Retention of Problem-Solving Performance in School Contexts

John A. Ross and Florence J. Maynes
the ontario institute for studies in education[1]

Previous research into retention issues has been limited by the narrow range of learning outcomes investigated, by the preference for laboratory contexts rather than naturalistic settings, and by the short duration of the retention periods studied. This article reports data on the retention of problem-solving performance after one month, four months, and six months, in three separate studies conducted in school settings involving a variety of grades and subjects.

Les recherches antérieures sur les questions de rétention ont été limitées par l'étroit éventail des résultats cherchés, par la préférence pour le contexte de laboratoire plutôt que l'environnement naturel, et par les courtes périodes de rétention étudiées. On présente dans cet article les données sur la rétention de la possibilité de résoudre des problèmes après un mois, quatre mois, et six mois, dans trois études séparées entreprises dans une école, à des niveaux et pour des sujets variés.

The erosion of learning over time is a fundamental problem facing teachers and curriculum designers. How can learning accumulate if its outcomes are ephemeral?

This article reviews evidence from previous studies of retention and predicts that instruction in problem-solving skills is likely to have enduring effects on student achievement. The prediction is tested in a series of instructional experiments conducted in elementary and secondary schools in eastern Ontario. Two studies provide evidence that there is virtually no loss of learning one month after instruction for students in Grade 6 science and Grades 7–10 geography. A third study, involving students in various grades and subjects, indicates that only modest declines occur after four months and six months. These results have implications for teacher decisions in allocating classroom time. The conception of problem solving presented is an information-processing account in which the construct of framework (an externally represented schema) is central.

PREVIOUS RESEARCH ON RETENTION OF LEARNING

Although academic interest in the problem of retention has waned in recent years, the body of literature on the topic is vast. Its utility for school investigators tends to be limited in three ways.

First, most inquiries have been conducted in laboratory contexts rather than in naturalistic settings; applying the findings of these studies to learning in schools is not a matter of simple extrapolation. Learning experiments conducted in the school, in contrast with those undertaken in the lab, are characterized by distractions: interruptions from the office, student absences, and classroom management problems, to name but a few. They are influenced in unforeseen ways by existing classroom routines which can sustain (or more likely impede) innovation, and they are administered in imperfect and often unintended ways. The classroom is an inhospitable place to study learning but it is the place that ultimately matters: researchers who limit their inquiries to laboratory contexts fall prey to "ecological invalidity" (Cole, Hood, & McDermott, 1979).

Second, the bulk of previous research has been limited to exploring the retention of low level cognitive outcomes (Gagné, 1978). The intentions of the school curriculum, reflected for example in governing curriculum guidelines, are much broader than this (Robinson, 1978) and typically give more emphasis to higher order objectives, a category of intended learning outcomes that includes, among others, complex intellectual skills such as the development of problem-solving competencies.

The third limitation of previous retention studies is that most are relatively short. Retention is typically measured over a few days or a week; Tyler's (1934) exemplary year-long study of retention has rarely been emulated. Such short timelines are of only peripheral concern in school contexts, being analogous to an interest in "cramming." Of much greater import are the enduring effects of instruction in which retention is measured in months and years.

Despite these limitations, several conclusions directly relevant to the retention of problem-solving performance can be drawn from the literature on retention of knowledge.

(1) There is a rapid loss of detailed information.
(2) The main ideas (e.g., of a prose passage) are retained for a longer period; (see for example the data reviewed by Ausubel & Robinson, 1968, and more recently, Riding & McQuaid, 1977, and Spiro, 1977). These findings are based upon the well-substantiated but still controversial theory that new information is stored by the learner at

the highest level of generality in schemata which govern retrieval (Calfee, 1981; Chase, 1978; Rumelhart & Norman, 1975).

(3) Retention is enhanced when the material to be learned is well organized (Shimmerlik, 1978), when the organizational structure is made explicit to students (Sherris & Kahle, 1984), when the learner has relevant prior knowledge appropriate to the new material (Gagné, 1978; Stein & Trabasso, 1980), and when the learner is required actively to process new information in various ways (Mackenzie & White, 1982).

This distillation of the literature on the retention of knowledge suggests some ground for optimism: the loss-of-learning curve for problem-solving performance may be much flatter and less dramatic than the pattern of rapid decline found for low level cognitive objectives.[2]

CONCEPTION OF PROBLEM SOLVING

Some evidence to help validate optimism concerning the retention of problem solving is provided in a series of instructional improvement experiments conducted in selected elementary and secondary schools in Ontario. Since a summary of the scope of this thrust is provided elsewhere (Robinson, Ross, & White, 1985; Ross & Maynes, 1983c), only a brief account will be given here.

