

Eliciting Self-Explanations Improves Children's Performance on a Field-Based Map Skills Task

Kim A. Kastens

*Lamont-Doherty Earth Observatory
Department of Earth & Environmental Sciences
Columbia University*

Lynn S. Liben

*Department of Psychology
The Pennsylvania State University at University Park*

To investigate children's ability to translate between the environment and an abstract representation, fourth graders were asked to indicate the location of colored flags by placing similarly colored stickers on a map. In the explaining condition, students wrote down what clues they had used; in the baseline condition, they placed stickers without explanation. The explaining students significantly outperformed the baseline students, especially with respect to egregious errors indicative of failure to understand basic representational correspondence. The hypothesized interpretation is that children who generated explanations were more likely to notice and then correct discrepancies between their answers in progress and the referent space and that they did so by activating existing spatial and symbolic competencies.

Becoming a skilled user of maps in the real world is not an easy cognitive task. When people look around them, they see a world that is constantly changing: day to night, season to season, and minute to minute. They perceive a landscape of millions of tiny details: blades of grass, cracks in the sidewalk, ripples on the pond. They look horizontally out across the landscape, from a vantage point 4 or 5 ft (ap-

Correspondence should be addressed to Kim Kastens, Lamont-Doherty Earth Observatory of Columbia University, Route 9W, Palisades, NY 10964. E-mail: kastens@ldeo.columbia.edu or Lynn S. Liben, Department of Psychology, The Pennsylvania State University, 450 Moore Bldg., University Park, PA 16802–3106. E-mail: liben@psu.edu

proximately 1.2 to 1.5 m) above the ground. Somehow, they need to learn to “translate” mentally from the intricate, constantly changing, horizontally viewed world that they see around them into a schematic, unchanging, vertically viewed map as summarized in Figure 1.

Anecdotal encounters with map-challenged adults, plus a large body of cognitive-developmental research literature, tell us that many people never master this “translation” skill (Liben, 2006). In a review of the research literature on children’s acquisition of map literacy, Liben and Downs (1989) concluded that many young children show competence at simple map tasks, such as interpreting a floor plan of a single room containing a few items. However, when children are asked to perform tasks that are more akin to those facing adults in either practical or professional map-using situations, their performance is much less impressive. For example, when asked to select the bird’s-eye view of their school building from among six choices, only 20% of first graders and 29% of second graders could correctly iden-

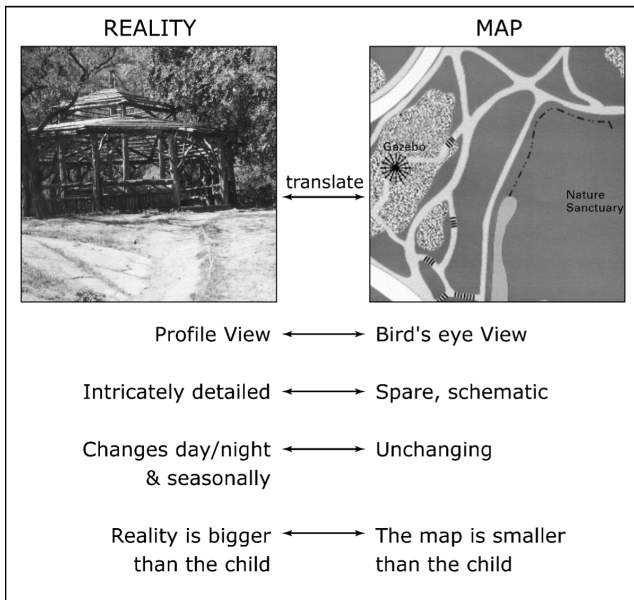


FIGURE 1 Locating oneself or an observed object on a map is a complex cognitive task. The map user must translate information seen in profile view to a spatial representation drawn in bird’s-eye view; from an intricately detailed, visually perceived landscape to an abstract and schematic spatial representation; from a view that changes over time to an unchanging representation; from a large environment to a small representation. From “Development and Evaluation of ‘Where Are We?’ Map-Skills Software and Curriculum,” by K. A. Kastens, D. Kaplan and K. Christie-Blick, 2001, *Journal of Geoscience Education*, 49, p. 249. Copyright 2001 by of *Journal of Geoscience Education*. Reprinted with permission.

tify the U-shaped plan view (Downs & Liben, 1990). When asked to place colored stickers on a map in locations that corresponded to colored flags on a three-dimensional model of the same terrain, only 27% of the first graders' stickers and 40% of the second graders' stickers ended up in approximately the correct location (Liben & Downs, 1989). These and other research results lead to the conclusion that skilled map use is not an ability that develops naturally and inevitably in all children, like walking or talking. Instead, it is a complex ability that must be taught and practiced if it is to be fully developed.

In thinking about the mistakes that people make in using maps, it is important to distinguish between understanding some general aspect of the map-space correspondence and being able to actually use that understanding to locate oneself or an object in an unfamiliar, realistically complex terrain. Working with adults experienced in using topographic maps and who could thus be presumed to understand the general concept of correspondences between map and space, Pick et al. (1995) found that only 1 of 17 could find their position on a topographic map when "dropped" into an unknown location and not allowed to move around. Liben, Kastens, and Stevenson (2002) reported a similar finding from a study with college students who were new to campus. Participants were taken to five locations on campus and asked to place stickers on a campus map to indicate where they thought they were. Performance covered the full range, with some participants correct on no locations and others correct on all. What was striking was the dramatic nature of many of the errors: Stickers were often placed far from the correct spot, and often on or near map symbols representing buildings or locations with very different qualities than the correct ones.

In summary, past research has shown that neither children nor adults are universally successful when asked to translate information obtained from the large-scale environment to an abstract plan map or vice versa. Our work has been aimed at trying to develop methods to enhance these skills in elementary school children and to understand more fully where children go astray. We have developed a set of field-based map skills assessments specifically to examine participants' ability to translate information from map to referent space and vice versa. These were originally developed to evaluate the *Where Are We?* software and curriculum materials (Kastens, Kaplan, & Christie-Blick, 2001; Liben et al., 2002) but have also proven of value for investigating children's strategies and mistakes while using maps more generally. The methods used in this study are notable for their use of a large-scale outdoor environment of sufficient size and complexity that the participants could not see the entire field area from any single vantage point. This spatial scale is understudied in the research literature on map-comprehension and spatial cognition, yet it is a crucial scale for the use of maps in everyday life and in map-using professions such as geology, ecology, architecture, or urban planning.

To better understand the sources of errors children make while using a map within the represented environment, we designed a method to ask children di-

rectly how they were solving location problems. The current research contrasts the overall performance and type of mistakes made by students who were required to explain their reasoning while performing the field-based map task versus those who did the task without explaining. As described in detail later, although the groups did not differ in spatial abilities, the students who were asked to explain performed significantly better than the students who completed the task without explanation.

Although our major goal was to learn more about children's spatial thinking, the data obtained also bear on questions that derive from research on metacognition, in particular, research on the *self-explanation effect*. Prior research has found that under some circumstances (e.g., Alevén & Koedinger, 2002; Bielaczyc, Pirolli, & Brown, 1995; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, DeLeeuw, Chiu, & LaVancher, 1994; Lin & Lehman, 1999), but not all circumstances (e.g., Hausmann & Chi, 2002; Mwangi & Sweller, 1998), requiring students to generate and articulate their own explanations of their reasoning or understanding enhances deep learning. Our study differed from most prior studies of self-explanation in that, first, the task called on spatial thinking rather than verbal, analytical, or quantitative skills, and second, participants had to develop a solution to a novel problem rather than applying learned problem-solving strategies. Our study shared with much of the prior work on metacognition the goal of understanding what assignments and learning environments can help students move beyond ability to repeat or paraphrase learned information to ability to tackle unfamiliar questions, tasks, and problems (Donovan & Bransford, 2005; Kintsch, 1994).

