Developmental Psychology
1999, Vol. 35, No. 1, 127–145

0012-1649/99/$3.00

How Children Change Their Minds: Strategy Change Can Be Gradual or Abrupt

Martha Wagner Alibali
Carnegie Mellon University

This study investigated patterns of change in children's strategies for solving mathematical equivalence problems. The strategies children expressed in speech and in gesture were assessed both before and after an instructional intervention. In the intervention, children received either no input, accuracy feedback, or feedback plus instruction about a principle, an analogy, or a procedure. From pretest to posttest, many children changed both the variability of their strategy use and the content of their strategy repertoires. Patterns of change depended on type of instruction and on children's initial level of variability. Children who received instruction were especially likely to generate new strategies, and children with high variability were especially likely to abandon prior strategies. Gradual change was most common; however, many children modified their repertoires abruptly. Abrupt strategy change was especially prevalent among children who received procedure-based instruction and among children with low initial variability.

Developmental researchers typically seek to identify commonalities in children's behavior. The implicit goal is to identify consistent patterns that hold across children and across contexts, in an effort to characterize children's predominant way of thinking or behaving at a given age. From this perspective, the "signal" of interest in developmental data must be discerned from the "noise" caused by variability across or within children. Consequently, much empirical work about development emphasizes *consistency* in children's thinking.

Portions of the data were presented at the annual meeting of the Jean Piaget Society, Chicago, Illinois, June 1994; the biennial meeting of the Society for Research in Child Development, Indianapolis, Indiana, March 1995; and the Seventh European Conference on Developmental Psychology, Krakow, Poland, August 1995.

This research was supported by a National Science Foundation Graduate Fellowship, a William Rainey Harper Dissertation Fellowship, and a P.E.O. Scholar Award to Martha Wagner Alibali and by Grant R01 HD31185 from the National Institute of Child Health and Human Development to Susan Goldin-Meadow.

I thank Susan Goldin-Meadow for her contributions at all stages of this project. I also thank Karen Adolph, Erin Alexander, Pete Alibali, Larry Barsalou, Miriam Bassok, Jeff Bisanz, R. Breckinridge Church, Philip Garber, Lisa Haverty, Janellen Huttenlocher, Spencer Kelly, James McClelland, David McNeill, Jill Morford, Bethany Rittle-Johnson, and Robert Siegler for helpful discussions and for constructive feedback on previous versions of the article. I am also grateful to Lucia Flevares, Michael Kang, Spencer Kelly, Martha Scott, and Sharon Syc for assistance with data collection, transcription, and coding, and to Brian Junker, Aidan Palmer, Stella Salvatierra, and Zdenek Valenta for assistance with statistical analysis and interpretation. Finally, I am especially grateful to the students, teachers, and administrators at St. Ambrose, St. Ann, St. Bede, St. Gall, St. Michael, and St. Rita elementary schools in Chicago, Illinois, who made this work possible.

Correspondence concerning this article should be addressed to Martha Wagner Alibali, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to alibali@andrew.cmu.edu.

Nevertheless, many theoretical accounts of developmental change accord *variability* a central explanatory role. Indeed, a variety of theoretical formulations, including Piagetian, neo-Piagetian, information-processing, connectionist, and dynamic systems frameworks, posit variability as either a prerequisite for or an indicator of developmental change. Within some theoretical frameworks, variability is causally related to change. For example, in the Piagetian view, the impetus for change comes from disequilibrium, which can be caused by "internal conflict" between competing approaches to a problem (Piaget, 1975/1985). Within other frameworks, variability is a concomitant of change but is not the causal force driving change. For example, in the dynamic systems framework, behavioral variability is manifested at transition points (Thelen, 1989), but variability appears to be an indicator of ongoing change rather than a cause of change.

Despite its theoretical importance, until recently, surprisingly little empirical research has examined variability per se. However, in the past several years, along with the shift toward more process-oriented accounts of developmental change, researchers have begun to take variability seriously (Siegler, 1996; Thelen & Smith, 1994). When the potential importance of variability is recognized, what was previously considered noise becomes data of central interest. Accordingly, developmentalists using many different methodologies have begun to document and quantify variability in children's behavior (e.g., Acredolo & O'Connor, 1991; Acredolo, O'Connor, & Horobin, 1989; Crowley & Siegler, 1993; Siegler & Jenkins, 1989). Variability is found not only in strategic behaviors but also at many other levels, such as problem representation and synaptogenesis (see Siegler, 1988). At all levels, variability allows organisms to adapt to local environmental conditions. As such, it makes sense that children's behavior is variable.

Across several different theoretical frameworks and content domains, one consistent empirical finding has been that periods of conceptual transition are marked by high variability. Knowledge change has been associated with variability in patterns of motor performance (Adolph, 1997; Thelen, 1989; Thelen & Ulrich, 1991), variability in the methods children use to arrive at problem solutions (Strauss & Rimalt, 1974), variability across problems in children's verbally reported strategies (Siegler, 1995), and variability within individual problems in the strategies children express in speech and in the accompanying gestures (Church & Goldin-Meadow, 1986; Perry, Church, & Goldin-Meadow, 1988). Thus, there is strong empirical support for the view that variability is heightened during periods of knowledge transition.

In general, the available data suggest that conceptual change is a cyclical process that involves moving from periods of lesser to greater and then back to lesser variability. This view suggests that, at a minimum, the process of change must involve adding new structures (e.g., representations, productions, or strategies) to the repertoire, thereby increasing variability, and then removing or integrating old structures, thereby decreasing variability. For example, within a strategy choice framework, the process of change can be viewed as a process of expansion and then contraction of the strategy repertoire (see, e.g., Siegler, 1995).

Thus, there is growing evidence that there are systematic and cyclical changes in the variability of children's behavior over the course of learning. However, relatively little is known about the sources of these changes or about the overall shape of change. The present study addressed two specific questions about the process of strategy change in children learning a mathematical concept. First, what leads children to generate or abandon strategies? Two potential sources of change are considered: the child's initial level of variability and the environmental input (in this study, instruction) provided to the child. Second, is strategy change gradual or abrupt? These questions are addressed in children learning the concept of mathematical equivalence, which is the principle that the two sides of an equation represent the same quantity.

Mathematical equivalence was chosen as the content area for the present study for two main reasons. First, most children in American elementary schools do not fully understand mathematical equivalence, as shown by their incorrect solutions to problems of the form $3 + 4 + 5 = \_\_ + 5$ (Baroody & Ginsburg, 1983; Perry, 1985). However, many children can learn the concept of equivalence in a brief lesson (Perry et al., 1988; Rittle-Johnson & Alibali, in press). Thus, it seemed likely that children would change their understanding of equivalence in the course of a single experimental session. Second, previous studies have shown that children use a variety of both incorrect and correct strategies when attempting to solve equivalence problems (Alibali & Goldin-Meadow, 1993; Perry et al., 1988). For example, many children add all the numbers in the problem (an incorrect strategy), whereas many others add the numbers on the left side and subtract the number on the right (a correct strategy). For this reason, mathematical equivalence problems are an ideal task with which to address issues of strategy change.

Potential Sources of Change: Initial Variability and Instruction

One goal of the present study was to examine potential sources of strategy change. Given the work reviewed earlier, it seems likely that strategy change depends on children's initial level of variability. If strategy change involves a cyclical process of increasing and decreasing variability, then children's prior level of variability should be an important predictor of the type of strategy change they will make. Children who start with low levels of variability should generate new strategies, thereby increasing variability, and children who start with high levels of variability should abandon strategies, thereby decreasing variability. In this framework, strategy generation can be viewed as a process of variation, and strategy abandoning can be viewed as a process of selection operating on variability.

It also seems likely that the surrounding environment could induce changes in children's strategy mix. In particular, environmental contexts that involve instruction (such as school) are a highly salient source of new problem-solving strategies. Instruction may serve to introduce strategies into children's repertoires, either directly or by setting up conditions that facilitate strategy discovery. Instruction may also serve to eliminate strategies from children's repertoires, either directly or by implication. Thus, instruction could serve both to heighten and to diminish variability.

The present study examines instruction as a source of strategy change, by assessing how children modify their strategy repertoires in response to several different types of instruction. Specifically, the study examines how children change their strategies in response to (a) feedback about whether solutions are correct, (b) instruction about a specific problem-solving procedure, (c) instruction about a principle presented in verbal form, and (d) instruction about a principle presented through an analogy. These particular types of instruction were chosen on the basis of prior findings about knowledge change and instruction. First, feedback about correctness has been shown to promote strategy change (Ellis, Siegler, & Klahr, 1993). Second, analogies have been shown to be a source of new approaches to problems (Chen & Daehler, 1989; Gentner, Ratterman, Markman, & Kotovsky, 1995). Third, elementary school teachers have been shown to draw on principles, procedures, and analogies in teaching children about mathematics (Stigler, Fernandez, & Yoshida, 1996; Stigler & Stevenson, 1992). Indeed, a previous experiment about mathematical equivalence showed that elementary school teachers draw on principles, procedures, and analogies when they teach fourth-grade students about mathematical equivalence (Syc, 1992). Fourth, there is empirical evidence that principle-based, procedure-based, and analogy-based lessons are effective at teaching children to solve equivalence problems (Perry, 1991; Perry, Berch, & Singleton, 1995; Rittle-Johnson & Alibali, in press).

Different types of instruction may have different effects on levels of variability and on strategy change. For example, a combination of feedback about correctness and information about alternative strategies may promote strategy generation more than feedback alone. Similarly, instruction about a specific problem-solving strategy may lead to more change or to different types of change than instruction that sets up conditions for strategy discovery. The present study examines how different types of instruction influence strategy change.

The Shape of Change: Gradual or Abrupt?

A second goal of the present study was to examine whether strategy change tends to be gradual or abrupt. Much past developmental research rests on the assumption that change is abrupt (Flavell, 1971). However, this assumption may depend on the "grain size" at which cognitive structures are studied. In the empirical tradition following from stage theories, cognitive structures are often defined at a fairly coarse grain size, such as the stage (Kohlberg & Kramer, 1969), the central conceptual structure (Case & Okamoto, 1996), or the "theory" (Carey, 1985). When examined at this coarse grain size, it is not surprising that conceptual change appears abrupt. In contrast, recent research that uses more fine-grained constructs, such as strategies, suggests that gradual change may be the rule. Several microgenetic studies of strategy change have documented gradual change in children learning in a variety of tasks, including simple addition problems (Siegler & Jenkins, 1989), Piagetian conservation tasks (Siegler, 1995), and isolation-of-variables problems (Kuhn & Phelps, 1982).

Depending on the grain size of the analysis, change can even be construed as simultaneously both gradual and abrupt. Change that is gradual at a fine-grained level may appear abrupt when analyzed at a coarser grain. For example, a recent connectionist model of change in the balance scale task showed that gradual changes in connection weights led to periods of relative stability followed by periods of rapid change (McClelland, 1995). At a high level of abstraction, the transitions from one period of stability to another could appear to be abrupt shifts.

