TARGET ARTICLE WITH COMMENTARY

Differentiation and integration: guiding principles for analyzing cognitive change

Robert S. Siegler¹ and Zhe Chen²

1 Department of Psychology, Carnegie Mellon University, USA
2 Department of Human and Community Development, University of California at Davis, USA

Abstract

Differentiation and integration played large roles within classic developmental theories but have been relegated to obscurity within contemporary theories. However, they may have a useful role to play in modern theories as well, if conceptualized as guiding principles for analyzing change rather than as real-time mechanisms. In the present study, we used this perspective to examine which rules children use, the order in which the rules emerge, and the effectiveness of instruction on water displacement problems. We found that children used systematic rules to solve such problems, and that the rules progress from undifferentiated to differentiated forms and toward increasingly accurate integration of the differentiated variables. Asking children to explain both why correct answers were correct and why incorrect answers were incorrect proved more effective than only requesting explanations of correct answers, which was more effective than just receiving feedback on the correctness of answers. Requests for explanations appeared to operate through helping children notice potential explanatory variables, formulate more advanced rules, and generalize the rules to novel problems.

Introduction

The importance of differentiation and integration have been highlighted by a variety of classic theorists: Baldwin (1896), G.H. Mead (1913), Lewin (1935), McGraw (1935), Werner (1948), Piaget (1952), Witkin, Lewis, Hertzman, Machover, Meissner and Wapner (1954), Gibson and Gibson (1955), and Bruner, Olver and Greenfield (1966), among them. Perhaps the most famous statement of the importance of these concepts was Werner's orthogenetic principle, which stated, 'Wherever development occurs, it proceeds from a state of relative globality and lack of differentiation to a state of increasing differentiation, articulation, and integration' (Werner, 1957, p. 126). Differentiation and integration were hypothesized to work in a hand-over-hand fashion, with initial differentiation of distinctive features allowing an initial integration of the features, which in turn allowed more nuanced differentiation and more comprehensive integration, and so on. The classic developmental theorists applied these concepts to a wide range of changes, including motor, perceptual, and conceptual development; changing understanding of the self; emergence of individual differences in field dependence/ independence; and acquisition of nuanced emotions.

One persuasive illustration of the usefulness of differentiation and integration for understanding cognitive change is provided by music perception (Miller, 2002). On first hearing a symphony, most listeners have only a vague impression of the piece, perhaps augmented by recollection of a melody or two. Subsequent exposures produce differentiation of the movements and their major themes and melodies into coherent and memorable units; the subsequent exposures also produce integration of the movements, themes, and melodies into a meaningful whole. Intuitively appealing examples such as this one, together with the widespread applicability of the concepts, were key contributors to the prominence of differentiation and integration within classic developmental theories.

More recently, however, the concepts have almost disappeared from use. For example, in the volume titled Cognition, perception, and language in the two most recent editions of Handbook of child psychology (Kuhn & Siegler, 1998, 2006), which had almost entirely non-overlapping authors, index entries for the terms 'differentiation' and 'integration' appeared in 0 of 19 and 0 of 22 chapters, respectively. In contrast, in the corresponding volume of the 1970 Handbook of child psychology, seven of the 19 chapters included index entries for the terms, and the concepts played a major role in at least four of them (Kagan & Kogan, 1970; Langer, 1970; Pick & Pick, 1970; Thompson & Grusec, 1970). This change cannot be explained by researchers substituting synonymous terms for the same concepts; for example, the related terms 'information integration', 'holistic perception', and 'analytic perception' also did not appear as index terms for any of the 41 chapters in the two Handbook volumes.

Address for correspondence: Robert S. Siegler, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA; e-mail: rs7k@andrew.cmu.edu

© 2008 The Authors Journal compilation © 2008 Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.
The relegation of differentiation and integration to the role of historical precursors occurred for several reasons. One was that the terms were used in many different ways, which led to their meaning becoming murky. For example, 'differentiation' was used to account for changes from 'primitive' to 'advanced' civilizations (Werner & Kaplan, 1956), the differentiation of the self into many distinct 'me's' (Mead, 1913), the differentiation of emotions from global to nuanced (Lewin, 1935), and the identification of distinctive features in letter-like forms (Gibson, Gibson, Pick & Osser, 1962). Some theorists used the terms as descriptions of real-time mechanisms, whereas others used them as descriptions of the outcomes of development. Another problem arose in testing whether the concepts were in fact involved in specific transitions. For example, Werner and Kaplan (1956, p. 866) suggested that differentiation and integration allowed 'coordinating within a single descriptive framework psychological phenomena observed in phylogenesis, ontogenesis, psychopathology, ethnopsychology, etc.' Although such a proposal was intriguing, it was unclear how to test whether the same processes were in fact involved in such diverse domains.

The demise of the concepts of differentiation and integration, and other broad integrative concepts such as assimilation, accommodation, and equilibration, is understandable, but it has had several adverse consequences. One is fragmentation of our understanding of development and learning. This fragmentation has been noted by a number of theorists, several of whom have traced the lack of coherence to a lack of integrative concepts (e.g. Butterworth & Bryant, 1990; Case & Okamoto, 1996; Demetriou & Kazi, 2001; Flavell, 1984; Miller, 2002). A further adverse consequence has been a lack of principles that allow prediction of developmental sequences for tasks and domains in which development has not yet been studied.

In the present article, we propose that the concepts of differentiation and integration may have a useful role to play as general principles for predicting the approaches that children and novices in a domain will use on unfamiliar tasks, the order in which different approaches will emerge, and the types of learning and instructional processes likely to be involved in acquisition of new approaches. Viewing differentiation and integration as guiding principles maximizes their strengths—broad applicability and ability to integrate diverse particular changes—and minimizes their weakness—lack of detailed specification of how the process works.

The argument is not that analysis at this relatively abstract level is preferable to analysis at the level of more specific real-time mechanisms. Rather, we believe that analysis at multiple levels, including relatively abstract ones, has something to offer above and beyond analyses of highly specific real-time mechanisms. The level of guiding principles allows predictions on many tasks where specification of real-time mechanisms remains well in the future. In addition, the relatively abstract analyses allow us to perceive commonalities across tasks that superficially have little in common. The present study is intended to illustrate these advantages in the context of the development of understanding of quantitative concepts.

### Differentiating and integrating quantitative concepts

We define differentiation as a process of distinguishing between variables that were not previously distinguished. Differentiation can occur through a variety of mechanisms. Consider three routes through which a young child might come to differentiate between her own memories and those of other people. One route would be by hearing a teacher say that different people remember things differently, understanding the point, and thereafter drawing the distinction. Another route would be by inducing the same lesson after many experiences with people remembering the same event differently. A third route would be by reflecting on a dramatic instance of differing memories and drawing the conclusion for oneself. Similarly, a child might come to differentiate the height of a liquid from its volume through someone explaining the difference, from the cumulative impact of many experiences, or through analyzing a single dramatic experience. The variety of mechanisms through which differentiation can occur is a major reason why treating differentiation as a specific, real-time mechanism has not worked out well. The pervasiveness of the phenomenon is the reason why the concept is useful.

