DOCUMENT RESUME

ED 234 344 CS 007 250

AUTHOR	Campione, Joseph C.; Armbruster, Bonnie B.
TITLE	An Analysis of the Outcomes and Implications of Intervention Research. Technical Report No. 283.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Education (ED), Washington, DC.
PUB DATE	Jul 83
CONTRACT	400-81-0030; NICHHD-HD-05951-12
NOTE	37p.; To appear in "Learning and Comprehension of Texts"; H. Mandl, N. Stein, and T. Trabasso, Eds.; Erlbaum; Hillsdale, NJ.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comprehension; Elementary Secondary Education; Evaluation Methods; Metacognition; Methods Research; *Reading Comprehension; *Reading Research; *Reading Skills; *Recall (Psychology); *Research Methodology; Training Methods
IDENTIFIERS	*Comprehension Monitoring; *Theory Practice Relationship

ABSTRACT

Aimed at improving the methodology used in comprehension research, this paper analyzes the designs and interpretations of intervention training studies and suggests the implications of that analysis for future research. It points out that the typical training approach, deriving data from three sources--comprehension tests administered to older students untrained in comprehension strategies, to younger untrained students, and to younger trained students--could be improved with additional data on how successfully trained students use their new comprehension strategies and on comprehension test results from older trained students. The paper also suggests that as findings and interpretations can be influenced by many factors, including the theoretical or practical motivation for the research, the criterion by which success is measured, and the difficulty of the task assigned, these factors must be considered carefully when formulating explanations of training studies. (MM)

CENTER FOR THE STUDY OF READING

Technical Report No. 283

AN ANALYSIS OF THE OUTCOMES AND IMPLICATIONS OF INTERVENTION RESEARCH

Joseph C. Campione and Bonnie B. Armbruster
University of Illinois at Urbana-Champaign

July 1983

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported in part by the National Institute of Education under Contract No. NIE-400-81-0030 and the National Institute for Child Health and Human Development under Contract No. HD 05951-12. To appear in H. Mandl, N. Stein, & T. Trabasso (Eds.), Learning and comprehension of texts. Hillsdale, N.J.: Erlbaum, in press.

An Analysis of the Outcomes and Implications of Intervention Research

Over the last decade or so, there has been a major change in the kinds of processes many investigators have begun to study and in the materials used in that research. From an emphasis on learning and recall of sets of words or sentences, we now see work investigating the comprehension and recall of larger segments of language, up to and including texts. Rather than being concerned with how people come to learn and remember bits of information provided in relative isolation, current interests emphasize to a greater extent the processes involved in the comprehension of material which is inherently meaningful, such as simple stories and more complex expository text segments.

We believe that some of the trends in this emerging area are similar to those which appeared in prior work in the broad area of memory development. As investigators have come to be more complete and confident in their accounts of the processes involved in text understanding, they have initiated research in which the goal is to teach students how to improve their comprehension capabilities. As in the earlier memory work, there are two distinct reasons investigators undertake training studies. One, primarily theoretical, is analogous to computer simulation approaches to the study of cognitive processes. If we are able to use a theoretical model to develop an instructional program to achieve some desired end, e.g., understanding a text, that result reinforces the theoretical approach adopted. If, according to some theory, activity A is an important component of comprehension, then teaching people who do not do so to employ A should enable them to improve their performance. If it does, we infer that the guiding theory was correct.

The second reason for conducting such research is more practical. Many students seem to have considerable trouble reading and comprehending texts independently. As such reading is an essential scholastic activity, it is worthwhile in its own right to attempt to develop curricula or programs which serve to improve the comprehension performance of academically poor students. Here theoretical niceties are less important. We do not mean to imply that these (theoretical and practical) approaches are independent. Adequate, specific theory can certainly help practitioners, and the fact that some program does promote comprehension provides important data for the theoretician. We simply mean that the emphases in the different types of research are different, that different experimental designs tend to be used, and that the interpretations which result are also likely to be of different kinds.

As interest in instructional research in the comprehension area increases, it seems worthwhile to review some of what we have learned from a decade or more of training studies aimed at evaluating some hypotheses about the nature of developmental and individual differences in memory performance. Keeping these lessons in mind should facilitate our attempts to use instructional methodologies to inform theory development in other domains, including comprehension. In our treatment here, we will be concerned with both an analysis of the design and interpretation of intervention studies in general and the implications of that analysis for research aimed at fostering comprehension.

In general, intervention research can be divided into two broad categories according to whether the <u>major</u> focus of the intervention is on the <u>learning materials</u> or the <u>activities of the learner</u>. In the first category, the approach to improving student performance is to modify the learning materials. For example, texts might be rewritten to clarify the organization and to call attention to the most important information. If students have difficulty identifying structure and determining main points, this modification should facilitate learning (e.g., Meyer, in press).

The second category of intervention research focuses on modifying the activities of the learner. Here the goal is to teach certain strategies or procedures that will help the student learn (e.g., Brown, Palincsar, & Armbruster, in press). In contrast to the <u>materials</u> emphasis aimed at facilitating the learning of particular text information, the <u>activities</u> approach is aimed at fostering learning to learn (see Brown, Bransford, Ferrara, & Campione, in press, for a more thorough discussion).