METHODS

Participants

The participants were fourth-grade students from a suburban elementary school. The school serves a diverse population, with 19% of the students eligible for free lunch (The University of the State of New York, State Education Department, 2004). All fourth graders in the school during the academic years 2001–02, 2002–03, and 2003–04 participated in the research, except for a handful who were absent during our field trip. No child or parent declined to participate. As reported by the school, the participants were 45% White, 35% Black, 10% Hispanic, and 9% Asian. As described later, we report on two groups: “baseline” and “explaining.” The baseline data were derived from pretests of 99 children (52 girls and 47 boys) who were participating in an evaluation of the *Where Are We?* software and curriculum during May–June 2002 and 2003. The explaining data were derived from 34 children (20 girls and 14 boys) who were given the mapping task in June 2004. The three teachers remained constant across the 3 years.

Field Area and Map

Our field area is the northern half of the campus of the Lamont-Doherty Earth Observatory in Palisades, NY. The campus is a former estate, with park-like grounds including paths and roads; a parking area; an equestrian statue; buildings of various sizes and shapes, including a greenhouse and a mansion; grassy areas; an orchard; isolated trees; massed shrubbery; and a walled garden with pond. The area used for the study is approximately 170×210 m. The area is sufficiently large and vegetated that it cannot be seen in its entirety from any single vantage point. During the years of data collection, no changes occurred to the field area of sufficient magnitude to be reflected on the map.

The map used was a plan-view architect's rendering, reproduced in Figure 2. A map key, north arrow, and scale bar were inset in a box in a corner of the map. The

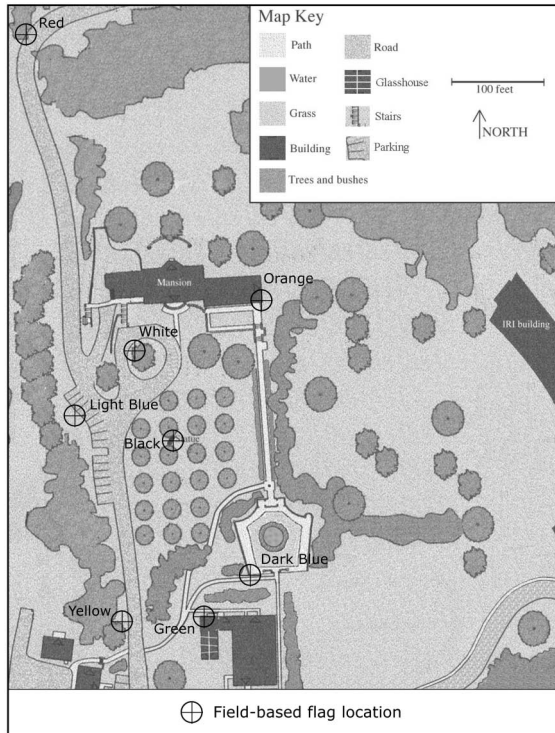


FIGURE 2 Black-and-white version of the map that the students used in the flag-sticker task. The students' maps were printed in color on $8\frac{1}{2} \times 11$ in. (approximately 20.3×27.9 cm) paper. The correct answer for each of the eight sticker locations is superimposed. The eight locations were chosen to represent a range of difficulties. The symbol indicating the correct flag locations shows the size of a sticker relative to the mapped features.

maps given to the children were color-photocopied onto $8\frac{1}{2} \times 11$ in. (approximately 20.3×27.9 cm) paper at a scale close to 1:1,000. Grass, bushes, and shrubs were shown in green, the pond and fountains in blue, and the paths and roads in shades of gray. Because the map scale was so large, the mapped width of the symbolized roads and paths was at the same scale as the map, and the mapped shape of the building symbols matched the real-world building footprints.

The Flag-Sticker Field-Based Map Skills Test

The field-based maps skills assessment used here was modeled after a test developed by Liben and Downs (1986, 1989). In the original task, seven colored flags were placed in various locations on a tabletop three-dimensional model of the local region, and children were asked to place similarly colored stickers on a topographic map of the same region to show the flags' locations. In our life-size version, we placed eight flags around the field area. We gave each student a paper map and a set of eight colored stickers. The students were asked to place each sticker on their map to indicate the position of the similarly colored flags. The stickers were round, and the diameter ($5/16$ in., or approximately 8 mm) corresponded to approximately 7 m on the ground. It was possible, with care, to peel off a placed sticker and reposition it elsewhere on the map.

The flag positions were selected to span a range from very easy to quite difficult. We made this judgment based on the recorded comments of students who did the flag-sticker activity in pairs during the pilot stage of the project and on prior research showing that locations on unique map symbols are easier than those on repeated symbols (e.g., Blades & Spencer, 1994; Liben & Downs, 1989), and locations proximal to landmarks are easier than locations remote from landmarks (Siegel, 1981).

Students were introduced to the activity one class at a time (14 to 20 students) as they sat on the steps of the mansion. As part of the introduction, each item on the map key was pointed out on a poster-sized version of the map held up in front of the group. After being given an opportunity to ask questions, students were called up one by one to begin the task. Individually, students were handed a map that was correctly oriented with the terrain, shown where they were standing and which direction they were facing, and launched on the task. Students could explore the field area and place their stickers in any order they chose. Adults were strategically positioned around the periphery of the field area to keep children from wandering out of bounds. There was no time constraint.

Baseline and Explaining Conditions

In the baseline condition, the students completed the task exactly as just described. In the explaining condition, students performed the flag-sticker task as described earlier, but with one modification: After they had found a flag and placed its sticker

on the map, they were asked to "write down [on a paper form] what clues you used to decide where that color sticker should go on the map" before going on the next flag. The forms on which the students recorded their clue answers had prelabeled response areas for each color; the order of responses was not recorded.

The map, flag locations, script for describing the sticker-placement task, and procedure for briefing and launching the children were identical in the two conditions. In both experiments, the investigator emphasized, "This is not a race, and there is no time limit."

Spatial Abilities Assessments

We administered three paper-and-pencil assessments of students' spatial skills to allow us to determine whether the two groups were well-matched in underlying spatial abilities. The three tests were selected to represent the three component spatial abilities identified by Linn and Petersen (1985): imagining figures or objects rotating in two- or three-dimensional space ("mental rotation"), identifying stable positions or axes such as vertical or horizontal in spite of a conflicting visual environment ("spatial perception"), and performing multistep spatial tasks using various strategies ("spatial visualization").

Mental rotation. Skill in mental rotation was assessed by the Spatial Relations Sub-Test of the Primary Mental Abilities test (Thurstone, 1962) as modified by Liben and Downs (1986). For each of the 21 items in this task, participants are shown a two-dimensional line drawing of a figure as a model and then asked to mark which of five drawings show the "same" figure—in other words, a figure that can be formed by rotation of the model within the plane of the paper. Scores are reported as the number of correct figures marked minus the number of incorrect figures marked within the allowed 6 min.

Water level. Ability to identify a stable axis in spite of a conflicting visual environment was assessed by a variant of the water level task, originally developed by Piaget and Inhelder (1956). The task was designed to tap respondents' ability to recognize the invariant horizontality of water in the face of the conflicting oblique cues provided by the sides of the tipped container that holds it. Following the procedure used by Liben and Golbeck (1980), children were asked to draw lines to show where water would be in each of six tilted bottles. No time limit was given for the task, which was typically completed within a few minutes. Performance was scored as the number of lines drawn parallel to the line indicating the tabletop, within a margin of error of 5°.

Hidden pictures. "Spatial visualization" (as defined by Linn & Petersen, 1985) was assessed by the Hidden Pictures test, designed by Liben as a children's

equivalent of the Group Embedded Figures Task (Witkin, Oltman, Raskin, & Karp, 1971). In this task, respondents are asked to find a simple figure within the pattern of a complex drawing. Scores are reported as number of simple figures correctly outlined, out of a maximum of 13, within the 6 min allowed.

RESULTS

Groups' Spatial Skills

As described earlier, to determine whether the baseline and explaining groups had different spatial abilities, we administered three spatial skills measures. Scores on the mental rotation, water level, and Hidden Pictures tests were analyzed with *t* tests for each measure. There were no significant differences between the groups on any of the measures (all *ps* > .25), and thus there is no evidence that the groups' spatial abilities differed.