In this approach, problem solving is conceptualized as a complex intellectual skill which requires the assembly and coordinated performance of a set of simpler component skills. The core skill in this conception is developing a framework. A framework is an external representation of the problem in a structure conducive to problem solution. It is a visible schema constructed by the problem solver that identifies problem elements and their relationship. A sophisticated problem solver works through a series of steps that constitute a generic model of problem solving: the problem solver develops a framework for the problem, fills the framework with relevant data, processes data within the framework, extrapolates beyond the framework (if appropriate to the situation), and reports the findings to an audience. Each step in the generic model can be decomposed into a series of discrete skills (11 in our current conception) which can be programmed across school divisions.

The generic model applies whenever a novel problem is confronted. But the problems presented to students in schools can also be classified into recurrent types. Such problematic situations as comparative, experimental, decision-making, and correlational inquiries tend to be encountered in many grades and subjects. They are distinguishable in that the frameworks for each type have distinctive features.

Problem-solving competence develops slowly through the grades. By analyzing student and adult behaviour, growth schemes (or learning hierarchies) can be constructed. These growth schemes describe the maturation of student problem-solving performance in a structure that facilitates the design of instruction.

An example of a growth scheme for problem solving is displayed in Figure 1. The figure shows increasing levels of sophistication in developing a framework for a comparative problem. The essence of each level is summarized in the first column of the figure; examples of student frameworks (in this instance, comparing living in the country to living in the city) are provided in the second column, and the third column describes an increasingly compact set of propositions or rules outlining how students might construct their frameworks. Similar growth schemes have been developed for other recurrent problem situations for various component skills (e.g., Ross & Maynes, 1981).

Instructional strategies have been designed to improve student problem-solving performance. These strategies involve making students' underlying cognitive structures explicit, modifying them in the direction specified by the growth schemes, and reinforcing the new levels of performance with practice activities. These strategies have been incorporated into instructional materials developed within existing elementary and secondary school courses, and have also been included in inservice programs. Various student assessment devices have been constructed to place students on the target growth schemes in order to measure student growth (and the effectiveness of instructional materials). These instructional improvement projects typically involve extensive collaboration between academics and teachers in designing materials and careful pilot testing in volunteer classrooms before broader dissemination is attempted.

STUDIES OF RETENTION OF PROBLEM SOLVING AFTER ONE MONTH

Science Study

The first study for which data on the retention of problem-solving performance could be obtained involved an attempt to improve the experimental design skills of Grade 6 science students (N = 265). An instructional program was field tested in a design in which students in the delayed treatment condition served as controls for students in the early treatment condition. Although the research design was a true experiment, data on retention of learning were available only for the early treatment group, rendering the design of this portion of the study pre-experimental.

Student performance was measured at one-month intervals on three occasions using multiple-choice and open-ended assessment instruments described in Ross and Maynes (1983a). The results of the study (reported in detail in Ross & Maynes, 1983b) indicated that the program positively affected student performance. Virtually no loss of learning existed for the early treatment group on the retention test administered one month after the posttest. When retention test scores were expressed as a percentage of posttest performance, the proportion of learning retained ranged from 93% to 105%.

Geography Study

The second study involved geography students in Grades 7, 9, and 10 ($N = 195$). This replicated the science study involving a different type of problem (in this case decision making) at different grade levels in a different school system. Instructional programs were field tested using a delayed treatment design. Classes of students were randomly assigned to treatment conditions, students were randomly assigned to test orders, and assessment involved multiple-choice and open-ended instruments (described in Ross & Maynes, 1981). The results of the study (reported in detail in Maynes & Ross, 1984) indicated that the program had a positive effect on student problem-solving performance. Again there was virtually no loss of learning for the early treatment group on the retention test administered one month after the posttest. The proportion of learning retained ranged from 96% to 107%.