As noted earlier, many studies have documented gradual change at the level of the strategy. It makes sense that strategy change is gradual for concepts that children typically learn without direct instruction, such as conservation (Siegler, 1995) or scientific reasoning (Kuhn & Phelps, 1982). In such tasks, children might persist in using incorrect or inadequate strategies because they do not recognize that alternative strategies could eventually lead to consistently better performance (see also Miller & Seier, 1994). For the same reason, one might also expect gradual change in tasks in which potential strategies differ primarily in efficiency, rather than accuracy, as Siegler (1995) pointed out. However, it is not logically necessary that change is gradual at the level of the strategy. It is possible that strategy change in response to feedback or instruction could be more abrupt. Similarly, for tasks in which potential strategies differ greatly in accuracy, abrupt change might be the rule.

The present study examines the shape of strategy change for a mathematical concept, learned through instruction, in which potential strategies differ greatly in accuracy. Because strategies are a fine-grained construct, one might expect strategy change to be gradual. However, because the concept is learned through instruction, and because potential strategies differ in accuracy, it is also possible that strategy change could be abrupt.

Using Gesture and Speech to Assess Strategy Use

To accurately assess the shape of change, one must carefully measure children's strategy use. In the present study, strategy use was measured on a trial-by-trial basis, using a highly sensitive method of measurement. This method involved examining both the speech and the gestures that children produce when they explain their problem solutions. Although gestures are often redundant with speech, speakers sometimes express knowledge in their gestures that they do not express in speech, as previous research with a wide variety of tasks has shown (e.g., Crowder, 1996; McNeill, 1992). In problem explanation contexts, children's gestures sometimes "mismatch" speech, in the sense that the two modalities convey different strategies for solving the problems (Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993; Perry et al., 1988). Thus, by examining gesture as well as speech, one can obtain a more complete picture of a child's strategy repertoire.

Gestured strategies that differ from verbal ones are of special interest because they indicate variability within a single problem explanation. For example, consider a child who provides an answer of 12 for the problem $3 + 4 + 5 = \_\_ + 5$. In speech, the child might explain that she added the numbers before the equal sign to get the solution. However, in the accompanying gesture, she might indicate all four of the numbers in the problem. In this case, the child's speech and gestures suggest that she is considering more than one strategy for solving this particular problem. A recent experimental study suggests that such gestures that mismatch speech reflect implicit knowledge about the concept being explained (Garber, Alibali, & Goldin-Meadow, 1998; see also Goldin-Meadow & Alibali, 1994). Furthermore, children's gestures reveal the leading edge of their knowledge, in the sense that newly generated strategies are typically expressed in gesture before they are expressed in speech (Alibali & Goldin-Meadow, 1993; Church, 1987).

Because gestures provide a window onto newly emerging strategies, they can reveal gradual changes in strategy use. For example, consider a child who used Strategy A in speech across all problems on a pretest and Strategy B in speech across all problems on a posttest. At first glance, such a child appears to have changed abruptly. However, the child's gestures might reveal that the child was already considering Strategy B at pretest, showing the change to be gradual rather than abrupt. Alternatively, the child's gestures might reveal that the child was still considering Strategy A at posttest, again showing the change to be gradual rather than abrupt. In the present study, children's gestures as well as their speech were used to assess their strategy repertoires. Thus, strategy generation and strategy abandoning were examined with a technique that is sensitive to the gradual emergence and disappearance of strategies.

Goals of the Present Study

In brief, the present study investigated two central questions about the process of strategy change in children learning a mathematical concept. First, how do instruction and initial variability influence patterns of strategy use and strategy change? Second, under what circumstances is strategy change gradual, and under what circumstances is it abrupt? To address these questions, I used a training study design that varied the type of instruction children received. Children's strategy repertoires were assessed both before and after the instructional intervention.

Method

Participants

One hundred seventy-eight third- and fourth-grade students from six urban parochial schools were screened for participation in the study. Children who solved at least one of six pretest problems correctly ($n = 25$) were excluded from the sample following the pretest.[1] Each of the remain-

[1] These included 5 third-grade children tested between April and June, 11 fourth-grade children tested between October and December, and 9 fourth-grade children tested between April and June.

Table 1
Problems Used on the Explanation and Paper-and-Pencil Tests

Equivalent operands?	Operation	Position of blank	Example	No. of problems	Success category
			Explanation tests		
Yes	Addition	After =	4 + 3 + 5 = ___ + 5	3	Learning
Yes	Addition	Final	5 + 7 + 6 = 5 + ___	3	Near transfer
			Paper-and-pencil tests		
Yes	Addition	After =	4 + 3 + 5 = ___ + 5	3	Learning
Yes	Addition	Final	5 + 7 + 6 = 5 + ___	3	Near transfer
Yes	Multiplication	After =	4 × 2 × 3 = ___ × 3	2	Near transfer
Yes	Multiplication	Final	3 × 4 × 5 = 3 × ___	2	Near transfer
No	Addition	After =	7 + 5 + 8 = ___ + 4	3	Far transfer
No	Addition	Final	4 + 3 + 7 = 6 + ___	3	Far transfer
No	Multiplication	After =	2 × 5 × 3 = ___ × 6	1	Far transfer
No	Multiplication	Final	3 × 4 × 2 = 6 × ___	1	Far transfer

ing children was randomly assigned to one of five intervention groups (described later). Four children were eliminated from the final sample because of irregularities in the administration of the experimental protocol, and an additional 6 children were eliminated because of coding difficulties (most of their gestured responses were uncodable, either because their body blocked their gestures on the videotape or because their gestures were produced too far away from the problem to be reliably interpreted). The final sample consisted of 143 children (76 girls and 67 boys, distributed approximately equally across the five groups). They included 43 third-grade children tested between April and June (late third graders), 54 fourth-grade children tested between October and December (early fourth graders), and 46 fourth-grade children tested between April and June (late fourth graders). Children ranged in age from 8 years 7 months to 11 years 3 months, with a mean age of 9 years 8 months.

Procedure

Overview. Children were tested individually in a quiet room during school hours. The experimental session included five segments, which are described in detail below: (a) a paper-and-pencil pretest, (b) an explanation pretest, (c) the instructional intervention, (d) an explanation posttest, and (e) a paper-and-pencil posttest. All parts of the session except the paper-and-pencil tests were videotaped. Each child's session lasted approximately 40 min.

Paper-and-pencil pretest. Children completed the paper-and-pencil pretest by working independently at a desk. The test included eight types of equivalence problems, which were constructed by varying three features of the problems: the operation (addition or multiplication), the position of the blank (in final position or immediately following the equal sign), and the presence of equivalent operands on both sides of the equation (yes or no). Sample problems are presented in Table 1.

Explanation pretest. After the paper-and-pencil pretest, children solved and explained a set of six equivalent-operands addition problems (three with the blank immediately following the equal sign, and three with the blank in final position) at the blackboard. An experimenter wrote the first problem of the set on the blackboard and asked the child to solve the problem. After the child wrote the solution in the blank, the experimenter asked, "How did you get that answer?" The child then explained how he or she solved the problem. The experimenter then erased that problem and wrote up the next problem of the set. The session continued in this manner for a total of six problems. Children who solved any of the six problems correctly did not continue in the study after this segment.

Intervention. Children then participated in one of five different instructional interventions, conducted by a second experimenter. In all of the interventions, children were asked to solve and explain three equivalent-operands addition problems with the blank immediately following the equal sign (e.g., 6 + 3 + 7 = ___ + 7). In one group (*control*), children received no feedback or instruction. In a second group (*feedback*), children received feedback about whether their solutions were correct but received no instruction. In the remaining three groups (*principle, analogy,* and *procedure*), children received both feedback and some type of direct instruction (see later). In all groups, if children asked questions, the experimenter simply repeated the training statements and told them that all questions would be fully answered at the end of the session.

The structure of the intervention was comparable in the four groups that received either feedback or instruction. For each of the first two intervention problems, after the child's explanation, the experimenter told the child whether his or her problem solution was correct, saying either, "That's a good try, but it's not the right way to solve this problem" or "Very good! That's the right way to solve this problem." If the child's solution was correct, the experimenter proceeded to the next problem. If the child's solution was incorrect, the experimenter then provided the appropriate training statements, which differed in each intervention (see later). All instruction was provided in speech; that is, the experimenter did not gesture with any of the training statements. After the training statements, in the three instruction groups (principle, analogy, and procedure), the experimenter asked the child, "Does that give you any ideas about what the right answer to this problem might be?" If the child was unable to offer another solution, the experimenter said, "Let's try one more," and proceeded to the next problem. If the child offered an incorrect answer, the experimenter said, "OK, that's another answer," and proceeded to the next problem. If the child offered a correct solution, the experimenter said, "Very good! That's the right answer to this problem," and then proceeded to the next problem. No instruction was provided following the third intervention problem.

In the feedback intervention, for each of the first two problems, after providing feedback, the experimenter simply asked the child if he or she could "think of another way" to solve the problem. If the child offered another way to solve the problem, the experimenter said, "OK, that's another way," and proceeded to the next problem.[2] If the child said that he or she could not think of another way, the experimenter simply said, "OK," and proceeded to the next problem.

[2] All children in the feedback group who suggested "other ways" to solve the problems suggested incorrect strategies.

In the principle intervention, for each of the first two problems, after providing feedback, the experimenter explained the principle of equivalence to the child. She told the child that the correct way to solve the problem is to "make both sides of the problem equal." She then said, "That means that you need to make what comes *before* the equal sign add up to the same amount as what comes *after* the equal sign." She then continued as described earlier.

In the analogy intervention, the experimenter made an analogy between the equation and an object that was familiar to all of the children—a teeter-totter (also called a see-saw). This particular analogy was chosen because prior work had shown that children successfully learned the principle of equivalence in a lesson that emphasized making the two sides of the equation "weigh the same" or "balance" (Perry et al., 1995). In the analogy group, for each of the first two problems, after providing feedback, the experimenter said that she would give the child "a hint about how to solve the problem the right way." She then asked the child if he or she knew what a teeter-totter was, and if the child was not sure, she described it to the child. The experimenter then said, "You know that when the person on one side of the teeter-totter is heavy, and the person on the other side is light, then the teeter-totter can't balance. But if the two people on the teeter-totter are the same size, or the same weight, then the teeter-totter can balance. Well, you can think about this math problem as kind of like a teeter-totter. You need to find a number to put in the blank to make the teeter-totter balance." She then drew a teeter-totter under the problem, with the fulcrum at the equal sign (see Figure 1), and continued as described earlier.

