DEVELOPMENT AND EVALUATION OF "*WHERE ARE WE?*" MAP-SKILLS SOFTWARE AND CURRICULUM

Kim A. Kastens — Lamont-Doherty Observatory and Department of Earth & Environmental Sciences, Columbia University, 109F Oceanography, Palisades, NY 10964, kastens@ldeo.columbia.edu
Danielle Kaplan — Teachers College, 525 W. 120th St., New York, NY 10027
Kottie Christie-Blick — Cottage Lane School, South Orangetown Central School District, New York

ABSTRACT

The *Where are We?* software and lessons are designed to help children in grades two through four learn to "translate" between the visually-perceived world that they sense around them, and the schematic representation of that landscape on a map. Formative evaluation of a prototype version of *Where are We?* resulted in the following improvements in the instructional materials: more and prompter feedback for students, additional assessment tools for teachers, development of lessons to model successful map-using strategies, development of lessons to overcome common misconceptions, and replacement of text-based instructions with a voiceover demo. To evaluate whether the skills practiced in *Where are We?* transfer to a real-world setting, we have developed field-based tests of the ability to absorb information in the real world and transfer it onto a ma and, conversely, the ability to absorb information from a map and transfer it into an action in the real world. In our reality-to-map test, students place colored stickers on a map to indicate the location of similarly colored flags in the real world; in the map-to-reality test, students place large, colored, numbered disks on the ground to indicate the location of similarly-colored, numbered stickers on a map. Average performance for a class of 24 fourth graders improved on both the map-to-reality and reality-to-map tests after using *Where are We?* Some children classified with learning disabilities performed extremely well both with the software and on the field-tests, supporting the idea that spatial skills are not closely connected to the verbal skills usually emphasized by school tasks. The persistence of certain kinds of misconceptions on the post-test, and the uneven improvement of subscores representing different aspects of map-to-reality correspondence, suggest directions for improvement in *Where are We?*

Keywords: Education — Computer-assisted; Education — precollege; Education — geoscience; Education — Testing and Evaluation

INTRODUCTION

Becoming a skilled map user is not an easy cognitive task. When children look around them, they see a world which 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 four feet or so above the ground. To use a map, 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 (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 and Downs (1989) review the research literature on acquisition of map literacy. They show that many young children show competence at simple map tasks, such as finding their way around a floor-plan of a single room containing a few items. However, if children are asked to perform tasks that are more akin to those facing adults in either practical or professional map-using situations, the children's performance is much less impressive. For example, only 20% of first graders and

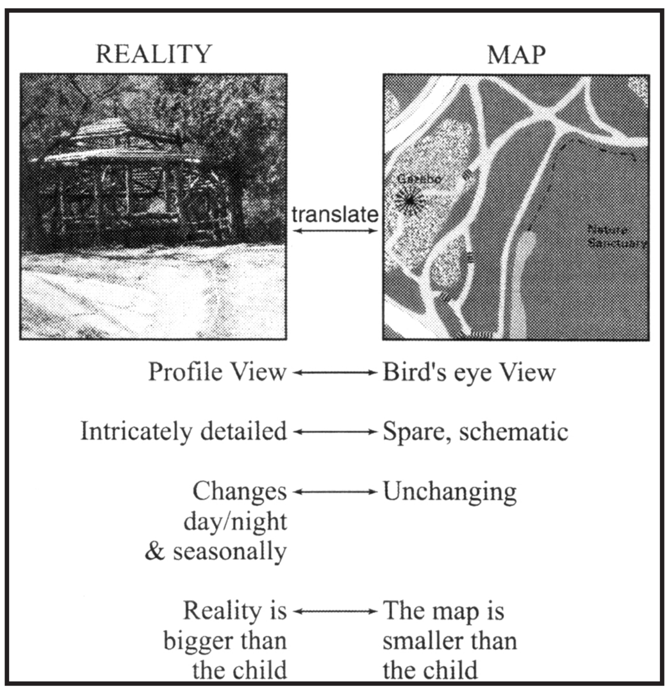

Figure 1. Becoming a skilled map user requires mastering the skill of *translating* back and forth between reality, as it is perceived by the human senses, and a representation of that reality, i.e. a map. This is a multifaceted cognitive task. Reality is seen in profile-view, is intricately-detailed, changes over time, and surrounds the child, whereas the map shows a bird's eye view, omits many details in favor of a schematic representation, doesn't change through time, and is small relative to the child.

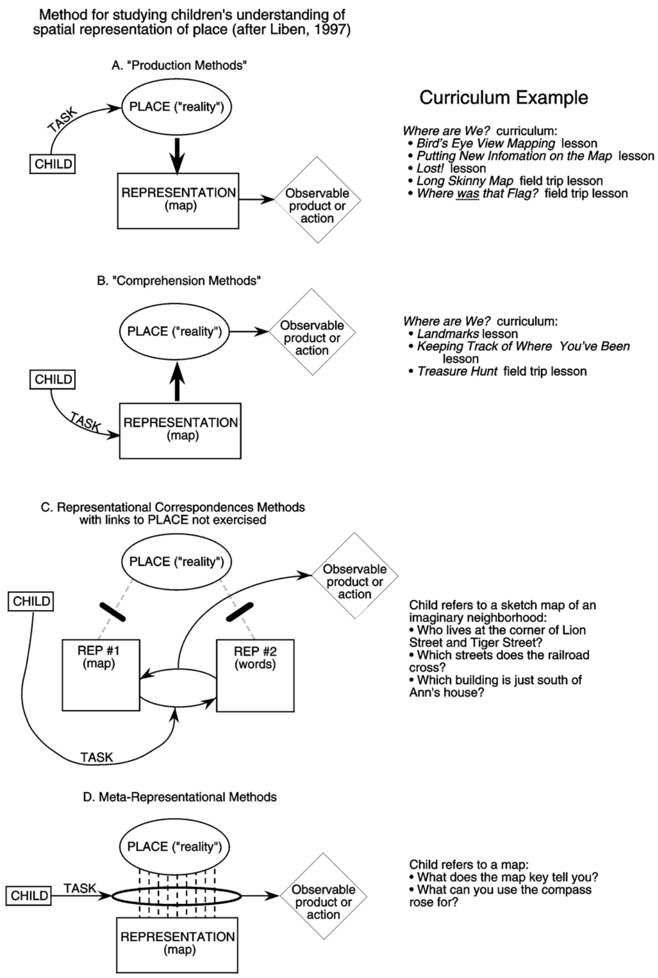

Figure 2. Liben's (1997) classification scheme for methods of studying children's understanding of spatial representations of place, annotated with curriculum examples of each classification. We think that most existing elementary map skills curricula overemphasize tasks in category C, in which children need not "translate" between the place and the representation (i.e. map). *Where are We?* is designed to exercise and strengthen the child's ability to translate information from reality to the map (Category A tasks), and from the map to reality (category B) tasks.

29% of second graders could correctly identify the U-shaped plan view of their school building, when asked to select the birds-eye view of their school from among six choices (Downs & Liben, 1990). When asked to place colored stickers on a map in locations which corresponded to colored flags on a 3-D 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 and 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 most children, like walking or talking. Instead, it is a complex ability that needs to be taught and practiced if it is to be fully developed.

In our examination of existing map skills workbooks and curriculum, we found a strong emphasis on skills that could be called "map analysis", answering questions about the map. There is little emphasis on the skill of relating the symbols on the map, and the spatial relationships among those symbols, to a realistic visually-perceived landscape. Liben (1997) has developed a classification scheme for the methods that cognitive/development researchers have used to study children's understanding of spatial representations of place (Figure 2). Adopting Liben's terminology, we would say that most existing elementary map skills curricula have a shortage of "production methods" of assessment (Figure 2A), in which children first experience a real space, and then translate some aspect of that experience onto a spatial representation of the space (i.e. a map), and demonstrate their understanding by doing something on or to or with a map. Similarly, most existing map curricula lack "comprehension methods" of assessment (Figure 2 B), in which children are first given a map, and then asked to translate some aspect of the mapped information to the corresponding real space, demonstrating their understanding by performing some action in the real space.

These two types of translation, from real space to map, and from map to real space, are the skills most needed by a person using a map as a tool for navigation, or as a tool for recording spatial information about an environment. Instead, existing map skills curricula emphasize "representational correspondence methods" of assessment, in which children translate from one representation of the place to another representation, most commonly from a map to words. In many cases, such tasks can be completed without referring to the link between either representation and the real place (Figure 2C). As Liben (1997) points out, such tasks are of limited utility for testing children's competence in understanding place representations. Existing map skills curricula also feature "meta-representational methods of assessment" (Figure 2D), in which the child is asked to reflect on the relationship between the place and the map, and articulate their conscious understanding of that relationship. Such tasks can play a valuable role in acquiring facility with maps, but they are no substitute for tasks in which children are explicitly required to translate from map to reality, and from reality to map.

We have developed a teaching and learning tool, called *Where are We?*, which helps children learn the skill of "translating" between the environment they see around them and a map of that environment. In this paper, we will first describe *Where are We?*, then detail the process of formative evaluation through which we developed and improved it, and finally describe the results of a set of field-based tests by which we have evaluated the effectiveness of *Where are We?*

DESCRIPTION OF *WHERE ARE WE?*

Where are We? (*WAW?*) comprises a software application, a Teachers' Guide including classroom and field-based lessons, a poster-sized paper version of the map and map key, and a figurine scaled to the size of the poster-map.

The Software - The *Where are We?* software displays a map of a park on one side of the screen, and live-action video, filmed within the park, on the other side (Figure 3). Students can "move" though the scene by clicking "turn left", "turn right" or "move forward" buttons at each path intersection. The video responds appropriately. The software provides four modes, each emphasizing a different skill or situation:

- In *Exploring the Park*, a red dot and arrow on the map continuously indicate the user's position and view direction, moving and rotating to track the route followed. The red dot plays the role of the finger of a parent or other skilled map-using mentor, showing a child her location on a map and helping her track her route during a walk.
- In *Are We There Yet?*, the student picks a destination and then must find his way to the destination based on visual information within the video, keeping track of where he has been and where he is going. In this mode, the red dot and arrow are available to indicate position and direction only if the student clicks a 'hint' button. *Are We There Yet?* simulates the most common real-world map task: using a map to find one's way from a known starting point to a desired destination.
- In *Add to the Map*, the student finds objects in the video that are not on the map (such as lampposts and fire hydrants), figures out where these objects should be located on the map, and adds this information to her map by dragging appropriate symbols onto the map. Symbols can only be added to the map at points where the object actually exists in the real world. *Add to the Map* introduces students to the concept of the map as a tool for organizing spatial information, as contrasted with the map as a tool for personal navigation. *Add to the Map* simulates a landscape architect's mapping task, and is analogous to field-mapping tasks undertaken by geologists, hydrologists, ecologists, and other field scientists.
- In *Lost!*, the computer places the student in an undisclosed location on the map. He must figure out where he is on the map by comparing what he sees in the video with the symbols on the map. *Lost!* simulates the real-world situation in which walkers or motorists realize that they are lost, pull the map out of the backpack or glove compartment, and try to figure out where they are on the map, in order to find their way home or to a desired destination.