RETENTION OF PROBLEM-SOLVING PERFORMANCE AFTER MORE THAN ONE MONTH

K-13 Study Description

The third study involved a series of eight in-service sessions for 22 volunteer teachers (four to seven from each of the primary, junior, intermediate, and senior divisions), drawn from 13 schools in a school system not involved in the science or geography studies. The first in-service session provided an overview of problem-solving instruction. Subsequent pairs of sessions provided detailed input on comparative, decision-making, and experimental thinking; in each case a cycle of input, classroom practice, and joint reflection on classroom experience was followed. The two final sessions addressed correlational thinking, without a classroom teaching component. The classroom teaching experience

<table>
<tr><th>Level</th><th>Example of Framework</th><th>Rule Structures*</th></tr>
<tr><td>1. No comparison</td><td>
<table>
<tr><th>Country Life</th><th>City Life</th></tr>
<tr><td>–must ride bus to school
–lots of farms
–have to have a car to go anywhere</td><td>–high taxes
–houses close together
–lots of factories</td></tr>
</table>
</td><td>1. Make a chart with 2 columns label each column with things to be compared
2. Fill in the chart</td></tr>
<tr><td>2. Comparison in one category</td><td>
<table>
<tr><th>Category</th><th>Country Life</th><th>City Life</th></tr>
<tr><td>Type of Work</td><td>farming</td><td>factories</td></tr>
</table>
</td><td>1. Make a chart with 2 columns
2. Fill in the chart
–put one fact in one column
–put matching fact in next column
3. Add category column
–invent name for each pair of matching facts</td></tr>
<tr><td>3. Comparison in many categories</td><td>
<table>
<tr><th>Categories</th><th>Country Life</th><th>City Life</th></tr>
<tr><td>Work</td><td>farming</td><td>factories</td></tr>
<tr><td>Travel to School</td><td>must take bus</td><td>walk or bicycle</td></tr>
<tr><td>Other Travel</td><td>must be driven by parents</td><td>can take bus, bike or walk</td></tr>
<tr><td>Houses</td><td>houses far apart</td><td>houses close together</td></tr>
<tr><td>School</td><td>required to go</td><td>required to go</td></tr>
</table>
</td><td>1. Make a chart with 2 columns
2. Fill in the chart
3. Add category column
–invent name for matching facts
–repeat until all information included</td></tr>
</table>

FIGURE 1

Example of a growth scheme for problem solving: frameworking a comparative problem

4. Summary of comparisons

Categories	Country Life	City Life	Summary
School	must attend school	must attend school	same
Work	farming	factories	different
Houses	far apart	close together	different
Postal Delivery	every day	every day	same

1. Make a chart with 2 columns
2. Fill in the chart
3. Add category column
4. *Add summary column*
   –indicate same or different on each category

5. Classification of comparisons

Categories	Country Life	City Life	Summary
Transportation			
–to school	must take bus	walk or bicycle	different
–other	must be driven by parents	can take bus, bike or walk	different
Housing			
–type	single family	single or multiple family	different
–density	houses far apart	houses close together	different
–maintenance	must cut grass	must cut grass	same
Postal Delivery			
–when	every day	every day	same
–who	mailman	mailman	same
–how	by car or truck	on foot	different

1. Make a chart with 2 columns
2. Fill in the chart
3. Add category column
4. Add summary column
5. *Classify categories*

–put categories into groups
–invent name for each group
–invent new categories for each group
–invent new groups

*Italics indicate what is new in each level.

FIGURE 1 (*Continued*)

required teachers to develop lesson plans for their subject-grade context by adapting model instructional materials for each problem type (provided in Ross & Maynes, 1982) and implementing these lesson plans in their own classrooms.

Each model unit contained detailed lesson plans that integrated knowledge and skill goals in a sequence of instructional activities requiring six to eight 30-minute periods. The instructional design of each unit had students perform the target skill; the lowest level of performance was displayed; the inadequacies of the lowest level were demonstrated; students were stimulated to invent the next highest level of performance in the growth scheme, and ample practice in various contexts was provided. Subsequent lessons contained a similar format to introduce and consolidate new levels of performance.

Teachers were also asked to test student performance before and after each instructional episode, and at the end of the school year. Each test consisted of a single open-ended item that required students to construct a framework for a particular instance of a problem type. Pre- and posttests were teacher adaptations of model test items using content situations considered appropriate for their classes. The final summative test was supplied by the investigators: the same instrument was administered in all classes within a divisional grouping.