We define integration as the combining of variables into a problem-solving approach (e.g. a rule, algorithm, or strategy). Such integrations can be either correct or incorrect. For example, a child might estimate the relative areas of rectangles by adding the height and width, thus integrating the correct variables in an incorrect way (Wilkening, 1982). Alternatively, a child might integrate the wrong variables; for example, 11-year-olds often integrate the beginning and end points of travel to judge the relative time of travel of two objects moving along parallel paths from the same beginning point (Siegel, 1983). Differentiating the relevant variables is necessary but not sufficient for correctly integrating them.

The importance of differentiating among variables is evident on many of Piaget's classic problems. For example, on conservation of liquid quantity, children need to differentiate the quantitative dimensions of height and volume and to realize that the glass with the taller liquid column does not necessarily have the greater volume of water. These tasks are challenging in large part because the relevant quantitative dimensions are highly correlated in the everyday environment, and because the relative values on the most salient dimension (e.g. height) suggest an incorrect answer. Table 1 provides a list of 10 classic Piagetian and neo-Piagetian tasks that pose similar challenges.
Table 1  Classic developmental tasks requiring differentiation of variables

<table>
<thead>
<tr>
<th>Task</th>
<th>Dimension to be differentiated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number conservation (Piaget, 1952)</td>
<td>Number of objects vs. length of row</td>
</tr>
<tr>
<td>Liquid conservation (Piaget, 1952)</td>
<td>Amount of water vs. height of liquid column</td>
</tr>
<tr>
<td>Class inclusion (Ahr &amp; Youniss, 1970)</td>
<td>Number of objects in set and in larger subset</td>
</tr>
<tr>
<td>Probability (Inhelder &amp; Piaget, 1958)</td>
<td>Ratio of objects and number of desired-color objects</td>
</tr>
<tr>
<td>Balance scale (Inhelder &amp; Piaget, 1958)</td>
<td>Weight and torque</td>
</tr>
<tr>
<td>Shadow projection (Inhelder &amp; Piaget, 1958)</td>
<td>Length of object and length of shadow</td>
</tr>
<tr>
<td>Distance (Acredolo &amp; Schmid, 1981)</td>
<td>Endpoint of travel vs. distance traveled</td>
</tr>
<tr>
<td>Fullness (Bruner, 1964)</td>
<td>Height of liquid and ratio of filled to empty space</td>
</tr>
<tr>
<td>Sweetness (Story et al., 1982)</td>
<td>Amount of sugar in each glass vs. ratio of sugar to water</td>
</tr>
<tr>
<td>Happiness (Marini &amp; Cuse, 1989)</td>
<td>Number of marbles received vs. ratio of number of marbles received to number of marbles desired</td>
</tr>
</tbody>
</table>

Piaget explained these and a wide range of other changes in terms of assimilation, accommodation, and equilibration. These constructs are so general, however, that it is unclear what they imply, beyond the undisputed facts that existing knowledge influences learning and that learning influences future thinking. In contrast, differentiation and integration specify the goals that learners need to accomplish on the tasks listed in Table 1 and many others: Differentiating among positively correlated dimensions and integrating the dimensions into an appropriate problem-solving approach. Thus, differentiation and integration provide a more specific framework than assimilation and accommodation for thinking about commonalities among the goals that children need to accomplish on a variety of superficially different, cognitive developmental tasks. Unlike assimilation and accommodation, they do not apply to all cognitive changes, just a substantial number.

Performance on the quantitative reasoning tasks listed in Table 1 and many others shows a common developmental sequence: Very young children (e.g., 3-year-olds) do not rely consistently on any dimension and answer correctly at a chance level; somewhat older children (e.g., 5-year-olds) answer consistently incorrectly, because they base responses on a misleading but salient, quantitative dimension; yet older children (e.g., those 9 years and older) consistently solve the problems correctly or at least take into account multiple relevant quantitative dimensions. On a number of the problems, children also progress through intermediate knowledge states, in which they consider one or more relevant dimensions under some but not all circumstances. For example, on probability problems (Inhelder & Piaget, 1958), the number of desired-color marbles needs to be differentiated from the ratio of desired to undesired-color marbles. In the everyday environment, the two dimensions are positively correlated, but on this task they are not. Most 3-year-olds do not respond in accord with any apparent rule; most 5-year-olds predict that they are more likely to get a marble of the desired color from the set with the larger number of marbles of that color, regardless of the number of undesired color marbles; and most children 8 years and older consider both the number of desired and the number of undesired color marbles and answer correctly (Siegler, 1981).

The water displacement task

The task that was presented to children in the present study was one that had received only minimal study previously: Inhelder and Piaget's (1958) water displacement task. On each trial, first through fourth graders were presented two identical beakers containing equal amounts of water. Then, the children were shown two objects, told their relative sizes and weights and that either both would float or both would sink, and asked which object would cause the water to rise higher if one object was placed in each beaker. The complexity of the problem stems from the fact that one variable (weight) determines water displacement when objects float, whereas a different variable (volume) determines displacement when they sink. To state the principle more formally: Sunken objects displace a quantity of water equal to their volume, floating objects displace a quantity of water equal to their weight.

This problem was of interest for several reasons. First, it requires differentiation of weight and volume, two quantitative dimensions that are highly correlated in the everyday environment and that even adolescents often confuse (Piaget, 1952). It was also of interest because it addresses the concept of interactions among variables in a particularly direct way. In water displacement, the relevance of all variables depends on the states of other variables. Weight matters when the objects float; volume matters when the objects sink; no variable matters across all displacement problems. A third source of interest was that the task is related to a milestone in the history of science: Archimedes' principle of buoyancy. This principle states that a body immersed in a fluid, either wholly or partially, is buoyed up by a force equal to the weight of the displaced fluid. A floating object displaces an amount of fluid equal to the object's weight, whereas an object that is totally immersed displaces an amount of fluid equal to the object's volume (as illustrated in the proverbial tale of Archimedes' insight in the bathtub, in which the mathematician realized that he could determine whether the king's crown was made of pure gold by examining the amount of water it displaced).

Following Inhelder and Piaget's (1958) pioneering study, several other studies (Berzonsky, 1971; Hobbs, 1973; Linn & Pulos, 1983; Rowell, 1983) examined children's and adolescents' understanding of the water displaced by
sunken objects (though no previous studies have examined understanding of the full principle of buoyancy, which includes floating as well as sunken objects). The basic finding from the prior studies was that even young adolescents rarely solved these problems accurately. For example, Linn and Pulos (1983) and Hobbs (1973) found that less than half of 13-year-olds relied on volume alone to predict which of two sunken objects would displace more water. A second finding was that 12- to 14-year-olds' accuracy in predicting the outcomes of water displacement improved after feedback (Burbules & Linn, 1988). The present study was designed to build on these findings by examining how children differentiate and integrate the quantitative dimensions relevant to water displacement and by observing how children acquire more advanced understanding.

The present study

In the present study, we assessed first through fourth graders' initial knowledge about water displacement, then examined their ongoing learning in response to three potentially instructive experiences that were intended to promote differentiation and integration of the relevant variables, and finally assessed their knowledge following the experiences. The learning experiences involved providing all children with feedback, also providing some children with encouragement to explain why the observed outcome occurred, and providing other children with encouragement to explain both why the observed outcome occurred and why another plausible outcome did not occur. A componential analysis made it possible to analyze the real-time processes through which each of these instructional procedures exercised its effects, how some conditions led to greater learning than others, and how developmental and individual differences in learning arose among children who received the same experimental procedure.