The Classroom/Computer Lab Lessons - The lesson plans in the *WAW?* Teacher's Guide use teacher modeling, student problem-solving and practice of successful map-using strategies, both in the classroom/computer lab and in the field. Classroom/computer lessons include:

- *Exploring Map*s: Students examine a variety of paper maps and establish a common understanding of what a map is and what it is used for.
- *Bird's-Eye View Mapping*: Students draw a simple paper map of the objects on their desks and use that map to convey information to a classmate. This lesson introduces the idea of a map as a planview representation, and a map as a tool for conveying information.
- *Map Symbols*: Students use the key on the *WAW?* poster map to identify objects on the map, and to imagine what would be seen by the *WAW?* figurine if she were standing at a particular location on the map.
- *Landmarks*: Through guided use of *Are We There Yet?* mode, students discover the characteristics that make specific landmarks useful (permanent, distinctive) or not useful (mobile, changeable, overly common) for personal navigation.
- *Keeping Track of Where You've Been*: Using *Are We There Yet?* mode, the teacher models, and students practice, keeping continual track of their position on the map.
- *Planning a Route*: Students plan a route to a destination and anticipate what they will see along that route. Using *Exploring the Park* mode, they test their predictions and verify their plan.
- *Map Scale*: By contrasting the rate of motion of the ground in the video and the rate of advance of the red dot across the map, students gain an intuitive appreciation of the contrast in size (i.e. scale) between the map and the represented landscape. Then, they use the map scale to estimate sizes and distances in the *Where are We?* landscape.
- *The Compass Rose*: In *Are We There Yet?* mode, students practice using a compass rose to figure out what direction they are facing or moving relative to a map, using *WAW?*'s hint button to check their work. The teacher's attention is drawn to two common sources of confusion: north as a region versus north as a direction, and right/left versus north/south/east/west.
- *Putting New Information on the Map*: Using *Add to the Map* mode, students use a map to compile and convey information about the spatial distribution of features (fire hydrants, water fountains, etc.) in the *WAW?* park.
- *Lost!*: In *Lost!* mode, students use their knowledge of map symbols, landmarks, and compass directions to obtain clues about their location by observing the landscape around them; then they combine multiple observations to infer their location on the map.
- *Summing Up: Comparing Maps with the Real World*: Students problem-solve on their own, and then brainstorm with others as they articulate their understanding of the similarities and differences between a map and the space represented by the map.

Field-Based Lessons - In the field-based lessons, students create, interpret, or modify a paper map in a real-world environment. These lessons cement and reward the students' growing map skills and help them build confidence in their ability to use these skills. The field-based lessons can also be used for assessment.

- *Long Skinny Map*: Children take a short walk along a straight path, observing the location of objects and their physical relationship to each other. They then create

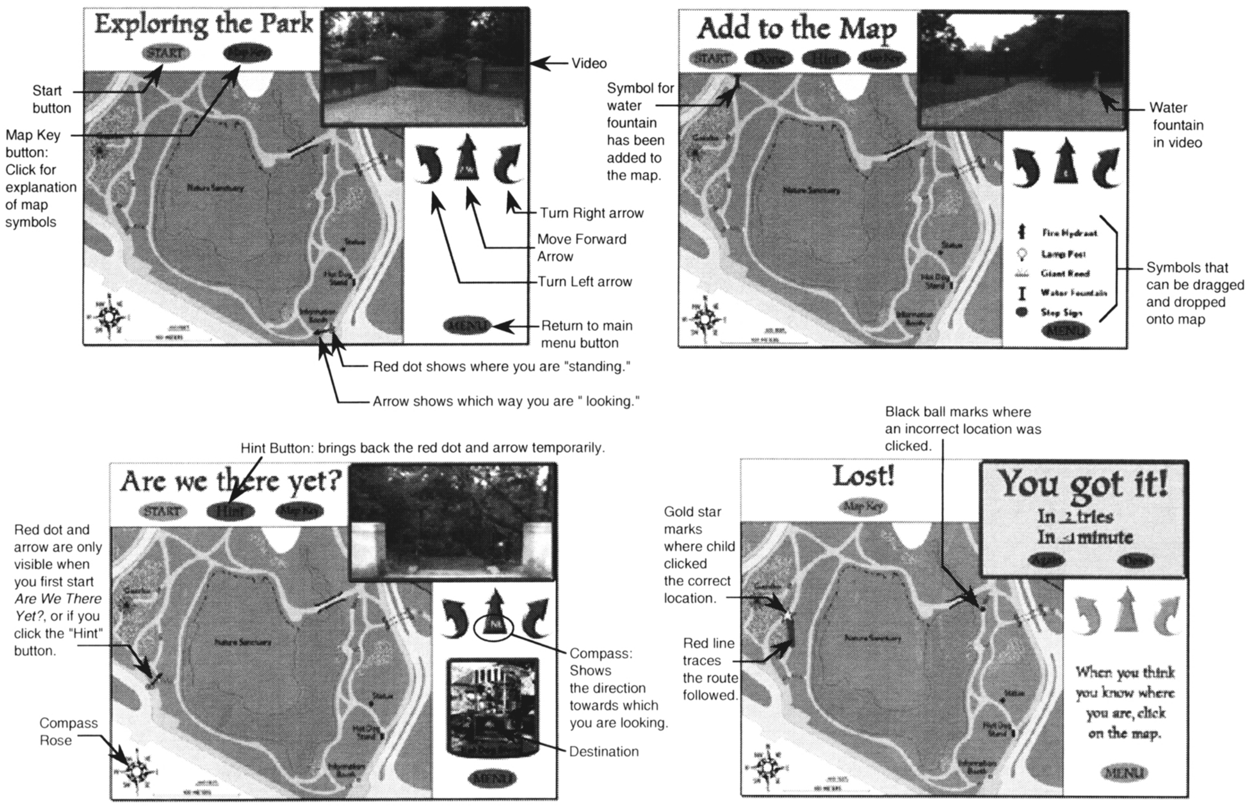

Figure 3. Screen shots of the four modes of use of *Where are We? Exploring the Park* familiarizes students with the software interface, and helps develop a sense of map scale. *Are We There Yet?* simulates the real-world task of navigating to a desired destination from a known starting point. *Add to the Map* introduces the map as a tool for organizing spatial information. *Lost*! challenges the student to combine all previously-learned map skills, including use of the compass rose, map scale, symbols and landmarks.

symbols for the objects, with a key, in order to transfer what they see in the real world to paper in the form of a one-dimensional map.

- *Map to the Real World: A Treasure Hunt*: Children are given a paper map with a starting point and a destination ("treasure") marked on it. Their task is to find their way to the location of the treasure.
- *Real World to Map: Where Was that Flag?*: Children find colored flags placed throughout the field area and place a matching sticker on a paper map at the location they think corresponds to the real-world position of each flag.
- *Integrating Map Skills with other Field Trips*: Children work with a map of a site which will be visited for some other purpose, and/or a map of the route from the school to the site. They analyze the map before the trip and anticipate what they will see, use a map to navigate at, or *en route* to, the site, and use a map to communicate the highlights of their trip to another person.

FORMATIVE EVALUATION

In the context of instructional materials development, "formative" evaluation (Stevens et al., 1993) is a methodical investigation of the usability, acceptability, and effectiveness of what has been developed, carried out during the development cycle, with the goal of improving the instructional materials. We carried out a formative evaluation of *Where are We?* involving six teachers and six classes at three schools: an urban private school, an urban public school, and a suburban public school, all in the New York metropolitan area. The students spanned *WAW?*'s target age range from second through fourth grades. Most of the formative evaluation was carried out on a prototype of *Where are We?*, the version described in Kastens et al. (1996).

In-school techniques used during the formative evaluation included observations in the classroom and computer room, taped and transcribed interviews with teachers, debriefing discussions with classes of students and teacher, and videotaping of selected computer room sessions.

In addition to the in-school observations, we also observed the students in four of our classes using paper maps in real-world settings. Two second-grade classes completed a "treasure-hunting" task in Central Park, in the same area where the video in the software had been filmed (Figure 4). One class of fourth graders visited the campus of Lamont-Doherty Earth Observatory, and completed a task in which they placed colored stickers on a map to indicate the position of colored flags (additional information about this task below). With another class of fourth graders, we piggy-backed a map-using activity onto an environmental education nature walk. During the nature walk, one of us (DK) circulated among the children asking them where they thought they were, and recording their positions and answers. For the second graders in Central Park, and the fourth graders at Lamont-Doherty, a map-literate adult shadowed each pair of children, keeping a running track chart of the route followed by the children, and recording their comments and actions.

In general, the reaction of formative evaluation participants to *Where are We?* was very positive. After the first session, the children approached subsequent *WAW?* lessons with enthusiasm; they scrambled for access to the computers; they maintained a high level of interest throughout each session; they began looking at maps in their free time; and they pleaded for copies of the software to take home with them. The teachers also reacted positively, citing *WAW?*'s obvious appeal to the children, its link to the geography standards, and its applicability to practical real-life situations.

A common thread throughout the teachers' comments was that the students who excelled with *Where are We?* were not necessarily the same students who typically score well on paper and pencil tasks. Four of the six teachers interviewed spontaneously revealed that some of their lowest-performing children had been among the quickest to catch on to how to use *Where are We?*, had demonstrated exceptional work during field evaluations, or had taken an unprecedented leadership role when paired with a normally high-performing child.

Our observations and discussions revealed a number of ways *Where are We?* could be improved. Based on our formative evaluation, we made the following changes:

(1) *Lowered the literacy requirement*: The prototype software had instructions to the user as text on the screen. The vocabulary and sentence structure were beyond the reading level of many children in the target age range, and most teachers and students simply ignored the instructions. We replaced the text instructions with an animated demonstration of each mode, accompanied by voice-over instructions in a child's voice.

(2) *Reinforced successful performances*: In *Add to the Map* and *Lost!* modes, the program now replays the route followed by the students as a bold red line, after they succeed in their objective. Watching their route progress across the map stimulates the students to think about and articulate what they did and why they did it: "Oh, that's where we weren't sure which way to turn, and I said we should turn left to go towards the lake, and it worked." Seeing , analyzing and verbalizing reinforces their successful map-reading strategies.