Types of Flag-Sticker Responses

All of the children's sticker maps were scanned, and the sticker positions were digitized. The digitized sticker positions were entered into a geographic information system (GIS) to give us flexibility in scoring the responses. After some pilot coding, we selected two ways of scoring each sticker location. The first was a simple numerical measure of the linear distance offset between a student's sticker location and the correct location. We used "sticker diameters" as the unit of measure, because the sticker size conveys an implicit message to the students about how much "slop" is allowed in mechanically placing the sticker onto the paper. The sticker size relative to the mapped features is shown in Figure 2. This offset measure is easy to understand and completely objective.

However, the offset measure fails to capture important aspects of the map-using process. For example, consider the orange flag, on the southeast corner of the mansion (see Figure 2). An orange sticker on some other building, even a building distant from the mansion, seems to show a higher level of understanding of the map-environment correspondence than a sticker placed in the middle of the lawn immediately north of the mansion, even though the latter position is metrically closer to the correct location. To capture these differences, we developed a categorical scheme of response types (see Table 1 and Figures 3 and 4), based in part on the types of errors observed in previous studies. Each sticker for each participant was categorized as falling into one of five error types or a "no error" category.

Error Types 1, 2, and 3 are the most egregious errors (see Figure 3). In Error Type 1, the sticker is placed on a map symbol that resembles the sticker in size

TABLE 1
The Categorical Scheme of Response Types Used
to Classify Sticker Placements

<i>Response Type</i>	<i>Description</i>	<i>Vulnerable Flags</i>	<i>Interpretation</i>
1	Sticker is centered on a map object that resembles the sticker in size and shape	All except white	Failure to understand or employ the assigned symbolic “stand-for” relationship
2	Sticker is placed on symbol for grass, trees or bushes, and not on or adjacent to any built object	All flags	Failure of representational correspondence
3a	Sticker should have been on structure (building, statue, or wall); placed instead on or adjacent to road or path	black, dark blue, orange, green	
3b	Sticker should have been adjacent to a road or path; placed instead on or adjacent to a structure	red, white, light blue, yellow	
4a	Sticker should have been on structure; placed on or adjacent to the wrong structure	black, dark blue, orange, green	Success of representational correspondence; substantial failure of configurational correspondence
4b	Sticker should be adjacent to a road or path; placed on or adjacent to the wrong road or path segment ^a	red, white, light blue, yellow	
5a	Sticker should be on structure; placed on correct structure, but outside of no-error ring ^b	dark blue, orange, green	Success of representational correspondence; subtle failure of configurational correspondence
5b	Sticker should be adjacent to a road or path; placed on correct path or road segment, but outside the no-error ring	red, white, light blue, yellow	
No error	Center of sticker is placed within two sticker diameters of the exact location	All flags	Good enough for personal navigation

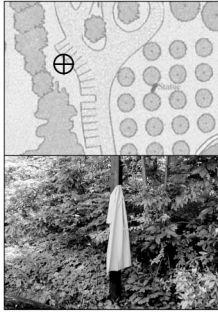
Note. For each color flag, the map area was divided into regions, each characterized by one of these response types. The assignment of stickers to regions was done via a geographic information system.

^aThe roads on the Lamont campus are connected to each other. We have divided them into segments based on where the observer in the field area perceives a marked change in character, such as a sudden change in width or gradient of the road. ^bThe black flag was placed on a statue, a structure that is small relative to the no-error ring; thus it is not vulnerable to error type 5a.

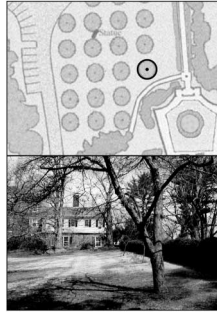
Representational Correspondence Errors

Category 1 Error: Sticker is centered on a map object that resembles sticker in shape and size.

Correct Answer Location

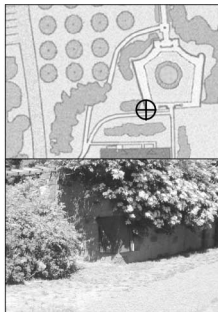


Type 1 Error



Category 2 Error: Sticker is placed on symbol for grass, trees, or bushes; not on or adjacent to any built object.

Correct Answer Location



Type 2 Error

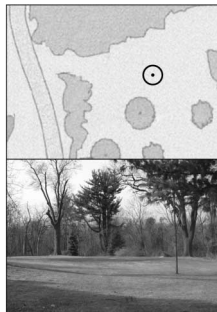


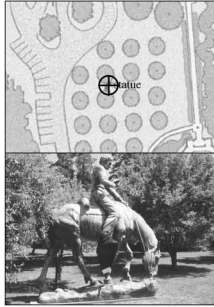
FIGURE 3 We interpret errors in Categories 1, 2, and 3 as indicative of failure to correctly exploit *representational correspondence*. Representational correspondence means that an object in the real world corresponds to a symbol on the map, and all instances of the same type of object in the real world are shown on the map with the same symbol.

and shape. Although this seems like an improbable thing to do, Liben and Yekel (1996) reported a similar finding in a study in which preschoolers placed round stickers on a classroom map that happened to contain a wastebasket symbol the same size and shape as the sticker. In Error Type 2, the sticker is placed out in the grass or trees, far away from any built feature, even though all of the flags were along roads, along paths, or on built structures. This is analogous to students in the Liben and Yekel study who placed stickers in the middle of the floor rather than on the furniture symbols. In Error Type 3, the sticker is placed on a built object, but the wrong sort of built object: for example, on a road when it should have been on a building. Error Categories 1, 2, and 3 are intended to capture circumstances in which the student has failed to understand the basic

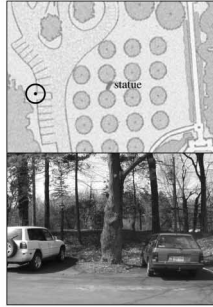
Representational Correspondence Errors

Category 3a Error: Sticker should have been on a structure (building, statue, or wall); placed instead on or adjacent to road or path.

Correct Answer Location

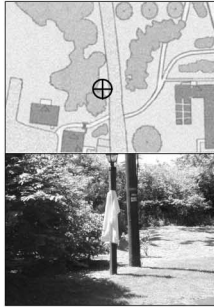


Type 3a Error



Category 3b Error: Sticker should have been adjacent to a road or path; placed instead on or adjacent to a structure.

Correct Answer Location



Type 3b Error



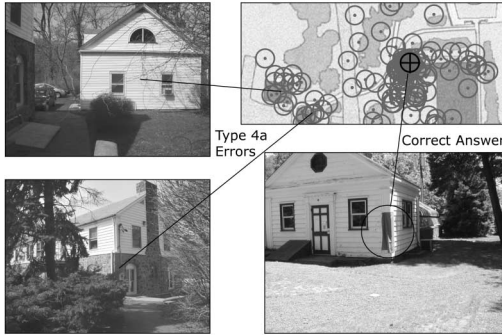
FIGURE 3 *Continued.*

“stand-for relationship” or “representational correspondence” (Liben & Downs, 1989, 1993) between something in the real world and the type of symbol that is used to represent that type of thing on the map.

Error Types 4 and 5 are more subtle errors (see Figure 4). In each case, the student has placed the sticker on the correct type of symbol: For example, a sticker corresponding to a flag on a building is placed on a building symbol on the map. However, in Error Type 4, the sticker is on the wrong symbol, for example, the wrong building. In Error Type 5, the sticker is on the right symbol, but on the wrong part of the symbol. Error Types 4 and 5 are intended to capture circumstances in which the student has correctly understood the representational correspondence between map and environment but has failed to use the configuration of features on the map to successfully disambiguate multiple occurrences of the same symbol on the map or to pinpoint location within a large map symbol.

Configurational Correspondence Errors

Category 4a Error: Sticker should have been on structure; placed on or adjacent to the wrong structure.



Category 4b Error: Sticker should have been adjacent to road or path; placed on or adjacent to wrong road or path segment.