These two approaches represent different <u>emphases</u> and are neither independent not mutually exclusive. For example, providing clearly structured texts could itself result in modifying the students' learning activities. Having learned from exposure to well-written texts to appreciate the effect of clear organization on understanding, students may search out structure in less well written texts. As another example, students taught an array of comprehension strategies.

aimed at discovering or imposing structure on poorly prepared prose may benefit even more than untrained students from well written materials. We believe that the most impressive learning outcomes will result from programs involving both high quality materials and students prepared with the strategies necessary to take maximal advantage of them. Because of space constraints, we will limit our analysis in this report to research emphasizing learning activities. However, the approach should apply as well to intervention studies focusing on the learning materials.

An Analysis of Intervention Studies: A Modal Approach

A typical intervention study found in the literature begins with a demonstration of performance differences between two groups of students, whom we will designate as less successful (L) and more successful (M). The L and M groups could be children of different ages, retarded and nonretarded groups, normally achieving students and students with a specific reading disability, etc.; the argument is essentially the same in all cases. To provide a more concrete example, younger children often perform more poorly than older children on memory tasks. To account for the difference, the researcher frequently tenders two hypotheses. The first is in the form of a theoretical task analysis, a specification of the components of adequate performance. In many cases, the task analysis indicates several learning activities or strategies that are critical to adequate memory performance. The second hypothesis is of the form that the observed differential performance is due to differences in the availability or use of one or more of the essential components; as an example, the researcher may assume that the memory differences are attributable to differences in the use of a "rehearsal" strategy.

The researcher then trains some of the L (here younger) students in the hypothesized missing component(s) and compares their post-training performance with that of untrained L students and with untrained M students. In our example, a group of younger students is trained to use a rehearsal strategy, and then their performance is compared with that of untrained younger students and untrained older students.

If performance of the trained group then increases significantly, the researcher may infer support for both of the guiding hypotheses. First, rehearsal is inferred to be an important component of task performance, for if it were not, performance would not have improved. Second, it is concluded that the differential use of rehearsal was responsible, at least in part, for developmental differences on this task, since the group of students who were performing poorly to begin with are now performing more similarly to the initially more proficient.

A comparison of the trained L and untrained M students provides some further information about the quality or completeness of the task analysis. If the trained L students' performance is still significantly below that of the M group, this is a clear sign that there are other factors associated with efficient performance and involved in the developmental differences, i.e., there are other as yet undetermined sources of developmental differences.

An excellent example of the "modal" approach can be found in Butterfield, Wambold, and Belmont (1973). In that work, retarded adolescents were trained to use a cumulative rehearsal strategy; they would repeat several times the first item after it was presented, the first two after presentation of the second, the first three after presentation of the third, etc. The trained subjects improved but not to the level of an untrained M group (in this case, nonretarded adolescents). This result indicated that the task analysis was incomplete. These researchers were also in an enviable position in that we have come to know a considerable amount about the determinants of memory performance; and in their work, the specific patterns of the subjects' responses provided hints about the other components which might be important. Without going into detail, we will simply report that additional training attempts centering on a specific retrieval plan were then undertaken by Butterfield et al. (1973), with the eventual outcome of bringing the retarded subjects' performance to a level comparable to that of nonretarded adolescents, i.e., comparative differences between the groups were "eliminated" via the specific training program.

An Evaluation of the Modal Approach

Given that the group differences have been eliminated in this wav following instruction, we miaht wish to claim that we have therebv documented the imoortance of the trained activities to adequate performance on the task at hand and have demonstrated that we have a very strong theory about the nature of L-M differences on that task. That is, we want to argue that this result reinforces both our task analysis and our view of individual or group differences. The question is how valid those claims are likely to be. We argue below that neither conclusion is appropriate without additional data. However, before dealing with the evaluations of the theoretical task analysis and the nature of group differences, we will mention briefly one other issue.

Practically- vs. Theoretically-motivated Research

Researchers can differ in terms of their initial motivation for doing the research. If the aim were the practical one of improving performance to some desirable level, much of what we have to say below would be largely irrelevant. If the training program resulted in the hoped-for gains, further theoretical niceties would be of limited interest. Similarly, if the major goal of the research was simply to demonstrate a degree of plasticity in L learners, the research would already be successful. Additional analyses would be nice but not necessary. In fact, some of the issues we raise below might be almost impossible to implement in many practical situations. However, if the research goal were to develop and evaluate theories about the components of adequate or excellent performance and about individual differences in those components, the results of the modal approach cannot by themselves enable strong endorsements of either of the guiding theories.

The Task Analysis

Returning to the case where the instruction has brought the L subjects' performance up to that of the M group, the first conclusion we may wish to draw is that the instructed activity (rehearsal, for example) is an important component of performance on the task. The argument is that if it were not important, teaching students to use it would not improve their performance. The problem is that it is possible for the rehearsal training to result in improved performance even if the specific activity taught were not itself important. The training could be effective because it influences some other cognitive process that is in fact responsible for the improved performance. For example, training could lead to increased attention to the task or to heightened motivation; and these could be the factors mediating the improved performance. As this issue has been dealt with in a number of other sources (e.g., Butterfield, Siladi, & Belmont, 1980), we shall be brief here and note that in the memory area, this has not been an enormous problem, as our theories of many of the experimental tasks employed are quite detailed.