(3) *Corrected mistakes immediately*: In the prototype software, a student could place an *Add to the Map* symbol at any point on the map. In the current version, the symbol will only "stick" on the map if it is placed at a location where a corresponding real object actually exists in the

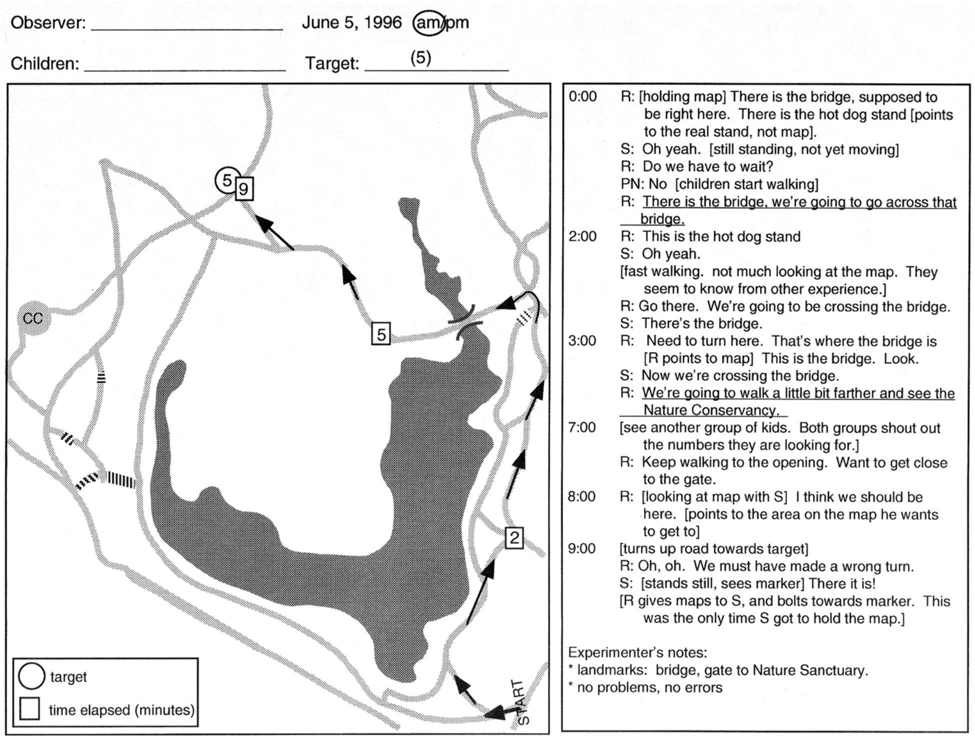

Figure 4. Example of data collected during the formative evaluation stage of *Where are We?* development. Working in the field area where *Where are We?* was filmed, pairs of children used a paper map to find a "hidden treasure." Each pair of children was trailed by an adult, who marked the path followed by the children on a map (left) and recorded the children's conversation (right). The numbers indicate time elapsed since leaving 'START.' 'R:' and 'S:' indicate which child was speaking. This pair of children demonstrated two powerful map-using strategies: focusing on good landmarks (the bridge, the hotdog stand, the Nature Center), and anticipating what they would see (see underlined sections of narrative.)

real world and can be seen in the video. Our original idea was that students would compare maps with each other and discover discrepancies, thus learning, among other things, that a map can look complete and official and yet contain errors. But we found that allowing students to place symbols anywhere on the map led to confusion and positive reinforcement for students who did not, in fact, know where they were. One of our second-grade teachers pointed out that, "Children this age need to learn to trust maps, before they learn to doubt them." This modification encourages the teacher to focus on the power of maps rather than on their limitations.

(4) *Reduced reliance on domain-specific knowledge*: In the prototype software, the user began each mode at the same entrance to the park, facing the same direction. As users became familiar with that corner of the park, they began to rely on their visual memory to find their way around, rather than on the map. The current version of the software has six different starting position/azimuth combinations, which launch users towards different areas of the map.

(5) *Increased "non-aligned" map situations*: In the prototype software, the user always started at the southwest corner of the map. Because almost all the features of interest were north or northeast of the starting point, students spent most of their time moving northwards, eastwards or northeastwards and little time travelling southwards. Our observations showed that when they did travel southwards, they tended to get much more confused than when travelling northwards. This observation is in accord with a wealth of psychological data (e.g. Liben & Downs, 1993; Breuer & Marzolf, 1999; Levine et al., 1984) showing that spatial relationship tasks involving a representation of a physical environment are much easier if the map/aerial photo/diagram is aligned with the referent space than if it is unaligned. For a map which follows the convention of north at the top, the easier "aligned" condition occurs when the map-user is facing towards the north. When facing north, a turn towards the map-user's left hand results in a turn towards the left side of the map. But when facing south, a turn towards the map-user's left results, confusingly, in a turn towards the right side of the map. In the current version of *Where Are We?*, the student begins near the north end of the map about half of the time, and thus is forced to travel southwards to get to the destination.

(6) *Provided assessment tools for teachers*: We added the capability of producing hard-copy artifacts, from which the teacher can assess the students' progress. In *Are We There Yet*?, students can print out a map showing the route they followed from starting point to destination, annotated with their names and number of hints used. Such maps tell the teacher the number of destinations the children reached, how heavily they relied on the "hint" button, and how convoluted or direct was their route to the destination. In *Add to the Map*, students can print out their map with added symbols superimposed, again annotated with the names and number of hints used. In *Lost!*, students can print out a "Navigator's Certificate", documenting from which spots the students have succeeded in finding themselves, plus the number of tries and the amount of time that was required to find themselves. The students seem to take great pride in these professional-looking maps and official-looking certificates.

(7) *Developed lessons to model successful strategies*: During our field-based map activities, we noted wide variation in map-reading performance, in accordance with work by previous investigators on individual differences on spatial skills tasks (Liben and Downs, 1991; 1993). Children skilled in map-reading used specific strategies that less successful map-users did not (Figure 4). Pairs of second graders who succeeded efficiently on the treasure hunting activity tended to discuss objects that were useful landmarks: the lake, the steps, an intersection. In contrast, the narratives for less competent pairs indicate attention to things that were temporary (e.g. blooming flowers), or moveable (e.g., squirrels), or non-distinctive (e.g. a park bench). In addition, two pairs of children demonstrated an ability to anticipate what they would see farther down the path, and plan accordingly (Figure 4). We built a lesson around each of these two strategies ("Landmarks" and "Planning your Route"), so that teachers can model, and all children can practice, the strategies that had proven so successful for the few children who had figured them out spontaneously.

(8) *Developed lessons to overcome common misconceptions*: We observed two widespread misconceptions. First, students and some teachers had difficulty understanding that north, south, east and west can refer to a direction in which a person or thing faces, points, or moves, and that the same words can also refer to a location or region. This confusion manifests itself when children read "SE" (for example) on the view direction compass and point determinedly to the lower right hand quadrant of the map when asked where they are, regardless of their actual position. Although the compass is referring to southeast as a direction; the student understands southeast as a region or location. The second confusion is between relative direction and absolute direction, i.e. right/left versus N/S/E/W. This confusion manifests itself when children interpret the "turn right" button as a "move east" button. The revised Teachers' Guide describes symptoms by which these misconceptions can be recognized, and suggests teaching strategies and student activities to overcome the misconceptions.

One modification repeatedly requested by participants in the formative evaluation was to be able to rotate the map on the screen. Although we appreciated the pedagogical validity of this request, we were not technically able to accomplish this change within the constraint imposed by school-caliber hardware.

SUMMATIVE EVALUATION

Goal - "Summative" evaluation (Stevens et al., 1993) differs from "formative" evaluation, in that the goal is to ascertain and document what children are actually learning by using the educational tool in question. For our

	Map to World (Task is posed on the map; student answers in the visually-perceived environment)	World to Map (Task is posed in the visually-perceived environment; student answers on the map)
Computer	*Are We There Yet?*	*Add to Map; Lost!*
Field Test	Place numbered marker in the real world at a position corresponding to similarly-numbered sticker on the paper map.	Place colored stickers on the paper map at positions corresponding to similarly colored flags in the real world

Table 1. "Translation" in both directions.

summative evaluation, we developed a field-based test of map skills. We wished to avoid the possibility that students might be merely learning to beat the computer game rather than mastering the target skills. In addition, the literature on spatial cognition suggests that moving through a large-scale space that surrounds the individual contributes to the development of spatial knowledge (Cohen and Cohen, 1985) in ways that differ from merely observing that space. Thus we wished to be sure that the skills developed indoors, using *Where are We?*, would transfer into a large-scale, real-world setting.

Just as *Where are We?* is designed to exercise students' ability to go from the visually-perceived real world to the map and vice versa, we wished to assess children's translation skills in both directions: their ability to absorb and process a query presented on a map and produce an appropriate response in the real world, and conversely, their ability to absorb and process a query presented in the real world and provide an answer on the map (Table 1). We wanted to use a reasonably complex map of a reasonably complex terrain, a first step towards a field geologist's basemap, rather than a simplistic schematic of a classroom. And we wished to do this in a way that was quantitative and reproducible.

The Subjects - The subjects were 24 nine- and ten-year-olds in a fourth grade class at a suburban public school. They took our field-based test of map skills in October (the "pretest"), worked with the *Where are We?* curriculum throughout the school year interspersed with their other geography lessons, and then repeated the field-based test in June (the "post-test"). All students were present for both pretest and post-test. The class was approximately evenly split by gender (13 M, 11 F). This class contained six children classified with learning disabilities. All children had normal color vision, and none had any mobility limitations that could have interfered with their ability to traverse the field area. Although most of the children were high-spirited during the trip and obviously enjoyed the tasks, one child, student #3, broke down in tears during the post-test and was barely able to complete the task. This child was described by her teacher as "emotionally-troubled", and results from student #3 have been separated from the rest of the class in some of the discussion below.

In addition to the pretest/post-test group, 132 children visiting the Lamont-Doherty Open House performed the flag-sticker test (see below). Of these, 56 reported that they had no adult assistance. This group, which ranged in age from 5 to 15, provides some context within which to interpret the results from the pretest/post-test group.

The Tests

Field area: The field area for both pretest and post-test was the northern half of the campus of Lamont-Doherty Earth Observatory. A former estate, the visually-diverse grounds include paths and roads; a parking area; an equestrian statue; buildings of various sizes and geometries, including a greenhouse and a mansion; grassy areas, an orchard, isolated trees, and massed shrubbery; and a walled garden with pond.