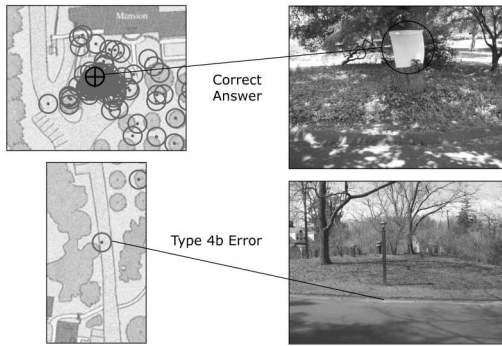


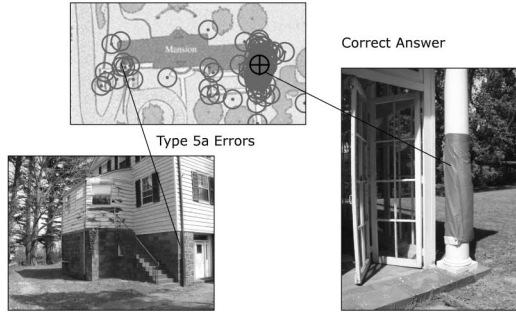
FIGURE 4 We interpret errors in Categories 4 and 5 as indicative of failure to correctly exploit *configurational correspondence*. Configurational correspondence means that the relative placement or arrangement of objects in the real world (i.e., their “configuration”) is the same as the arrangement of the corresponding symbols on the map.

A sticker was classified as a “no error” response if it fell within two sticker diameters of the correct placement, which corresponds to 16 mm on the map or approximately 15 m on the ground. If map readers were using the map for navigation to reach a destination, by the time they had come this close to the target, they would likely switch from using the map to scanning the terrain for the destination; thus, this two-sticker-diameter criterion can be read as “close enough” for personal navigation.

This categorization was implemented by defining fields within the GIS system, so that we were able to iterate through multiple definitions of the response types until we felt satisfied that each was internally consistent. The final set of fields incorporated a 0.5-sticker-diameter buffer around each built object (struc-

Configurational Correspondence Errors

Category 5a Error: Sticker should have been placed on a structure (building, statue, or wall); placed on correct structure, but outside of "correct" ring.



Category 5b Error: Sticker should have been adjacent to a road or path; placed on correct road or path segment, but outside of the "correct" ring.

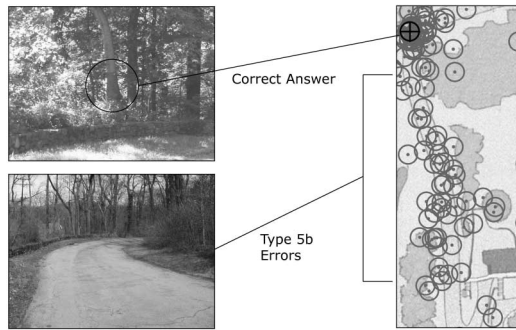


FIGURE 4 *Continued.*

ture, road, or path) so that a sticker was counted as being “on” the object if the center of the sticker was within half a sticker diameter of the object, that is, if the edge of the sticker overlapped the object. The GIS fields for each response type for each flag are shown in Figure 5. This coding scheme is similar in spirit and intent, but differs in detail, from the manual coding scheme described in Kastens et al. (2001).

The clue answers written by the explaining students were typed up and coded with respect to the following questions: (a) Is the clue answer an accurate or inaccurate description of the environment? (b) Does the clue answer describe one or more objects that exist both on the map and in the real world? (c) Does the clue answer contain sufficient detail to pinpoint the flag location? Details of the students' clue answers will be reported in a separate paper.

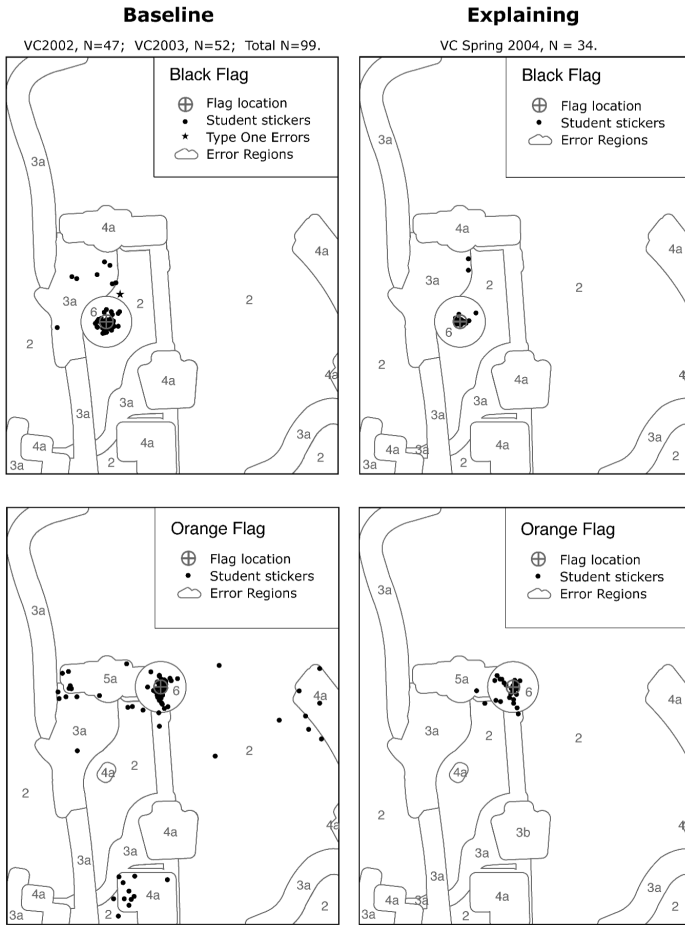


FIGURE 5 Maps showing the locations of all student stickers relative to the correct location and relative to the response type regions. For each flag, the stickers of the baseline students (left-hand map of each pair) are more scattered than the stickers of the explaining students (right-hand map). To show all of the data, the dots on these maps are smaller than the stickers placed by the students. The numbered regions show the GIS-based fields corresponding to each response type for each flag. VC refers to the name of the school.

Explaining Versus Baseline Students' Flag-Sticker Performance

Maps showing sticker placements for baseline and explaining groups are presented in Figure 5. Visual comparison of the left and right panels of the figure shows that the group of students who explained their reasoning placed their stickers far more accurately than did the group who did not.



FIGURE 5 Continued.

To quantify the students' responses, we first calculated the offset distance between each student sticker and the corresponding correct location, and then averaged across all eight stickers to compute an "average offset score" for each student, in units of sticker diameters. Note that on this measure, higher scores indicate worse performance. To visualize the magnitude of these sticker-diameter offsets, refer to Figure 2. Figure 6 provides histograms for sticker-diameter offset score. The baseline group includes many extremely poor scores (very high sticker-diameter offsets), whereas this tail of poor performers is absent in the explaining group.

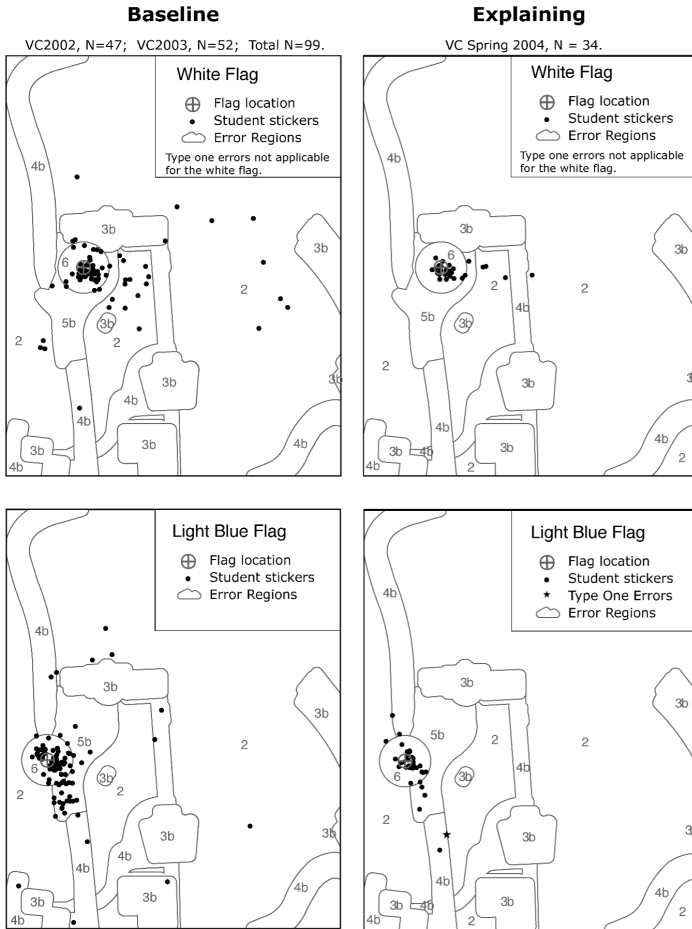


FIGURE 5 Continued.