In the procedure intervention, for each of the first two problems, after providing feedback, the experimenter explained the grouping procedure to the child. She told the child that because one of the addends appeared on both sides of the equation, the child needed only to add the other addends to obtain the correct answer. She then continued as described earlier.

Explanation posttest. Immediately following the intervention, each child solved and explained a second set of six problems to the first experimenter, working at the blackboard. No feedback was provided on this set of problems.

Paper-and-pencil posttest. Each child then completed a paper-and-pencil posttest, working independently at a desk. This test was comparable to the paper-and-pencil pretest.

Debriefing. Following the experimental session, children in the control and feedback groups received a brief lesson about the principle of mathematical equivalence. Thus, every child received a lesson, either as part of the experiment proper or during debriefing. Before returning to class, each child chose a brightly colored pencil as a prize for participating in the study.

Coding Solutions

Solutions to the paper-and-pencil pretest and posttest problems were scored as correct or incorrect.

Coding Explanations

Verbal and gestured explanations produced during the explanation sessions were coded in terms of the problem-solving strategy that they conveyed. Verbal explanations were coded by listening to the audio portion of the videotape only, without reference to the video portion (i.e., with the picture turned off) except in cases in which deictic pronouns had to be disambiguated (e.g., "Then I added *this one*"). Gestured explanations were coded using the video portion of the videotape only, without reference to the audio portion (i.e., with the sound turned off). The verbal and gestured components of each response were coded independently by different coders. Finally, the relationship between gesture and speech was evaluated by comparing the codes for the verbal and gestured components of each response.

Coding strategies expressed in speech. Each verbal explanation was transcribed and coded according to a system established in previous research (Alibali, 1994; Perry et al., 1988). As seen in Table 2, the children in this study expressed many different strategies in speech, including some that yielded incorrect solutions and some that yielded correct solutions.

Spoken explanations that described operations on specific subsets of numbers that did not conform to any of the strategies in Table 2 were coded as idiosyncratic strategies and were considered incorrect (e.g., one child said that she subtracted the addend on the right side of the equation from the third addend on the left side of the equation; another child added the numbers on the left side of the equation and divided that sum by the addend on the right). Such idiosyncratic strategies accounted for 8% of the 1,710 verbal explanations produced by children in this study. Spoken explanations that could not be assigned to a strategy category and that did not clearly specify an idiosyncratic strategy were classified as ambiguous. Fewer than 4% of verbal explanations were classified as ambiguous. Children's verbal explanations were coded in terms of the strategy they described for solving the problem, even when that strategy did not yield the solution the child provided for that problem. However, discrepancies between verbal strategies and solutions were rare (fewer than 3% of responses).

Coding strategies expressed in gesture. Each gestured explanation was transcribed and coded by another coder, using a system established in previous research (Alibali, 1994; Perry et al., 1988). Each of the verbal strategies had a counterpart in gesture, as seen in Table 2. In addition, gestures to specific subsets of numbers or symbols that did not conform to any of the patterns in these strategies were coded as idiosyncratic strategies and were considered incorrect (e.g., one child pointed to the third addend on the left side of the equation, then to the addend on the right side of the equation, and then to the solution). Such idiosyncratic strategies accounted for 10% of the 1,056 gestured explanations produced by children in the study. Gestured explanations that could not be assigned to a strategy category or that did not clearly specify an idiosyncratic strategy were classified as ambiguous. Approximately 4% of the gestured explanations produced by children in the study were classified as ambiguous. Children's gestured explanations were coded in terms of the strategy they conveyed, even when that strategy did not yield the solution the child provided for that problem. Such discrepancies between gestured strategies and solutions typically occurred when gesture conveyed a different strategy from speech.

Coding the relationship between gesture and speech. Finally, the relationship between gesture and speech was assessed. Any response that did not include gesture was coded as a *speech-alone* response. Any response in which speech and gesture expressed the same strategy was coded as a gesture–speech *match.* Any response in which speech expressed one strategy and gesture expressed a different strategy was coded as a gesture–speech *mismatch.* Any response in which one modality conveyed a distinct strategy and the other modality conveyed an ambiguous strategy was also coded as a mismatch.[3]

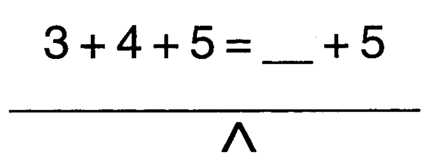

Figure 1. Drawing used in the analogy instruction.

[3] For responses in which both modalities conveyed idiosyncratic strategies, the relationship between speech and gesture was coded on a case-by-case basis. If both modalities conveyed the same idiosyncratic strategy, the response was coded as a gesture–speech match. If the two modalities conveyed different idiosyncratic strategies, the response was coded as a gesture–speech mismatch.

Table 2
Examples of Strategies Expressed in Speech and Gesture for the Problem 3 + 4 + 5 = 3 + ___

Strategy	Definition	Spoken explanation	Gestured explanation (Handshape: Referents)
		Incorrect strategies	
Add all	Add all the numbers in the problem.	"I added 3 plus 4 plus 5 plus 3 and I got 15."	Right hand point: Left 3, 4, 5, right 3, solution.
Add to equal sign	Add the numbers that appear to the left of the equal sign.	"3 plus 4 made 7 and 7 plus 5 made 12."	Right hand point: Sweep across under 3 + 4 + 5, point to solution.
Add to equal–Add all	Add the numbers that appear to the left of the equal sign, write that sum in the blank; then add that sum to the number on the right side, and write the total to the right of the problem.	"3 plus 4 plus 5 made 12, so I wrote the 12 down, and then 12 plus 3 made 15, so I wrote 15."	Right hand point: Left 3, 4, 5, first solution (12), right 3, second solution (15).
Carry	Take one number from the left side of the equation and place it in the blank.	"There was a 4 on this side so I put 4."	Right hand point: 4, solution.
Add two	Add the first two numbers on the left side of the equation.[a]	"I added 3 and 4 and it was 7."	Right hand point: Left 3, 4.
Read all elements	Read the elements in the problem without describing a specific strategy.	"3 plus 4 plus 5 equals 3 plus 12."	Right hand point: Left 3, first +, 4, second +, 5, =, right 3, right +, solution.
		Correct strategies	
Grouping–equivalent addends	Note that one number appears on both sides of the problem, and group the other numbers.	"All I did was add 4 plus 5, because there was 3 on each side."	Right hand two finger "V" point: Span 4 + 5, hand down. Right hand point: Right 3, left 3.
Grouping	Group the numbers that do not appear on both sides of the equation.	"I added 4 and 5."	Right hand point: 4, second +, 5.
Add–subtract	Add the numbers on the left side of the equation and subtract the number on the right.	"3 plus 4 equals 7, 7 plus 5 equals 12, and 12 minus 3 equals 9."	Right hand point: Left 3, 4, 5, hand down. Left hand point: Downward stroke under right 3.
Equalize	Make both sides of the equation sum to the same total.	"3 and 4 and 5 made 12, so to make 12 over here, I had to put 9 with the 3."	Right hand point: Sweep back and forth under 3 + 4 + 5, hand down. Left hand point: Sweep back and forth under 3 + solution.

[a] The add-two strategy was coded only on problems with the blank in final position (e.g., 3 + 6 + 7 = 3 + ___). For problems with the blank immediately following the equal sign (e.g., 3 + 6 + 7 = ___ + 7), adding the first two numbers yielded a correct solution and was therefore categorized as an instance of the grouping strategy.

Evaluating Children's Strategy Repertoires

Each child's repertoire of strategies was evaluated for the explanation pretest and the explanation posttest. Strategies that children expressed in speech and in gesture were identified for each test. For example, during the pretest, if a child described the add-all strategy in speech for all six problem explanations, and the same child described the add-all strategy in gesture for four problems and the grouping strategy in gesture for two problems, that child's repertoire for that session would consist of two strategies: add-all (expressed in both speech and gesture) and grouping (expressed in gesture only).

Evaluating Whether Children Generated and Abandoned Strategies

For each child, the strategies that were maintained, generated, and abandoned were identified. Strategies that children expressed in gesture or speech during both the pretest and the posttest explanations were considered to be *maintained* in the repertoire. Strategies that were expressed in either modality on the posttest but were not expressed in either modality on the pretest were considered to be newly *generated.* Strategies that were expressed in either modality on the pretest but were not expressed in either modality on the posttest were considered to be *abandoned.*

Because the explanation pretest and posttest included only six problems, one cannot be absolutely certain that strategies were actually newly generated or completely abandoned over the course of the study. It is possible that strategies that appeared to be abandoned were still present in the child's repertoire at posttest but were simply not sampled at posttest, or that strategies that appeared to be generated were actually present in the child's repertoire at pretest but were simply not sampled at pretest. However, given 12 opportunities to sample each strategy (6 in speech and 6 in gesture), the probability of not sampling a particular strategy on a given test is only .0002 for a child with a repertoire of two strategies, .008 for a child with a repertoire of three strategies, and .03 for a child with a repertoire of four strategies. Because the likelihood of having a strategy in the repertoire but not sampling it is low, the terms *abandon* and *generate* are used throughout this article. However, bear in mind that it is impossible to demonstrate conclusively that a particular strategy has been either entirely abandoned or newly generated.

Reliability

Interrater reliability was assessed by having a second coder assess the explanations produced by a randomly selected subset of the participants (approximately 10% of the participants for strategy coding and 15% of the participants for gesture–speech relationship and repertoire coding). Agreement between coders was 90% for assigning strategy codes to verbal responses (n = 269 responses), 83% for assigning strategy codes to gestured responses (n = 170 responses), 95% for categorizing responses as matches, mismatches, or speech-alone responses (n = 334 responses), and 94% for identifying individual strategies within children's pre- and posttest repertoires (n = 142 strategies).

Results

The results are organized into four sections. The first two sections present basic findings about changes in variability and patterns of learning and transfer. The third section examines changes in the content of children's strategy repertoires, focusing on the relations between instruction and initial variability, on the one hand, and strategy generation and strategy abandoning, on the other. Finally, the fourth section evaluates whether strategy change in children learning about equivalence is gradual or abrupt.