For example, in the case of rehearsal strategies, the problem is relatively minor because, whereas attentional or motivational mechanisms can be expected to produce enhanced performance, the increase should be a somewhat general one. Improvements due to rehearsal, in contrast, can be predicted to take a much more specific form. It is possible to specify in some detail the patterns of accuracy and latency which should emerge following training, rather than simply to predict that performance will increase. For example, rehearsal-produced improvements should be particularly large on items presented earlier in a series, rather than later. It is also possible to predict that rehearsing subjects will differ from non-rehearsing ones in terms of their patterns of self-pausing during study (Belmont & Butterfield, 1971), their overt production of the strategy (Flavell, Beach, & Chinsky, 1966), and the extent to which their accuracy and speed of response should be affected by variations in list structure (Brown, Campione, Bray, & Wilcox, 1973). In the Brown et al. (1973) experiment, all but one of these measures were used; and they all converged on the same conclusions regarding the importance of rehearsal processes, both in leading to excellent performance and in being partly responsible for differences between ability groups. Butterfield et al. (1980) provide a detailed discussion of the process of relating performance variations to specific changes in processing activities.

The modal training study is simple: students who do not do so spontaneously are told to carry out some specific activity, and their performance after instruction is compared with their pre-training accuracy. In the best studies, we have information not only about what the subjects are told to do, but also direct evidence that they have in fact been doing that correctly (e.g., Belmont & Butterfield, 1971, Brown et al., 1973). We also have evidence that the quality of execution of the strategy is strongly related to the level of recall. In addition, we have evidence that the improvements in recall accuracy are precisely what would be expected theoretically from a rehearsing subject. As such, the conclusion that the trained activity is an important component of performance on the task is considerably strengthened.

Our reason for empahsizing this point here is that the same problem exists in situations where instruction is aimed at improving comprehension processes. In the area of comprehension, in fact, the problem of attributing improvement to the wrong factor is much more acute than in the memory examples simply because we know much less about comprehension than about deliberate memorization. The general point we would make (see also Campione & Armbruster, in press) is that assessments of both strategy execution and the sequelae of instruction be as detailed as possible.

To cite examples of the ways in which more detailed evaluations can facilitate our analyses, consider the following cases. The first involves the importance of data on the quality of strategy execution. Brown and Smiley (1978) were interested in the extent to which students who underlined or took notes while reading a story would show better recall of that story than those who did not. As it turned out, students who carried out these activities did outperform those who did not, but only if the underlining and notetaking were done reasonably. Students who underlined randomly, for example, did not perform any better than those who did not underline at all. As those who underlined randomly were primarily those who underlined in response to instructions to underline, one might have inferred from a simple instructional study that underlining is not a useful comprehension-fostering activity. Information about the quality of underlining and its relation to learning and recall provided a much clearer picture of its role in influencing learning than would have been obtained otherwise.

A slightly different type of example is also relevant. Many of the studies involving instruction in comprehension activities have emphasized target processes more complex or multifaceted than has been the case in the memory research. It is thus possible that a "single" intervention could affect any of a number of different component processes. To illustrate, consider the series of studies reported by Palincsar and Brown (1982) and summarized in Brown, Palincsar, and Armbruster (in press). They sought to increase students' comprehension scores by teaching them to summarize what they had just read, predict the type of questions a teacher might ask on a subsequent test, note inconsistencies and ambiguities, etc. Training was clearly successful, as performance improved dramatically on ten-question comprehension tests administered after students had read a passage independently.

In addition, the experimental design of Palincsar and Brown allowed them to describe the nature of the process changes underlying this improvement in some detail. The design allowed them to monitor the extent to which students actually improved on the target processes throughout training, and there was correspondence between those measures and comprehension scores. Also, they administered a number of transfer tests following the experiment to obtain additional assessments of the extent to which specific processes had been influenced by the intervention. The instructed students showed reliable (pretest to posttest) improvements in summary writing, question prediction, and their ability to detect incongruities, but not in their ability to judge relative thematic importance. The overall package offered by

Palincsar and Brown then not only indicates that the training was effective in bringing about substantial improvement, but it also allows an accurate accounting of the more specific changes underlying the overall improvement. It also indicates some areas where the instruction appears to be less effective, thus leading to suggestions about how it might be improved.

Sources of Group Differences

In our opinion, the more interesting interpretive question associated with the modal training study concerns the inference that group differences were due completely or in part to differential use of the instructed activity. This inference rests on the assumption that training was unnecessary for the M students. Consider again the case where the trained L group performs as well as the untrained M group. Presumably, this is because the only difference between the groups had been due to variations in use of the instructed activity, a difference eliminated by instructing the L group. The implicit assumption here is that the M group is already using the instructed activity; as a result, they would not improve if training were provided. Training only the L group is sufficient to "equate" the groups' learning activities.

To evaluate this questionable assumption, we need to provide the same instruction to the M group as we did to the L group; that is, we need to employ an age/ability x instruction factorial design. As we shall argue, the use of such a design permits stronger conclusions about developmental/comparative differences, it indicates areas where M students can also benefit from training; and it can also facilitate our attempts to account for some situations where instruction is ineffective. In the next section, we explore possible outcomes of training studies using such a factorial design and the implications of these outcomes for theory and practice.

The Age/Ability x Instruction Factorial Design: Some Outcomes and their Implications

To reiterate, the proposed factorial design involves four groups: an L untrained group, an L trained group, an M untrained group, and an M trained group. The design can result in several possible patterns of outcomes, as shown in Figure 1.