Pre- and posttests were all marked by the teachers themselves following procedures provided at in-service sessions. The marking task involved recapturing the framework implicit or explicit in the student's response and assessing it by placing it on the appropriate growth scheme. The use of a particular format was irrelevant to the scoring. For example, students could describe their thinking in many different ways – in a point outline, in diagrams of various sorts or in a paragraph – and receive the same interpretation of their response. The central concerns were not mode of representation but the underlying cognitive structure: the dimensions of the problem and the relationships established among these dimensions identified by the student. Scoring of all final tests was done by one marker, previously trained and experienced, who used the same procedures as the teachers used with pre- and posttests. A previous investigation (Ross & Maynes, 1983b) found that the interrater reliability between a trained tester and classroom teachers ranged between 0.80 and 0.86.

These data collections provided information on retention after six months for comparative problem solving, after four months for decision making and after one month for experimental thinking. All data collection activities in the study were voluntary and many teachers exercised their option at various points to delete one or more test administrations. The reasons for opting out varied with individuals. The most common reasons concerned competition for data collection activities

TABLE 1
Structure of the data sets available for analysis

Available Scores	*Problem Type*	*Number of Classes*				
		Primary	*Junior*	*Intermediate*	*Senior*	*Total*
Pre-, post- and final test scores available	Comparative	0	5	4	0	9
	Decision-making	0	6	4	0	10
	Experimental	1	5	3	0	9
Pre- and post-test scores only available	Comparative	4	1	2	4	11
	Decision-making	4	0	0	1	5
	Experimental	2	0	0	0	2

from other agents in the school system, a reluctance (particularly on the part of secondary teachers), to allocate more instructional time to testing, and many teachers' not unreasonable belief that their own classroom observations provided sufficient information for them to judge the effectiveness of the in-service program. There were no overt indications that teachers opted out of testing because they believed students were not performing well; further evidence will be presented below in Table 5. The structure of data sets available for analysis is displayed in Table 1.

K-13 Study Findings

Tables 2, 3, and 4 show the analyses of variance of class means where testing was done on all three occasions. It should be noted that, although the scores in each case are growth scheme levels, the levels do not correspond exactly across problems, and scores are not directly comparable across problem types. However, the increments between one level and another represent roughly equivalent amounts of learning regardless of problem type.

The tables show that students' skill performance improved substantially from pre- to posttest with some loss from the posttest to the final test occasion for all three problem types. The extent of the loss is discussed below. Linear contrasts displayed in Tables 2 to 4 show that both the retention scores and the posttest scores were significantly greater than the pretest for all problem types.

The retention of learning (expressed as a percentage of posttest performance) was reduced for all problem types in this study. For comparative problem solving, retention after six months was 83%; for

TABLE 2
Analysis of variance of class means: Comparative problems

Source of Variation	*S.S.*	*d.f.*	*M.S.*	*F*
Between Classes	6.36	8	0.80	
Within Classes	33.22	18	1.85	
Occasions	21.59	2	10.80	14.86*
Residual	11.62	16	0.73	
Total	39.58	26		
	$\bar{X}$	S.D.	*n*	
Occasion 1 (pretest)	1.79	0.84	9 classes	
Occasion 2 (posttest)	3.93	0.74	9 classes	
Occasion 3 (retention)	3.25	0.94	9 classes	

Linear Contrasts of Means

Occasion 2 *vs* Occasion 1: $F = 28.45$*
Occasion 3 *vs* Occasion 1: $F = 13.26$*

*$p < .01$

TABLE 3
Analysis of variance of class means: Decision-making problems

Source of Variation	*S.S.*	*d.f.*	*M.S.*	*F*
Between Classes	10.33	9	1.15	
Within Classes	37.20	20	1.86	
Occasions	32.60	2	16.30	63.72*
Residual	4.61	18	0.26	
Total	47.53	29		
	$\bar{X}$	S.D.	*n*	
Occasion 1 (pretest)	1.13	0.30	10 classes	
Occasion 2 (posttest)	3.40	0.82	10 classes	
Occasion 3 (retention)	3.19	0.88	10 classes	

Linear Contrasts of Means

Occasion 2 vs Occasion 1: $F = 101.43$*
Occasion 3 vs Occasion 1: $F = 89.35$*

*$p < .01$

TABLE 4
Analysis of variance of class means: Experimental problems

Source of Variation	*S.S.*	*d.f.*	*M.S.*	*F*
Between Classes	1.28	8	0.16	
Within Classes	23.67	18	1.32	
Occasions	20.02	2	10.01	43.82*
Residual	3.65	16	0.23	
Total	24.95	26		

	$\bar{X}$	S.D.	*n*
Occasion 1 (pretest)	1.30	0.48	9 classes
Occasion 2 (posttest)	3.40	0.30	9 classes
Occasion 3 (retention)	2.07	0.72	9 classes

Linear Contrasts of Means

Occasion 2 vs. Occasion 1: $F = 87.53^*$
Occasion 3 vs. Occasion 1: $F = 24.52^*$

*$p < .01$

decision making, retention after four months was 94%; for experimental thinking, retention after one month was 62%. No clear relationship is visible between the proportion of performance retained and the duration of the retention period. Teachers reported their estimates of the time devoted to instruction in each problem type (data not displayed); no clear relationship was found between retention and these estimates.