Average offset scores served as the dependent variable in a Group (baseline, explaining) \times Sex analysis of variance (ANOVA). The main effect of group was significant, $F(1, 129) = 23.20, p < .001$, with children in the baseline group having larger average errors than children in the explaining group, means (standard deviations), respectively, of 4.9 (3.1) versus 2.2 (1.5) sticker diameters. There was neither a significant main effect for sex nor a significant interaction between sex and group.

The GIS-derived data are summarized in Table 2, which shows the mean number of stickers that the students placed in each response category. Category 1, 2, and 3 responses are the most serious kinds of errors. These errors were collapsed into a total representational correspondence error score, which was then analyzed

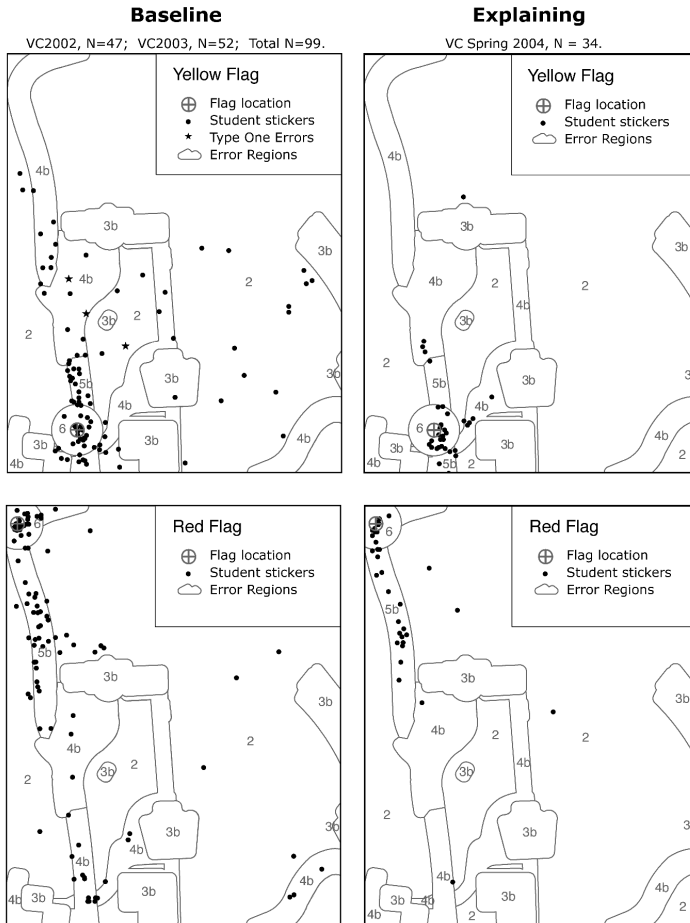


FIGURE 5 Continued.

with a Group (baseline vs. explaining) \times Sex ANOVA. There was a significant main effect of group, $F(1, 129) = 8.12, p = .005$, with a greater number of representational correspondence errors in the baseline group than in the explaining group, means (standard deviations), respectively, of 1.79 (1.59) versus 0.88 (1.04) stickers. There was also a significant main effect for sex, $F(1, 129) = 5.18, p = .025$, with boys making a greater number of these errors than girls, with means (standard deviations), respectively, of 1.87 (1.78) versus 1.29 (1.21). There was no significant interaction between group and sex.

The intermediate category of configurational correspondence errors also contains fewer errors among children in the explaining group than among those in the

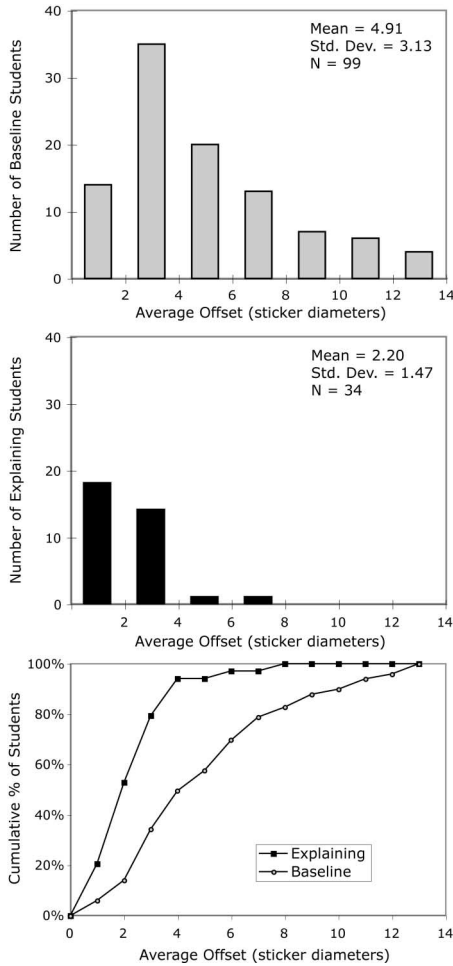


FIGURE 6 Histograms comparing performance of the baseline group and the explaining group on the average offset measure. The scale is in sticker diameters; see Figure 2 for size of stickers relative to the map features. On this measure, lower scores represent better performance. The baseline group has a long tail of poor performers that is not present in the explaining group, even accounting for the smaller sample size in the explaining group. Std. Dev. = standard deviation.

baseline group (see Table 2). However, this measure is ambiguous and is not further analyzed because an intervention that successfully improved map-using performance would move some answers into this category from the representational correspondence category and other answers out of this category into the no-error category.

TABLE 2
Mean Number of Stickers Per Student in Each Response Category

<i>Response Group</i>	<i>Response Type</i>	<i>Baseline</i>		<i>Explaining</i>	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Representational Correspondence Errors	1	0.06	0.24	0.06	0.24
	2	1.17	1.36	0.38	0.65
	3a & 3b	0.56	0.80	0.44	0.70
Configurational Correspondence Errors	4a & 4b	0.97	0.99	0.47	0.66
	5a & 5b	1.20	0.91	1.29	1.09
No error (within 2 sticker-diameters)	no error	4.04	1.70	5.35	1.43

Note. Relative to the baseline group, the explaining group placed fewer of their stickers in positions attributed to representational correspondence errors, and more of their stickers in the no error circle.

At the other end of the response spectrum were responses that fell within two sticker diameters of the correct location. The numbers of no-error responses were analyzed with a parallel ANOVA. Again, there was a significant main effect of group, $F(1, 129) = 14.71, p < .001$, with the number of no-error responses being significantly lower in the baseline than the explaining groups, means (standard deviations), respectively, of 4.04 (1.70) versus 5.35 (1.43). There was neither a main effect of sex nor an interaction between group and sex.

Both groups of students were told that there was no time limit and encouraged to take as long as they needed with each flag, but the explaining students spontaneously elected to spend longer at each flag. We did not time the students individually, but at the group level we noted that the time from the launch of the first student to the return of the last student was approximately 45 min for the explaining classes versus 15 to 20 min for baseline classes. The extra time was longer than would be required just to write down the fairly short clue answers, which averaged 11 words each. Students were observed to use this extra time to look around, pause, write a few words, look back at the flag, look around some more, and write some more.

DISCUSSION

Possible Causes for Superior Performance of Explainers: Motivation and Metacognition

Children who were required to explain what clues they had used to place each sticker performed significantly better, as a group, on our field-based map skills task than did children who performed the task without explaining their reasoning.

The study was designed to avoid confounding factors that could favor the explaining group. The map was the same and the flags were in the same locations. The flag-sticker task was explained to both groups using the same script and props. The participants began in the same location (sitting on the steps of the mansion facing south). The map was oriented individually for students in both groups. Both groups seemed to find the task interesting and enjoyable. The students were drawn from the same school, taught by the same teachers, and tested at the same point in the school year. They did not differ detectably in underlying spatial abilities.