We hypothesized that both older and younger children would use systematic rules to predict water displacement. The reasoning underlying this hypothesis was that the water displacement task met all four criteria proposed by Siegler (1996a) regarding the types of tasks on which children are most likely to use systematic rules. The hypothesized criteria were that the task (a) be unfamiliar, (b) require a quantitative comparison, (c) require a choice among discrete response alternatives, and (d) include one or more salient but misleading dimensions. The water displacement task possessed all four features: It was almost certainly novel to young elementary school children; it required a comparison of the relative heights of the water following addition of an object to each glass; it required a choice among three discrete response alternatives (this displacement greater, that displacement greater, or equal displacement); and it included two salient dimensions (weight and volume) that suggested incorrect rules regarding water displacement (the larger object always displaces more water, the heavier object always displaces more water). In all four respects, the water displacement task resembled other problems on which children have been found to use systematic rules, such as the balance scale, shadows projection, probability, and liquid quantity conservation tasks.

Predictions about rule use

What rules would children use to solve water displacement problems, and what would be the developmental ordering of the rules? Differentiation and integration provided principles for predicting the likely rules and their developmental ordering. The principles suggest that conceptual understanding should progress from undifferentiated earlier forms to progressively more differentiated and complexly integrated ones. Figure 1A describes an end-state rule, a structure that integrates the influences of weight, volume, floating, and sinking in a way that produces 100% correct performance.

This structure, together with the principle of differentiation and integration, suggested five approaches that children might adopt for solving water displacement problems. We labeled these five approaches: Correct, Partially Correct, Weight/Volume, More-Is-More, and Guessing. Schematic descriptions of these rules are provided in Figures 1A to 1D; here we provide verbal descriptions.

Children who used the Correct Rule (Figure 1A) would predict that on problems involving floating objects, the heavier object would cause a greater increase in the water level; on problems involving sunken objects, the larger object would cause the greater increase. This approach reflects full differentiation of the relevant dimensions and their correct integration.

Children who used the Partially Correct Rule (Figure 1B) would possess one-half of the hierarchical structure depicted in Figure 1A but not the other half. Thus, they would proceed in the same way as children who used the Correct Rule either on problems involving floating objects or on problems involving sunken objects but not on both. On the other type of problem, children who used the Partially Correct Rule would guess or use some other unsystematic approach. This approach reflects full differentiation of the relevant variables—children would respond differently to weight, volume, sinking, and floating—but only partial integration of them into a single structure.

Children who used the Weight/Volume Rule would rely entirely on one quantitative dimension, either weight or volume, regardless of whether the objects sank or floated (Figure 1C). A child who used the W version of this rule would consistently predict that the heavier objects would make the water rise more, regardless of whether the object floated or sank; a child who used the V version of the rule would consistently predict that the larger object would make the water rise more, again regardless of whether the object floated or sank. This would represent
differentiation of the weight and volume variables but not of floating and sinking, and without any integration.

The More-Is-More Rule involved reliance on whichever dimension (weight or volume) was unequal on problems on which the other characteristic was equal, and no consistent approach on problems on which one object was bigger and the other heavier (Figure 1D). This type of rule has been found with other quantitative tasks, including time, number conservation, and balance scale tasks (Levin, 1982; Richards & Siegler, 1981; Siegler, 1981). The rule reflects not only a lack of differentiation of floating and sinking events but also incomplete differentiation of weight and volume, in that the two dimensions are treated as interchangeable; whichever dimension has unequal values determines the child's prediction. It closely resembles the approach labeled Rule I' on the balance scale and number conservation tasks (Siegler, 1981) in its reliance on a dimension only when values on the dimension are unequal. Like Rule I', it was hypothesized to emerge before consistent reliance on a single dimension (the W/V Rule).

Finally, some children might guess or respond unsystematically. This could occur if children entirely failed to differentiate the relevant dimensions from irrelevant ones, or if children switched their response criteria from trial to trial.

Predictions regarding the ordering of these rules were based on the general principle that children first need to differentiate the relevant variables from the irrelevant ones, then need to differentiate the relevant dimensions from each other, and then need to form an integrated structure for representing the conditions under which each variable is relevant.

This principle suggested the following developmental sequence. First, children would respond unsystematically, because they had not differentiated the relevant variables from the irrelevant ones. Then, they would use the More-Is-More Rule, which requires identification of weight and volume as relevant but does not require differentiating between them. Instead, children use each when values on the other dimension are equal. Next, they would use the Weight/Volume Rule, which requires differentiating between the two relevant variables and relying on one of them on all types of problems, but that does not require any integration of variables. Then, children would rely on the Partially Correct Rule, which demands differentiating between weight and volume and also constructing half of the hierarchical relation – either the relation between floating and weight or the relation between sinking and weight, but not both. Finally, children would rely on the Correct Rule, which requires differentiation of weight and volume and construction of the entire hierarchical structure.

**Rule assessment method**

Several methods have proved useful for assessing whether children use rules: latent class analysis (e.g. Jansen & van
nder Maas, 2002), information integration (e.g. Anderson, 1981; Wilkening, 1982), and rule assessment (e.g. Siegler, 1981). As noted in previous discussions (e.g. Siegler & Chen, 2002), these approaches have distinctive strengths and weaknesses. Latent class analysis offers precise and continuous measures of the statistical fit between the data and each rule model. Information integration techniques are particularly sensitive for identifying partial or inconsistent reliance on variables that other techniques might not reveal. The rule assessment approach is especially useful for identifying the type of stable and robust reliance on rules that most influences learning.

This last property was especially important in the present context, because a primary goal of the study was to use assessments of existing knowledge to predict and analyze learning of a challenging scientific reasoning problem. The assessments of initial knowledge yielded by the rule assessment approach have proved useful for predicting children's learning on a variety of problems: balance scales (Siegler, 1976; Pine & Messer, 2000; Siegler & Chen, 1998), time (Siegler, 1983), fullness (Siegler & Vigo, 1978), living things (Opfer & Siegler, 2004), and Tower of Hanoi (Klahr & Robinson, 1981). This track record of successful prediction of learning, together with the usefulness of the technique with relatively small samples and with relatively small numbers of trials on each type of problem, led to adoption of the rule assessment approach as the knowledge assessment method in the present study.

The predicted developmental ordering of the hypothesized rules was similar to the percentage of accurate predictions that the rules would yield on the problems in Table 2. Guessing predicts 33% correct answers, the Weight/Volume Rule and the More-Is-More Rule predict 50%, the Partially Correct Rule predicts 67%, and the Correct Rule predicts 100% (see Appendix for explanations). These levels of accuracy deviated from the hypothesized ordering of the rules in one way: The Weight/Volume Rule was hypothesized to be used by older children than the More-Is-More Rule, despite both approaches yielding 50% correct. The reason is that the Weight/Volume Rule reflects complete differentiation of weight and volume, whereas the More-Is-More Rule does not. Thus, the predicted rule sequence went beyond the usual prediction that accuracy improves with age and experience.