Table 5 compares the pre- and posttest scores of classes that completed the final testing with those that did not. In only one case, the posttest of experimental problem solving, is there any significant difference between the means of two groups of classes. Only two classes, both in the primary division, actually did this pre- and posttesting without also completing the retention test. On the whole it appears that the teachers who failed to administer the summative test did not do so because they thought their students were in difficulty and would not succeed on it.

DISCUSSION

The findings from these three studies provide some support for earlier optimism about the retention of problem-solving performance. There was virtually no loss of learning in the first study after one month, and this finding tended to be maintained in the second study when variations were introduced in the subject, grade level, problem type, and geographic

TABLE 5
Pre- and posttest means and standard deviations in classes that did and did not complete the final test

Test	*Final Test Completed*	*n*	*Final Test Not Completed*	*n*	*t*
Comparative					
Pretest	1.79 (0.84)	9	2.41 (0.95)	11	0.15 ns
Posttest	3.93 (0.74)	9	3.78 (0.73)	11	−0.46 ns
Decision-Making					
Pretest	1.13 (0.30)	10	1.26 (0.50)	5	0.67 ns
Posttest	3.40 (0.82)	10	2.87 (1.87)	5	−0.76 ns
Experimental					
Pretest	1.29 (0.48)	9	1.00 (0.00)	2	−0.83 ns
Posttest	3.40 (0.30)	9	1.64 (0.52)	2	−6.82*

NOTE: All *t*-tests were two-tailed tests of probability
$*p < .01$

locality of the investigations. In the third study, there was some loss of learning of experimental problem-solving skills after one month, but only very moderate loss of comparative and decision-making skills after four and six months.

In each study, the instructional treatment modified and consolidated students' mental structures, enabling them to develop a relatively stable executive schema for problem solving. This schema was introduced in one curriculum-specific context and practiced in other contexts. For example, in the geography study, the decision-making schema was initially developed in a real-life context; it was repeated with a geographic decision, and student practice involved working through additional decision problems drawn from the same geographic topics. The instructional treatment thereby focussed student learning on the schema – the commonalities discovered in working through related problem instances – rather than on the particular details of the problem instances themselves. Since previous research has indicated that main ideas (in these studies the problem-solving schema) are less likely to be forgotten, the high rates of retention are understandable.

The same focus on main ideas (the schema) was maintained in the testing procedures. In every instance the items required performing the target skills in novel contexts not encountered by students in the instructional treatment. Each item tested specific transfer where detail retention of the content topic was valueless.

Although conducted in field settings, the one-month studies were only partially naturalistic in their intervention. Teachers were provided with highly detailed lesson plans which served as virtual scripts. Such curriculum support is rarely available to classroom teachers who are more typically required to develop their own lesson plans. Furthermore the number of teachers involved was very small, facilitating attempts by the investigators to ensure high-fidelity implementation. Wise and Okey (1983) found in a meta-analysis of 160 science studies that the effect sizes of treatments declined significantly as the number of teachers involved in the studies increased. One could therefore anticipate the finding from the K-13 study that retention of problem-solving performance would be lower in studies involving a broader segment of the teacher population.

The K-13 study provided a more naturalistic intervention in teachers' practices. The data indicate that retention of problem solving is not limited to programs based on high-fidelity use of prescribed lesson plans. On the contrary, the data suggest the viability of requiring teachers to develop personal lesson plans by adapting exemplary materials developed for other curriculum contexts. This finding suggests that a highly prescriptive curriculum need not be produced; this is particularly important given the deep-seated preference of teachers to maintain a preeminent role in program decision making at the lesson level. For example, Wade (1983), in a meta-analysis of social studies programs, found that the developmental role of the key teacher was central to program effectiveness. Programs which were entirely controlled by the instructional designer requiring high-fidelity implementation from teachers tended to be far less effective than those that allowed teachers to modify instructional models and share a collaborative role in development.