Map: The map (Figure 5) was an architect's rendering, shown in true plan view, with no obliquely-viewed elements. The map was colored, with blue for the pond, dark green for trees and shrubbery, light green for grass, black for buildings, grey for roads, and light grey for paths. We added a map key, north arrow, and scale bar to the map. The maps given to the children were color-photocopied onto 8 1/2" x 11" paper, at a scale close to 1:1000. Unfortunately, between our pretest and post-test, a building was torn down and a replacement building begun in the southeastern corner of the field area. On the post-test map, we indicated the construction area with brown diagonal hatching and the words "Construction Area", and its surrounding fence with a brown line. To keep the area and the length of road that the students could explore approximately the same size in the post-test as in the pretest, we moved the northern boundary of the field area northwards, and extended the map to the north as well. The pretest map and post-test map were printed at exactly the same scale.

Reality to Map test: The inspiration for our reality-to-map test was a test developed by Liben and Downs (1986; 1989), involving a topographic map, a tabletop 3-D model, colored flags, and colored stickers. In our open-air variant of their test, we placed eight brightly-colored flags around the field area. We gave each student a map, and a set of eight colored stickers. The students were asked to place each sticker on their maps, at the point that they thought coincided with the position of the similarly-colored flag. The flag positions were selected to span a range from very easy to quite difficult, as judged by our own map-using experience, and by considering the recorded comments and performance of pairs of students

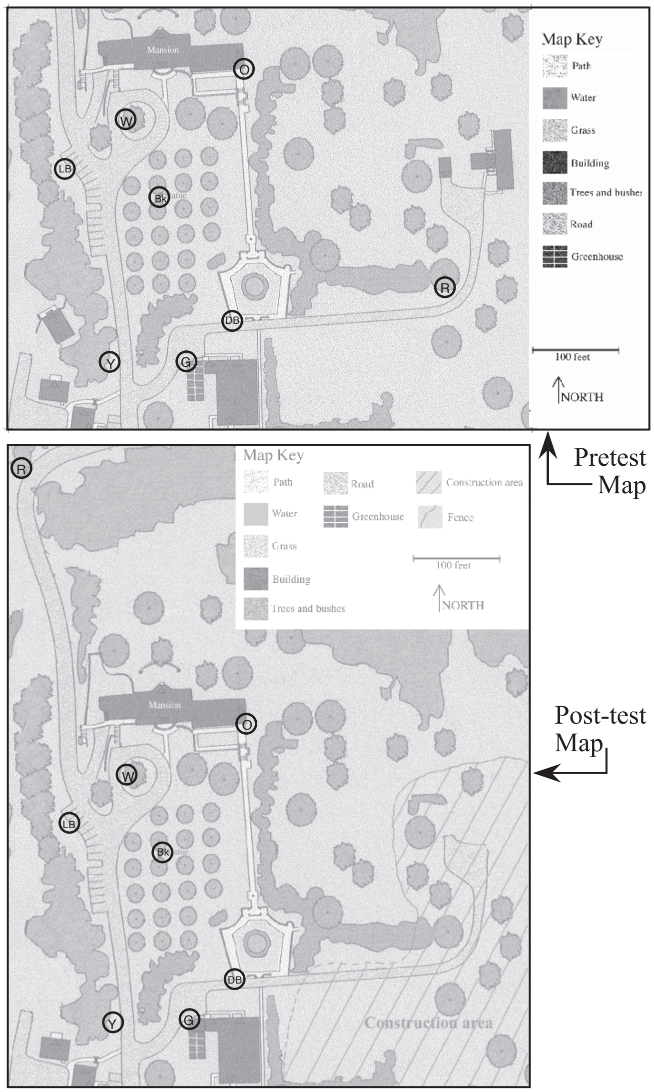

Figure 5. Map used for the field-tests of map skills. The students had colored photocopies of the maps. The circled letters indicate the true location of each flag on the flag-sticker test: O: orange; W: white; LB: light blue; Bk: black; DB: dark blue; G: green; Y: yellow, and R: red. The locations of the flags were chosen to present a range of difficulties, and to include both "bounded" and "unbounded" conditions.

who did the flag-sticker activity during the formative evaluation stage of the project. The stickers were round, and the diameter (5/16" or approximately 8mm) corresponded to approximately 7m on the ground. Students were introduced to the activity as a group while sitting on the steps of the mansion (i.e. facing southwards). One-by-one, each student was handed a correctly-oriented map, shown where he/she was standing and which direction he/she was facing, and launched on the task.

Map to Reality test: As each student returned from the Reality-to-Map (Flag-sticker) task, we took away the flag-sticker map, and handed him/her a fresh copy of the map. These new maps already had a numbered, colored sticker or stickers on them (for the pretest there was one sticker; for the post-test, two stickers). Together with their new maps, we gave the students large markers (~25cm diameter plastic disks), colored and numbered to match the stickers on their maps. Each student's map had the sticker(s) in different place(s). The students were directed to place each large marker on the ground at the point that they thought coincided with the position of the similarly-colored, numbered sticker on their map. The students were not reshown their starting position on the map nor their starting orientation when they began the Map-to-Reality test.

Scoring strategy: We created a four point scoring rubric for each flag-sticker or marker: One point was awarded for being in the correct *region*, for example, in the orchard, or around the circular driveway. A second point was awarded for being on the correct category of *object*, for example, on a building, or a tree, or a parking lot. The third point required being on the specifically correct object: the correct building, or correct tree, or correct parking lot. The fourth point was awarded for being in exactly the correct place, within one half sticker diameter (~4mm on the map; ~3.5m on the ground) of the precisely correct location. The region, object, and place criteria form a nested set, in that the correct place is always within the correct specific object, which in turn is always within the correct region. We favor this scoring strategy over a straight measure of distance between sticker location and flag or marker location, because it can detect students' mastery of representational correspondences (Liben and Downs, 1989) in the absence of mastery of spatial correspondences (as manifested, for example, by putting a sticker or a marker on a building, but not the correct building).

Results - Reality-to-Map (Flag-sticker) Test

Scope: This report focuses on the comparison between pretest and post-test results, with an eye towards evaluating whether the intervening use of *Where are We?* helped children improve their map skills. Additional insights about the nature of the children's mistakes and misconceptions have been gleaned from both the pretest/post-test group and the larger Open House group, and will be reported separately. The pretest/post-test discussion of the flag-sticker activity compares performance on seven of the eight flags; the site of the red pretest flag was obliterated by the construction.

Statistical Significance: Aggregating scores across all 24 students and all 7 flags, the average score improved from 18.4 to 20.3, out of a maximum possible score of 28 (Table 2). To test whether this difference is statistically significant, we used a Wilcoxon signed rank test (Table 3). This is the nonparametric equivalent of the paired t-test. We used a nonparametric test because the data are not normally distributed, and because our scoring scale is an ordinal rather than an interval scale. The results are significant at the 5% level (P=0.035, or if student #3 is omitted, P = 0.016, Table 3).

	All Students	Omitting Student #3
N	24	23
Pretest mean score	18.35	18.35
Post-test mean score	20.33	20.65

Table 2. Mean Scores: Flag Sticker Task

	All Students	Omitting Student #3
#0 Differences	4	4
# Ties	5	5
Z-value	-2.109	-2.555
P-value	0.0349	0.016

Table 3a. Wilcoxon signed rank sest for pretest vs. post-test on Flag-sticker task

	Count	Sum Ranks	Mean Rank
# Ranks < 0	15	161.5	10.767
# Ranks > 0	5	48.5	9.700

Table 3b. Wilson rank info (all students).

Results by flag: The relative difficulty of the flag locations, as judged from the average score for each flag, is consistent from flag to flag among the three trials: pretest, post-test, and Open House. The flag on the one and only statue (the black flag, Figure 5) always received the highest average score (Figure 6). The flag on a corner of the mansion (orange), and the flag inside the circular driveway (white), were next easiest. These were followed by the flag on a corner of the greenhouse (green) and the flag on a corner of the wall surrounding the garden (dark blue). The two most difficult flag locations were along the side of a road away from distinctive buildings or intersections (yellow and light blue flags). Although the ordinal nature of our scoring system prevents a quantitative comparison of flag location difficulty, it is clear that we have achieved our experimental design goal of presenting flag locations that range from quite easy to quite hard, and that this range of difficulty is persistent.

Average scores on six out of seven of the flags improved between the pretest and post-test (Figure 6, Table 4). To compare the amount of improvement on the flags of varying difficulty, we use the Hake factor of Redish and Steinberg (1999), which compares actual improvement with the possible improvement:

$$\text{Hake factor} = \frac{\text{observed gain}}{\text{possible gain}} = \frac{(\text{avg post \%}) - (\text{avg pre \%})}{100\% - (\text{avg pre \%})}$$

The greatest improvement (Hake factor >20%; Table 4) was seen on the easiest flags (black and orange) and most difficult flags (light blue and yellow). For example, all the children's post-test black stickers, representing the very easy flag on the statue, were tightly clustered around the correct location (Figure 6a). A majority of the post-test children placed the light blue and yellow stickers, representing the most difficult flags, alongside undistinguished sections of road, on the correct section of road on the correct side; the pretest results showed far more outliers (Figure 6b). The least change (Hake factor < 10%) was seen on the intermediate-difficulty flags: scores for the green flag (on the corner of the building containing the greenhouse) improved slightly, while scores on the blue flag (on the corner of the wall around the garden) got slightly worse.

The yellow flag revealed a persistent mistake on the part of a substantial group of students. Five students placed their post-test yellow sticker in almost exactly the same spot they had placed their pretest yellow sticker, but that spot was substantially north of the correct location (Figure 7). We revisited this location, student maps in hand, seeking some second-order visual clue that could have led so many students to persist in their belief that the flag was so far north, but could find no such clue. Among the rest of the data set (six flags times 24 children, or 144 child-sticker placements), there was only one other occurrence when a child's erroneous pretest and post-test stickers overlapped.

Results by scoring criteria: As a group, averaged across all flags, the children improved on all the scoring criteria (Figure 8 (left), Table 5). As a percentage of possible gain (Hake factor), the largest gain came in the "region" criterion (Hake factor: 43%). The least improvement, as a percentage of possible gain, came on the "place" criterion (Hake factor: 11%). To test the possibility that manual dexterity in the ability to place a small sticker could be artificially limiting the maximum possible score on the place criterion, we asked the 132 Open House children to place an extra sticker on top of a cross marked on their copy of the map. Very few of the children misplaced the sticker onto the cross by more than a millimeter, suggesting that manual dexterity is not an important limiting factor.

The interplay between the Category and Object criteria is interesting. On the pretest, these scores differed substantially (category average score 5.13, object 4.67), reflecting numerous cases in which students placed a sticker on the correct category of object, but not on the specific correct object. Such an error implies that the student grasps representational correspondences but has a weaker command of spatial correspondences. By the post-test, the gap between category and object scores was smaller (category 5.50, object 5.29), perhaps reflecting an improvement in students' ability to work with spatial correspondences.