Table 2 Problem types

<table>
<thead>
<tr>
<th>Problem type</th>
<th>Outcome</th>
<th>Volumes</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Float</td>
<td>Equal</td>
<td>Unequal</td>
</tr>
<tr>
<td>2</td>
<td>Float</td>
<td>Unequal</td>
<td>Equal</td>
</tr>
<tr>
<td>3</td>
<td>Float</td>
<td>Unequal</td>
<td>Unequal</td>
</tr>
<tr>
<td>4</td>
<td>Sink</td>
<td>Equal</td>
<td>Unequal</td>
</tr>
<tr>
<td>5</td>
<td>Sink</td>
<td>Unequal</td>
<td>Equal</td>
</tr>
<tr>
<td>6</td>
<td>Sink</td>
<td>Unequal</td>
<td>Unequal</td>
</tr>
</tbody>
</table>

Note: On Problem types 3 and 6, one object is heavier and the other is larger

Promoting differentiation and integration

What types of experiences promote differentiation and integration? With frequently presented stimuli, such as letters of the alphabet, sheer exposure seems sufficient (e.g. Gibson, 1969). When encountering novel problems, however, it seems likely that more active analysis is necessary. One means of eliciting such active analysis is to request explanations of why unanticipated outcomes occurred. Such requests lead students to identify variables that would have led to the accurate predictions and to differentiate those variables from variables that would have predicted different outcomes. Past empirical studies have shown that being asked to explain why events occur often leads to faster and more frequent discovery of superior rules and strategies than does making the same observations but not explaining them (Calin-Jageman & Ratner, 2005; Chi, de Leeuw, Chiu & LaVancher, 1994; Pine & Messer, 2000; Renkl, 2002; Renkl, Atkinson, Maier & Staley, 2002; Siegler & Chen, 1998).

Almost all previous studies of self-explanation have focused on the effects of requests to explain why the observed outcome occurred or why the experimenter said that a given answer was correct. However, overlapping waves theory (Siegler, 1996b, 2006) implies that adoption of a new approach should depend not only on the degree to which experiences encourage generation of that approach but also on the degree to which experiences discourage use of other approaches. The same assumption is reflected within recent computer simulations of strategy discovery, such as SCADS ( Shrager & Siegler, 1998) and SCADS* (Siegler & Araya, 2005). Within these models, the two learning processes -- generation of new approaches and reduction of use of older ones -- are distinct both conceptually and empirically. Experiences that elicit generation of new approaches, even approaches that are clearly superior to previous ones, often lead to only small reductions in use of previous approaches, at least initially (Kuhn, 1995; Miller & Coyle, 1999; Miller & Seier, 1994). Considerable use of the new, superior strategy often is needed before it predominates. Conversely, when an existing approach leads to negative feedback, children often reduce their use of the existing approach without having discovered superior alternatives; they guess or oscillate among alternatives (Kuhn, Garcia-Mila, Zohar & Anderson, 1995; Perry & Lewis, 1999; Siegler & Chen, 1998). In all of these studies, many children eventually generated more advanced approaches, but they did not do so until after a period of inconsistent responding.

Recognizing that reducing reliance on previous approaches is a separate process from discovering new strategies suggests a prediction for how differentiation and integration of dimensions can be enhanced beyond the level produced by explaining the correct answer: Ask children to explain not only why observed outcomes occurred but also why other plausible outcomes, that would have been predicted by prior approaches, did not.
To the extent that such questions lead children to differentiate the dimensions that were relied on previously from the ones that are central in the new approach, the questions should reduce the strength, and thus the future use, of the prior approaches.

Empirical data from the one prior study that examined the effects of explaining why incorrect answers are incorrect supported this prediction. Requesting such explanations, as well as explanations of correct answers, led to greater learning of mathematical equality problems than only asking peers to explain correct answers (Siegel, 2002).

In addition, encouraging children to explain why incorrect answers are incorrect is a major feature of the Japanese approach to math education, and that system and the correct answers are correct did not learn more than the group that was just given feedback. Thus, the greater effectiveness of explaining both correct and incorrect answers than of explaining only correct answers in that study may have been produced by the group that only explained correct answers performing at an idiosyncratically low level. These concerns, together with a desire to understand the factors that promote differentiation and integration of dimensions, were motivations for the present study.

A componential analysis of rule learning

How might children generate new rules or strategies for solving problems? Three information processing components have been found to play particularly important roles: (1) noticing potential explanatory variables that are not included in existing approaches; (2) formulating more advanced approaches that incorporate the new, potentially explanatory, variables; and (3) generalizing the discoveries to new problems (Alibali, 1999; Chen & Klahr, 1999; Siegel & Chen, 1998). Each of these processes poses challenges for learners; failure to successfully execute any one of them reduces learning.

The first component in the sequence, noticing potentially relevant variables, involves recognizing that variations along a previously ignored dimension could account for observed outcomes. After people realize that a variable is relevant, it often is impossible for them not to perceive variations in it. However, noticing relevant variables is often easier in retrospect than in prospect. Studies of chess (Chase & Simon, 1973), gear motion (Perry & Elder, 1997), balance scales (Siegel & Chen, 1998), and mathematical equality (Alibali, Ockuly & Fischer, 2005) indicate that children and adults often fail to notice relevant dimensions. Children seem especially likely to notice additional variables when unexpected outcomes lead them to broaden their search for an explanation beyond the variables they already are considering (Cain-Jageman & Ratner, 2005).

A second essential process in discovering new rules and strategies is formulating a predictive approach that includes new variables. Noticing a new variable does not automatically lead to incorporation of the variable into a predictive rule or strategy. For example, in a study of how children learn about balance scales (Siegel & Chen, 1998), many children who noticed the distance variable, as indicated by their citing it as a possible explanation for observed outcomes, did not immediately utilize the variable to predict the balance scale's actions. Similarly, in a study of how children learn to solve matrix completion problems (Siegel & Svetina, 2002), children often noticed relevant variables but did not incorporate them into their choices of how to complete the matrix. Thus, formulating an approach that makes specific predictions based on the values of the noticed variable is a second essential step in rule discovery.

The third key process involves generalizing the new approach to novel problems. As noted previously, children often continue to use less advanced approaches even after formulating more advanced ones. This lack of generalization occurs even when the two problems are structurally isomorphic and the new approach is logically superior to earlier ones. For example, on analogical reasoning and conservation tasks, children often fail to generalize solutions onto parallel problems (Chen & Daehler, 1989; Siegel, 1995). Similarly, on number conservation tasks, children who already rely on the type of transformation to solve some problems continue to rely on relative length of the rows on other, similar problems.

An earlier application of the present componential analysis (Siegel & Chen, 1998) pointed to noticing as the key variable in learning how to solve balance scale problems. Failure to learn balance scale rules was most often due to failure to notice distance of the weights from the fulcrum and to differentiate that variable from the overall configuration of weights on pegs. Asking children to explain observed outcomes improved learning largely because it increased the likelihood of noticing the distance variable. The present study tested whether noticing would also prove to be the key process in learning on a quite different type of task, one in which interactions among variables were crucial. More generally, the componential analysis provided a means for determining whether self-explanation exercises its effects by increasing the likelihood of successful execution of a single component or whether it operates by somewhat increasing the likelihood of successful execution of multiple components.

To summarize, the present study involved three main goals: to examine the usefulness of the concepts of differentiation and integration for predicting the rules that young school-age children use to solve water displacement problems and the developmental sequence of the rules;
to determine the effects of explaining correct and incorrect answers on children's learning; and to identify the component processes underlying learning through self-explanation.

