Council of Teachers of Mathematics (1996). Justification and Reform. *Journal for Research in Mathematics Education*, Vol. 27, No. 5, pp. 516-520.

Richland, L.E., Zur, O., & Holyoak, K.J. (2007). Cognitive supports for Analogies in the Mathematics Classroom. *Science, 316*, 1128-1129.

Rittle-Johnson, B, & Star, J.R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*, 561-574.

Ross, B.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629–639.

Ross, B.H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 456–468.

Stevenson, H. W. and Stigler, J. W. (1992). *The learning gap: Why our schools are failing, and what we can learn from Japanese and Chinese education*. New York: Summit Books.

Stigler, J. W., & Fernandez, C. (1995). Learning mathematics from classroom instruction: Cross-cultural and experimental perspectives. In C. Nelson (Ed.), *Contemporary perspectives on learning and development*. Hillsdale, NJ: Erlbaum. Pp. 103-130.