The three studies generated retention data for a broad array of subjects and grades. Although the findings are somewhat sketchy because of small cell sizes, the K-13 study data in particular tentatively indicate that problem-solving retention does not vary significantly between subjects and grade divisions.

The K-13 study also provides evidence of retention of problem solving for periods longer than one month. Although retention was greater for four months than for six, it should be noted that the lowest retention rate was for the shortest period, suggesting that the relationship between retention and time is not simply linear. The design of the study confounds length of retention time and type of problem presented, thus precluding any overall statements regarding the proportion of performance retained and length of retention period. These data suggest a complex relationship in which length of retention period is only one variable among many; other variables (such as the extent and quality of practice, characteristics of the learning task, and characteristics of the learner) are likely to be equally salient.

The K-13 study provides information lacking in the one-month studies about the effects of seriating problem-solving episodes in that a different type (requiring a slightly different schema) is addressed in each instructional segment. The K-13 study data provide some evidence that introducing a multiplicity of problem types leads to interference rather than facilitation as proposed by Ausubel. For example, students tended to confuse one schema with another in responding to the final test. In fairness to Ausubel's theory it should be noted that, contrary to his intention, the program as a whole did not provide advance organizers or other devices to help students overtly distinguish between different problem-solving schemata developed over the course of the intervention. This is particularly important in terms of retention because Luiten, Ames, and Ackerson (1980), in a meta-analysis of 135 studies of advance organizers, found that the effect of advance organizers increased over time (although the retention period in most of the studies cited was less than four weeks and never exceeded 15 weeks). Furthermore, teachers' comments during the intervention suggest that teacher fatigue should not be eliminated as a possible explanation for the poor retention of experimental-design performance (since experimental thinking was the final instructional episode in the series).

These three studies are limited in several ways. First, no data exist on the role played by learner characteristics in the retention of problem-solving. Since there is evidence that prior knowledge structures of learners positively and negatively affect problem-solving performance (Champagne & Klopfer, 1981; Chi, Glaser, & Rees, 1982; Clement, 1981) and evidence that cognitive styles (i.e., field independence) also affect problem solving (McCurdy & Ronning, 1982; Ross & Robinson, in press; Shymansky & Yore, 1980), it is reasonable to anticipate that these learner characteristics would be reflected in differential rates of problem-solving retention. Gagné (1978) identifies other learner characteristics that have been found to affect retention of prose learning; she particularly cites general ability, learner self-confidence, and the extent to which individuals tend to separate new information from old. One would also expect these to affect retention of problem solving.

A second major limitation of the K-13 study concerns lack of knowledge of the quality, frequency, and duration of the practice activities provided to students during the instructional episodes and retention periods. Attempts to obtain detailed information on practice were largely unsuccessful. Teacher self-reports suggested that the amount of massed practice provided during the instructional episodes was lower in the K-13 study than in the science and geography studies, but a valid quantitative measure could not be obtained. Data on distributed practice through the retention periods were even sketchier, with most teachers providing ambiguous reports alluding to incidental reinforcement, a vague construct that could not be defined with any confidence. There were indications that teachers found it easier to provide incidental practice for comparative problem solving than for the other problem-solving schemata, but it is not known what this practice entailed; for example, practice could have taken the very weak form of displaying information about a new topic within a comparative framework, or it could have taken a strong form in which students were required to develop a comparative framework using information from a new topic.

Finally, the three studies are limited by their pre-experimental research designs: there were no control groups. The most serious threats to internal validity in single group pre-post comparison designs are testing effect, instrument decay, history, and maturation (Campbell & Stanley, 1963). However, there is reason to believe that in the one-month studies, at least, it is unlikely that any of these forces provides a powerful alternative explanation for the findings.

Consider testing effect. The concern is that a pretest might influence scores on a posttest, thereby confounding the interpretation of treatment effects. The rationale is that the first administration sensitizes students to what will be taught (e.g., a pretest might serve as a weak advance organizer); similarly the first administration provides practice (a weak form of instruction) on the objectives contained in the program (Wilson & Putnam, 1982). It is reasonable to presume that, in each case, testing effects would be strongest between the first and second administration, gradually diminishing in subsequent administrations. Now, recall that in the one-month retention studies the third test administration provided the retention data. We would anticipate that any testing effect on the third occasion would be less than the testing effect on the second occasion. Since we have control group data for the earlier portions of these studies, we can establish that the testing effect was small and on most indicators non-significant between the first and second administration (Maynes & Ross, 1984; Ross & Maynes, 1983b).

