



Supporting Online Material for

Cognitive Supports for Analogies in the Mathematics Classroom

Lindsey E. Richland,* Osnat Zur, Keith J. Holyoak

*To whom correspondence should be addressed. E-mail: l.e.richland@uci.edu

Published 25 May, *Science* **316**, 1128 (2007)
DOI: 10.1126/science.1142103

This PDF file includes:

Materials and Methods
Figs. S1 to S4
References and Notes

Cognitive Supports for Analogies in the Mathematics Classroom

Lindsey E. Richland, Osnat Zur, and Keith J. Holyoak

Online Supplemental Materials

METHODS

Video Data

Video-data used for the analysis were randomly sampled from the corpus of data collected as part of the Trends in International Mathematics and Science Study (TIMSS 1999 Video Study; *S1*). In the full dataset, specific teachers and lessons for videotaping were identified through a random probability sample of all lessons available to eighth-grade students in a given country throughout one school year. Each selected classroom and teacher was videotaped on one occasion during a normal class period. Lessons were typically approximately 50 min long. For the current analysis, ten lessons were randomly selected from the full sample collected in each of the United States, Hong Kong, and Japan. We thus analyzed thirty lessons total, or approximately twenty-five hours of videotaped data (*S2*).

Development of Codes

Lessons were analyzed using V-Prism 3.045 (*S3*), a computer software system designed to allow simultaneous viewing of a digitized video and its typed transcript on a computer screen. In this software, the transcript moves temporally with the video and can be marked with research codes. Qualitative codes were used to generate quantitative information about patterns of analogy usage that relate to dimensions previously identified, using experimental paradigms, as important to relational reasoning. The concepts of the codes were developed before the observational phase of the current project. Many of the codes were conceived during a prior study of thirty U.S. videotapes collected as part of an earlier video TIMSS study (*S4*). These

were not the same videos analyzed in the current study. Importantly, because the seeds for these codes emerged from the U.S. video data, the present study did not begin from the Asian videotapes or from a deficit model of teaching practices in the U.S.

We refined and honed the codes to be broadly relevant to practices that were typical in the Japanese and Hong Kong classrooms as well as those in the U.S. As noted by experts in the field of cross-cultural video survey analysis, “we may be blind to some of the most significant features that characterize teaching in our own culture because we take them for granted as the way things are and ought to be. Cross-cultural comparison is a powerful way to unveil unnoticed but ubiquitous practices” (S5).

Coding Procedure

The methodology used in this study was based on that employed in previous cross-cultural surveys of classroom instruction (S1, S6). Coders were undergraduates without training about relational reasoning (although they knew the nation of origin for each videotape). They were blind to the hypotheses in that they were not instructed about which codes would be deemed “better or worse” practice. In research meetings, we framed our research aims as “describing instructional differences,” rather than as “assessing comparatively better or worse practices.” Many coders were not even psychology majors. Further, many of the coders were native to Asian countries and had come to the U.S. for their college education. If anything, these students described to us a general predisposition to believe that teaching in the U.S. would be of higher quality than in their home countries.

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 qualitative measures that produced quantitative frequency data. At least

two coders analyzed the data on each pass, and reliability was calculated for all coders on approximately 20% of the data. All disagreements were resolved through discussion. When enough information was not available to make a decision on a particular code for a specific unit of analysis (e.g., the camera did not capture a section of a worksheet), that relational comparison was excluded from analysis of that particular code. Such occurrences were not common but did result in some eliminated items per code, as noted below.

Identification of Relational Comparisons

Two expert coders (researchers in the field of analogy), the first two authors Richland and Zur, separately identified all units of relational comparisons within every lesson. All disagreements were resolved by discussion. The definition of a relational comparison was derived from Gentner's structure-mapping theory (S7). Relational comparisons were defined as a higher-order relationship between a source (or base) object and a target object, which could highlight alignable similarities or differences (S8). Objects were defined as entities that function as wholes at the given level of analysis. Several criteria were used to mark an interaction as a relational comparison. We used a conservative identification strategy. First, a source and a target had to be clearly identifiable. For example, an instance in which the teacher stated: "let's 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 intention to compare the source and target items. This could be an explicit verbalization (e.g., "X is *just like* Y") or a less explicit verbalization signifying a comparison (e.g., "We just finished with X, now lets do *another one* (Y)"). Serial or parallel spatial positioning on the board or on a worksheet or textbook page could also evidence intention to compare. Finally, gesture that motioned between two or more representations could signify a

comparison. Mathematical comparisons that were the heart of a problem were not included (e.g., in a lesson on proportions, every time a teacher used a mathematical proportion was not considered an instructional relational comparison). When a single relational comparison was repeated with different students, it was counted as a single comparison and was only coded once. See Table S1 and Figure S1 for a verbatim transcript and illustration of a sample relational comparison from the U.S. dataset.