The Outcomes

(1) One possible outcome is that training will improve performance of the L group but have no effect on the (nonceiling) performance of M students (see Figure 1, Panel A). This outcome resembles the outcome of the successful modal study discussed above but with the factorial design, the interpretation is more straightforward and the conclusion sounder.

Clear examples of the pattern of results represented in Panel A occur in Brown (1973) and Brown, Campione, and Gilliard (1974). In these studies, the tasks involved a judgment of relative recency. Students were shown a series of single pictures followed by a test trial. On the test, two of the previously seen pictures were presented, and the students' task was to indicate which of the two had been seen more recently. If background cues to anchor the temporal series were not provided, younger and older students performed alike. If background cues were provided, however, the older subjects outperformed the younger, presumably because the older, but not the younger, subjects used the background cues to their advantage. Instruction in how to use the background cues did not change the excellent (but not ceiling) performance of the older subjects, but it did succeed in bringing the younger ones up to a comparable level. This outcome is the strongest possible evidence that differential use of the trained component was a major, if not sole, determinant of developmental differences and that training was largely unnecessary for the older subjects.

(2) Another outcome is displayed in Panel B. In this case, training also affects the performance of the L group, but after training their performance is still not up to the level achieved by the M group. Training does not improve the performance of the M group. Such a result would indicate that the M subjects were in fact competent with regard to the instructed activity <u>and</u> that there are other sources of group differences still to be determined.

Another example of the pattern of results depicted in Panel B comes from research on teaching reading comprehension skills. Hansen and Pearson (1982) trained classroom teachers to provide instruction designed to improve the inferential comprehension ability of good and poor fourth grade students. One dependent measure was performance on worksheets of literal and inferential questions which accompanied the stories in which the instruction was embedded. Results indicated that the training enhanced the inferential comprehension of poor readers but not of good readers.

Studies reported by André and Anderson (1978-79) provide yet another example of the pattern of results in Panel B. High school students were taught to generate comprehension questions while studying textbook-like prose. The performance of trained students on a constructed response achievement test over a 450-word passage was compared to the performance of untrained students who used a read-reread studying method. Verbal ability, as measured by the Wide Range Vocabulary Test, was used to assign subjects _ex post facto_ to three levels. Results revealed a significant treatment x verbal ability interaction: the low ability trained group scored higher than the low ability untrained group, while the high ability students scored about the same in both the trained and untrained groups.

(3) A third possibility is depicted in Panel C. Both the L and M groups improve following training, but the L group profits from instruction to a greater degree than the M group. One set of possible conclusions from this pattern of results is: (a) the M group was not entirely proficient in the use of the target process (otherwise training would not have helped); (b) differential use did not contribute to the original developmental differences (because equating use did reduce those differences); and (c) other sources of performance variations exist.

(4) A fourth possible pattern, illustrated in Panel D, is that training has the _same_ effect on both developmental levels; that is, both the L and M groups exhibit the same increment in performance after training. While several explanations are possible for such a result, a simple interpretation is that the trained activity was important for

performance on the criterion task, but that it did not contribute to developmental differences.

As one example, Huttenlocher and Burke (1976) evaluated the hypothesis that developmental differences in digit span were due to the fact that older children grouped the input into richer "chunks." In a standard condition, they found the usual developmental differences. In a grouped condition, in which the input string was grouped by the experimenter to simulate the chunking presumably done by older subjects, both the younger and older subjects improved, and to about the same degree. Thus, the intervention which might have been expected to reduce the developmental difference by being more effective or necessary for the younger group was equally effective for all subjects. Similar effects have been obtained by Lyon (1977) using college students who differed in memory span. Interventions designed to reduce individual differences by providing "expert help" to the lower scorers improved everyone's performance and had no effect on the magnitude of individual differences.

Note that without training the mature students, the results might have been interpreted in the same way as the "modal" training study. That is, developmental differences would be attributed to differential tendencies to chunk the input; and inducing mature subjects to engage in such chunking would not be deemed necessary or helpful. Both of these conclusions obviously need to be re-evaluated. The opinion that the mature students would not benefit from chunking interventions is certainly incorrect, as the effects of the intervention were equal for the mature subjects. Also, if the grouping manipulation does in fact simulate the kinds of organizational processes which are presumed to underlie developmental differences, the parallel improvement result is strong evidence against the chunking hypothesis. Indeed, Huttenlocher and Burke (1976) argued that developmental differences were more likely due to differences in the efficiency with which subjects identified incoming items and/or to the ability to maintain information about order.

The Hansen and Pearson (1982) study mentioned earlier also provides an example of the Panel D pattern of results. Besides worksheets, another dependent measure was performance on literal and inferential questions over a transfer story at a level that could be read by both good and poor readers. For the inference questions, results revealed significant effects for ability and treatment, but not for their interaction. In other words, the experimental treatment of inferential comprehension instruction was about as effective for both the good and the poor readers, at least on one type of criterion task.