We consider two possible explanations of the explaining students' better performance: motivation and metacognition. First, with respect to motivation, it is possible that the task in the baseline condition was not taken as seriously. Perhaps the activity of hunting for colored flags in a park-like setting seemed more like a game than a cognitive challenge. In both conditions, the children's own teacher introduced the activity and the experimenter and encouraged the students to do their best. However, the act of writing answers on paper in the explaining condition may have made the task seem more school-like and serious, which in turn could have motivated some students to spend more time and try harder.

Second, with respect to metacognition, or awareness of one's own thought processes, we note that activities that foster metacognition are associated with improved learning in children (Bransford, Brown, & Cocking, 2000; Donovan & Bransford, 2005; White & Frederiksen, 1998). The requirement that students write down the clues they used may have triggered more metacognitive awareness than occurred when students were merely asked to put the stickers on the map. Previous studies have found that explaining what one thinks or understands can improve learning or problem solving, a phenomenon sometimes called the *self-explanation effect*. The term *self-explanation* (Chi et al., 1989) or *self-generated explanation* (Chi & Bassok, 1989) refers to explanations generated by the learner, as contrasted with explanations provided by the instructor, textbook, or other external source. Compared to the modest gains typical in education research, the improvements attributed to self-explaining are stunning; self-explainers often do twice as well as non-self-explainers on the same task. The self-explanation effect has been documented under a wide range of conditions, on tasks ranging from text reading (e.g., Chi et al., 1994; Coleman, Brown, & Rivkin, 1997; Kintsch, 1994) to physics problem solving (Chi & Bassok, 1989) to computer programming (Pirolli & Recker, 1994) to spreadsheet use in accountancy (Reimann & Neubert, 2000) to experimental design (Lin & Lehman, 1999), and has been found for students ranging from middle school through college. The explanation may be spoken out loud or in one's own head (e.g., Didierjean & Cauzinille-Marmèche, 1997) or written or typed (e.g., Alevin & Koedinger, 2002). Self-explanation helps both when feedback is given on the correctness of the explanation (Alevin & Koedinger, 2002) and when it is not (e.g., Chi et al., 1989, 1994).

In light of the strong contrast in performance between our explaining and baseline groups, and in the context of these previous studies documenting the self-explanation effect, we attribute at least part of the superior performance of the explaining group in our map skills study to increased metacognitive awareness fostered by the requirement that students articulate their thought processes. We cannot, however, disambiguate the extent to which increased motivation may also have played a role. It is striking that the mere act of thinking and writing about their thought processes improved student' thought processes. There was no intervening step of external feedback or formative evaluation through which students could learn from their right and wrong answers.

Note that the students wrote down their clues after they placed each sticker. How, then, could this improve the accuracy of the stickers they had already placed? We see three possibilities: First, anticipation that they were soon going to have to provide a written self-explanation may have caused them to begin covert self-explanation (Chi, 2000) during the problem-solving step, while they were examining the landscape, planning their answer, and poised with sticker on finger above the map. Second, if they became aware of a discrepancy between their sticker placement on the map and the flag location in the terrain while writing their clue answer, they could peel off the sticker and reposition it on the map. Finally, self-explaining on earlier flags may have helped them figure out strategies that they could then apply to subsequent flags, such as using the pond as a landmark or using multiple landmarks to pin down their answer.

Hypotheses for Why Self-Explaining Improves Performance

There seems to be no single, simple answer concerning why students who explain their thinking perform better than those who do not. Table 3 summarizes some of the hypotheses in the literature. Note that these ideas are not mutually exclusive. One family of hypotheses (see Table 3a) centers around the concept that self-explaining helps students correct, repair, revise, or reorganize their existing knowledge, understanding, or mental model. For example, Chi et al. (1994) and Chi (2000) have suggested that for students learning by studying a biology textbook, self-explaining may help because it gives rise to multiple opportunities to see, and resolve, conflicts between one's evolving mental structure and the veridical description. The second family of hypotheses (see Table 3b) considers that self-explaining helps students construct something new: a new problem schema, a new situational model, new connections between broad scientific principles and individual actions. For example, Didierjean and Cauzinille-Marmèche (1997) inferred that some students developed a new abstract problem-solving schema when they were asked to self-explain while studying algebra problems.

The most parsimonious explanation for the superior performance of our explaining students falls in the correct–repair–revise family of hypotheses. Follow-

TABLE 3
Literature Review of Why Student Generation of Explanations Improves Learning, Understanding, or Problem Solving

<i>Reference</i>	<i>Name</i>	<i>Domain/Task</i>	<i>Interpretation</i>
(a) Correct, repair, revise, or reorganize			
Chi et al. (1989); VanLehn & Jones (1993)	Gap-detection and gap-filling	Studying worked examples of physics problems	Self-explaining causes students to detect, and seek to fill, gaps in their knowledge. Self-explanation causes students to detect, and seek to remedy, failures of their understanding.
Lin & Lehman (1999)		Experimental design using controlled variables in life sciences	
DeLeeuw & Chi (2003); Chi (2000)	Mental model revision	Studying biology textbook passage on blood circulation	Students begin with a mental model; as they read they (a) insert new information from text into model, (b) replace wrong knowledge with correct information, and (c) make inferences that address flaws in the model.
Chi et al. (1994)	Conflict-detection and resolution during knowledge integration	Studying biology textbook passage on blood circulation	Integration of new and existing knowledge during self-explaining gives rise to multiple opportunities to see [and potentially resolve] conflicts between one's evolving mental structure and the veridical description from the text.
Lin & Lehman (1999)	Organize thoughts	Experimental design using controlled variables in life sciences	Self-explaining helps students organize their thoughts, which helps them plan and monitor their activities.
Current study	Error detection & self-correction	Pinpointing map location of real world objects in a large-scale environment	Self-explaining causes problem-solvers to recognize that they are making a mistake and self-correct the mistake before finalizing their answer.

(b) Construct

Didierjean & Cauzinille-Marmèche (1997)	Schema formation	Studying worked examples of factoring in algebra	Self-explaining helps the learner to develop a new abstract problem-solving schema.
	Case-based reasoning		Self-explaining helps the learner to develop a library of annotated, cataloged examples, which are used in later case-based reasoning.
Reimann (1992)	Analogic enhancement ^a	Studying worked examples of physics problems	Self-explaining creates a richer, more-nuanced understanding of the example, which increases likelihood that student will retrieve the example and use it effectively when solving an analogous problem.
Aleven & Koedinger (2002)	Visual/verbal integration	Solving geometry problems	Helps integrate two modes of learning: implicit visual induction and explicit verbal knowledge acquisition.
Chi et al. (1994)	Construction of new declarative or procedural knowledge	Studying biology textbook passage on blood circulation	Self-explaining fosters construction of new knowledge, including inference rules that can be used in problem solving, or knowledge inferences that can be used to answer complex questions.
Chi et al. (1989); VanLehn et al. (1992)		Studying worked example of physics problems	
Lin & Lehman (1999)	Connect principles to action	Experimental design using controlled variables in life sciences	Self-explanation provides opportunities to make use of domain-specific knowledge to explain procedures used, and thus build connections between broad scientific principles and individual actions.
Kintsch (1994)	Situational model building	Reading text in science	Self-explaining helps student to combine prior and new knowledge to build a situational model (i.e., a working model with which one can make inferences, explanations, and predictions).

Note. Summary of published explanations for why student generation of explanations improves learning, understanding, or problem solving. Our study falls among those that attribute the self-explanation effect to the enforced opportunity to correct, repair, revise, and/or reorganize one's thinking.

^aThe term *analogic enhancement* is from VanLehn and Jones (1993).

ing Chi et al. (1994), we hypothesize that self-explaining may cause children to notice a conflict between the veridical location of the flag and their model of that location, that is, the sticker position on the map. They then self-correct this incipient mistake to the best of their ability. Given the short time on task and the absence of instruction or feedback, we further suggest that students achieve these self-corrections by activating their existing spatial and symbolic competencies, rather than by acquiring new competencies. Students who are not encouraged to explain are less likely to make the critical comparison between their model of the flag location and the incoming information (Chi, 2000) from the surrounding environment.

Where Was the Improvement Concentrated?