Table S1. Example of a relational comparison (Hong Kong data set).

The class solves two algebraic problems together by determining whether they can make the left hand side equivalent to the right hand side. The first problem leads to the solution ($x=2$) in which the left hand side and the right hand side are not equivalent. The second problem leads to the resolution $0 = 0$ and the left hand side can be made equivalent to the right hand side. The latter is used to illustrate the concept of “identity.” The teacher ends with the following summary:

Teacher: We have the conclusion, this equation [References problem ending with $x=2$] is not an

identity. Okay? If it is not identity, we can still call it an equation. Okay?

It is only an equation, not an identity. If it is an equation, it may be one solution only.

Because it is one unknown, one equation.

Therefore, you may find the solution for it. Just one. But for identity, you

have, in fact, infinite many solutions [References problem ending in $0=0$].

Okay?

It will be satisfied for all the values of X for identities. Understand? Know the difference between identity and equation.

And for identity, in between the two sides, you can use a new symbol

with three lines. And we read it as, is identical to. Or you can say that they are identically equal. Okay?

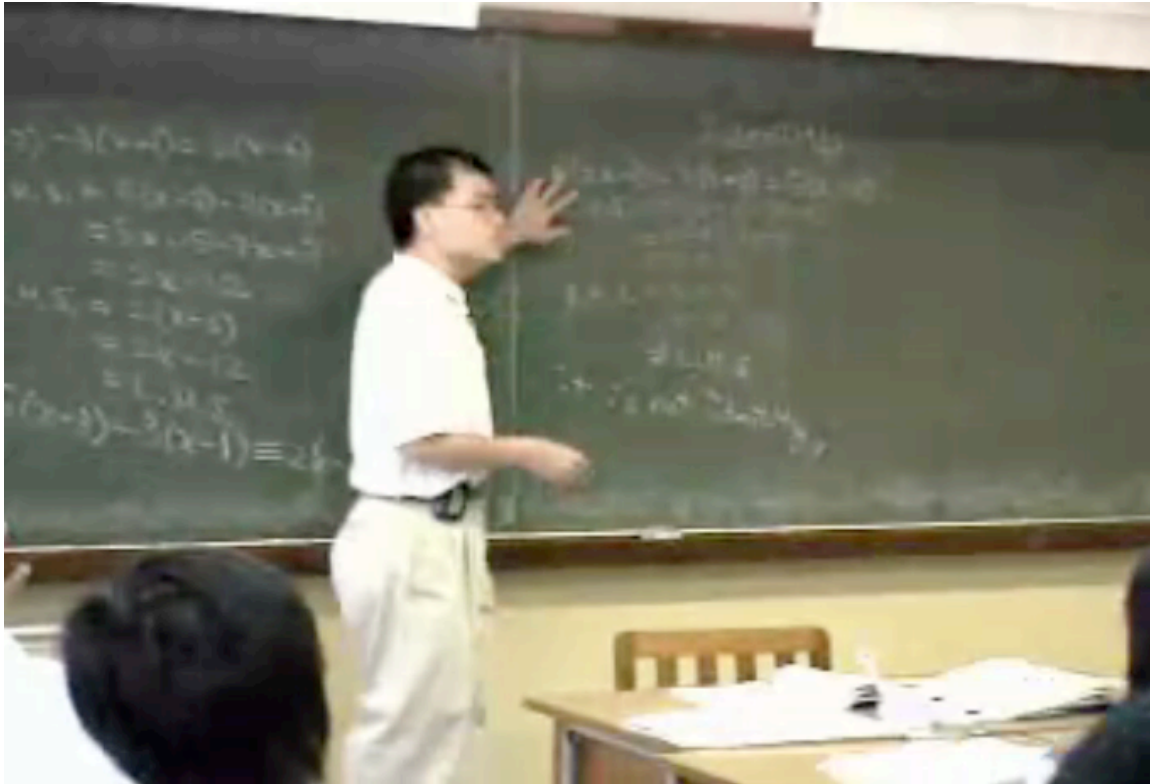


Figure S1. Teacher from Hong Kong compares two equations in which only one illustrates identity (right equation).