**Method**

**Participants**

The sample initially included 120 first, second, third, and fourth graders. Of the 120 children, five answered correctly on at least 15 of the 18 pretest problems, thus indicating that they already knew how to solve the problems, and seven other children did not complete both sessions because of illness or boredom. This left a sample of 108 children who completed the study. The sample included a younger group of 55 first and second graders (mean CA = 7.7 years, SD = .62) and an older group of 53 third and fourth graders (mean CA = 9.6 years, SD = .67). Originally, there were 20 children in each of the six age/condition groups; in the sample that completed the study, there were between 15 and 20 in each age/condition group. Children were recruited from Pittsburgh-area elementary schools that serve predominantly lower-middle-class and middle-class populations. Approximately equal numbers of boys and girls at each age level and from each school participated in the three conditions.

**Materials**

The apparatus consisted of two identical transparent glass containers, 14 cm in diameter and 18 cm tall; water that filled about two-thirds of each container; and 68 blocks that could be placed in the water. The blocks included 36 cubes and 32 cylinders. The cubes varied in size from 1.25 to 6 cm/side; the cylinders varied in diameter from 1.25 to 6 cm and in height from 2.5 to 9 cm. Both cubes and cylinders also varied in their type of material and therefore in their density and weight. Some blocks were made of materials that would sink when placed in water (steel, stone, and heavy plastic), and others of materials that would float (wax, wood, or light plastic). The cubes were presented on the pretest and posttest, the cylinders in the learning phase.

**Procedure**

The procedure was presented over two days. Day 1 included the pretest and the first half of the learning phase; Day 2 included the second half of the learning phase and the posttest.

**Pretest and posttest**

The pretest was an assessment of children's initial knowledge of the relative effects of different blocks in raising water levels. Children were presented 18 problems, three items each from the six problem types in Table 2. Items were ordered so that each successive set of six included one problem of each type. On each item, children were presented two identical beakers with equal amounts of water; the experimenter told the children that the containers and water levels were the same. Then, children were shown two blocks, heard descriptions of their relative sizes and weights (e.g. ‘this block is bigger than the other one, but they weigh the same’), and were encouraged to pick them up. Next, children were told, ‘Imagine that I put this block into this container and that block into that container, and that both float (or sink). Which container will have a higher water level, or will they be the same?’ After predicting, children were asked to explain their reasoning. In particular, they were asked, ‘Can you tell me why you think (this one will have a higher water level) (they will have the same water level)?’

The posttest was identical to the pretest, except for the particular blocks that were presented. No feedback was provided on the pretest or the posttest.

**Learning phase**

The learning phase included four items from each of the four problem types on which the blocks’ weight or volume, but not both, were equal (problem types 1, 2, 4, and 5 in Table 2). Each set of four successive items included one of each of the four types. The learning phase was like the pretest and posttest in that children were asked to predict which block, if either, would make the liquid rise more if put into the beaker. Unlike in the other phases, however, on each learning phase trial, children received both verbal and visual feedback regarding the correctness of their prediction. First, children predicted the effects of putting the blocks in the water; then, they observed what happened when the blocks were put in the water; then, they were told by the experimenter ‘You were right’ or ‘No, that’s not right.’

The three experimental conditions differed in what happened after children received this feedback. Children in the explain-correct-and-incorrect-answers condition were asked after the feedback to explain why the correct answer was correct and then why the answer suggested by the status of the variable that was irrelevant on the trial (weight when the object sank, volume when it floated) was incorrect. The exact question regarding why the correct answer was correct depended on whether the child’s prediction on the trial was accurate. If the child’s prediction was accurate, the experimenter said: ‘You were right. Now look at what happened carefully and see if you can figure out why.’ If the child’s prediction was inaccurate, the experimenter said: ‘No, that’s not right. Look carefully and see if you can figure out why that wasn’t right. Now tell me why the water level in this container is higher than in that one.’

Having explained the correct answer, children in this condition were then presented the wrong answer and
asked to explain why it was wrong. They were told, 'A child from another school thought that the water level in this container would be higher than in that one after we put these two blocks into the containers. Why do you think she thought this container would have a higher water level? Do you know why she was wrong?' The experimenter responded with a non-contingent 'very good' after the children's explanations in this and the other two conditions.

Children in the explain-correct-answers condition were presented a procedure that was identical to the first part of the procedure presented to children in the explain-correct-and-incorrect-answers condition. Children in the explain-own-answers condition only received the feedback described above, that children in all groups received; they were not asked to explain the outcome after seeing it.

Rule classification criteria

Criteria for rule use on the pretest and posttest paralleled those in previous studies that used the rule assessment approach (e.g. Siegler, 1981), and are described in the Appendix. Due to the likelihood that children's approaches would change during the learning phase, rule use was assessed only on the pretest and posttest.

Results

Reliability of classifications of the explanations and predictions data in all three phases of the study (pretest, learning, and posttest) was very high. Two raters who independently coded eight children's explanations and predictions (a total of 448 trials) agreed on more than 99% of classifications. All post-hoc comparisons are Newman Keuls tests, with \( p < .05 \), unless otherwise stated.

The results are organized into three parts. In the first, we examine pretest performance, with the goal of determining whether children used the hypothesized rules and whether the rules showed the predicted developmental ordering. In the second, we examine differences among the three experimental conditions in progress from pretest to posttest toward more differentiated and hierarchically integrated rules. In the third section, we present a microgenetic analysis of changes during the learning phase in each experimental condition and examine how three component processes contributed to developmental and individual differences in learning.

Pretest

The mean ages of children using each rule on the pretest provided evidence consistent with the hypothesized developmental ordering. The mean age of children who used no rule was 7.8 years, of children who used the More-Is-More Rule 8.6 years, of children who used the Weight/Volume Rule 8.9 years, and of children who used the Partially Correct Rule 9.5 years, \( F(3, 104) = 6.91, M_{H} = 9.72, \eta_{p}^2 = 0.17, p < .001 \). Fisher PLSD tests indicated that children who used no rule were younger than those who used any of the other three rules (\( ps < .01 \)), and that those who used the More-Is-More Rule were younger than those who used the Partially Correct Rule (\( p = .01 \)).

Pretest-posttest changes

Rule use

As shown in Table 3, each of the hypothesized rules fit the performance of substantial numbers of children. On the pretest, 14 children were classified as using the Partially Correct Rule, 34 the Weight/Volume Rule, 38 the More-Is-More Rule, and 22 no rule (five other children used the Correct Rule and therefore did not participate further in the study). On the posttest, 29 children were classified as using the Correct Rule, 23 the Partially Correct Rule, 12 the Weight/Volume Rule, 27 the More-Is-More Rule, and 17 no rule.