Looking at the data from the perspective of the types of errors made, the most dramatic improvement of the explaining students relative to the baseline students occurred in Error Categories 1, 2, and 3, those that we attribute to failure to fully understand and utilize the representational correspondence between map and represented space. A fourth grader, at age 9 or 10, would be expected to have a firm grasp of representational correspondence, given that even by age 3, children typically understand the general “stand-for” significance of representations (e.g., DeLoache, 1987). From this developmental perspective, then, the surprise is not that the explaining children avoided representational correspondence errors but that the baseline children made such errors. We infer that self-explanation activated underlying spatial and symbolic competencies pertaining to representational correspondence in map use and thus allowed the explaining children to avoid errors in representational correspondence.

If self-explaining is indeed activating existing spatial skills, it is interesting to ask whether it is doing so for both high-spatial and low-spatial children. Figure 7 shows regression plots for map skills performance as a function of mental rotation score, for both the baseline and explaining groups. Across the full range of spatial ability captured by the mental rotation test, the explaining group outperformed the baseline group. Likewise, on the water level and hidden picture spatial measures, both high- and low-spatial children in the explaining group outperformed children of similar spatial ability in the baseline group.

Consistency Between Self-Explanation Interpretation and Children’s Reported Clue Answers

A detailed analysis of the explaining students’ clue answers will be reported elsewhere as part of a study of the kinds of observations children make of map and referent space while using maps. For our current focus on the distinction between the baseline and explaining groups, the important finding about the clue answers is that these data are consistent with the self-explanation interpretation offered earlier. Good clue answers were associated with more accurate sticker placements,

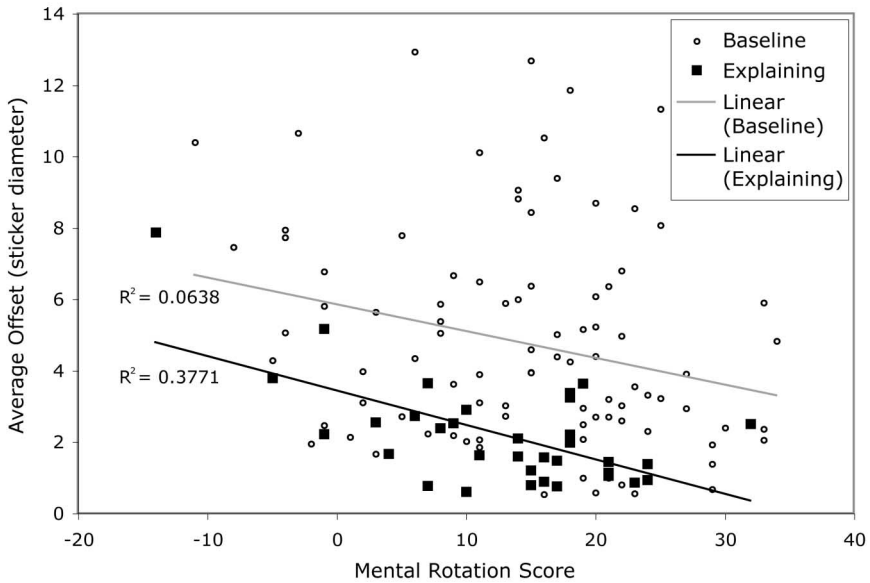


FIGURE 7 Regression plots for map skills performance (in sticker diameter offset) on mental rotation score, for the baseline and explaining groups. The downward shift of the entire regression curve from the baseline group to the explaining group shows that self-explanation improves performance across the full range of spatial abilities captured by the mental rotations measure. The tighter scatter around the explaining group regression line shows that mental rotation score is a stronger predictor of map skills performance among the explaining group ($R^2 = .38$) than among the baseline group ($R^2 = .06$). This is consistent with our interpretation that self-explanation is helping students activate their underlying spatial competencies.

where good clue answers are those that (a) mention at least one landmark that exists in the real world and is shown on the map and (b) describe the location of the flag accurately. This observed association is consistent with the hypothesis that task-relevant reflection on one's own reasoning improves sticker placement (although it could also mean merely that children who are good at one cognitive task are likely to be good at another).

Among the clue answers that mentioned at least one landmark that exists in the real world and is shown on the map, we distinguish between those that contain only topological information (e.g., “near the mansion” for white, “on a tree” for red) and those that contain information about the spatial configuration of multiple features (e.g., “on a tree in the middle of the circle in the road” for white or “on the edge of a building; inside is a pond” for dark blue). Clue answers that contain topological information are more abundant than clue answers that contain configurational information, which is consistent with our interpretation that self-explaining activated students' understanding of representational correspondence more so than configurational correspondence.

Finally, we note that many stickers were placed more accurately than would be possible if the child had been working from information in the associated clue answer alone. For example, one student wrote only “I walked and I spotted it [the green flag]” and yet placed that sticker perfectly. This is consistent with our view that self-explaining (externally, to an audience) serves as a thinking aid that can be useful for mitigating certain kinds of errors in certain circumstances, but the ability and willingness to recognize and articulate the reasoning behind one’s sticker placement is not a necessary part of the problem-solving process.

Is This Just a Gimmick?

If self-explaining on our task merely activates existing competencies rather than building new competencies, does this mean that self-explanation during problem solving is just a trick, a gimmick that improves tested performance without actually improving students’ knowledge or ability? In part that depends on one’s view of the pedagogical purpose of providing students with problems to solve. If problems are assigned primarily to assess students’ mastery of a body of content knowledge, then presence or absence of self-explanation could be viewed as a confounding factor that distorts the assessment’s validity.

If, however, it is an explicit goal of education to help students become skilled problem solvers (Bransford et al., 2000; Hassard, 2005), then our results suggest that all students would benefit from being taught to self-explain during and after problem solving. By doing so, students should be better able to apply their full range of competencies to any problem at hand—even to problems that bear no resemblance to problems they have previously encountered.

New Constraints on the General Question of How and Why Self-Explanation Works

Most prior work on the effectiveness of self-explanation has focused on verbal (e.g., Chi, 2000; Chi et al., 1994; Coleman et al., 1997; Kintsch, 1994) or quantitative tasks (e.g., Chi et al., 1989; Didierjean & Cauzinille-Marmèche, 1997). Together with the early study of Gagne and Smith (1962) and Alevén’s work on geometry (summarized in Alevén & Koedinger, 2002), our work shows that eliciting self-explanations can also increase performance in spatial thinking. As discussed in detail in a recent report by the National Research Council Committee on Support for Thinking Spatially (2006), spatial thinking is critical for a wide range of tasks encountered in both daily life (e.g., assembling a piece of furniture) and professions (e.g., remodeling a building or designing a garment).

Most prior work on self-explanation in the absence of feedback has examined spoken statements. An exception is the work by Hausman and Chi (2002), in which students typed their messages on a computer keyboard. Although Hausman and Chi used the same materials and tasks as in the explanation-rich study of Chi et al.

(1994), their participants generated very few self-explanations and achieved no more learning than the control group. Even when participants were explicitly prompted to explain, most statements typed were paraphrases rather than explanations. We note that in Hausman and Chi's study the presented material and the student response were identical in format and locale: words on the computer screen. This similarity may have encouraged or allowed the student to create the response from the presented material with minimal deep-level cognitive processing, that is, as a paraphrase. In our study, the presented material was nonverbal spatial information, and the response was verbal. To transform the presented spatial information into the required verbal response could not be done without significant processing of the information. It may be that requiring a transformation of format of the material (e.g., written verbal to oral verbal; spatial to written verbal) fosters the type of metacognition that leads to improved performance.

Previous studies documenting a self-explanation effect have usually involved junior high, high school, or college students. Our finding of a self-explanation effect among fourth graders (9 and 10 years old) raises the question of how young the self-explanation effect emerges. It seems plausible that the self-explanation effect documented among adolescents and adults is a continuation of the effect of "private speech," which has been shown to help 3- to 10-year-olds guide themselves through challenging tasks (Berk, 1994; Berk & Garvin, 1984; Bivens & Berk, 1990).