Coding Effective Use of Relations by Teachers

Codes (Figure 1) addressed students' familiarity with the source, teachers' use of imagery, teachers' visual presentation of source objects, and the visual availability of the source during comparisons to the target. Codes were developed for the mathematics classroom context by a team of international researchers. Researchers were fluent in English, English and Cantonese, or English and Japanese, and had attended school (including eighth grade) in these regions. The content for the codes was derived from basic experimental research on analogy, but was adapted to the mathematical classroom context by the research team. The international composition of the research team was considered important to ensure that codes were sensitive to

cultural practices and did not discriminate between functionally equivalent activities across regions. Reliability in the measures was calculated for each code on 20% of the coded lessons (five lessons). Lessons for reliability assessments were selected from across the three countries. Reliability was calculated between at least two members of the research team, and typically three (one per country) for each code.

Code A: Teacher uses a familiar source. All sources were coded for the level of students' presumed familiarity with the object. Coding decisions were based upon explicit information 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 math learned the prior year or before) or ‘new information’ (information taught in the current or immediately prior lesson). Coders native to each region (U.S., Hong Kong, and Japan) provided supplemental information about whether a source was a part of standard cultural knowledge for students in that region. 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.

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, $\chi^2(2) = 6.6, p < .05$. As shown in main text Figure 1, 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.

Code B: Teacher presents the source analog visually. Binary codes were used to assess whether or not the source was presented using a visual representation. Visual representations were defined as source objects that were written in text (e.g. on a chalkboard, an overhead projector, a worksheet, or a textbook), designed using physical manipulatives (mathematics instructional tools that can be handled by a teacher or students, e.g. blocks or cuisenaire rods), or shown on a dynamic computer or television platform. Eight units were excluded from analysis. Reliability between coders was calculated as 91%. There was a difference in the likelihood that the source was visually presented by country, $\chi^2(2) = 34.6, p < .001$. Although teachers in all countries presented most sources visually, U.S. teachers presented sources visually in a smaller percentage of analogies than did Hong Kong and Japanese teachers.

Code C: Teacher keeps the source visible during comparison with the target. A binary code categorized whether the source was visible to students during instruction of the target. A positive score was given if students could easily see the source while they were expected to be reasoning about the target. A visually available source generally meant that the source was written on the classroom chalkboard or overhead projector and this representation was left in front of the class during production of the target (see sample Figure S2). Sources could also be written on an open page of the textbook or on an active worksheet page, as have been shown to improve relational learning (S9).

Eight units were excluded from analysis. Reliability between coders was calculated as 87%. The frequency of source visual availability differed across countries, $\chi^2(2) = 28.4, p < .001$. U.S. teachers were least likely to make the source visually available during production of the target, while Hong Kong teachers were more likely, and Japanese teachers were even more likely to do so.