Results by students' attributes: The gains in performance between the pretest and post-test came at the top and bottom of the ability range (Figure 9), with the largest gains among those who had done least well on the pretest.

The children classified as having learning disabilities were split in their performance. Two students (#11 and #19 on Figure 9) were at the bottom of the class in their performance on both the pretest and post-test. But three (#8, #15, and #20) clustered near the top of the class. Five of the learning-difficulty classified children improved from pretest to post-test; one stayed the same.

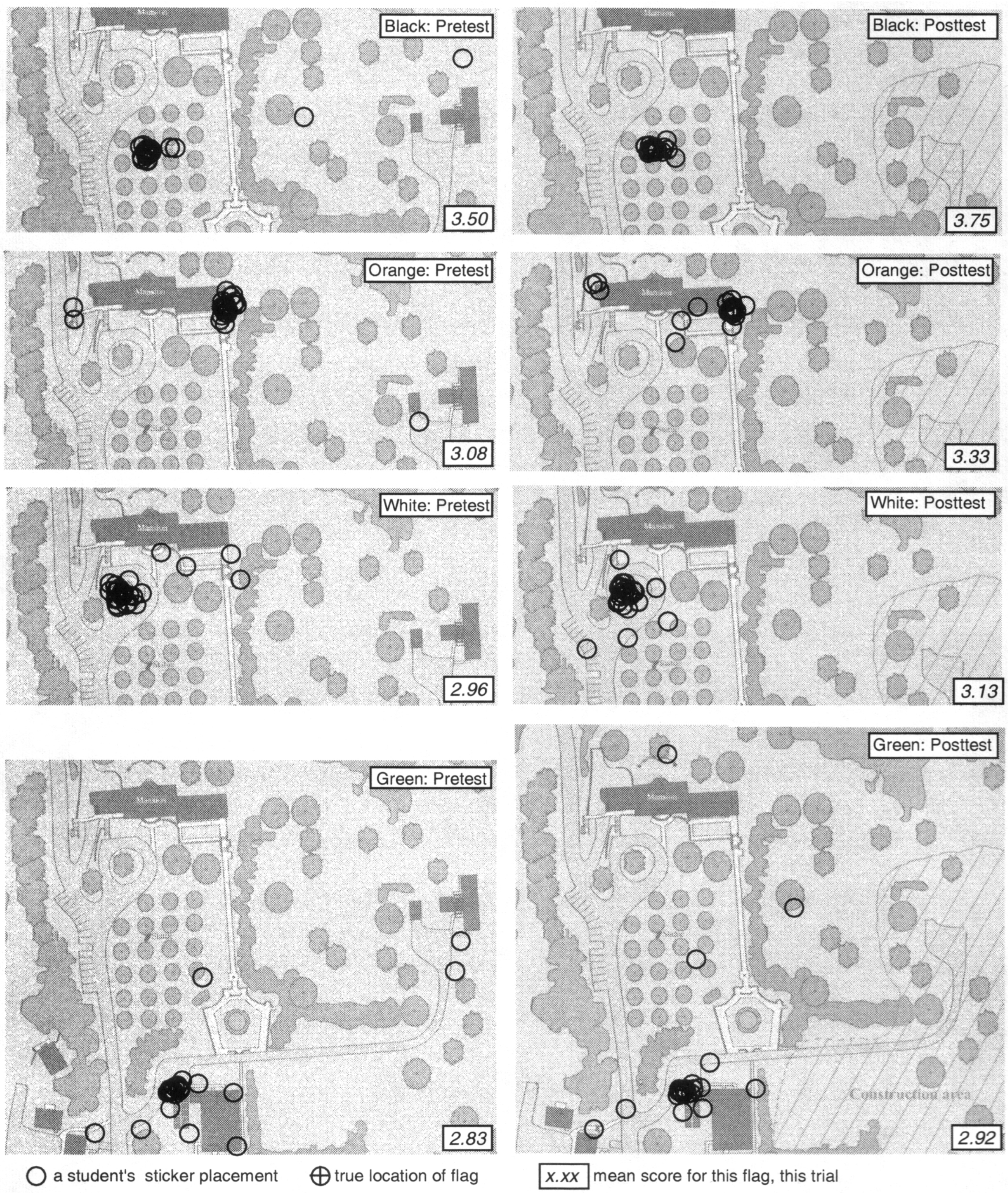

Figure 6a and b. Results from the Reality-to-Map task, by flag. The maps are arranged in pairs, with the pre-test results on the left, and the post-test results on the right. The students' sticker placements on the post-test tend to cluster more tightly around the correct flag location (circle with cross), than do the sticker placements for the same flag on the pretest. The pairs of maps are arranged from easiest-placed flags (black, orange) at the top to most difficult-placed flags at the bottom (light blue and yellow). The number in the lower right hand corner of each map shows the average of students' total score (sum of region, category, object and place scores) for that flag on that test, out of a maximum possible score of four. Note that maps are cropped to show only the area in which stickers were placed.

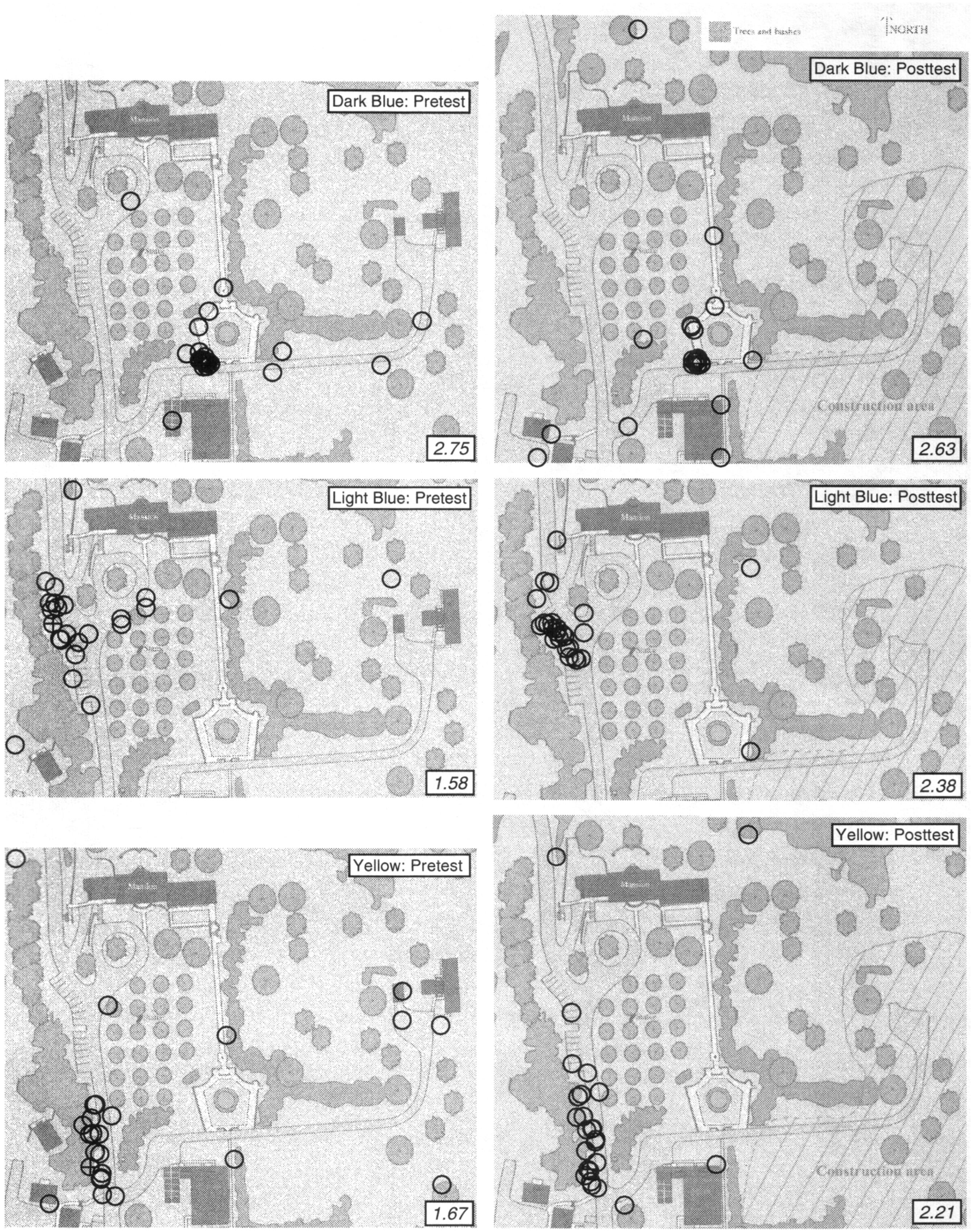

Figure 6b. Results from Reality-to-Map task, by flag.

	Pretest Mean. Score	Pretest Mean Score	Observed Gain	Possible Gain	Hake Factor*
black	3.5	3.75	6.3%	12.5%	50.0%
orange	3.08	3.33	6.3%	23.0%	27.2%
white	2.96	3.13	4.3%	26.0%	16.3%
green	2.83	2.92	2.3%	29.3%	7.7%
dark blue	2.75	2.63	-3.0%	31.3%	-9.6%
light blue	1.58	2.38	20.0%	60.5%	33.1%
yellow	1.67	2.21	13.5%	58.3%	23.2%

* Hake factor = $\frac{\text{observed gain}}{\text{possible gain}} = \frac{\text{(avg post \%) - (avg pre \%)}}{\text{100\% - (avg pre \%)}}$

Table 4. Results from Flag-Sticker test by Flag.