Most prior work on the effectiveness of self-explanation has focused on self-explanation that occurs while the student is studying or practicing (e.g., Bielaczyc et al., 1995; Chi et al., 1989; Chi et al., 1994; Coleman et al., 1997; Gagne & Smith, 1962; Lin & Lehman, 1999). The students' mastery of the material was then tested during a posttest phase when self-explanation was not monitored. Our study differs in that we examined self-explanation during problem solving, rather than during studying or practicing. We found a self-explanation effect when students were solving an unfamiliar problem in the absence of external feedback about either the quality of their answers or their explanations. The fact that self-explanation seems to help in solving novel problems without explicit instruction is a promising finding for a world in which novel problems abound and fostering young people's ability to innovate is seen as a national imperative (National Research Council Committee on Prospering in the Global Economy of the 21st Century, 2006).

Conclusions

We have found that requiring children to explain their reasoning as they complete a field-based maps skills task improved their overall performance and decreased the frequency of errors grounded in failure to understand the representational correspondence between map and referent space. We infer that the process of generating explanations caused children to detect discrepancies between a location in the environment and the proposed corresponding location on the map; they then self-corrected these incipient errors by drawing on their underlying spatial competencies.

These findings accord well with prior research showing that eliciting self-generated explanations from students can improve performance on tasks as diverse as reading a text about Darwinian evolution (Coleman et al., 1997), factoring algebra equations (Didierjean & Cauzinille-Marmèche, 1997), or designing an ecological experiment (Lin & Lehman, 1999). Our findings add to the literature on the self-explanation effect by showing such an effect on a task that is spatial rather than verbal or quantitative, with participants who are elementary-school age rather than adolescent or adult, and in a situation where the students self-explain while problem solving rather than while studying or practicing.

There is growing recognition of the need to foster spatial thinking among K–12 students (Liben, 2006; National Research Council Committee on Support for Thinking Spatially, 2006), in light of the importance of spatial thinking in science, engineering, and everyday life. Yet there is no consensus on how this fostering can be accomplished. Our findings suggest that one useful step would be for curriculum designers and teachers to structure opportunities for self-explanation into geography lessons and other spatial tasks, just as math curriculum materials now routinely ask students to “show your work.”

ACKNOWLEDGMENTS

This work was funded through National Science Foundation Grants ESI-01-01806 and ESI-01-01758 to Kastens and Liben, respectively; however, all opinions, findings, and conclusions or recommendations expressed are ours and do not necessarily reflect the views of the National Science Foundation. This is Lamont-Doherty Earth Observatory Contribution No. 6968.

We gratefully acknowledge the assistance of Jennifer Griffith, Daniel Heuer, Toru Ishikawa, Lauren Myers, Linda Pistolesi, Elisabeth Sultzbaugh, Margaret Turrin, and Brian Verdine in acquiring, coding, and analyzing the data reported in this article. We appreciate the insightful participation of Mrs. Davies, Mrs. Hankinson, Mrs. Kanthal, and the fourth graders of Valley Cottage Elementary School, Nyack School District, New York. We thank Director P. Michael Purdy for permission to use the grounds of the Lamont-Doherty Earth Observatory.

REFERENCES

- Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26, 147–179.
- Berk, L. A., & Garvin, R. A. (1984). Development of private speech among low-income Appalachian children. *Developmental Psychology*, 20, 271–286.
- Berk, L. E. (1994, November). Why children talk to themselves. *Scientific American*, 271, 78–83.
- Bielaczyc, K., Pirolli, P., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13, 221–252.

- Bivens, J. A., & Berk, L. A. (1990). A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly*, *36*, 443–463.
- Blades, M., & Spencer, C. (1994). The development of children's ability to use spatial representations. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 25, pp. 157–199). San Diego, CA: Academic.
- Bransford, J., Brown, A. I., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self explanations. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 251–282). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.
- Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.
- Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences*, *6*, 347–365.
- DeLeeuw, N., & Chi, M. T. H. (2003). Self-explanation: Enriching a situation model or repairing a domain model? In G. Sinatra & P. Pintrich (Eds.), *International conceptual change* (pp. 55–78). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*, 1556–1557.
- Didierjean, A., & Cauzinille-Marmèche, E. (1997). Eliciting self explanations improves problem solving: What processes are involved? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *16*, 325–351.
- Donovan, M. S., & Bransford, J. D. (Eds.). (2005). *How students learn: Science in the classroom*. Washington, DC: National Research Council.
- Downs, R. M., & Liben, L. S. (1990). Getting a bearing on maps: The role of projective spatial concepts in map understanding by children. *Children's Environments Quarterly*, *7*, 15–25.
- Gagne, R. M., & Smith, E. C., Jr. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, *63*, 12–18.
- Hassard, J. (2005). *The art of teaching science*. Oxford, England: Oxford University Press.
- Hausmann, R. G. M., & Chi, M. T. H. (2002). Can computer interface support self-explaining? *Cognitive Technology*, *7*, 4–14.
- Kastens, K. A., Kaplan, D., & Christie-Blick, K. (2001). Development and evaluation of "Where Are We?" map-skills software and curriculum. *Journal of Geoscience Education*, *49*, 249–266.
- Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist*, *49*, 294–303.
- Liben, L. S. (2006). Education for spatial thinking. In W. Damon & R. Lerner (Series Eds.) & K. A. Renninger & I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 197–247). Hoboken, NJ: Wiley.
- Liben, L. S., & Downs, R. M. (1986). *Children's production and comprehension of maps: Increasing graphic literacy: Final report to the National Institute of Education*. University Park: The Pennsylvania State University, Department of Psychology.
- Liben, L. S., & Downs, R. M. (1989). Understanding maps as symbols: The development of map concepts in children. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 22, pp. 145–201). New York: Academic.
- Liben, L. S., & Downs, R. M. (1993). Understanding person-space-map relations: Cartographic and developmental perspectives. *Developmental Psychology*, *29*, 739–752.

- Liben, L. S., & Golbeck, S. L. (1980). Sex differences in performance on Piagetian spatial tasks: Difference in competence or performance? *Child Development, 51*, 594–597.
- Liben, L. S., Kastens, K. A., & Stevenson, L. M. (2002). Real world knowledge through real-world maps: A developmental guide for navigating the educational terrain. *Developmental Review, 22*, 267–322.
- Liben, L. S., & Yekel, C. A. (1996). Preschoolers' understanding of plan and oblique maps: The role of geometric and representational correspondence. *Child Development, 67*, 2780–2796.
- Lin, X. D., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*, 837–858.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.
- Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self explanations. *Cognition and Instruction, 16*, 173–199.
- National Research Council Committee on Prospering in the Global Economy of the 21st Century. (2006). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academy Press.
- National Research Council Committee on Support for Thinking Spatially. (2006). *Learning to think spatially*. Washington, DC: National Academy Press.
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.
- Pick, H. L., Heinrichs, M. R., Montello, D. R., Smith, K., Sullivan, C. N., & Thompson, W. B. (1995). Topographic map reading. In P. A. Hancock, J. M. Flach, J. Caird, & K. J. Vicente (Eds.), *Local applications of the ecological approach to human-machine systems* (pp. 255–284). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12*, 235–275.
- Reimann, P. (1992). Modeling active, hypothesis-driven learning from examples. In E. DeCorte, M. Linn, H. Mandl, & L. Verschffel (Eds.), *Computer-based learning environments and problem solving* (pp. 129–148). Berlin, Germany: Springer-Verlag.
- Reimann, P., & Neubert, C. (2000). The role of self-explanation in learning to use a spreadsheet through examples. *Journal of Computer Assisted Learning, 16*, 316.
- Siegel, A. W. (1981). The externalization of cognitive maps by children and adults: In search of ways to ask better questions. In L. S. Liben, A. H. Patterson, & N. Newcombe (Eds.), *Spatial representation and behavior across the life span: Theory and application* (pp. 167–194). New York: Academic.
- Thurstone, T. G. (1962). *Primary mental abilities*. Chicago: Science Research Associates.
- University of the State of New York, State Education Department. (2004, March 10). NY State School Report Card: Comprehensive Information Report. Retrieved June 8, 2005, from <http://www.emsc.nysed.gov/reprcd2000/500304030005.pdf>
- VanLehn, K., & Jones, R. M. (1993). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polsen (Ed.), *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1034–1039). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences, 2*, 1–59.
- White, B. Y., & Fredrickson, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Science, 16*, 90–91.
- Witkin, H., Oltman, P., Raskin, E., & Karp, S. (1971). *A manual for the Embedded Figures Tests*. Palo Alto, CA: Consulting Psychologists Press.