Changes in Level of Variability

I first examined how children's level of variability changed from the explanation pretest to the explanation posttest. Table 3

Table 3
Mean Number of Strategies and Mismatches Produced by Children in Each Instruction Group on the Explanation Pretest and Posttest

	No. of strategies					
	Verbal		Gesture			
Condition/test	Incorrect	Correct	Incorrect	Correct	Total	No. of mismatches
Control (n = 26)						
Pre	1.58	0.00	0.62	0.50	2.69	2.15
Post	1.35	0.00	0.38	0.35	2.08	1.54
Change	−0.23	0.00	−0.23	−0.15	−0.62	−0.62
Feedback (n = 29)						
Pre	1.69	0.00	0.90	0.24	2.83	2.03
Post	1.55	0.00	0.52	0.17	2.24	1.38
Change	−0.14	0.00	−0.38	−0.07	−0.59	−0.66
Principle (n = 28)						
Pre	1.50	0.00	0.61	0.29	2.39	2.04
Post	1.29	0.11	0.82	0.25	2.46	1.96
Change	−0.21	0.11	0.21	−0.04	0.07	−0.07
Analogy (n = 28)						
Pre	1.61	0.00	0.68	0.36	2.64	1.93
Post	1.00	0.39	0.75	0.39	2.54	1.82
Change	−0.61	0.39	0.07	0.04	−0.11	−0.11
Procedure (n = 32)						
Pre	1.41	0.00	0.38	0.13	1.91	1.00
Post	0.63	0.84	0.38	0.09	1.94	0.78
Change	−0.78	0.84	0.00	−0.03	0.03	−0.22

Table 4
Mean Proportion Correct on the Explanation and Paper-and-Pencil Posttests

	Explanation posttest (6 possible)		Paper-and-pencil posttest (18 possible)		
Condition	*M*	*SE*	*M*	*SE*	*n*
Control	0.00	0.00	0.01	0.01	26
Feedback	0.01	0.01	0.00	0.00	29
Principle	0.11	0.06	0.12	0.05	28
Analogy	0.30[a]	0.08	0.25[a]	0.07	28
Procedure	0.58[a]	0.06	0.31[a]	0.03	32

[a] Differs significantly from the feedback group, Dunnett's procedure, $p < .01$.

presents the mean number of different strategies expressed by children in each condition, broken down by modality and correctness, and the mean number of gesture–speech mismatches produced by children in each condition. For statistical analysis, variability was operationalized in two dependent measures: total number of distinct strategies expressed and number of gesture–speech mismatches produced. Because both measures were expected to index behavioral variability, they were expected to be correlated, and indeed they were, $r(141) = 0.75, p < .001$, for the explanation pretest, and $r(141) = 0.71, p < .001$, for the explanation posttest.[4] As shown in Table 3, overall levels of variability were strikingly high throughout the study. On average, children expressed more than two strategies on each test (pretest, $M = 2.48$; posttest, $M = 2.24$) and produced between one and two mismatches on each test (pretest, $M = 1.80$; posttest, $M = 1.48$).

To assess the effects of instruction on changes in variability, I used repeated measures analysis of variance, with instruction group as a between-participants variable and test (pre or post) as a within-participants variable. Patterns of change were comparable for both measures of variability, which was expected given the high correlation between the measures. Overall, variability decreased from the explanation pretest to the explanation posttest, $F(1, 138) = 4.60, p < .05$, for total number of strategies, and $F(1, 138) = 4.51, p < .05$, for gesture–speech mismatches. For both measures, the decrease in variability was greatest in the two groups that did not receive direct instruction (control and feedback). However, the interaction between test and condition did not reach significance for either measure, $F(4, 138) = 1.76, p = .14$, for number of strategies, and $F(4, 138) = 0.63, p > .50$, for gesture–speech mismatches.

The overall decrease in variability does not imply that all children decreased in variability over the study. Indeed, many individual children increased their behavioral variability. Collapsed across instruction groups, 27% of the participants increased the number of strategies they expressed, whereas 37% decreased and 36% made no change. Similarly, collapsed across instruction groups, 29% of participants increased the number of mismatches produced, whereas 41% decreased and 30% made no change.

In summary, although many children decreased in variability, and many others increased in variability, patterns of change in variability did not depend on the type of instruction that children received. However, this does not imply that children's problem-solving performance was unchanged as a result of instruction. I next consider whether children began to solve the problems correctly after instruction.

Patterns of Learning and Transfer

Recall that the sample was made up of children who had solved the six problems on the explanation pretest incorrectly (this was the criterion for inclusion in the study). Most of the children also had little success on the paper-and-pencil pretest: 85% (121) solved none of the 18 problems correctly, and an additional 8% (12) solved only one problem correctly.

Overall levels of success. To evaluate progress in understanding equivalence, I first assessed the total number of problems children solved correctly on both the explanation posttest and on the paper-and-pencil posttest. Instruction had a significant effect on success, $F(4, 138) = 22.97, p < .001$, for the explanation test; $F(4, 138) = 11.77, p < .001$, for the paper-and-pencil test (see Table 4). Dunnett's post hoc tests showed that, on both tests, children in the analogy and procedure groups succeeded on more problems than children in the feedback group; however, children in the control and principle groups did not differ significantly from the feedback group. The feedback group was used as the reference (i.e., baseline) category, because each of the other groups differed from the feedback group in only one way: the control group differed in the absence of feedback, and the three instructed groups (principle, analogy, and procedure) differed in the presence of instruction (i.e., they included both feedback and instruction). Thus, comparing the control group with the feedback group provides a test of the effects of feedback alone. Similarly, comparing each of the instructed groups with the feedback group provides a test of the effects of instruction, above and beyond the effects of feedback.

Improvement on different types of problems. I next assessed improvement on each of the problem types on the explanation posttest and the paper-and-pencil posttest. As shown in Table 1, the explanation tests (six problems) consisted of equivalent-operands addition problems that varied the position of the blank. The paper-and-pencil tests (18 problems) included problems that varied the position of the blank, the operation, and the presence of equivalent operands. Only one type of problem was used during the instructional intervention: equivalent-operands addition problems with the blank immediately following the equal sign (e.g., $5 + 3 + 7 = \_\_ + 7$). Thus, improved performance on problems

[4] These correlations did not arise simply because children with only one strategy could not produce mismatches; they remained high and significant when the analyses were limited to children with more than one strategy in their repertoires, $r(109) = 0.67, p < .001$, for the explanation pretest, and $r(96) = 0.56, p < .001$, for the explanation posttest.

of this type on the posttest (relative to the pretest) reflected simple learning, because children in all the groups (except the control group) received either feedback or instruction about such problems. Improved performance on other types of problems reflected transfer to new problem types, because children did not receive feedback or instruction about such problems. I defined *near transfer* as improved performance on equivalent-operands problems that differed from those used in the intervention (i.e., multiplication problems or addition problems with the blank in final position) and *far transfer* as improved performance on nonequivalent-operands problems.

To statistically evaluate patterns of learning and transfer, I used logistic regression to model binary outcomes (i.e., improvement or not) on each problem type. For each dependent measure, the best-fitting model was selected using a forward stepwise procedure, using a significant ($p < .05$) change in the log-likelihood ratio statistic as the criterion for adding a given variable (described in Norusis, 1994). The potential predictor variables included instruction group, the two measures of pretest variability (total number of pretest strategies and total number of pretest mismatches), and the background variables of gender and grade level. Both measures of pretest variability were treated as continuous variables.[5]

For every measure of learning and transfer, the best-fitting model revealed a significant effect of instruction group. Figure 2 displays the proportion of children in each instruction group who showed learning, near transfer, and far transfer on the explanation test (A) and the paper-and-pencil test (B). (Note that far transfer was not assessed on the explanation test.) Improvement was limited to children who received direct instruction; almost none of the children in the control and feedback groups improved on any of the measures.

The results of the statistical analysis are summarized in Table 5. As described earlier, for all comparisons, the feedback group was used as the reference (i.e., baseline) category, because each of the other groups differed from the feedback group in only one way. Table 5 presents the odds ratios for each outcome measure in relation to the feedback group. For example, the relative odds for learning on the explanation posttest (i.e., the ratio of learning to not learning) were estimated to be 4.36 times higher in the principle group than in the feedback group (Table 5, column 2).

For both tests, *learning* was defined as improved performance on problems identical in form to those used in the intervention (e.g., $3 + 4 + 5 = \_\_ + 5$). As seen in Figure 2, on both the explanation test and the paper-and-pencil test, the rate of learning was highest among children in the procedure group, followed by the analogy group. For both tests, the relative odds of learning in the procedure and analogy groups were significantly greater than in the feedback group (see Table 5).[6]

Near transfer was defined as improved performance on equivalent-operands problems that differed from those used in the intervention. Both the explanation posttest and the paper-and-pencil posttest included equivalent-operands addition problems with the blank in final position (e.g., $4 + 6 + 8 = 4 + \_\_$). The paper-and-pencil posttest also included equivalent-operands multiplication problems with the blank either following the equal sign or in final position (e.g., $4 \times 2 \times 3 = \_\_ \times 3$, $2 \times 3 \times 5 = 2 \times \_\_$). On the explanation posttest, the rate of near transfer was highest among children in the procedure and analogy groups and was comparable in these two groups. On the paper-and-pencil posttest, the rate was highest among children in the procedure group, followed by the analogy group and the principle group. On both tests, the relative odds of near transfer in the procedure and analogy groups were significantly greater than in the feedback group (see Table 5).

Far transfer was assessed on the paper-and-pencil posttest only and was defined as improved performance on nonequivalent operands problems (e.g., $4 + 6 + 8 = 7 + \_\_$, $2 \times 3 \times 5 = 6 \times \_\_$). As seen in Figure 2, children in the analogy group showed the highest rate of far transfer, followed by children in the principle group. Children in the procedure group showed almost no far transfer. The relative odds of far transfer were significantly greater in the analogy group than in the feedback group (see Table 5). None of the other groups differed significantly from the feedback group.

In general, the pattern of results over the set of learning and transfer measures can be interpreted as a trade-off between general and specific learning across instruction groups. In the procedure group, many children learned how to solve problems like those used in the intervention; however, most were unable to transfer their knowledge broadly. In contrast, in the analogy group, although relatively few children learned, those who did learn were able to transfer their knowledge more broadly. Only 36% of the children in the analogy group improved on the paper-and-pencil test, compared with 91% in the procedure group. However, among these children, the average number of problems improved was 11.90 in the analogy group ($SE = 1.82$) and only 5.90 ($SE = 0.52$) in the procedure group, $t(37) = 4.37$, $p < .001$. Thus, instruction about an analogy led to more flexible learning than instruction about a problem-solving procedure, but only for a small number of children.

One surprising outcome across all the learning and transfer measures was the ineffectiveness of the principle instruction. Although the rate of learning and transfer was higher in the principle group than in the feedback group for every measure, the difference did not reach the conventional significance level for any of the measures. This finding contrasts with other intervention studies of mathematical equivalence, in which similar principle-based in-

[5] In comparisons in which there were sampling zeroes, the maximum likelihood estimates of the parameters were located on or near the boundary of the parameter space, making standard errors both difficult to calculate and hard to interpret. To estimate parameters for the individual instruction groups, I computed the parameter estimates for comparisons with sampling zeroes by adding 0.20 to each cell of the relevant contingency table (Condition × Grade Level × Outcome) and refitting the model (see, e.g., Bishop, Fienberg, & Holland, 1975, p. 407). This procedure moves the parameter estimates to the interior of the parameter space, in which the estimates and their standard errors are both more tractable computationally and more interpretable. This procedure adds some small-sample bias to the parameter estimates, which slightly suppresses the differences between cell means, making tests more conservative. However, the results were stable when the prior weight was reduced to as little as 0.05.