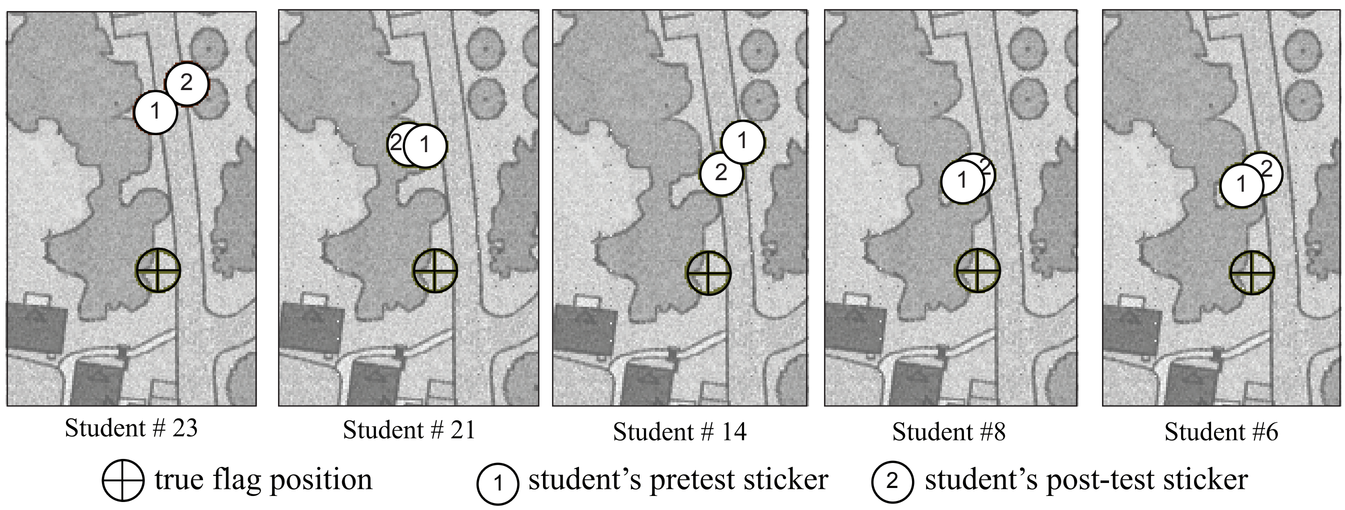

Figure 7. In general, individual students did not repeat the same mistakes on the post-test that they had made on the pretest. One exception was the yellow flag, located at a fairly nondescript position along the side of a road. Five students placed this flag well north of the correct location, making the same mistake on both pre- and post-test. Our interpretation is that students failed to utilize available information about map scale and distance, even after *WAW?*-based instruction.

On both the pretest and the post-test, boys performed better than girls (Table 6, Figure 9,10). Both boys' average and the girls' average rose from pretest to post-test. However, even after using *Where are We?*, the average girl's score was lower than the average boys' score had been before beginning the *WAW?* curriculum (Table 6, Figure 9). By the time of the post-test, nearly half (6 of 13) of the boys were performing at the ceiling level of the assessment, whereas only one girl achieved that level of performance and the remainder of the girls were widely scattered (Figure 10). Although suggestive, these male/female differences do not show up as statistically significant in nonparametric tests for similarity of distribution (Kolmogorov-Smirnov Test) or random selection from the same population (Wald-Wolfowitz Runs Test).

Results by child's age: Among the group of children who did the flag-sticker task at the Lamont Open House, and considering only those 56 children who reported that they had not been helped by an adult, age is a poor predictor of map skills performance ($r2=0.106$; Figure 11). The *WAW?*-using children (including learning disabled children) averaged well below the trend line for the Open House children (a self-selected group whose families chose to bring them to an event at a geoscience laboratory) on the pretest. But on the post-test, the *WAW?* children averaged well above the Open House trendline.

Results: Map-to-Reality (Place-marker) Test - Because individual students were assigned different marker locations on the pre- and post-test, and marker locations differ in their inherent difficulty, and the number of markers per student was small (one per student on the pretest and two per student on the post-test), the results for this test are not as robust as for the Reality-to-Map test. Figure 8 shows a comparison of pre- versus post-test results aggregated across students. As a percentage of the maximum possible

	Pretest Mean. Score	Pretest Mean Score	Observed Gain	Possible Gain	Hake Factor *
Region	5.92%	6.38%	6.6%	15.4%	43%
Category	5.13%	5.50%	5.3%	26.7%	20%
Object	4.67%	5.29%	8.9%	33.3%	27%
Place	2.67%	3.17%	7.1%	61.9%	11%
Total	13.38%	20.33%	7.0%	34.4%	20%

* Hake factor = observed gain / possible gain = (avg post %) - (avg pre %) / 100% - (avg pre %)

Table 5. Results from Flag-Sticker test by scoring criteria.

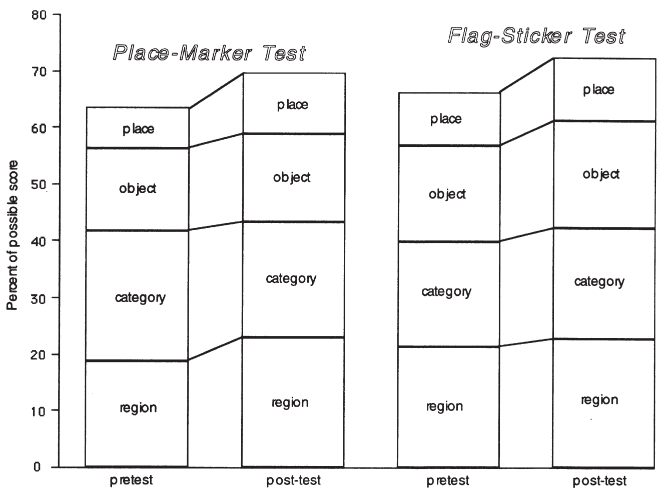

Figure 8. Comparison of results from Map-to-Reality (Place-Marker) and Reality-to-Map (Flag-Sticker) task. The total height of each bar represents the class' average performance on one task at one trial (pre- or post-test); subscores on each of the four scoring criteria (region, category, object and place) are summed together to get the total score. To permit comparison between the 1-marker task of the pre-test and the 2-marker task of the post-test, and between the flag-sticker and place-marker tasks, results are shown here as a percent of the maximum possible score. As a whole, the class improved approximately the same amount on the place-the-marker task as on the flag-sticker task. On the flag-sticker task, subscores improved for all scoring categories; on the place-marker task, "category" subscores did not improve.

score, the group improved approximately the same amount on the place-marker (map-to-reality) test as on the flag-sticker (reality-to-marker) test. However, this improvement does not register as significant on the Wilcoxon signed rank test. This may be because there were so few possible scores: integer values from 0-4 on the pretest, and from 0-8 on the post-test. In future tests, we will increase the number of markers per subject to four. With our four point scoring rubric, this will give a range of possible scores from 0-16, which should provide more sensitivity in the ranking technique of the non-parametric statistics. In any case, the improvement shown in Figure 8 is encouraging, if not conclusive.

DISCUSSION

On the Effectiveness of *Where are We?* - Although this represents only a first attempt at summative evaluation, with only one class of students, the evidence suggests that use of *Where are We?* has indeed helped children strengthen their ability to carry out real-world map tasks. Scores improved between pre- and post-test on both directions of translation (map-to-reality and reality-to-map). On the reality-to-map test (flag-sticker), where the test resolution permits some disaggregation of the data, we observed improvement between pre- and post-test for boys and for girls, for children with and without learning disabilities, for six out of seven flag locations, and for all scoring criteria.

On the Inherent Relative Difficulty of Map Skills Tasks- An interesting aspect of the flag sticker results is that the relative order of difficulty of the flags is so consistent from group to group. The relative easiness of the flag locations on the statue and circular driveway (black and white), and the relative difficulty of the flags on the side of the main road (light blue and yellow), are consistent with other research (Liben & Downs, 1986) showing that both adults and children can more accurately place themselves on a map in a "bounded" position (i.e. a location adjacent to a conspicuous landmark) than they can in an "unbounded" position. However, the bounded versus unbounded explanation does not account for the relative difficulty of the blue and green flags relative to the orange flag. All are on the corners of structures, and the structures occupied by the green flag (on a U-shaped building with one greenhouse wing) and the blue flag (on the pentagonal wall surrounding the formal garden and pond) are not obviously less distinctive than the mansion on which the orange flag sits. We suggest that there is a dynamic, time-sensitive, element to student's performance, in that they do better on the flag that is adjacent to their starting position (the orange flag), before they have had time to lose track of the

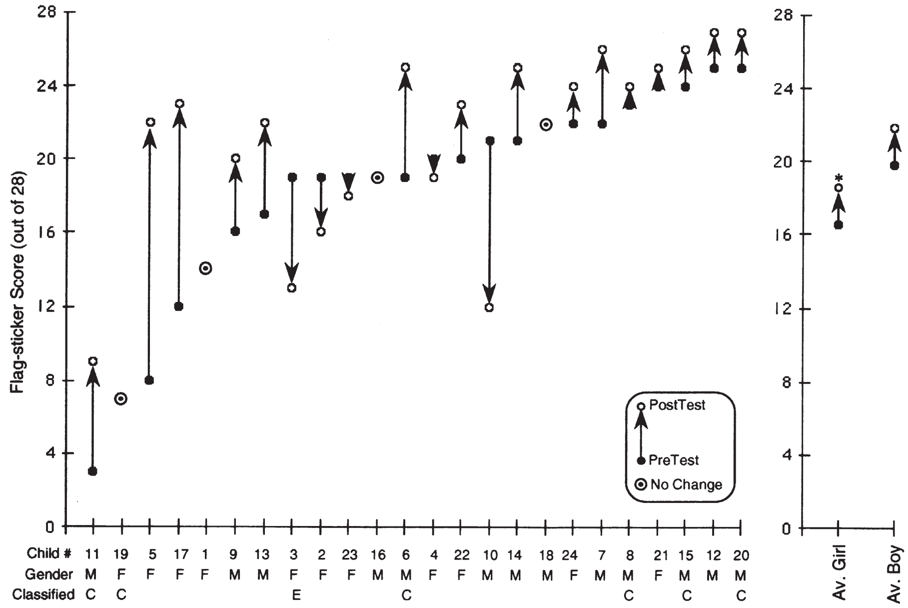

Figure 9. Results from Reality-to-Map (Flag-Sticker) task by student attribute. The vertical axis is individual student's score on the flag-sticker task, with closed circles indicating score on the pre-test and open circles indicating score on the post-test. The codes across the bottom indicate male/female, and students who are classified with learning difficulties ("C") or emotional difficulties ("E"). The student are ordered by score on the pre-test. Note that three out of the top five scores on the pre-test were obtained by students classified as having learning disabilities. On average, boys scored higher than girls on both the pretest and the posttest.

	All Girls	Girls omitting student #3	Boys
N	**11**	**10**	**13**
Pretest Mean (SD) *(out of 28 possible)*	16.73 (5.64)	16.50 (5.89)	19.77 (5.78)
Post-test Mean (SD) *(out of 28 possible)*	18.55 (5.61)	19.10 (5.59)	21.85 (5.67)
Post-test minus Pretest	1.82	2.60	2.08

Table 6. Average Results on flag-sticker task by gender.

initial position and orientation with which they were launched from the steps of the mansion.