(5) Panels E and F portray variations on another pattern of results, in which the developmental differences are greater _after_ training than _before_ training. This divergent pattern is rather common in the literature (Cronbach, 1967; Snow & Yalow, 1982). One interpretation of this pattern of results is that the trained routine was not employed efficiently, if at all, by the more advanced students prior to training, and that its use requires some additional skills or knowledge before it becomes maximally effective. The first conclusion is straightforwar If the advanced students were proficient when left unaided, instruc should not be particularly beneficial. The second point addresses the relatively weaker

effects of instruction on the initially poorer performers. The explanation we have offered is that the poorer students are also unlikely to have available or to produce other skills which are prerequisite to the one(s) being trained. From the point of view of instruction, this would indicate that the analysis of the task upon which the intervention was based was not sufficiently detailed. Without the inclusion of the older/more capable group, a different interpretation could easily have resulted, namely that the task analysis was in error and that the activities being taught or simulated were not important ones. Given this interpretation, the overall approach might then be abandoned rather than refined. That is, the outcome obtained with the older learners influences the interpretation of the null result with the younger ones.

As an example of this pattern of results, consider a number of experiments on the balance beam problem reported by Siegler (1976, 1978). Subjects are shown a series of weight arrangements and asked to predict whether the beam will balance or whether one side or the other will fall if support is withdrawn. Siegler has analyzed the problem in terms of a number of increasingly complex rules which represent a progression toward a full understanding of the principles involved. An early rule, Rule I in Siegler's taxonomy, is based on a consideration of only weight factors. If the amount of weight on either side of the fulcrum is the same, the scale will balance; otherwise, the side with more weight will drop.

An extremely simple type of instruction is to provide examples from which a rule can be inferred. Siegler adopted this approach with groups of three- and four-year-olds who had not yet acquired Rule I. Their predictions of balance beam performance were essentially random. Interested in how his subjects might attain that rule, Siegler administered a series of feedback trials. The subjects would first predict what would happed to the beam when supports holding it in place were removed; then the supports were withdrawn and the subjects were allowed to observe what actually happened. This method simulated the process of formulating hypotheses, obtaining data, and then re-evaluating those hypotheses. The main result was that the four-year-olds tended to induce Rule I, whereas the three-year-olds did not. Note that if only the young children were included, it would be possible to conclude that leading them to explore the domain in this way was an ineffective way of producing learning.

Subsequent experiments showed that four-year-olds did in fact encode the relevant weight dimension even though they predicted randomly prior to feedback; the three-year-olds, however, did not encode the weight dimension. In this sense, one might say that the older children know more about the balance problems (i.e., that weight is a relevant dimension) than the younger children; and that this knowledge or competence is necessary for the intervention to produce learning. This conclusion prompted a more detailed training procedure in which three-year-olds were taught to encode weight before receiving the feedback trials. In this situation, they showed an increased tendency to acquire Rule I.

A second example of this type of result comes from a study reported by Brown and Campione (1977). They were concerned with teaching two groups of retraded children to systematically deploy their study time in a list learning situation. The paradigm, based on a prior study by Masur, McIntyre, and Flavell (1973), involved studying and remembering the labels of a set of 12 pictures. On each trial after the first, the subjects could select only one-half (6) of the pictures for further study. The "ideal" pattern would appear to be to select for study those items which had not been previously recalled, i.e., ones which were causing particular problems for the learner; and in fact this is what college students do, both in a free recall task (Masur et al., 1973) and in a text studying situation (Brown & Campione, 1979).

The retarded adolescents did not show this strategic selection during a baseline phase of the experiment, and there was no age difference in recall prior to intervention. When both groups were required to study missed items, the older group (who had a mean mental age of 8 years) significantly surpassed the younger group (mean mental age of 6 years), again a divergent effect. The data here indicate that the study time apportionment strategy can help students learn more quickly, but that the young sample seemed to lack some other skills necessary for its use. Their recall pattern was informative in this regard. They tended to recall the studied items (one-half the total set), but not the unstudied but previously recalled set. The interpretation proferred was that they failed to continue to attend to, or rehearse, that set. The failure to produce this essential activity led to the failure of the overall approach. In this case, the pattern of recall provided clues about the specific additional components which needed to be taught to improve the effectiveness of the instructional package.

In a reading comprehension intervention study, Gordon and Pearson (in press) provide a third example of the divergent pattern of results depicted in Panels E and F. Fifth graders of high and average ability received eight weeks of instruction in one of two procedures designed to increase their ability to make inferences from stories. In one treatment (Content and Structure), students were taught to relate new information to prior knowledge within a structural framework for stories (a simplified story grammar). In the second treatment (Inference Awareness), students were taught, through modeling and feedback, a step-by-step procedure for drawing inferences from the text and evaluating the plausibility of those inferences. Higher ability students improved their story comprehension (as measured by both experimenter-designed and standardized-tests) more as a result of the instruction than did lower ability students. In addition, higher ability students showed greater improvement in ability to recall stories after content and structure training than did lower ability students. Gordon and Pearson speculated that complexity of training procedures or difficulty of training materials may have been responsible for the divergent pattern of results.

In the balance beam and study time examples, the divergent effect indicated that the approach taken was a reasonable one, and that more input would need to be provided to make the teaching packages more effective for the L children. As we know a considerable amount about determinants of performance on both domains, it was possible to develop more powerful procedures. These procedures were based on a detailed analysis of the younger children's response protocols. In the area of comprehension, where our models are not as detailed, this may be more difficult. But the presence of a divergent effect, for example, would at least provide information about the directions future remediation attempts might take, information which simply would not be available if only the younger or poorer groups were included in the research.

While there are other outcomes which are possible, this set is sufficient to show some of the types of additional information which can be obtained by the simple expedient of including instruction for older subjects in an age/ability x instruction factorial design. To add further to the analysis, we would also like to argue that a number of other factors -- specifically the criterion task used to assess the effects of training and task difficulty -- can influence the specific outcome obtained in a particular study.