[6] On the explanation posttest, 5 children in the procedure group used the grouping strategy on problems of the form $a + b + c = \_\_ + c$ and used the add-two strategy on problems of the form $p + q + r = p + \_\_$. This pattern of responses suggests that they succeeded on the problems, not because they learned a correct procedure (grouping), but because they inferred an incorrect procedure (add the first two numbers). The pattern of results does not differ if these children are reclassified as not demonstrating learning.

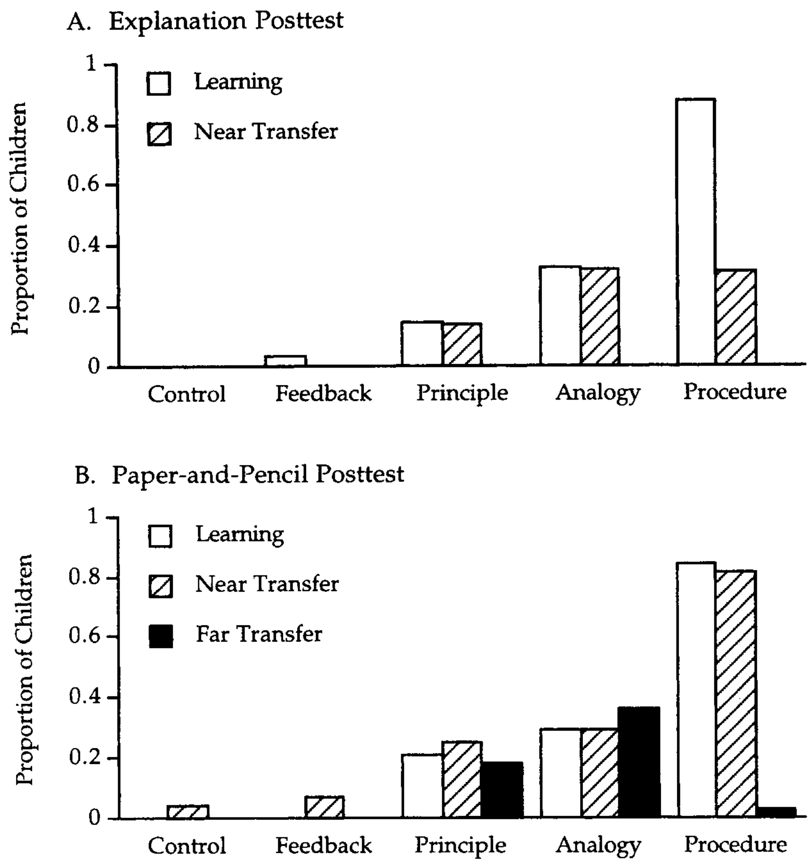

Figure 2. Proportion of children who showed learning and transfer (A) on the explanation posttest and (B) on the paper-and-pencil posttest. *Learning* was defined as improvement on problems like those used in instruction (i.e., left-blank equivalent-operands addition problems, e.g., 3 + 4 + 5 = ___ + 5). *Near transfer* was defined as improvement on equivalent-operands problems unlike those used in instruction (e.g., 4 + 3 + 6 = 4 + ___, 4 × 5 × 3 = ___ × 3). *Far transfer* was defined as improvement on nonequivalent-operands problems (e.g., 5 + 8 + 4 = ___ + 6). Far transfer problems were presented only on the paper-and-pencil posttest.

struction yielded much greater learning (Perry et al., 1988; Rittle-Johnson & Alibali, in press). I consider possible explanations for this finding in the Discussion section.

On the basis of prior findings by Perry et al. (1988), I expected a linear relation between pretest variability and success on the posttest and transfer problems. Specifically, I expected that higher levels of pretest variability would be associated with greater success overall, and in particular, with greater success at far transfer. Surprisingly, neither of the two measures of pretest variability was related to whether children succeeded on any of the learning or transfer measures. However, given the low levels of learning and transfer in most of the instruction groups, floor effects may have obscured any potential relation.

There were also no effects of gender on any of the learning or transfer measures, and grade level was associated with only one of the learning measures. The relative odds of learning on the explanation posttest problems were greater among children tested early in fourth grade than among children tested either late in third grade or late in fourth grade, Wald $z^2(1) = 9.35$, $p < .01$.

In sum, the different instruction groups showed different patterns of learning and transfer. However, as described in the preceding section, the different instruction groups did not display different patterns of change in variability. Taken together, these findings suggest that instruction did not influence the size of children's strategy repertoires, but did influence their content. I turn next to this issue.

Changes in the Content of Children's Strategy Repertoires

Changes in the content of children's strategy repertoires were expected to depend both on the type of instruction the children received and on their initial level of variability. Note that children could generate strategies either in speech (with or without matching gesture) or uniquely in gesture (i.e., in gestures that mismatched the accompanying speech). Similarly, children could abandon strategies that they had expressed on the pretest either in speech (with or without matching gesture) or uniquely in gesture. Table 6 presents the proportion of children in each instruction group who generated and abandoned correct and incorrect strategies in each modality. (Note that children could generate or abandon strategies in multiple categories.) On the basis of these data,

Table 5
Estimated Odds Ratios for Measures of Learning and Transfer, in Relation to Feedback Group

	Learning		Near transfer		Far transfer
Condition	Explanation posttest[a]	Paper-and-pencil posttest	Explanation posttest	Paper-and-pencil posttest	Paper-and-pencil posttest
Control	0.47	1.11	1.11	0.54	1.11
Feedback (reference)	1.00	1.00	1.00	1.00	1.00
Principle	4.36	14.39†	6.93	4.50†	11.69
Analogy	12.02**	20.57*	20.57*	5.40*	28.08*
Procedure	182.16**	242.82**	23.11*	58.50**	2.49

Note. *Learning* was defined as improvement on problems like those used in instruction (i.e., left-blank equivalent-operands addition problems, e.g., 3 + 4 + 5 = ___ + 5). *Near transfer* was defined as improvement on equivalent-operands problems unlike those used in instruction (e.g., 4 + 3 + 6 = 4 + ___, 4 × 5 × 3 = ___ × 3). *Far transfer* was defined as improvement on nonequivalent-operands problems (e.g., 5 + 8 + 4 = ___ + 6). Parameter estimates for comparisons with sampling zeroes were computed by adding 0.20 to each cell of the Condition × Grade Level × Outcome contingency table (see Footnote 5 in text).
[a] Reported odds ratios are adjusted for grade level.
† $p < .10$, Wald statistic. * $p < .05$, Wald statistic. ** $p < .01$, Wald statistic.

several binary outcome measures were developed to assess the effects of instruction group and initial variability. The data were analyzed using logistic regression, with the procedure described in the preceding section.

Generating strategies. To evaluate patterns of strategy generation, I developed three binary outcome measures. First, children were classified on the basis of whether or not they generated any new strategies in either modality (speech or gesture). Table 7, column 2, displays the proportion of children in each instruction group who generated at least one new strategy. Next, children were classified based on whether or not they generated any new strategies in speech. Table 7, column 4, displays the proportion of children in each group who expressed new strategies in speech. Finally, children were classified based on whether or not they generated any new, correct strategies in speech. Table 7, column 6, displays the proportion of children in each group who expressed new, correct strategies in speech. For each of the three measures of strategy generation, the best-fitting model revealed a significant effect of instruction group. As described earlier, the feedback group was used as the reference category for interpreting the effects of instruction.

First, consider how instruction influenced whether or not children generated any new strategies. In this analysis, a child was counted as generating a new strategy if that child expressed any new strategy in either speech or gesture during the posttest explanations. As seen in Table 7, column 3, on this measure, the relative odds of strategy generation (i.e., the ratio of generating to not generating) were significantly higher in the procedure group than in the feedback group. None of the other groups differed significantly from the feedback group. The overall pattern suggests that procedure-based instruction makes strategy generation more likely; however, feedback and instruction are not essential for strategy generation for this concept. Indeed, 35% of the children in the control group spontaneously generated new strategies.

Next, consider how instruction influenced whether or not children generated new strategies that they expressed in speech. A child was counted as generating a new strategy in speech if that child expressed any new strategy (correct or incorrect) in speech during the posttest explanations. For this analysis, children who expressed strategies uniquely in gesture on the pretest problems, and then expressed those same strategies in speech on the posttest problems, were also classified as generating strategies in speech. As seen in Table 7, the relative odds of strategy generation were significantly higher in the procedure group than in the feedback group, and the relative odds in the principle and analogy groups did not differ from the feedback group. However, for this measure,

Table 6
Proportion of Children in Each Instruction Group Who Generated and Abandoned Strategies

	Generate				Abandon			
Condition	Verbal incorrect	Verbal correct	Gesture incorrect	Gesture correct	Verbal incorrect	Gesture incorrect[a]	Gesture correct[a]	*n*
Control	0.00	0.00	0.19	0.23	0.15	0.91 (11)	0.82 (11)	26
Feedback	0.24	0.00	0.14	0.14	0.38	0.60 (15)	1.00 (7)	29
Principle	0.29	0.07	0.32	0.21	0.36	0.75 (12)	0.75 (8)	28
Analogy	0.07	0.29	0.21	0.29	0.43	0.62 (13)	0.67 (9)	28
Procedure	0.09	0.75	0.22	0.03	0.63	0.80 (10)	0.33 (3)	32
Overall	0.14	0.24	0.22	0.17	0.40	0.72 (61)	0.76 (38)	143

[a] Children who did not produce gestured strategies on the explanation pretest could not abandon such strategies. Therefore, *n* is reduced for the abandon-gesture categories. For each cell, the appropriate *n* is noted in parentheses.