On the Persistent Misplacement of the Yellow Flag — A Scale Problem? - Five children placed their pre- and post-test yellow stickers in the same, but erroneous, position (Figure 7). Three strategies were possible to determine the correct N/S position of the yellow flag: first, one could notice that the flag was exactly aligned with the northern edge of the building containing the greenhouse; second, one could visually estimate the distance (~30 feet) from the flag to the small path south of the flag and use the map scale; third, one could visually estimate the relative distance of the flag between the small path to the south and the parking lot to the north (about a quarter of the way from the path to the edge of the parking lot). None of these strategies was taught in the *WAW?* curriculum. Our tentative explanation is that these children saw the yellow flag as "between the path and the parking lot" on both the pre- and post-test, and stuck the sticker mid-way between the two landmarks without fully considering either the relative or the absolute distances between the features. This interpretation is consistent with Gerber and Kwan's (1994) finding that long roads with a limited number of intersections caused confusion in a wayfinding task for students with a "neophytic restricted approach" to wayfinding, an approach characterized by (among other things) inability to use the map scale.

Insights for Software and Curriculum Development - The relatively less impressive percentage gain on the "place" scoring criteria suggests that the skills tested by this criteria are an area where future improvements could be sought. The *WAW?* activity most analogous to the precise placement in the flag-sticker task is *Add to the Map*, with its requirement for figuring out where you are on the map, and then deriving and marking the position of a seen object. Since this test group used *WAW?*, we have modified the software to make it less easy to guess the correct object location, which we think will increase the effectiveness of this mode. The persistence of an apparent scale error on the children's placement of the yellow flag suggests that we should build into the curriculum more opportunities to translate between sizes and distances measured on a map and sizes and distances perceived visually in the real world.

About Learning Disabilities and MapSkills - During our formative evaluation, several teachers spontaneously

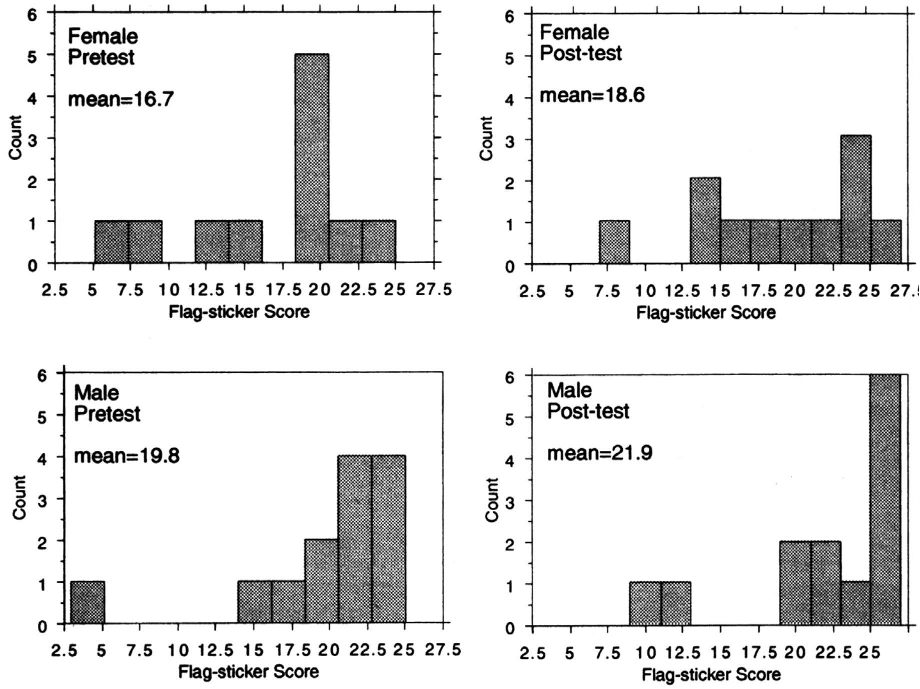

Figure 10. Histograms of students' scores on the Reality-to-Map (Flag-sticker) task, grouped by gender. On both pretest and posttest, most boys performed near the ceiling of the test, with a sparsely-populated tail of low-performers. Girls' scores had a lower mean and were more scattered.

pointed out students (all boys) who were doing exceptionally well with *Where are We?*, and commented that these were individuals who typically don't do well in school. Then, during our field tests, three of the top six scorers on both the pre- and post-test flag-sticker task, were students classified with learning disabilities. These observations are consistent with the theory of multiple intelligences (Gardner, 1983), which postulates a "spatial intelligence" independent from the "verbal intelligence" rewarded by most school tasks. Both sets of observations suggest that *Where are We?* and other map skills activities can be of value in providing an opportunity to excel for a subset of children with poor verbal but strong spatial skills.

About Gender and Mapskills - Boys outperform girls in many (but not all) tasks related to either spatial relationships or to geography. Males outperformed females on the geography component of the International Assessment of Educational Progress (Lazer, 1992), the National Assessment of Educational Progress (especially items dealing with space and place; Persky et al., 1996), and the National Geography Bee (Liben et al., 1995; Liben, 1996). Males of all ages outperform their females peers on tests requiring mental rotation and spatial perception (Linn and Petersen, 1985; Newcombe, 1982). Boys outscored girls at every age from 7 to 18 on the "Road Map Test of Direction Sense," in which participants were asked whether turns along a route marked on a city map were towards the left or right (Harris 1981). Male college students' spatial knowledge of their campus after 3 weeks in residency, and male 4-to-6 year olds' spatial knowledge of their classroom, were more accurate than those of their female classmates (Harris, 1981). An exception to this generalization is that females often outperform males on tasks involving memory for the location of objects (Dabbs et al., 1997). Although our male/female results do not register as statistically significant, the marked qualitative differences in the histograms between the boys and girls in our study (Figure 10), with most boys' scores clustering near the ceiling of the test while the girls' scores scatter, suggest that we may be recording a gender effect as well. Without entering into the controversy about the cause of this gender gap, we recommend that schools seek out and adopt curriculum materials that permit girls (as well as boys) to develop their spatial abilities. *Where are We?* shows promising signs of being such a tool.

On the Sparse Improvement Over Age - The results from the Open House attendees (Figure 11) suggest that mastery of the map skills targeted by *Where are We?* does not develop spontaneously with maturation, at least not in all children. Instruction is needed, at least for some children. Almost all of the Open House participants answered affirmatively the question "Have you studied maps in school?", and it seems that this classwork is also not effectively teaching certain real-world map-using skills.

Additional Work Planned - During 2001-2003, we will be extending the field-based map skills investigation to a larger group of subjects, and following up on several questions, including: To what extent do map skills, as measured by our field-based tests, improve with maturation, or with experience on the tests, in the absence of *Where are We?* instruction? Is there really a significant gen-

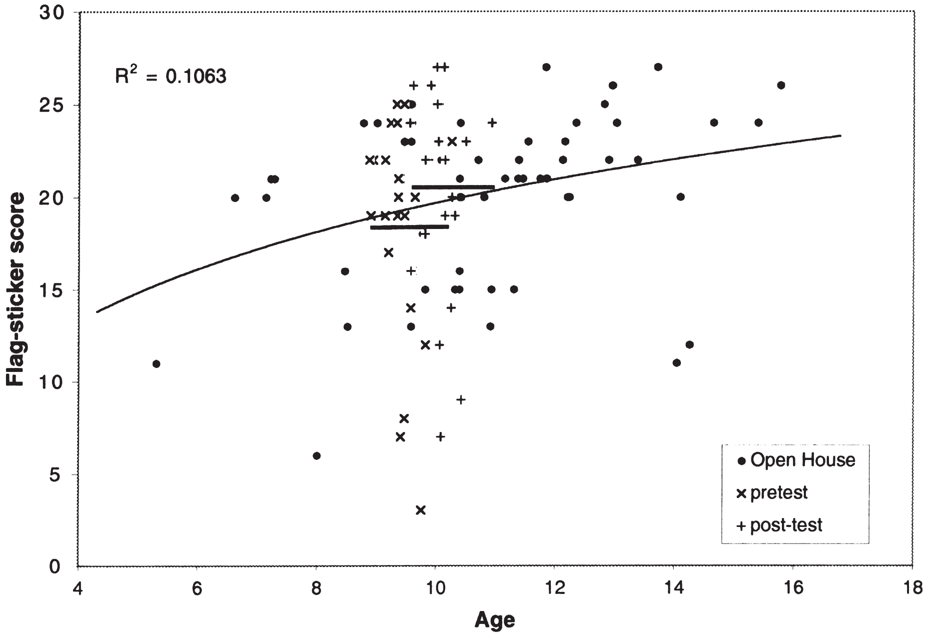

Figure 11. The flag-sticker test was also administered as a volunteer "map skills challenge" to children attending the Lamont Open House. The black dots and trendline show the scores of the Open House attendees as a function of age. Age is a poor predictor of performance (R^2 = 0.1063). In other words, maturation alone will not necessarily lead to mastery of the tested skill. Results from the pretest/post-test group are superimposed for comparison (x = pretest; + = post-test; horizontal lines = average score for each trial.)

der difference in performance on our field-based tests, and/or on the effectiveness of *WAW*? Do reality-to-map translation skill and map-to-reality translation skill occur consistently in the same individuals? Which of the artifacts produced during classroom and computer-room use of *WAW?* are good predictors of student performance on real-world map skills tasks? How do the teachers' experiences with, and attitudes towards, maps impact the way they present this material to students? Additional questions for future study include: how does *WAW?*'s technology-assisted approach compare in effectiveness with orienteering training? What kinds of professional development are most useful to help teachers help children learn map skills and related spatial skills?

THOUGHTS FOR COLLEGE TEACHERS

Geoscientists will recognize the skills of *WAW?*'s Add to the Map mode, and the Flag-sticker field-based test, as a precursor skill to geological field mapping, in which one places symbols or colors onto a basemap to indicate the real-world location of geological phenomena of interest. Similarly, the skills of *WAW?*'s *Are We There Yet*? mode, and the Place-marker field-based test, are skills needed to carry out a systematic sampling scheme in nature, in which one must return repeatedly and efficiently to a predetermined map location in order to take a sample or measurement of an environmental process or phenomena. The overarching learning objective of *WAW?*, the ability to make a facile and accurate mental translation between a terrain and its mapped representation, is needed every time a geoscientist uses a published map to learn about a field area he or she has not visited.

Our work shows that these abilities and skills are unevenly distributed. This is keeping with previous work showing extremely wide range of individual performance in mapping and spatial tasks, to the extent that among children of the same age in the same class in the same school, some children will get every item correct on a test of map skills while other children get no items correct (Liben and Downs, 1986). These skills improve only slightly throughout the middle school and high school years (Figure 11), under the combined influence of maturation, experience, and social studies education. Lynn Liben has adopted a variant of our map-to-reality and reality-to-map field-based tests for college students, and reports a similarly wide range of performance. Downs and Liben (1991) report that many adults have difficulty on spatial tasks designed for children.