The Criterion Measure

To demonstrate the effects of this variable, consider a series of experiments on teaching self-monitoring skills to mildly retarded children (Brown & Barclay, 1976; Brown, Campione, & Barclay, 1979). The children were required to study a set of items larger than their memory span for as long as they wanted until they were sure they could recall all the items. Baseline performance was poor, and instruction was undertaken. In some conditions, the children were taught both procedures for learning the items and methods for checking on their state of learning. The effects of this strategy plus regulation training for the older (MA = 8 years), but not the younger (MA = 6 years), children were: immediate beneficial effects of the instruction; maintenance of the strategy over a one-year period;

and evidence for generalization to a quite different task -- studying and recalling prose passages. The younger group showed only immediate effects of training; on maintenance probes given a few days after training, they reverted to baseline levels of performance, although mild prompts were sufficient to elicit the trained activities even one year later.

If we consider this age x instruction experiment, which of the various outcomes illustrated in Figure 1 best typifies the results? Note that if we adjust for memory span differences, the MA 6 and MA 8 groups did not differ significantly prior to training. Immediately after training, the subjects were given a prompted posttest (on which they were told to continue executing the trained activities); both groups improved significantly, and there was still no reliable difference between them. Given these data, parallel improvement (Panel D) could be said to be the result. When unprompted tests were given a day later, however, the younger group abandoned the trained routines, and their performance reverted to baseline levels. The older subjects, in contrast, continued to perform well, and for the first time, there was a significant difference between the groups. If degree of independent (unprompted) learning is the criterial task, a divergent pattern (Panel F) is obtained. If we add to that the fact that the older children demonstrated transfer to a prose recall task, the divergent pattern becomes even more pronounced. Thus, when initial response to instruction is the metric, studies which produce convergent patterns (Panels A-C) might turn out to produce a divergent effect (Panels F and G) if more demanding criteria, such as maintenance and transfer, are included.

A similar example within the area of comprehension can be found in the Hansen and Pearson (1982) work mentioned earlier. Recall that they obtained either relative convergence (Panel B) or parallel improvement (Panel D), depending upon the criterion measures used to evaluate the results of training. Relative convergence was the result when the criterion measure was performance on worksheets accompanying the stories used during instruction, while parallel improvement was the result when the dependent measure was performance on a transfer task.

Difficulty of Trained Activity

To illustrate this issue, we can consider an experiment by Day (1980) aimed at teaching junior college students strategies for summarizing expository prose passages. The instruction consisted of teaching a set of rules of varying difficulty which could be used to generate adequate summaries (adequate in the sense that they would include the main points of the text and be judged acceptable by college rhetoric teachers). Day also worked with students of varying ability levels: those with no diagnosed reading or writing problems; some with writing problems; and a final group who were receiving remedial help in both writing and reading. Ignoring the details of the different rules, we can classify them into three difficulty categories: easy, intermediate, and difficult. The ability x instruction interaction took different forms depending upon this variable. Prior to instruction, the groups did not differ with regard to use of any of the rules. All were proficient when the easiest cases were investigated; hence, training produced no improvement. For the intermediate rules, a pattern of parallel improvement was found; all groups improved, and to about the same extent. With the most difficult rules, however, a divergent pattern was obtained. The most proficient students showed the largest improvement; those with only writing problems showed some but significantly smaller gains; and the poorest students' rule use was unaffected by instruction.

While Day's experiment was more complicated than described here, we can summarize the main point for our purposes fairly simply. When we restrict our attention to one of her teaching procedures, featuring both a detailed description of the various rules and explicit instruction in the management of those rules, the relative effects of that general approach on the different ability groups was systematically related to the difficulty parameter. As the complexity of the specific rule under scrutiny increased, the tendency toward, and magnitude of, the divergent effect increased.

Summary

In this paper, we discussed the training approach frequently used in the developmental/instructional literature. This approach involves data from three different conditions. Younger and older (or L and M) students are tested under unprompted conditions to assess the presence and magnitude of some developmental or comparative difference. The L are then instructed, and after a suitable intervention, their performance may improve to the level of the contrast group. We might then infer that (a) the activities manipulated during training were important components of adequate task performance, (b) the differential use of those activities was responsible for the original group differences, and (c) with suitably older or more proficient students, the same training programs would have been redundant with what those students were already doing and hence relatively unnecessary.

We argued that while such conclusions were possibly (even probably) correct, more stringent analysis would require additional data of two sorts. First it would be highly desirable to have data on the quality and extent of production of the target activities by the students during and following instruction; telling students to do something does not guarantee that they do it well, or at all. Such data can help in a number of ways. Obviously, if students do not use the activities at all, or produce only marginal approximations of what is intended, we would not expect training to be effective. More interestingly, if we do have measures of the topography of students' productions of the activities, we may be able to use that information to refine our approach. For example, we may find that students who do not improve markedly produce different or less complete examples of the target activities than do more successful tutees. The specific ways in which the groups' actual activities differ can then be used to modify instruction for those who are not benefitting as much as hoped.