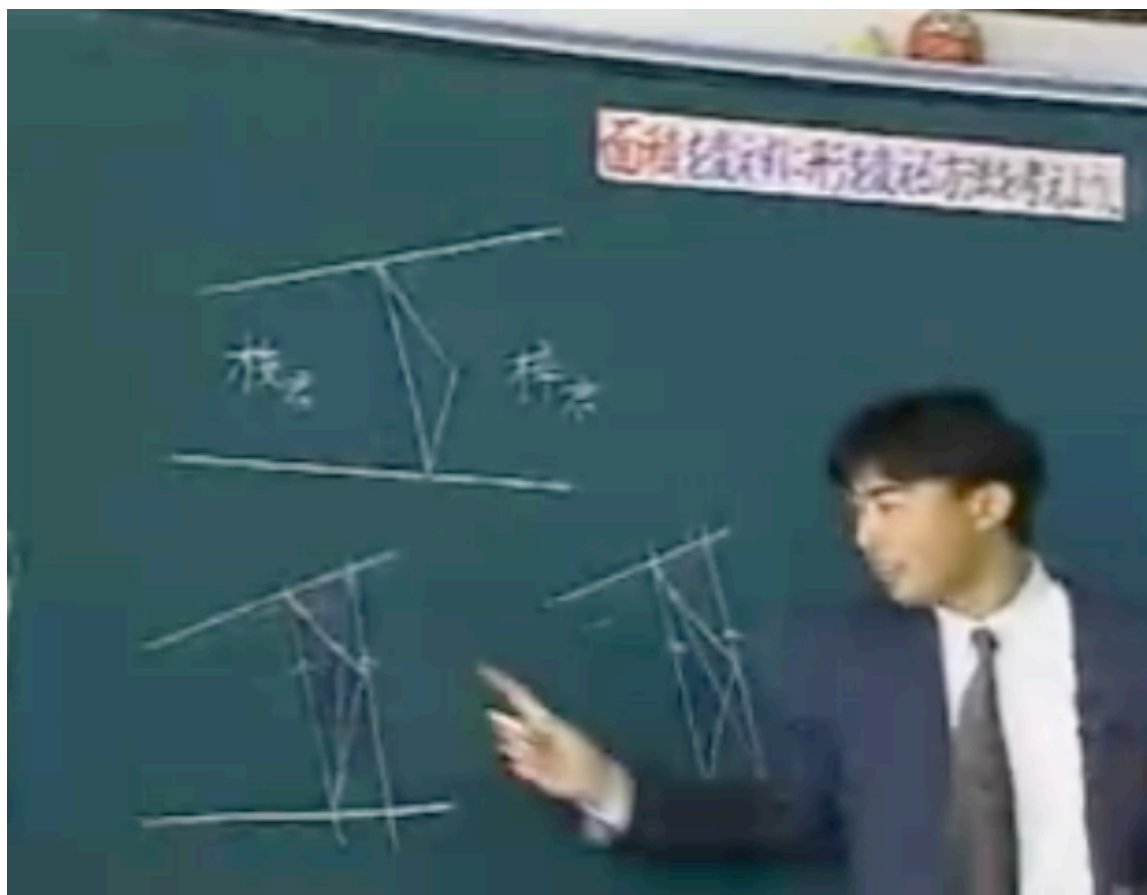


Figure S2. Teacher from Japan keeps the source strategy (bottom right drawing) visible while explaining the target solution strategy (bottom left drawing).

Code D: Teacher uses spatial cues to highlight the alignment between corresponding elements of the source and target. A binary code assessed presence or absence of spatial cues to highlight the structural alignment between corresponding elements within the source and target objects. A comparison was coded as providing spatial cues if the source and target objects were written on the board, on textbook or worksheet pages in a way that emphasized the structural similarity and alignment between the two objects. Cases included instances where the teacher drew a parallel diagram for two problems, aligned sequential steps in solving two problems, or put objects in a table that highlighted property correspondences (see sample Figure S3).



Figure S3. Teacher from Hong Kong uses spatial cues to highlight the similarities and differences between the terms “equilateral” and “equiangular.”

Four units were excluded from analysis. Reliability was 91% between coders. Frequency of visual alignment differed across countries, $\chi^2(2) = 30.1, p < .001$, with teachers in the United States using visual or spatial mapping cues least frequently, while Hong Kong and Japanese teachers were more than twice as likely to provide these cues.

Code E: Use of Comparative Gesture. Coders first marked presence or absence of any hand or arm gesture. Gesture was considered any hand or arm movement that was part of instruction (e.g. touching a desk while talking would not be counted) and that could be interpreted as helping students to interpret the instruction. We followed researchers who have described gestures as providing additional information to support instruction (*S10, S11*). If

present, gesture was categorized according to the focus of the gesture. Gestures were designated as either referring only to the source or the target analog (not Comparative) or as comparative between the source and target analog (Comparative). To be designated as comparative, the gesture had to mark the relationship between the source and target analogs. Typically, this meant a continuous gesture was first directed at one analog and then moved to the second analog. This action might coincide with a statement such as the following: “remember how we subtracted x from y here in the last problem? *Gesture to the source*. Now look at this new problem, we can do the same thing. *Gesture moves from the source to the target*.” If a teacher used this type of gesture that moved between source and target objects, the unit was marked as comparative gesture. Twelve units were excluded from analysis. Coders were 91% reliable.

The overall frequency of using gesture during relational comparisons did not differ reliably across countries, $\chi^2(2) = 4.9, p = .09$. Although U.S. teachers were least likely to develop an analogy with gesture (means U.S. = 83%, Hong Kong = 90%, Japan = 90%) all countries used gesture in most relational comparisons. The countries differed in the type of gesture used in comparisons with gesture, $\chi^2(2) = 36.0, p < .001$. Specifically, as shown in Figure 1, teachers in the U.S. used comparative gesture least often, whereas those in Hong Kong were much more likely to do so, and those in Japan were more than three times as likely.