Both the Weight/Volume Rule and the Partially Correct Rule had two versions: the weight version and the volume version. Children tended to use the same version on the posttest as on the pretest. Of the children who used the Weight/Volume Rule, the large majority used the weight version, in which they consistently predicted that the heavier object would cause the water to rise more. This was true on both the pretest (85%) and the posttest (75%). Of the children who used the Partially Correct Rule, most answered consistently correctly on the sunken objects problems, where volume was the key variable, and answered neither consistently correctly nor consistently incorrectly on the floating objects problems. This was true of 79% of those who used the Partially Correct rule on the pretest and 57% of those who used it on the posttest.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Explain both</th>
<th>Explain correct</th>
<th>Explain own</th>
<th>Explain both</th>
<th>Explain correct</th>
<th>Explain own</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>49</td>
<td>28</td>
<td>6</td>
</tr>
<tr>
<td>PC</td>
<td>9</td>
<td>13</td>
<td>17</td>
<td>22</td>
<td>34</td>
<td>8</td>
</tr>
<tr>
<td>W/V</td>
<td>40</td>
<td>34</td>
<td>22</td>
<td>6</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>MIM</td>
<td>33</td>
<td>28</td>
<td>44</td>
<td>18</td>
<td>10</td>
<td>46</td>
</tr>
<tr>
<td>No Rule</td>
<td>18</td>
<td>26</td>
<td>17</td>
<td>6</td>
<td>18</td>
<td>22</td>
</tr>
</tbody>
</table>

Note: Correct = Correct Rule, PC = Partially Correct Rule, W/V = Weight/Volume Rule, MIM = More-Is-More Rule

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd.
The rule classifications were strong statements about the children's judgments, as evidenced by the close fit between predicted and observed accuracy patterns. Children who were classified as using the Correct Rule answered 98% of posttest problems correctly (predicted value = 100%). Children classified as using the Partially Correct Rule correctly answered 70% of pretest problems and 71% of posttest problems (predicted value = 67%). Those classified as using the Weight/Volume Rule correctly answered 50% of pretest problems and 51% of posttest problems (predicted value = 50%). Those classified as using the More-Is-More Rule correctly answered 51% of pretest problems and 49% of posttest problems (predicted value = 50%). Finally, those classified as not using any rule correctly answered 34% of pretest problems and 32% of posttest problems (predicted value = 33%). Thus, the hypothesized rules provided adequate descriptions of the children's judgments.

Use of the Correct Rule

Because predictor and outcome variables were categorical, a sequential multinomial logit analysis was performed using SPSS NOMREG to examine the relation of children’s pretest rule, experimental condition, and age to their use of the Correct Rule on the posttest. For the pretest rule variable, use of no rule was coded as 1, use of the More-Is-More Rule as 2, use of the Weight/Volume Rule as 3, and use of the Partially Correct Rule as 4. For the experimental condition variable, explain own answers was coded as 1, explain correct answers was coded as 2, and explain correct and incorrect answers was coded as 3. For the age variable, first or second grade was coded as 1 and third or fourth grade as 2. The dependent variable, adherence to the Correct Rule on the posttest, was coded as 1 for children who did not use the rule and as 2 for those who did.

All indicators converged on the conclusion that pretest rule and experimental condition were highly predictive of Correct Rule use on the posttest. The model fit well using a deviance criterion, \( X^2(20, N = 108) = 12.42, p = .90 \), and yielded a high log likelihood ratio, \( X^2(3) = 43.85, p < .0001 \). Logistic likelihood ratio tests showed substantial effects of both experimental condition, \( X^2(2, N = 108) = 15.94, p < .001 \), and pretest rule, \( X^2(2, N = 108) = 15.60, p < .001 \), but not of age, \( X^2(2, N = 108) < 1 \). \( R^2_N \), a measure of strength of association analogous to \( R^2 \) in a standard linear regression analysis (Nagelkerke, 1991), provided converging evidence. The value of \( R^2_N \) was .49, indicating that experimental condition and pretest rule together accounted for half of the variance in posttest rule use.

Finer grain analyses also indicated that the pretest rules hypothesized to be more advanced, and the experimental conditions hypothesized to be more impactful, had the expected effects. Use of the Correct Rule on the posttest was highest among the 14 children who used the Partially Correct Rule on the pretest (57%), next highest among the 34 children who used the Weight/Volume Rule (41%), next highest among the 38 children who used the More-Is-More Rule (18%), and lowest among the 22 children who used no rule on the pretest (0%). Chi-square tests showed that all differences except that between use of the Partially Correct Rule and the Weight/Volume Rule were significant. More children who were asked to explain both correct and incorrect answers acquired the Correct Rule than did children asked to explain their own answers (49% versus 6%, \( X^2(1, N = 69) = 16.46, \Phi = .49, p < .001 \)), more children asked to explain correct answers only acquired the Correct Rule than did children asked to explain their own answers (28% versus 6%, \( X^2(1, N = 75) = 6.70, \Phi = .30, p < .01 \)), and there was a trend toward children who were asked to explain both correct and incorrect answers more often acquiring the Correct Rule than children asked to explain only correct answers (49% versus 28%, \( X^2(1, N = 73) = 3.14, \Phi = .21, p < .08 \)).

The sample size, and presence of '0's in several cells of the condition × pretest rule × age matrix, precluded inclusion of interaction terms within the logistic regression analysis. However, the data pattern suggested that type of explanation had a larger effect on older children's adoption of the Correct Rule than on younger children's. Among the older children, those who were asked to explain both correct and incorrect answers were more likely to use the Correct Rule on the posttest than were children who were asked to explain correct answers (62% versus 37%), and children who were asked to explain correct answers adopted the Correct Rule more often than children asked to explain their own answers (37% versus 5%), \( X^2(2, N = 53) = 14.20, \Phi = .46, p < .01 \). Among the younger children, frequency of Correct Rule use on the posttest showed the same trend, but the difference among the three conditions was not significant (33%, 20%, and 6%, \( X^2(2, N = 55) = 4.12, \Phi = .26, p > .10 \)).

Changes in the learning phase

Predictions

A 3 (Condition) × 2 (Age) × 4 (Trial block: first, second, third, or fourth quarter of the 16 learning phase problems) ANOVA revealed main effects for trial block, \( F(3, 306) = 16.41, \Phi_S = 10.75, \eta^2_p = 0.17, p < .0001 \), condition, \( F(2, 102) = 9.03, \Phi_S = 17.63, \eta^2_p = 0.15, p < .001 \), and age, \( F(1, 102) = 7.88, \Phi_S = 15.31, \eta^2_p = 0.07, p < .01 \). All paralleled the main effects in the previous analysis. The interaction between condition and trial blocks was also significant, \( F(6, 306) = 3.22, \Phi_S = 1.97, \eta^2_p = 0.06, p < .01 \), and again paralleled the pretest–posttest changes. Percent correct predictions of children asked to explain both correct and incorrect answers increased from 48% to 73% over the four trial blocks of the learning phase, versus changes from 44% to 66% among those asked to explain only correct answers and 44% to 48% among those who only received feedback.
Components of learning

Three processes were hypothesized to be central to learning: noticing, formulating, and generalizing. Children were credited with noticing both relevant variables if they appropriately cited each variable in explaining their prediction on at least one successive pair of problems where different variables were relevant during the learning phase. That is, they needed to cite one object's greater weight to explain why they predicted that that object would raise the water level more on a problem where the objects floated, and they needed to cite one object's greater volume to explain why it would raise the water level more on a problem where the objects sank. Children were credited with formulating a rule that incorporated both the weight and volume variables if they generated correct predictions and explanations on all four problems in at least one trial block during the learning phase or generated three correct predictions and explanations in at least two trial blocks. Children were credited with generalizing if their predictions on the posttest met the criteria for the Correct Rule, including extension of the correct approach to conflict problems that had not been encountered during the learning phase.

Figure 2 illustrates the path that led to children's learning of the Correct Rule. It shows that each component represented a substantial source of difficulty in the learning process. The criterion for noticing the potential roles of both weight and volume was met by 63% of children. Of those who noticed, 66% formulated a rule that consistently included weight as the only relevant variable on problems involving floating objects and that consistently included volume as the only relevant variable on problems involving sunken objects. Of those who formulated such a rule, 67% generalized the rule to solve new problems, including conflict problems, on the posttest.