Second, we advocated the addition of data from the fourth cell of a hypothetical factorial design--the performance of M students following the same instruction afforded the L students. From that factorial design, a number of different patterns could and do emerge, ranging from complete or partial convergence through parallel improvement to various degrees of divergence. While we do not wish to claim that any particular outcome leads to a unique interpretation, we do argue that the different outcomes can preclude the strongest interpretation suggested by the modal package and do succeed in constraining significantly the possible interpretations which can be made.

The addition of the fourth cell helps us to evaluate in much more detail hypotheses about the source and nature of developmental or comparative differences in task performance, estimate the presumed competence of more mature subjects, assess the appropriateness and completeness of our task analysis, and derive hints about the directions in which instructional packages need to be modified to increase their power.

We also noted some data which make it clear that the outcomes we obtain and the resulting interpretations can be influenced by other factors, including the criterion measure against which "success" is measured. The implication is that we need to consider these factors carefully when we formulate our explanations of training studies, and that in some cases it may be necessary to include these variables directly in our research programs before a clear picture can emerge.

While the interpretation of training studies is not a simple matter, we believe that they represent a significant methodology for attempts to understand the nature of active comprehension and to design instructional programs which can aid students to become more proficient comprehenders. More to the point here, we believe that we have learned a considerable amount about the strengths, weaknesses, and interpretation of training studies from work in the areas of memory and problem-solving, along with more recent attempts in the area of comprehension.

As these lessons are noted and become applied to the area of comprehension, we believe that the instructional approach will yield valuable insights into both theories of comprehension and methods of teaching critical comprehension skills. Finally, on a very global level, we regard "comprehension" as a more difficult task than "remembering." If the general conclusions about the effects of task difficulty we have drawn are correct, we should find that divergent effects are likely to be the modal outcome in research addressing the teaching of comprehension-fostering activities. Essentially, this would suggest that advanced students are not nearly as proficient gleaners of meaning as we might assume them to be, and that their performance can be enhanced considerably by the kinds of detailed training procedures which have been developed in the "simple" memory tasks upon which we have lavished so much attention.

References

André, M. D. A., & Anderson, T. H. The development and evaluation of a self-questioning study technique. *Reading Research Quarterly*, 1978-79, *14*, 605-623.

Belmont, J. M., & Butterfield, E. C. Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, 1971, *2*, 411-420.

Brown, A. L. Mnemonic elaboration and recency judgments in children. *Cognitive Psychology*, 1973, *5*, 233-248.

Brown, A. L., & Barclay, C. R. The effects of training specific mnemonics on the metamnemonic efficiency of retarded children. *Child Development*, 1976, *47*, 71-80.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Carmichael's manual of child psychology* (Vol. 1). New York: Wiley, in press.

Brown, A. L., & Campione, J. C. Training strategic study time apportionment in educable retarded children. *Intelligence*, 1977, *1*, 94-107.

Brown, A. L., & Campione, J. C. The effects of knowledge and experience on the formation of retrieval plans for studying from texts. In M. M. Gruneberg, P. E. Morris, & R. M. Sykes (Eds.), *Practical aspects of memory*. London: Academic Press, 1979.

Brown, A. L., Campione, J. C., & Barclay, C. R. Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. *Child Development*, 1979, *50*, 501-512.

Brown, A. L., Campione, J. C., Bray, N. W., & Wilcox, B. L. Keeping track of changing variables: Effects of rehearsal training and rehearsal prevention in normal and retarded adolescents. Journal of Experimental Psychology, 1973, 101, 123-131.

Brown, A. L., Campione, J. C., & Gilliard, D. M. Recency judgments in children: A production deficiency in the use of redundant background cues. Developmental Psychology, 1974, 10, 303.

Brown, A. L., Palincsar, A. S., & Armbruster, B. B. Inducing comprehension-fostering activities in interactive learning situations. In H. Mandl, N. Stein, & T. Trabasso (Eds.), Learning and comprehension of texts. Hillsdale, N. J.: Erlbaum, in press.

Brown, A. L., & Smiley, S. S. The development of strategies for studying texts. Child Development, 1978, 49, 1076-1088.

Butterfield, E. C., Siladi, D., & Belmont, J. M. Validating process theories of intelligence. In H. W. Reese & L. P. Lipsitt (Eds.), Advances in child development and behavior (Vol. 15). New York: Academic Press, 1980.

Butterfield, E. C., Wambold, C., & Belmont, J. M. On the theory and practice of improving short-term memory. American Journal of Mental Deficiency, 1973, 77, 654-669.

Campione, J. C., & Armbruster, B. B. Acquiring information from texts: An analysis of four approaches. In S. Chipman, J. Segal, & R. Glaser (Eds.), Thinking and learning skills: Relating instruction to basic research. Hillsdale, N. J.: Erlbaum, in press.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. M. Gagné (Ed.), Learning and individual differences. Columbus, OH: Charles E. Merrill, 1967.

Day, J. D. Training summarization skills: A comparison of teaching methods. Unpublished doctoral dissertation, University of Illinois, 1980.

Flavell, J. H., Beach, D. H., & Chinsky, J. M. Spontaneous verbal rehearsal in memory tasks as a function of age. Child Development, 1966, 37, 283-299.

Gordon, C., & Pearson, P. D. The effects of explicit inference training on students' comprehension and recall (Tech. Rep.). Urbana: University of Illinois, Center for the Study of Reading, in press.

Hansen, J., & Pearson, P. D. An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers (Tech. Rep. No. 235). Urbana: University of Illinois, Center for the Study of Reading, March, 1982.