More optimistically, our work and the related literature also show that these field-skills can be improved by teaching and by practice. Some suggestions for those who involve college or secondary students in field-based activities that involve map-using, way-finding, and moving through space:

- Don't assume that all your students already know how to use a map effectively. Don't assume that your students' spatial abilities are similar to your own; as someone who self-selected a career in geosciences, you are likely to be towards the talented tail of the distribution. "What is logical, reasonable, and self-evident from the perspective of an expert geographer may not match the psychological realities of a [college] student" (Downs and Liben, 1991, p. 304).

- Model the technique of aligning the map with the terrain. Of the cognitive skills involved in map-using, the skill of "perspective-taking," or mentally visualizing what something would look like if seen from another viewpoint, is among the last to develop (Piaget and Inhelder, 1956; Downs and Liben, 1990) and remains problematic for many adults (Downs and Liben, 1991).
- Consciously identify and articulate the factors that contribute to your own "feel for the terrain" in your field area. Gerber and Kwan (1994) have shown that knowledge of what they call "spatial structure" can promote efficient wayfinding strategies. Examples of articulating "spatial structure" would be "notice that the ground generally slopes down towards the east, towards the river," or "which way do the ridges and valleys trend around here?" Try to convey both the narrow message of the spatial structure of today's specific field area, and the general message that terrains have a fabric, a structure, which is useful for orientation and may contain clues about formative processes.
- When giving directions for wayfinding, vary the type of information you provide. There is some evidence that, on average, women favor directions given in terms of landmarks, whereas men favor directions given in terms of Euclidean descriptors such as mileage and compass directions (Choi and Silverman, 1997; Dabbs et al., 1997). Both strategies have value; stretch your students' ability to use both.
- Demonstrate, and require your students to demonstrate, correspondences between aspects of the map and aspects of the terrain, as you stand within the terrain. The easiest kind of correspondence is the representational correspondence (Liben and Downs, 1989), i.e. blue stands for water; black squares stand for buildings. The correspondences that remain difficult for older children and adults are "geometric correspondences" (Liben and Downs, 1993, 1989, in press), such as correspondences that involve the shape of mapped objects or the relative location of multiple mapped objects. An example of a question that requires understanding of both geometrical and representational correspondence is: "there are several ponds shown on the map; which one are we standing by?"
- Encourage your students to "eyeball" distances, heights, and slopes of the terrain, and then measure them on the map or with an instrument. This should help develop their ability to mentally visualize a terrain they have not visited by examining a map of that terrain. In the development of spatial abilities, "Euclidean" concepts such as those involving scale and angle measured with respect to a specified frame of reference are among the last to develop (Liben and Downs, 1986) and remain problematic even for college students and adults (Merriwether and Liben, 1997).

SUMMARY & CONCLUSIONS

- We have developed *Where are We?*, a multimedia software application and associated curriculum designed to help children learn to use maps.
- We conducted a formative evaluation of *Where are We?* including classroom and computer-room observation and observation of student performance in authentic, field-based map-using situations.
- As a consequence of formative evaluation, we have improved *Where are We?* by making the software easier to use for people with low literacy level, by providing more feedback to both students and teachers, and by developing lessons to model successful map-using strategies and to overcome common misconceptions.
- We have developed field-based procedures for testing a subject's ability to translate from reality-to-map (the flag-sticker test) and from map-to-reality (the place-marker test).
- We have used these field-based tests to evaluate whether use of *Where are We?* improved the map skills of one class of 24 fourth graders, and to explore the improvement of map skills with age in the absence of instruction.
- Student performance improved on both the reality-to-map and map-to-reality tests. The map-to-reality improvement was statistically significant at the 5% level according to the Wilcoxon signed rank test. The reality-to-map improvement did not register as statistically significant, perhaps because of lack of sensitivity in the way the test was administered and scored.
- Three students classified as learning disabled scored among the highest scores in the class, on both the pretest and post-test, suggesting `that the spatial ability required to excel at map-using is not tightly coupled with the verbal abilities required by most school tasks.
- Over the age span from six to sixteen, age is a poor predictor of skill on our Reality-to-Map (Flag-sticker) test ($R^2 = 0.11$). It seems that maturation alone does not foster mastery of these skills in all children; targeted instruction is needed.
- The skills and understanding targeted by *Where are We?* and our field-based tests are precursor skills for higher education or professional work in geosciences. We have extrapolated from our own work and the literature to offer suggestions for fostering these skills in college students during field-based activities.

ACKNOWLEDGEMENTS

The development and evaluation of *Where are We?* was funded by an Oracle Media Objects Challenge Grant and by grant ESI-96-17852 from the Instructional Materials Development program of the National Science Foundation. We thank the teachers and students who participated in the school-based testing of *Where are We?*: Karen Katz, Melissa Dollinger and their students at the Dalton School; Jen Goodwin, Nace Jones, Wendy Smith and their students at the Manhattan School for Children; and the students of co-author Kottie Christie-Blick at the Cottage Lane School. We thank the many volunteers who helped with the administration of the field-based components of the formative and summative evaluations: Doug Brusa, Martha Bryan, Dale Chayes, Christa Farmer, Bruce Huber, Zvi Karcz, Jean Kennedy, Marianne Malinconico, Julie Nichols, Linda Pistolesi, James Rubenstone, Debbie

Shepard, Linda Sohl, Margie Turrin, David van Esselstyn, Michael West, Stewart Wills. Linda Pistolesi and Pamela Gourley prepared some of the illustrations. We thank Michelle Hall-Wallace, Joseph Kerski and an anonymous reviewer for insightful critiques of the manuscript. This is Lamont-Doherty Earth Observatory contribution number 6143.

REFERENCES

Breuer, S., and Marzolf, D. P., 1999, Spatial encoding in young children's use of scale models: Biennial meeting of the Society for Research in Child Development, Albuquerque, NM, presented paper.

Choi, J. and Silverman, I., 1997, Sex dimorphism in spatial behaviors: applications to route finding: Evolution and Cognition,v. 2,p. 165-171.

Cohen, S. L. and R. Cohen, 1985, The role of activity in spatial cognition. The Development of Spatial Cognition: R. Cohen. Hillsdale, NJ, Lawrence Erlbaum Associates, p. 199-223.

Dabbs, J. M., Chang, E. L., Strong R. A., and Milun, R., 1997, Spatial Ability, Navigation Strategy, and Geographic Knowledge among Men and Women: Evolution and Human Behavior, v. 19, p. 89-98.

Downs, R. M. and Liben, L. S., 1990, Getting a bearing on maps: The role of projective spatial concepts in map understanding by children: Children's Environment Quarterly, v. 7, p. 15-25.

Downs, R. M. and Liben, L. S., 1991, The development of expertise in geography: A cognitive-developmental approach to geographic education: Annals of the Association of American Geography, v. 8, p. 304-327.

Gardner, H., 1983, Frames of Mind: The Theory of Multiple Intelligences: New York, Basic Books, 440 p.

Gerber, R. and Kwan, T., 1994, A phenomonographical approach to the study of pre-adolescents' use of maps in a wayfinding exercise in a suburban environment: Journal of Environmental Psychology, v. 14, p. 265-280.

Harris, L. J., 1981, Sex-related variations in spatial skills: Spatial Representation and Behavior across the Life Span: Theory and Application. L. S. Liben, A. H. Patterson and N. Newcombe (editors), New York, NY, Academic Press, p. 83-125.

Kastens, K. A., VanEsselstyn, D., and McClintock, R. O., 1996, An Interactive Multimedia tool for helping students "translate" from maps to reality and vice versa: Journal of Geoscience Education, v. 44, p. 529-534.

Lazer, S., 1992, Learning about the World: Educational Testing Service, Princeton NJ, Center for the Assessment of Educational Progress.

Levine, M., Marchon, I., and Hanley, G., 1984, The placement and misplacement of You-Are-Here maps: Environment and Behavior, v. 16, p.139-158.

Liben, L., 1996, Psychology meets Geography: Newsletter of the American Psychological Association Science Directorate.

Liben, L. and R. M. Downs, 2001, Geography for Young Children: Maps as Tools for Learning Environments: Psychological Perspectives on Early Childhood Education, S. L. Golbeck. Mahwah, NJ, Lawrence Erlbaum Associates (in press).

Liben, L. S., 1997, Children's Understanding of Spatial Representations, of Place: Mapping the Methodological Landscape, *in* N. I.. Foreman and R. Gillet, editors, Handbook of Spatial Research Paradigms and Methodologies, Vol I: Spatial Cognition in the Child and Adult: East Sussex, United Kingdom, The Psychology Press, p. 41-82.

Liben, L. S.and Downs, R.M., 1986, Children's production and comprehension of maps: Increasing graphic literacy, Final Report to National Institute of Education, 84 p.

Liben, L. S. and Downs, R M., 1989, Understanding maps as symbols: The development of map concepts in children: Advances in Child Development and Behavior, H. W. Reese, Academic Press, v. 22, p. 145-201.

Liben, L. S. and Downs, R. M., 1991, The Role of Graphic Representations in Understanding the World, *in* R. M. Downs, L. S. Liben and D. S. Palermo: Visions of Aesthetics, the Environment and Development: The Legacy of Joachim F. Wohlwill, Hillsdale, NJ, Lawrence Erlbaum Associates, p. 139-180.

Liben, L. S. and Downs, R. M., 1993, Understanding Person-Space-Map Relations: Cartographic and Developmental Perspectives: Developmental Psychology, v. 29, p. 739-752.

Liben, L. S., Downs, R M., and Signorella, M. L., 1995, Sex Differences in adolescents' success on an academic competition in geography: Explanations and implications: Biennial meeting of the Society for Research in Child Development, Indianapolis, IN.

Linn, M. C. and Peterson, A. C., 1985, Emergence and characterization of sex differences in spatial ability: A meta-analysis: Child Development, v. 56, p. 1479-1498.

Merriwether, A. M. and Liben, L. S., 1997, Adults' failures on Euclidean and projective spatial tasks: Implications for characterizing spatial cognition: Jour. Adult Development, v. 4, p. 57-69.

Newcombe, N., 1982, Sex-related differences in spatial ability: Problems and gaps in current approaches: Spatial Abilities: Development and Physiological Foundations: New York, Academic Press, p. 223-250.

Persky, H. R., C. M. Reese, O'Sullivan, C. Y., Lazer, S., Moore, G., and Shakrani, S.., 1996, NAEP 1994 Geography Report Card: Findings from the National Assessment of Educational Progress: Washington DC, U.S. Department of Education, 79 p.

Piaget, J. and Inhelder, B., 1956, The Child's Conception of Space: London, Routledge & Kegan Paul.

Redish, E. F. and Steinberg, R. N., 1999, Teaching Physics: Figuring Out What Works, Physics Today, p. 24-30.

Stevens, F., Lawrenz, F., and Sharp, L., 1993, User Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering, and Technology Education, National Science Foundation, 104 p.