Comparisons of performance of children in the three experimental conditions indicated that the two groups of children who were asked to explain correct answers were more likely to notice the role of both variables than were children who only explained their own answers (79% and 69% versus 42%, $X^2(2, N = 108) = 11.20, \Phi = 31, p < .005$). Of those children who noticed both variables, children in the same two groups were more likely to formulate a correct rule (81% and 59%, versus 40%, $X^2(2, N = 69) = 6.04, \Phi = .29, p < .05$). Of the children who formulated a correct rule, children in the same two groups were more likely to generalize the rule on the posttest (76% and 69% versus 33%). Despite the large percentage difference, this last difference was not statistically significant; the likely reason was that only those children who formulated a correct rule were included in the analysis, which resulted in only six children from the explain-own-answers condition being included.

Discussion

The experimental findings attest to the usefulness of viewing the development of understanding of quantitative dimensions as a process of differentiation and integration. The findings also supported predictions from overlapping waves theory regarding the effects on learning of explaining correct and incorrect answers, and demonstrated that the componental analysis could help specify the processes through which generating self-explanations enhance learning. In this concluding section, we examine each of these conclusions in turn.

Differentiation and integration

Viewing development of understanding of quantitative dimensions as a process of differentiation and integration produced accurate predictions regarding which rules children would use to solve water displacement problems and the order in which the rules would be used. Despite the criteria for rule use being quite stringent – more than 80% of the three-choice responses conforming to the rule, and additional criteria specific to the rule also being met – 81% of children were found to use a rule other than guessing on the pretest and 85% on the posttest. The hypothesized rules closely fit children's performance; for all approaches, the observed percent correct for children said to use a rule closely matched the theoretical prediction for percent correct yielded by that rule. Moreover, the performance of children who did not meet the rule criteria was consistent with the hypothesis that the children were guessing or relying on an approach uncorrelated with the correct answer. These children answered correctly on 34% of pretest problems and 32% of posttest problems, closely akin to the 33% that would arise through guessing. Thus, the statement that a child was using a rule was a strong statement about the child's performance.

The present approach also yielded accurate predictions regarding the order in which the rules would emerge. The hypothesized ordering of the rules exactly matched the mean ages of the children who used them, with each...
successive rule in the hypothesized ordering being used by older children than its predecessor.

Perhaps the strongest evidence for the hypothesized developmental sequence came from the learning data, which indicated the probability of each pretest-posttest transition. The more advanced the child's initial rule within the hypothesized ordering, the more likely that the child would learn the Correct Rule. This pattern was clearly evident for the two rules that yielded identical percentages correct, with use of the Correct Rule on the posttest being more frequent among children who used the Weight/Volume Rule on the pretest than among children who used the More-Is-More Rule on the pretest—41% versus 18%. Overall, almost four times as many transitions were upward in the hypothesized rule progression as downward.

The nature of the learning components fit well with the emphasis on differentiation and integration. In the present context, noticing involved differentiating between weight and volume; formulating the Correct Rule required integrating weight, volume, sinking, and floating into an appropriate hierarchy. Similar correspondences are evident in previous demonstrations of the importance of these components in learning about balance scales, separation of variables, mathematical equality, analogical reasoning, and other problems (Aiball, 1999; Chen & Klahr, 1999; Perry & Lewis, 1999; Siegler & Chen, 1998).

This approach to cognitive change has the potential to re-focus attention on many of the fascinating phenomena uncovered by Piaget and co-workers without the stage theory assumptions that proved incorrect. Many classic Piagetian and neo-Piagetian tasks are difficult precisely because they require children to differentiate among quantitative dimensions that are highly correlated in the environment. In addition to the 10 tasks listed in Table 1, these include (with the dimensions whose differentiation they require in parentheses) time (endpoints of travel vs. time of travel), speed (endpoint of travel vs. speed of travel), solid quantity conservation (length of a clay sausage vs. its volume), length conservation (length vs. endpoint of sticks), life status (animacy vs. being a living thing), and fairness (amount of reward vs. ratio between reward and contribution).

The present framework suggests several predictions. One is that experiences that require children to differentiate highly correlated dimensions should be especially effective in promoting solutions to problems on which such correlations are present. Existing evidence is consistent with this prediction. One of the most effective training techniques ever devised for teaching children conservation concepts was Gelman's (1969) learning set procedure. This procedure presented problems that discriminated between number and length and provided feedback as to which dimension was relevant. Participants showed rapid learning on the original task and impressive transfer to related problems. More generally, feedback on problems on which reliance on the relevant dimension and confusable other dimensions produce different answers has been found to produce substantial learning on a wide range of problems, as have verbal rules that distinguish between relevant and irrelevant dimensions (Bavel, 1977, Brainerd, 1978; Field, 1987).

Another implication of this perspective is that failure to differentiate correlated physical dimensions may interfere with learning of science concepts in school. For example, if children fail to differentiate clearly between time of travel and distance traveled, speed and acceleration, or mass and weight, their learning of integrative problem-solving rules will be hindered. Teaching children to differentiate clearly between these concepts should improve their ability to learn to solve problems in each domain.

Explaining correct and incorrect answers

One means for promoting differentiation and integration of variables is requesting self-explanations. Such explanations can focus on how procedures produce their effects, how structural aspects of a system influence its functioning, how characters' motivation in a book or movie lead to their actions, and so on. In short, self-explanations are causal inferences regarding 'how' and 'why' events occur.

The present findings replicate a growing body of findings indicating that asking children to generate self-explanations enhances their learning (Calin-Jageman & Ratner, 2005; Chi et al., 1994; Pine & Meser, 2000; Renkl, Stark, Gruber & Mandl, 1998; Siegler, 1995, 2002). The present findings go beyond the previous ones in providing clear evidence that explaining why incorrect answers are incorrect contributes to learning, above and beyond the effects of receiving feedback on the correctness of one's own answer and explaining why the correct answer is correct.

In the only previous study to examine this issue (Siegler, 2002), children who explained both correct and incorrect answers learned more than children who explained correct answers only, and more than children who only received feedback. However, in that study, explaining correct answers alone did not have any effect. That anomalous finding opened the possibility that the superiority of the group that explained both correct and incorrect answers in that study merely reflected the well-established positive effect of explaining correct answers. The present findings, in which children who explained both correct and incorrect answers learned more than children who just explained correct answers, and the latter group learned more than children who only explained their own answers, indicates more clearly that explaining why incorrect answers are incorrect enhances learning beyond the effect of explaining why correct answers are correct.

The value of explaining incorrect as well as correct answers was predicted by overlapping waves theory. Within that theory, probability of using a new strategy reflects the strength of prior strategies as well as the strength of the new approach. Therefore, experimental manipulations that weaken prior strategies should increase the use of new
approaches. However, use of a new strategy also requires discovery of that approach. Our hypothesis was that self-explanations would promote discovery through increasing the likelihood of noticing the relevant variables, which was essential for formulating them into a correct rule. Although this hypothesis was consistent with findings from both our own previous studies and those of other investigators, it proved only partially correct in the present instance. Consistent with the hypothesis, explaining both correct and incorrect answers increased the likelihood of noticing both relevant variables. However, explaining correct answers alone increased the likelihood of noticing both relevant variables almost as much. The superior learning in the group asked to explain both correct and incorrect answers seemed to derive from such requests somewhat increasing the likelihood of successful execution of each of the three components rather than the effect deriving from any one of them.