For the one-month studies, a similar argument can be made with respect to the other threats to internal validity. In each case, control-group data are available for the first and second testing periods which indicate that instrument decay, maturation, and history had minimal effects. It is reasonable to extrapolate to the third testing period when the retention findings were produced. Of course no such argument can be made for the K-13 study.

The acknowledgment of design deficiencies should not lead researchers to discard the results of the study (although many would do so automatically). Educational researchers are becoming increasingly sensitive to one of the fundamental dilemmas of conducting research in schools: pursuit of internal validity impedes external validity and the well-controlled study contaminates natural processes. This recognition has been stimulated by ideological attacks (e.g., Schon, 1983) on the dominant positivist paradigm that governs most empirical research and has been encouraged by the discovery, through meta-analysis, that when research studies are aggregated, characteristics of research design are less salient than previously anticipated (Glass, McGaw, & Smith, 1981). From this perspective, the limitation of these three studies is not that they have design deficiencies, but that they have the same design deficiency.

Cronbach (1982) is particularly open to non-experimental designs. He proposes that general experience can provide a frame of reference in a single treatment study to judge the satisfactoriness of a program. Even stern exponents of the true experiment like Stanley support, in some circumstances, studies without any explicit control group. Stanley relies upon general experience as the basis for applying what he calls the "inter-ocular traumatic test of statistical significance" (Stanley, 1983, p. 36), that is, results so obviously contrary to ordinary experience that fresh data from control samples are not needed. Gilbert, Light, and Mosteller (1975) also talk about "slam-bang effects" for which control group data are unnecessary. The findings concerning problem-solving retention in our discussion are clearly less powerful than these, but the reasoning is the same: our everyday experience in schools shows that children forget what they have learned. When they do not do so it is noteworthy, even if the findings are not presented with new control group data.

IMPLICATIONS FOR SCHOOLS

The most important practical implication of this article's series of studies concerns the high rates of problem-solving retention after one month, four months, and six months. These retention rates far exceed comparable data for low level knowledge objectives, and they speak directly to the curriculum choices made by teachers and program designers. The scope of objectives contained in governing curriculum guidelines is so vast that teachers and local planners must set priorities in allocating instructional time. Given the long-term effects of teaching for various types of objectives, it makes sense for teachers and others to select disproportionately from the classes of outcomes likely to endure. The data reported here suggest that teachers who invest scarce classroom time in the pursuit of problem-solving objectives are likely to find their investment amply rewarded.

NOTES

[1] The research reported in this article was funded in part by the Ontario Ministry of Education, Transfer Grant Project 4028.

[2] The representation of loss of learning curves and their interpretation is a matter of some controversy (see for example Bahrick, 1965). The pattern of rapid decline for replicative knowledge is not, however, in dispute. Incidentally a similar pattern of decay has also been reported in the affective realm; a review of adult personal encounter groups found that the personal growth achieved over the course of the treatment does not tend to endure (Berman & Zimpfer, 1980).

REFERENCES

Ausubel, D. P., & Robinson, F. G. *School learning: An introduction to educational psychology.* New York: Holt, Rinehart & Winston, 1968.

Bahrick, H. P. The ebb of retention. *Psychological Review,* 1965, *72*(1), 60–73.

Berman, J. J., & Zimpfer, D. G. Growth groups: Do the outcomes really last? *Review of Educational Research,* 1980, *50*(4), 505–524.

Calfee, R. Cognitive psychology and educational practice. Pp. 3–73 in D. C. Berliner, *Review of research in education* (Vol. 9). Washington, DC: A.E.R.A., 1981.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental design for research on teaching. Pp. 171–246 in W. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand-McNally, 1963.

Champagne, A. B., & Klopfer, L. E. Student beliefs about gravity and motion. *Problem-Solving,* 1981, *3*(1), 12–13.

Chase, W. G. Elementary information processes. Pp. 19–90 in W. K. Estes (Ed.), *Handbook of learning and cognitive processes (Vol. 5): Human information processing.* Hillsdale, NJ: Erlbaum, 1978.

Chi, M. T. H., Glaser, R., & Rees, E. Expertise in problem solving. Pp. 7–76 in R. Sternberg (Ed.), *Advances in the psychology of human intelligence.* Hillsdale, NJ: Erlbaum, 1982.