Table 7
Proportion of Children Who Generated Strategies, and Odds Ratios in Relation to the Feedback Group

	Generate any verbal or gestured strategy		Generate any verbal strategy		Generate correct verbal strategy	
Condition	Proportion of children	Odds ratio	Proportion of children	Odds ratio[a]	Proportion of children	Odds ratio[a,b]
Control	0.35	0.65	0.00	0.05*	0.00	1.31
Feedback (reference)	0.45	1.00	0.31	1.00	0.00	1.00
Principle	0.61	1.90	0.39	1.41	0.11	8.60
Analogy	0.57	1.64	0.43	1.63	0.32	27.32*
Procedure	0.88	8.60**	0.84	10.58**	0.78	248.27**

[a] Parameter estimates for comparisons with sampling zeroes were computed by adding 0.20 to each cell of the Condition × Grade Level × Outcome contingency table (see Footnote 5 in text). [b] Reported odds ratios are adjusted for grade level.
* $p < .05$, Wald statistic. ** $p < .01$, Wald statistic.

the relative odds of strategy generation were significantly lower in the control group than in the feedback group. The pattern suggests that feedback about correctness is essential to generating verbal strategies. Children in the control group, who did not receive such feedback, were unlikely to generate verbal strategies. Indeed, all of the children in the control group who generated strategies expressed those new strategies uniquely in gesture, rather than in speech.[7]

Finally, consider how instruction influenced whether or not children generated correct verbal strategies. A child was counted as generating a correct verbal strategy if that child expressed a new, correct strategy in speech during the posttest explanations. For this analysis, children who expressed correct strategies uniquely in gesture on the pretest problems, and then expressed those same correct strategies in speech on the posttest problems, were also classified as generating correct strategies in speech. As seen in Table 7, children in the control group were just as unlikely as children in the feedback group to generate correct verbal strategies, and children in the groups that received instruction were more likely to generate correct verbal strategies. Indeed, the relative odds of strategy generation were significantly higher in both the procedure and analogy groups than in the feedback group. In the principle group, the relative odds were slightly higher than in the feedback group, although not significantly so. The pattern suggests that simple feedback is not enough for children to generate correct verbal strategies for equivalence problems. Some type of instructional input appears to be essential.[8]

Overall, the results show that feedback and instruction play an important role in strategy generation for this concept. In general, children in the control group generated strategies uniquely in gesture, whereas children who received feedback (with or without instruction) generated strategies in speech. With feedback but without instruction, children tended to generate incorrect verbal strategies. With both feedback and instruction, children tended to generate correct verbal strategies. In brief, for this concept, feedback and instruction were not essential for strategy generation per se, but feedback was essential for generation of new verbal strategies, and instruction was essential for generation of correct verbal strategies. Not surprisingly, children in the procedure group tended to adopt the grouping procedure, which was provided in instruction. Children in the analogy group tended to generate either the equalize procedure or the grouping procedure. Children in the principle group tended to generate either the equalize procedure or an incorrect procedure.

Initial variability was also expected to be related to patterns of strategy generation. A cyclical model of strategy change would imply an inverse linear relation between initial variability and strategy generation. Children with low variability should be most likely to generate new strategies, and children with high variability should be least likely to generate strategies. Surprisingly, however, in this study, neither measure of pretest variability was related to whether or not children generated strategies, for any of the three measures of strategy generation.

It is possible the effects of initial variability depend on the nature of the instructional input provided to the child. That is, initial variability may be related to strategy generation for some types of instruction but not others. Indeed, it is possible that no overall effect of initial variability was observed in this study, because the effects of initial variability differed across conditions. To test this possibility, I ran the stepwise logistic regression again, this time including terms for the interaction between condition and initial variability (both Condition × Number of Pretest Strategies and Condition × Number of Pretest Mismatches) as potential predictors. The interaction terms did not enter the model for any of

[7] Of course, it is possible that the strategy generation observed in the control group may be illusory. That is, it is possible that the strategies children appeared to "generate" at posttest were actually present in their repertoires at pretest but were simply not sampled. The present data cannot conclusively rule out this interpretation. However, even if this interpretation is correct, the data suggest that there may have been alterations in the "strengths" of particular strategies among children in the control group, which led to their being sampled more frequently at posttest than at pretest. In this regard, it is worth noting that at least one formal model of strategy choice, the adaptive strategy choice model (Siegler & Shipley, 1995), predicts that simply solving problems can affect the strengths of various strategies.

[8] The pattern of results does not differ if children who used the grouping/add-two pattern are reclassified as not generating a correct strategy (see Footnote 6).

the three dependent measures. However, because there may have been insufficient power to detect an interaction, no firm conclusions can be drawn on this issue.

There were no effects of gender on any of the strategy generation measures, and grade level was associated with only one of the measures. The relative odds of generating a correct verbal strategy were greater among children tested early in fourth grade than among children tested either late in third grade or late in fourth grade, Wald $z^2(1) = 8.25, p < .05$.

Abandoning strategies. The process of strategy change involves not only generating new strategies but also abandoning old ones. To statistically evaluate patterns of strategy abandoning, I developed two binary outcome measures. First, children were classified on the basis of whether or not they abandoned any strategies, from either modality (speech or gesture). Table 8, column 2, indicates the proportion of children in each instruction group who abandoned a strategy from either modality. Next, children were classified on the basis of whether or not they abandoned any strategies from speech. Table 8, column 4, indicates the proportion of children in each instruction group who abandoned a verbal strategy.

First, consider the effect of instruction on whether or not children abandoned any prior strategies. As seen in Table 8, the overall rate of strategy abandoning was slightly higher in the procedure and principle groups than in the other groups. However, the relation between instruction and abandoning any strategies was not significant. Instead, instruction was related to abandoning verbal strategies. More children in the procedure group than in the other groups abandoned verbal strategies, and the relative odds of abandoning a verbal strategy were higher in the procedure group than in the feedback group.

A cyclical model of strategy change would suggest that children with high initial variability should be especially likely to abandon strategies. Indeed, as seen in Figure 3 (lines with squares), children with higher levels of initial variability were more likely to abandon at least one strategy. When both gestured and spoken strategies were considered, the relative odds of strategy abandoning increased as number of strategies increased, Wald $z^2(1) = 22.71$, $p < .001$. This pattern also held for number of pretest mismatches, Wald $z^2(1) = 15.57, p < .001$, if number of pretest strategies was omitted from the model.

When verbal strategies alone were considered, the relations between initial variability and strategy abandoning were more complex. Surprisingly, both number of pretest strategies and number of pretest mismatches were related to abandoning verbal strategies, but in opposite ways. As seen in Figure 3 (line with black circles), as total number of strategies increased, the rate of strategy abandoning increased, Wald $z^2(1) = 14.83, p < .001$. However, as total number of mismatches increased (line with white circles), the rate of abandoning verbal strategies decreased, Wald $z^2(1) = 10.52$, $p < .01$. The overall pattern suggests that children who produced many mismatches at pretest tended to abandon gestured strategies rather than verbal ones.

Indeed, it is worth noting that, in general, children were very likely to abandon strategies that they expressed uniquely in gesture on the pretest. As seen in Table 6, overall, 72% of the children who expressed one or more incorrect strategies in gesture on the pretest ($n = 61$) abandoned at least one of those strategies on the posttest, and 76% of the children who expressed one or more correct strategies in gesture on the pretest ($n = 38$) abandoned at least one of those strategies. In contrast, only 40% of children abandoned an incorrect strategy that they expressed in speech on the pretest.

To sum up thus far, patterns of change in the content of children's repertoires depended both on children's initial level of variability and on instructional input. Children who received instruction tended to generate new, correct verbal strategies and to abandon incorrect verbal strategies. Children who started with many strategies tended to abandon strategies. Thus, both initial variability and instruction contributed to changes in the content of children's repertoires.

Gradual and Abrupt Strategy Change

I next examined whether children modified their strategy repertoires gradually or abruptly. To address this issue, I first categorized

Table 8
Proportion of Children Who Abandoned Strategies or Changed Abruptly, and Odds Ratios in Relation to the Feedback Group

	Abandon any verbal or gestured strategy		Abandon any verbal strategy		Change abruptly (Abandon all + generate new)	
Condition	Proportion of children	Odds ratio[a]	Proportion of children	Odds ratio[a]	Proportion of children	Odds ratio[a,b]
Control	0.62	1.34	0.19	0.39	0.00	0.15
Feedback (reference)	0.62	1.00	0.38	1.00	0.10	1.00
Principle	0.71	3.08	0.39	1.65	0.14	1.35
Analogy	0.61	1.00	0.43	1.44	0.21	2.15
Procedure	0.72	4.96	0.66	5.02**	0.47	5.03*

[a] Odds ratios are adjusted for those measures of pretest variability (i.e., number of mismatches or number of strategies) that were significantly related to each outcome measure (see text). [b] Parameter estimates for comparisons with sampling zeroes were computed by adding 0.20 to each cell of the Condition × Grade Level × Outcome contingency table (see Footnote 5 in text).
* $p < .05$, Wald statistic. ** $p < .01$, Wald statistic.

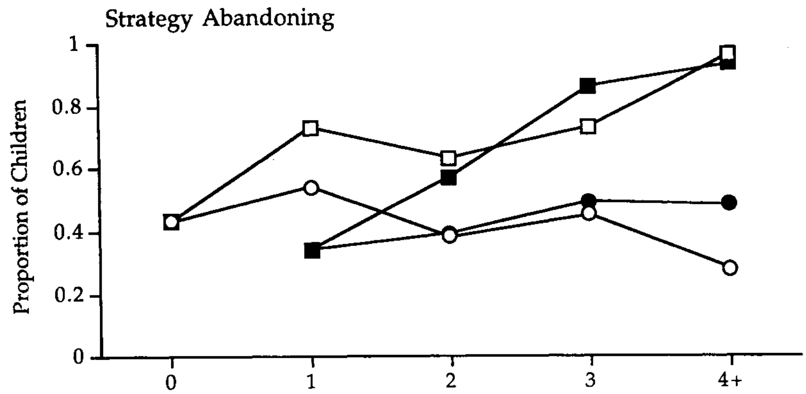

Figure 3. Proportion of children at each level of initial variability who abandoned at least one strategy from speech or gesture (squares) or from speech (circles).

children as changing abruptly or not. Abrupt change was defined as change that involved abandoning all of the strategies expressed on the explanation pretest and generating one or more new strategies on the explanation posttest. In essence, abrupt change involved replacing the repertoire of strategies used on the pretest with an entirely new repertoire of strategies on the posttest. I found that a surprisingly large proportion of the children modified their strategy repertoires abruptly—24% of the 117 children who changed their strategies over the course of the study did so abruptly. Although gradual change was more common than abrupt change in the sample as a whole, abrupt change was also relatively frequent.

Logistic regression was also used to assess the effects of initial variability and instruction group on whether or not children changed abruptly. The best-fitting model included both instruction group and initial variability. As seen in Table 8, column 6, children in the procedure group were more likely to change abruptly than were children in the feedback group. Indeed, the relative odds of abrupt change were greater in the procedure group than in the feedback group. Thus, direct instruction about a problem-solving procedure led children to modify their strategy repertoires abruptly.

As seen in Figure 4, the rate of abrupt change was inversely related to initial variability. As number of pretest mismatches increased, the relative odds of abrupt change decreased, Wald $z^2(1) = 5.16$, $p < .05$. This pattern also held (but was only marginally significant) for total number of strategies, Wald $z^2(1) = 2.90$, $p = .09$. Note that it was clearly possible for children with low initial variability to change gradually. For example, among children who produced no mismatches on the pretest and who generated strategies, 56% did so gradually, by adding a strategy to the repertoire, whereas 44% did so abruptly, by replacing their pretest repertoire.