Code F: Teacher uses mental imagery or visualizations. Coders used a binary code to assess presence or absence of imagery. This code was marked as present when the teacher invoked a familiar mental image from outside the domain of mathematics that provided substantive information to help clarify the mathematical concept (e.g. “adding negative numbers is like when you lose football yards and then lose more yards”). The imagery contributed to providing an enriched analog from which students could draw inferences about the target. This

code was also marked as present when the teacher used a visual representation to make the analogy more vivid (e.g. a computer animation, manipulatives, or a diagram) (see Figure S4).

Problem setting was not included in this code. For instance, a word problem set in a non-mathematical context was not coded as containing imagery. Rather, a non-mathematical context had to provide mathematically relevant insights to be considered imagery. For example, a scale used to describe balancing two sides of an equation provided specific relational information that supported students' understanding of the process of balancing equations.



Figure S4. U.S. teacher uses blocks to create a visualization of the mathematical concept exponential growth.

Coders were 94% reliable on imagery codes. No units were excluded from analysis. The difference in use of imagery differed by country, $\chi^2(2) = 21.1, p < .001$, such that teachers in the

United States was least likely to use imagery, those in Hong Kong used imagery almost twice as often, and those in Japan used imagery almost three times as often. (*S12*)

REFERENCES and NOTES

- S1. J. Hiebert et al, *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study* (NCES 2003-013) (U.S. Department of Education, National Center for Education Statistics, Washington DC, 2003).
- S2. Four lessons that were selected and coded from each country are publicly available for purchase at (<http://www.llri.org/html/TIMSS/index.htm>). These lessons are identified as U.S. Public Use Lessons 1-4, Hong Kong Public Use Lessons 1-4, and Japanese Public Use Lessons 1-4. We also randomly sampled and coded six lessons per country from the corpus of TIMSS 1999 video data (*S1*). The corpus was labeled from 1-100 (or 1 to the maximum number of videotaped lessons) per country. Of the total U.S. sample, we randomly selected the following U.S. lessons and transcripts for analysis: #21, #31, #34, #69, #74, #75. Of the total Hong Kong sample, we randomly selected the following Hong Kong lessons and English transcripts for analysis: #18, #21, #39, #55, #72, #88. Of the total Japanese sample, we randomly selected the following Japanese lessons and English transcripts for analysis: #2, #5, #35, #53, #55, #46.
- S3. S. Knoll, J.W. Stigler, *International Journal of Education Research*, 31, 725 (1999).
- S4. L. E. Richland, K.J. Holyoak, J.W. Stigler, *Cognition and Instruction*. 22, 37 (2004).
- S5. J. Stigler, R. Gallimore, J. Hiebert, *Educational Psychologist*. 35, 87 (2000), p.88.
- S6. Markman, A., D. Gentner, *Memory and Cognition*, 24, 235 (1996).
- S7. J.W. Stigler, P. Gonzales, T. Kawanaka, S. Knoll, A. Serrano. *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. (Department of Education, National Center for Education Statistics, Washington, DC, 1999).

S8. D. Gentner, Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155 (1983).

S9. B. Rittle-Johnson, J.R. Star *Journal of Educational Psychology*, (2007).

S10. S. Goldin-Meadow, *Hearing gesture: How our hands help us think* (Harvard University Press, Cambridge, MA, 2003).

S11. S. Goldin-Meadow, S. Kim, & M. Singer. *Journal of Educational Psychology*, 91, 720 (1999).

S12. Useful Internet links for further information:

Lessonlab Research Institute. Links are provided to research articles and reports from the TIMSS 1995 and 1999 Video Studies, as well as to public release videotaped lessons that can be purchased: <http://www.llri.org/>

Public use videotaped lessons from the TIMSS 1999 video study can be purchased at this site in the form of a box set with four lessons from eight countries: <http://www.rbs.org/catalog/pubs/pd57.php>

Results from the TIMSS 1999 and 2003 achievement tests can be viewed and downloaded at: <http://www.nces.ed.gov/timss/>

For related research conducted in Richland's Learning and Cognition laboratory at University of California, Irvine, please see: <http://www.gse.uci.edu/richland/>

For more information about related research on analogy and other aspects of human reasoning conducted in the laboratory of Keith Holyoak at the Department of Psychology, University of California, Los Angeles, please see: <http://reasoninglab.psych.ucla.edu/>.