The difference between the present and previous findings may be attributable to variations in the ways in which children were presented the tasks. In the studies in which noticing or encoding the key variables was the dominant source of differences among conditions and age groups, the problem presentation did not identify the essential variables. For example, in Siegler and Chen (1998), the description of the task did not refer to weight and distance; children were simply told that disks would be placed on pegs. On such a task, noticing that distance of the disks from the fulcrum varied, and that the variation was related to the balance scale's rotation, was a challenging task. In contrast, in the present study, children were explicitly told at the beginning of each trial which block (if either) was heavier and which block (if either) was larger. This form of presentation probably facilitated noticing the relation between the relevant dimension and the outcome. For example, being told 'This block is heavier but this block is larger', and then seeing that the water into which the heavier block was placed rose more, probably facilitated noticing that weight varied and might be influential. Consistent with this analysis, most children in the two groups that were asked to generate explanations met the criteria for noticing both weight and volume, whereas only a minority of children noticed both dimensions in Siegler and Chen (1998). Thus, the problem presentation may have reduced the variation in whether children noticed both variables and thus reduced its role as a potential impediment to learning.

The fact that no one component accounted for the greater effectiveness of requests to explain both correct and incorrect answers raised the question of why this instructional condition was more effective than the other two. One plausible explanation is that such requests stimulated greater depth of search of the child's prior knowledge base and observations within the experiment (Newell, 1990; VanLehn, 1999). Given that the phrasing of questions in all conditions cited the relative weights and sizes of the objects, it would be easy for a child to use 'more is more' logic to explain the correct answer without more than a superficial search of their knowledge base. They could simply cite the variable whose relative status matched that outcome. Thus, if one object was described as larger, and the water into which it was placed rose more, the child could easily explain the outcome by saying that the water rose more because the object placed in it was bigger (as most children did). This would allow children to believe they had explained the outcome without forcing them to confront the inadequacies in their way of thinking about the problem. Such shallow processing may explain the seeming discrepancy between children who used the More-Is-More Rule on the pretest and therefore relied on weight on some trials and volume on others—rarely progressing to the correct rule on the posttest. Such children may simply have been searching for a dimension whose values were unequal and that matched the dimensions cited in the question. In contrast, the question phrasing did not suggest an explanation for why the incorrect answer was incorrect. Hearing an object described as larger did not explain why that object did not raise the water level more; instead, the question highlighted the puzzling non-occurrence of the outcome. Why wouldn’t a larger object raise the water level more? This puzzle seems likely to have increased the depth of search for an explanation for the puzzling outcome. In particular, requests to explain incorrect answers may have led children to wonder whether the fact that the objects sometimes floated and sometimes sank (which the experimenter had also mentioned) had anything to do with the outcome. Such requests may also have led children to reason that a sunken object would displace the same amount of water no matter what its weight, but that a floating object would displace more water if it rode lower in the water, and that perhaps the amount of weight would influence how high the floating object rode in the water.

Findings from the only previous study of the effects of explaining incorrect as well as correct answers (Siegler, 2002) are consistent with this depth-of-search interpretation. In that study, explaining correct and incorrect answers was more likely than only explaining correct answers to stimulate adoption of conceptually advanced, transferable strategies in a situation in which simpler strategies also solved all training problems but not all posttest problems. Children who explained only correct answers more often adopted simpler, less transferable strategies. Also supporting the depth of search interpretation, explaining both correct and incorrect answers increased solution times on the first few trials (until a successful rule was adopted) to a greater extent than explaining only correct answers or simply receiving feedback. Future studies in which direct measures of depth of search, including verbal protocols, are included seem likely to be useful for testing this depth-of-search interpretation of why explaining incorrect as well as correct answers promotes differentiation and integration of variables.
Appendix: Rule classification criteria

The criteria for using a rule paralleled those in previous applications of the approach (e.g. Siegler, 1976, 1981; Siegler & Richards, 1979). For a child to be said to have used a rule, roughly 80% of responses had to be consistent with the rule. In cases where it was possible for 80% of responses to be consistent with two rules, additional criteria that discriminated between those rules also had to be met. Thus, on the Correct and Weight/Volume Rules, which made determinate predictions for all items, at least 7/9 of responses (78%) needed to fit the prediction on both subsets (sinking/ floating or weight/volume).

In past studies, such criteria have not only provided evidence for hypothesized rules but also have ruled out hypothesized rules and allowed discovery of unanticipated rules. Thus, on fullness problems, Siegler and Vago (1978) found that children did not use a hypothesized volume rule (the glass with more liquid is fuller), but did use an unanticipated empty space rule (the container with more unfilled space is less full).

It is worth noting that the rules identified here could be viewed as general approaches, within which sub-rules that differed in subtle ways were also present. In this article, we only focus on the five general approaches that reflect the hierarchical relation among the structures of the rules. Substantially larger sample sizes and number of trials for each type of problem would have been necessary to allow unambiguous assessment of the sub-rules. This goal was beyond the scope of the present article.

In the present study, the Correct Rule generated 100% correct responses. The criteria for being classified as using the Correct Rule on the pretest and posttest were at least 15 of 18 (83%) correct responses on the six types of problems in Table 2, including at least 7 of 9 trials on which the objects floated and at least 7 of 9 on which they sank.

The two versions of the Partially Correct Rule generated 100% correct answers on problems involving either sunken objects or floating objects and 33% of problems on the other type of problem – 67% correct overall. The criteria for using the Partially Correct Rule were at least 8 of 9 correct predictions on one type of event (sinking or floating) and no more than 6 of 9 predictions on the other type of event.

Children who used the Weight/Volume Rule on all trials would generate 100% correct answers on the nine problems on which the objects floated and 0% on the nine problems on which the objects sank, or vice versa – 50% correct overall. In addition, their errors would involve saying that the two displacements would be equal when the preferred dimension (weight or volume) was equal and the non-preferred dimension was not, and that the wrong object would raise the water level more when one object was larger and the other heavier. The criteria for using the Weight/Volume Rule were at least 15 predictions being based on a single variable (weight or volume), including at least 7 of 9 problems on which the objects floated and 7 of 9 on which they sank.

Children who invariably used the More-Is-More Rule also would generate 50% correct, but through a different route. They would respond 100% correctly on the two types of problems on which the values of the irrelevant dimension were equal (because on those problems, they would always rely on the relevant dimension, the only dimension with unequal values). They would respond correctly on 50% of the problems on which both relevant and irrelevant dimensions were unequal (because they would choose arbitrarily between relying on relevant and irrelevant dimensions on those problems). Finally, they would respond correctly on 0% of the problems where values of the relevant dimension were equal and those of the irrelevant dimension unequal (because on those problems, they would always rely on the irrelevant dimension, because its values were unequal). The criterion for use of the More-Is-More Rule was that on at least 10 of the 12 problems on which values of one dimension were equal and values of the other dimension were unequal, the child would predict that the object that was larger or heavier would raise the water level more (the child would respond on the basis of the dimension on which the two objects were unequal).

Finally, children who could not be classified as using any rule were expected to respond correctly on about 33% of items on each type of problem and on 33% of items overall. The logic was that they were not basing responses on weight, volume, sinking, floating, or any other variable that led to systematic responses on the problems they encountered and therefore should be at a chance level of accuracy on three-choice problems. All children who did not fit one of the other four rules were classified as using no rule.

References


© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd.