In general, the results indicate that abrupt change was especially likely in children who were low in variability, and in response to instruction that provided children with an easily accessible problem-solving procedure. However, abrupt change was observed in children at all levels of variability, and in all the instruction groups except the control group. These findings are particularly striking because in this study strategy use was measured using gesture as well as speech. Gesture has been shown in previous work to be sensitive to the early and gradual emergence of new strategies (Alibali & Goldin-Meadow, 1993). The fact that in this study, change often appeared to be abrupt, even when a highly sensitive measure was used, lends strength to the conclusion that strategy change is frequently abrupt. The present findings show that abrupt change is especially likely for children who receive instruction about a procedure and for children who are low in initial variability.

Taken together, the results indicate that strategy change in children's learning of equivalence is most often gradual, but can also be abrupt. This finding contrasts with other reports of strategy change, in which gradual change appears to be the rule (e.g., Siegler, 1995). In this study, the rate of change depended both on the nature of the instructional input that children received and on their prior level of variability.

What Types of Transitions Were Made Abruptly?

Because correct strategies are logically superior to incorrect strategies on the mathematical equivalence task (rather than being simply more efficient), it was natural to ask whether children who acquired a correct strategy changed abruptly and whether children who generated an incorrect strategy changed gradually. To address this issue, I compared the likelihood of abrupt change for children who generated verbal strategies that were either correct or incorrect. Although the effect was in the predicted direction, the association between type of change and the correctness of the newly

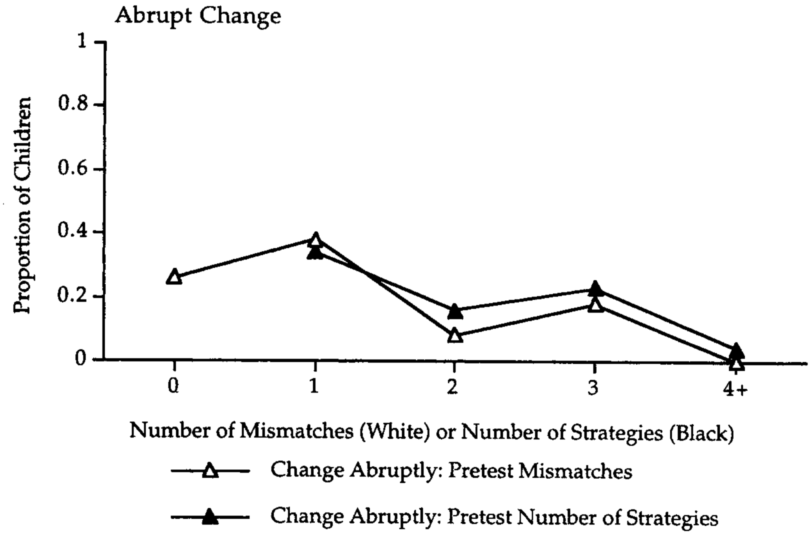

Figure 4. Proportion of children at each level of initial variability who modified their strategy repertoires abruptly.

acquired strategy was not significant, $\chi^2(1, N = 53) = 0.35, p > .50$. Collapsing across experimental groups, 56% of the 34 children who acquired a correct verbal strategy changed abruptly, whereas 47% of the 19 children who acquired an incorrect verbal strategy changed abruptly. Thus, even children who added incorrect strategies to their repertoires often changed abruptly.

I next considered whether abrupt change was associated with certain types of changes in variability. For example, it is conceivable that decreases in variability might occur abruptly (e.g., by replacing a set of incorrect strategies with a single, correct strategy), whereas increases in variability would occur more gradually (e.g., by adding a strategy). To address this issue, I examined how changes in the size of children's strategy repertoires were related to changes in content of the repertoires. Each child was classified into one of three groups based on change in the number of strategies used: (a) increase, (b) decrease, or (c) no change. Next, I assessed the types of repertoire changes made by children in each group.

As seen in Table 9, most change was consistent with a gradualist account. That is, when children increased the size of their repertoire, they typically did so by maintaining at least one strategy and generating another (59% of cases). Similarly, when children decreased the size of their repertoire, they typically did so by maintaining at least one strategy and abandoning another (64% of cases). Finally, when children did not change the size of their repertoire, they typically did so by maintaining their repertoire intact (51% of cases).

However, the results in Table 9 highlight that, to understand change, one must consider not only the size of the repertoire but also the content. Many of the observed changes were not consistent with a gradualist account. Even when changes in the size of the repertoire might suggest a smooth change (e.g., an increase in size of the repertoire suggests the gradual addition of strategies), close examination sometimes revealed a more abrupt change (e.g., abandoning a prior strategy for a new set of strategies). Abrupt change was observed for all types of changes in number of strategies (i.e., increase, decrease, no change). Furthermore, there was no significant association between type of change in number of strategies and the shape of change (gradual vs. abrupt), $\chi^2(2, N = 117) = 3.58, p > .10$. Thus, although change often involves a smooth and gradual modification of the strategy repertoire, change also fre-

Table 9
Proportion of Children Who Increased, Decreased, or Did Not Change in Number of Strategies Who Made Each Type of Repertoire Change

Change in no. of strategies	No change	Gradual change			Abrupt change	
	Maintain only	Maintain + generate	Maintain + abandon	Maintain + abandon + generate	Abandon + generate	*n*
Increase	NA	0.59	NA	0.26	0.15	39
Decrease	NA	NA	0.64	0.11	0.25	53
No change	0.51	NA	NA	0.31	0.18	51

Note. NA = not applicable.

quently involves abrupt switches from one strategy (or set of strategies) to another.

Discussion

This study investigated variability and strategy change in children learning the concept of mathematical equivalence. Overall levels of variability were high throughout the study, and changes in levels of variability did not depend on instruction. However, changes in the content of children's strategy repertoires depended both on instruction and on initial variability. Children who received instruction were especially likely to generate new strategies, and children who started with high variability were especially likely to abandon at least one strategy. Children who started with low variability were most likely to change their strategy repertoires abruptly.

If strategy change involves shifts from low to high variability (Alibali & Goldin-Meadow, 1993), one might expect that low initial variability would be associated with strategy generation. However, one could also argue that children with high initial variability are most responsive to instruction (Church & Goldin-Meadow, 1986; Perry et al., 1988), so such children should be especially likely to adopt new, correct strategies. In this study, high initial variability was related to strategy abandoning but not to strategy generation. Together with previous findings, the present results suggest that the relation between initial variability and strategy generation may differ for different types of instruction. That is, initial variability may be important for strategy generation in response to some types of instruction but not others. When explicit instruction is provided, as in the present study, the effects of instruction may override potential effects of initial variability. In the present study, initial variability appeared not to be a prerequisite for or even a facilitator of strategy generation. However, the present findings leave open the possibility that variability (or lack of variability) may play a causal role in strategy generation under some circumstances.

A majority of the strategy changes observed in this study involved a smooth and gradual modification of the strategy repertoire. Children most often added or abandoned strategies while maintaining at least one strategy. However, abrupt strategy change was also surprisingly frequent—indeed, 24% of the children who modified their repertoires did so abruptly. Most of the children who changed abruptly did so after receiving direct instruction about a correct strategy. Thus, the shape of strategy change depended on the type of instruction that children received. The findings suggest that there is no single right answer to the question of how strategy change occurs. Instead, theories and models about the process of strategy change must be able to account for both gradual and abrupt change. Such theories and models must also be able to respond differently to different types of environmental input.

The frequency of abrupt strategy change documented in this study is striking, because several features of the design might have been expected to promote gradual rather than abrupt change. First, the instruction provided was very brief. If it had been more intensive, one might expect even more abrupt change. Second, children's strategy use was observed over a fairly short period of time (approximately 40 min). If the interval between the pretest and posttest had been longer, changes that were truly gradual (as revealed in this study) might have appeared to be abrupt. Because children's strategy use was monitored over the entire interval, such changes were accurately identified as gradual in the present study.

It is possible that the abrupt strategy change observed in this study may be a result of more gradual change at still lower levels (e.g., at the level of problem encoding [Alibali, McNeil, & Perrott, 1998] or of strategy activation). Nevertheless, the present findings show that, even when assessed at a fine-grained level, change can be sudden and abrupt. The present results highlight an important question for further research: Under what circumstances do children change gradually, and under what circumstances do they change abruptly? In the present study, the rate of abrupt strategy change was highest among children who received instruction about a problem-solving procedure. This finding suggests that the gradual strategy change documented in previous studies may be due to the types of interventions used in the studies. The present research represents a first step toward identifying conditions that influence the shape of strategy change.

It might be argued that at least some of the present findings could be explained by simple regression to the mean. In particular, regression to the mean might account for the finding that children who started with many strategies were especially likely to abandon at least one strategy. However, regression to the mean would also suggest that children who started with few strategies should be especially likely to generate strategies. This was not the case. Thus, the overall pattern of results is not consistent with regression to the mean as an alternative explanation.

Insights About Strategy Change From Children's Gestures

In the present study, children's gestures revealed several important pieces of information about their strategy use. First, many children in the control group generated strategies that they expressed uniquely in their gestures (and never in speech). Without evidence from gestures, one might have concluded that strategy generation is highly infrequent when feedback and instruction are not provided. Evidence from gestures suggests a different conclusion, namely, children often consider new strategies, even without feedback or instruction. When children find that their strategies are incorrect or inadequate, they may draw on this "reserve capital" in constructing new strategies (see Siegler, 1984).

Second, strategy change was frequently abrupt, even when assessed by both gesture and speech. Because children's gestures can reveal both the early emergence of new strategies and the late disappearance of old strategies, evidence from gesture allows stronger conclusions to be drawn about the abruptness of strategy change.

Third, much of children's experimentation with new strategies was evident only in gestures and not in speech. Many of the strategies that children abandoned were expressed uniquely in gesture on the pretest, and many new strategies were expressed uniquely in gesture on the posttest. Thus, gesture provides a window on variability that would go unnoticed if strategy use were "diagnosed" through speech alone.

Fourth, initial variability, as assessed in gesture and speech, was an important predictor of strategy abandoning and abrupt change. In the present study, variability was operationalized in two different ways: total number of strategies expressed and number of gesture–speech mismatches produced. These two measures were highly correlated, as they should be if both measure the same construct. Furthermore, the pattern of results was comparable for both measures: Children with high initial variability were especially likely to abandon strategies, and children with low initial variability were especially likely to modify their strategy repertoires abruptly. Because children's strategy use was assessed in gesture as well as speech, the study provided converging evidence from two measures about the importance of initial variability.