Huttenlocher, J., & Burke, D. Why does memory span increase with age? Cognitive Psychology, 1976, 8, 1-31.

Lyon, D. R. Individual differences in immediate serial recall: A matter of mnemonics? Cognitive Psychology, 1977, 9, 403-411.

Masur, E. F., McIntyre, C. W., & Flavell, J. H. Developmental changes in apportionment of study time among items in a multitrial free recall task. Journal of Experimental Child Psychology, 1973, 15, 237-246.

Meyer, B. J. F. Text dimensions and cognitive processing. In H. Mandl, N. Stein, & T. Trabasso (Eds.), *Learning and comprehension of texts*. Hillsdale, N. J.: Erlbaum, in press.

Palincsar, A. S., & Brown, A. L. *Reciprocal teaching of comprehension-monitoring activities*. Unpublished manuscript, University of Illinois, 1982.

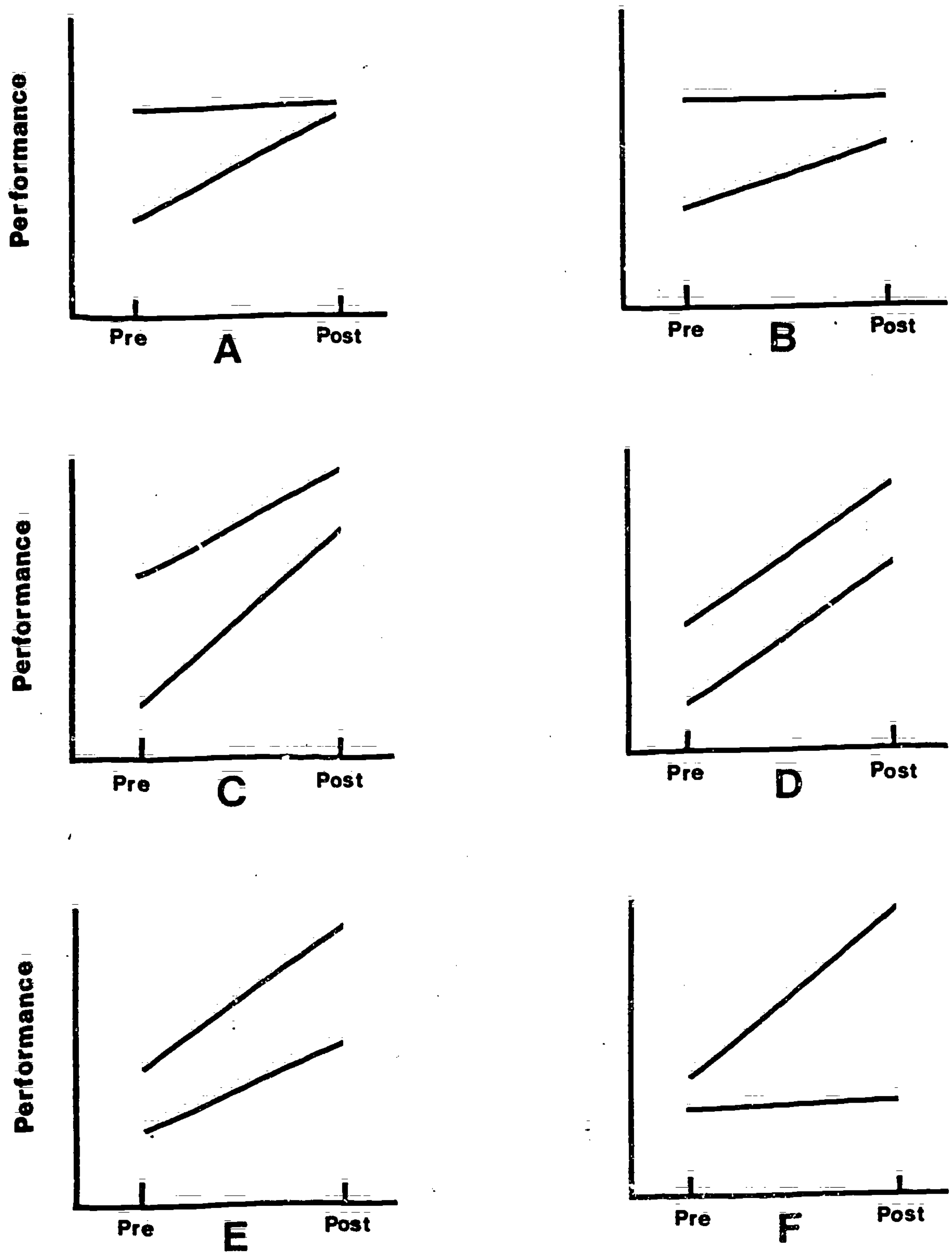
Siegler, R. S. Three aspects of cognitive development. *Cognitive Psychology*, 1976, *8*, 481-520.

Siegler, R. S. The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N. J.: Erlbaum, 1978.

Snow, R. E., & Yalow, E. Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. New York: Cambridge University Press, 1982.

Figure Caption

Figure 1. Possible outcomes from the ability x instruction design. The data points on the left of each panel represent performance prior to training; those on the right represent performance following training. The upper curve represents the data of the originally more proficient group; the lower curve depicts the performance of the originally less successful group.


Performance
Pre
A
Post
Pre
B
Post
Performance
Pre
C
Post
Pre
D
Post
Performance
Pre
E
Post
Pre
F
Post