Sources of New Strategies

This study provided several pieces of evidence about the sources of new strategies. First, a surprisingly large number of children in the control group spontaneously generated new strategies. Thus, it appears that simply explaining problems is enough to encourage children to generate new strategies. Furthermore, all of the strategies generated by children in the control group were expressed uniquely in gesture. This finding underscores the importance of using microlevel methods in studying processes of change. This finding also highlights that behavioral variability is pervasive, although it is sometimes manifested in subtle ways.

Second, feedback about whether solutions were correct also plays a role in strategy generation. Overall levels of generation were comparable in the feedback and control groups; however, children in the feedback group generated verbal strategies more frequently than children in the control group. Taken together, these results suggest that feedback did not necessarily help children to construct new strategies, but instead encouraged children to express their new strategies in speech, rather than uniquely in gesture. Feedback appeared to make children aware that they should *say* something different.

Finally, instruction both increased the likelihood that children would generate strategies and increased the likelihood that those new strategies would be correct. Levels of strategy generation were elevated in all three instructed groups, most strikingly in the procedure group. Furthermore, no children in the control or feedback groups expressed correct procedures in speech on the posttest problems, whereas many children in each of the instructed groups did so. However, it should be noted that the three types of instruction used in this study were not equally effective at promoting strategy generation or at promoting change from incorrect to correct strategies. Fewer children in the principle and analogy groups than in the procedure group generated correct strategies after the intervention; however, more children in the analogy group than in the procedure group solved the far-transfer problems correctly.

To solve equivalence problems correctly, children need to understand the fundamental notion that two quantities can be equal, as well as to understand cultural conventions such as the symbolic notation used in equations and the meaning of the equal sign (Rittle-Johnson & Alibali, in press). As Geary (1995) suggested, the acquisition of knowledge that is culturally based (in Geary's terms, "biologically secondary" [p. 25]) is likely to depend heavily on instruction. The present study shows that instruction is essential for learning a culturally based concept such as symbolic equivalence. However, for other types of knowledge that are not culturally based (in Geary's terms, "biologically primary" [p. 25] knowledge, such as some fundamental number concepts), explicit instruction is likely to be less important.

The present findings suggest that both procedural and conceptual instruction may be important for learning the concept of equivalence (see also Perry, 1991). However, because it is unclear whether most children in the principle and analogy groups actually understood the lessons, it is inappropriate to draw conclusions about which type of instruction is most effective. Nevertheless, at least two educational messages can be drawn from this study.

Educational Implications

One educational message has to do with the fact that both the principle and analogy interventions were relatively ineffective at promoting change from use of incorrect strategies to use of correct strategies. This stands in contrast to other studies of mathematical equivalence, in which brief, principle-based instruction was highly effective at promoting both learning and transfer (Perry, 1991; Rittle-Johnson & Alibali, in press).

One possible explanation for this finding is that, unlike in previous studies, the instructor in this study did not produce any gestures herself during the training interventions. Instead, the lessons were provided in strictly verbal form (in the principle group) or in verbal form along with a drawing (in the analogy group). It is possible that children in the principle and analogy groups had difficulty understanding the abstract language used in the interventions (e.g., "before the equal sign") because it was provided in verbal form alone, without gestures linking the abstract words to their concrete referents. In fact, anecdotal evidence from children's errors supports this interpretation (e.g., on the posttest, one child attempted to set the right side of the equation equal to the *first addend* on the left side). Thus, it may be crucial for students' learning that instruction is provided in multiple modalities, particularly if the ideas to be conveyed are abstract. An experimental study by Perry et al. (1995) provides some evidence for this idea; however, further research on this issue is needed.

A second educational message has to do with the effectiveness of procedure-based instruction. In this study, procedural instruction led to a great deal of learning, as well as to relatively frequent abrupt strategy change. However, procedural instruction also led to only limited transfer to problem types not used in instruction. Thus, strategy change due to procedural instruction may be superficial and without deep understanding.

The present findings corroborate those of Perry (1991) and Rittle-Johnson and Alibali (in press), who also found that principle-based instruction tended to support broader transfer than procedure-based instruction. As Perry suggested, if instruction provides an easily accessible procedure for solving a problem, the learner may not consider the rationale underlying that procedure. Specific patterns of transfer in the procedure group support this interpretation. When they transferred at all, children in this group tended to transfer only to problem types on which they could succeed by applying their newly learned procedure by rote. That is, they succeeded only on problem types in which the correct solution could be reached through a strategy of canceling equivalent operands and then performing the required operation. No child in this group successfully modified the procedure in a way that revealed a deep understanding of the rationale underlying the grouping procedure. For example, no child canceled part of an addend in a nonequivalent-operands problem (e.g., for the problem $4 + 3 + 6 = \_\_\_ + 5$, no child canceled the right 5 and the left 6, leaving 1 on the left side, and then summed 4, 3, and 1).[9]

From an educational standpoint, it is worth considering whether abrupt strategy change is a worthwhile goal, particularly if that change is not accompanied by transfer. Although it may seem best to help children learn as quickly as they can, abrupt change that occurs without deep understanding may be less desirable than no change at all. Gradual change may lead to more robust and flexible learning outcomes than abrupt change, particularly if the gradual change is accompanied by reflection about how old and new strategies relate to one another.

In sum, many factors contribute to shaping the path of strategy change. Among them are the level of variability of children's knowledge and the nature of the instruction that children receive. In this study, a majority of the observed strategy changes involved maintaining at least one strategy, while gradually adding or abandoning others. However, abrupt change was also frequent, particularly among children who received procedure-based instruction and among children who were low in variability. In characterizing the process of strategy change, one must consider not only the variability of children's behavior but also the content of their strategy repertoires, because children have many different ways to change their minds.

[9] This modification of the grouping strategy has been observed on transfer problems in other studies (e.g., Rittle-Johnson & Alibali, in press).

References

Acredolo, C., & O'Connor, J. (1991). On the difficulty of detecting cognitive uncertainty. *Human Development, 34,* 204–223.

Acredolo, C., O'Connor, J., & Horobin, K. (1989, March). *Children's understanding of conservation: From possibility to probability to necessity.* Poster presented at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.

Adolph, K. E. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development, 62*(Serial No. 251).

Alibali, M. W. (1994). *Processes of cognitive change revealed in gesture and speech.* Unpublished doctoral dissertation, University of Chicago.

Alibali, M. W., & Goldin-Meadow, S. (1993). Transitions in learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523.

Alibali, M. W., McNeil, N. M., & Perrott, M. A. (1998). What makes children change their minds: Changes in encoding lead to changes in strategy selection. In M. A. Gernsbacher & S. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 36–41). Mahwah, NJ: Erlbaum.

Baroody, A., & Ginsburg, H. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal, 84,* 199–212.

Bishop, Y. V. V., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*(Serial No. 246).

Chen, Z., & Daehler, M. W. (1989). Positive and negative transfer in analogical problem solving by 6-year-old children. *Cognitive Development, 4,* 327–344.

Church, R. B. (1987). *The mismatch between gesture and speech as an index of transitional knowledge.* Unpublished doctoral dissertation, University of Chicago.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23,* 43–71.

Crowder, E. (1996). Gestures at work in sense-making science talk. *Journal of the Learning Sciences, 5,* 173–208.

Crowley, K., & Siegler, R. S. (1993). Flexible strategy use in young children's tic-tac-toe. *Cognitive Science, 17,* 531–561.

Ellis, S., Siegler, R. S., & Klahr, D. (1993, March). *Effects of feedback and collaboration on changes in children's use of mathematical rules.* Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA.

Flavell, J. (1971). An analysis of cognitive developmental sequences. *Genetic Psychology Monographs, 86,* 279–350.

Garber, P., Alibali, M. W., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Development, 69,* 75–84.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50,* 24–37.

Gentner, D., Ratterman, M. J., Markman, A., & Kotovsky, L. (1995). Two forces in the development of relational similarity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 263–313). Hillsdale, NJ: Erlbaum.

Goldin-Meadow, S., & Alibali, M. W. (1994). Do you have to be right to redescribe? [Review of the book *Beyond Modularity* by A. Karmiloff-Smith]. *Behavioral and Brain Sciences, 17,* 718–719.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review, 100,* 279–297.

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development, 12,* 93–120.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior* (Vol. 17, pp. 1–44). New York: Academic Press.

McClelland, J. (1995). A connectionist perspective on learning and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Erlbaum.

McNeill, D. (1992). *Hand and mind.* Chicago: University of Chicago Press.

Miller, P. H., & Seier, W. (1994). Strategy utilization deficiencies in children: When, where, and why. *Advances in Child Development and Behavior, 25,* 107–156.

Norusis, M. J. (1994). *SPSS advanced statistics 6.1.* Chicago: SPSS, Inc.

Perry, M. (1985, April). *Children's strategies in interpreting symbolic equivalence.* Paper presented at the biennial meeting of the Society for Research in Child Development, Toronto, Ontario, Canada.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6,* 449–468.

Perry, M., Berch, D., & Singleton, J. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues, 6,* 213–235.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3,* 359–400.

Piaget, J. (1985). *The equilibration of cognitive structures.* Chicago: University of Chicago Press. (Original work published 1975)

Rittle-Johnson, B., & Alibali, M. W. (in press). Conceptual and procedural knowledge: Does one lead to the other? *Journal of Educational Psychology.*

Siegler, R. S. (1984). Mechanisms of cognitive growth: Variation and selection. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 141–162). Prospect Heights, IL: Waveland Press.
Siegler, R. S. (1988). Mechanisms of cognitive development. *Annual Review of Psychology, 40,* 353–379.
Siegler, R. S. (1995). How does change occur? A microgenetic study of number conservation. *Cognitive Psychology, 28,* 225–273.
Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York: Oxford University Press.
Siegler, R. S., & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.
Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Erlbaum.
Stigler, J. W., Fernandez, C., & Yoshida, M. (1996). Cultures of mathematics instruction in Japanese and American elementary classrooms. In T. Rohlen & G. LeTendre (Eds.), *Teaching and learning in Japan* (pp. 213–246). Cambridge, England: Cambridge University Press.
Stigler, J. W., & Stevenson, H. W. (1992). *The learning gap.* New York: Simon & Schuster.
Strauss, S., & Rimalt, I. (1974). Effects of organizational disequilibrium training on structural elaboration. *Developmental Psychology, 10,* 526–533.
Syc, S. E. (1992). *Teacher responsivity to children's gesture and speech communication.* Unpublished doctoral dissertation, University of Chicago.
Thelen, E. (1989). Self-organization in development processes: Can systems approaches work? In M. Gunnar & E. Thelen (Eds.), *Minnesota symposium on child psychology: Vol. 22. Systems and development* (pp. 77–117). Hillsdale, NJ: Erlbaum.
Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press.
Thelen, E., & Ulrich, B. D. (1991). Hidden skills. *Monographs of the Society for Research in Child Development, 56*(Serial No. 223).

Received September 20, 1996
Revision received June 17, 1998
Accepted June 30, 1998 ■