Clement, J. Students' preconceptions in physics and Galileo's discussion of falling bodies. *Problem-Solving,* 1981, *3*(1), 3–5.

Cole, M., Hood, L., & McDermott, R. *Ecological niche-picking: Ecological invalidity as an axiom of experimental cognitive psychology.* New York: Rockefeller University, 1979.

Cronbach, L. S. *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass, 1982.

Gagné, E. D. Long term retention of information following learning from prose. *Review of Educational Research,* 1978, *48*(4), 629–665.

Gilbert, J. P., Light, R. J., & Mosteller, F. Assessing social innovations: An empirical base for policy. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment.* New York: Academic Press, 1975.

Glass, G. V., McGaw, B. V., & Smith, M. L. *Meta-analysis in social research.* Beverly Hills, CA: Sage, 1981.

Luiten, J., Ames, W., & Ackerson, G. A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal,* 1980, *17*(2), 211–218.

Mackenzie, A. A., & White, R. T. Field work in geography and long-term memory structures. *American Journal of Educational Research,* 1982, *19*(4), 623–632.

Maynes, F. J., & Ross, J. A. Integrating school curricula with social concerns: A cognitive approach to a progressive ideal. *Alberta Journal of Educational Research,* 1984, *30*(3), 179–193.

McCurdy, D. W., & Ronning, R. R. *The role of cognitive styles in problem-solving.* Paper presented at the National Association for Research in Science Teaching, Lake Geneva, WI, 1982.

Riding, R., & McQuaid, D. G. The effect of detail importance and passage structure on the retention of prose by twelve-year-old children. *Educational Studies,* 1977, *3*(2), 145–151.

Robinson, F. G. *The scope of guideline aims and objectives* (Information Bulletin No. 12). Toronto: Commission on Declining School Enrolments in Ontario, 1978.

Robinson, F. G., Ross, J. A., & White, F. *Curriculum development for effective instruction.* Toronto: OISE Press, 1985.

Ross, J. A., & Maynes, F. J. *Geographic thinking skills.* Toronto: OISE Press, 1981.

Ross, J. A., & Maynes, F. J. *Teaching problem-solving.* Toronto: OISE Press, 1982.

Ross, J. A., & Maynes, F. J. The development of a test of experimental problem-solving. *Journal of Research in Science Teaching,* 1983a, *20*(1), 63–75.

Ross, J. A., & Maynes, F. J. Experimental problem-solving: An instructional improvement field experiment. *Journal of Research in Science Teaching.* 1983b, *20*(6), 543–556.

Ross, J. A., & Maynes, F. J. Teaching problem-solving: An instructional design strategy. *Canadian Journal of Education,* 1983c, *8*(2), 155–173.

Ross, J. A., & Robinson, F. G. The use of rule structures in teaching experimental design to secondary school students. *Science Education* (in press).

Rumelhart, D. E., & Norman, D. A. The active structural network. Pp. 33–64 in D. A. Norman, D. E. Rumelhart & the L.N.R. Research Group (Eds.), *Explorations in cognition.* San Francisco: W. H. Freeman, 1975.

Schon, D. A. *The reflective practitioner: How professionals think in action.* New York: Basic Books, 1983.

Sherris, J. D., & Kahle, J. B. The effect of instructional organization and locus of control orientation on meaningful learning in high school biology students. *Journal of Research in Science Teaching,* 1984, *21*(1), 83–94.

Shimmerlik, S. M. Organization theory and memory for prose: A review of the literature. *Review of Educational Research,* 1978, *48*(1), 103–120.

Shymansky, J. A., & Yore, L. D. A study of teaching strategies, student cognitive development, and cognitive style as they relate to student achievement in science. *Journal of Research in Science Teaching,* 1980, *17*(5), 369–382.

Spiro, R. J. Remembering from text: The "state of schema" approach. Pp. 137–165 in R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, NJ: Erlbaum, 1977.

Stanley, J. C. Education in the fast lane: Methodological problems of evaluating its effects. *Evaluation News,* 1983, *4*(1), 28–46.

Stein, N. L., & Trabasso, T. What's in a story: An approach to comprehension and instruction. Pp. 213–267 in R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2). Hillsdale, NJ: Erlbaum, 1980.

Tyler, R. W. Some findings from studies in the field of college biology. *Science Education,* 1934, *18,* 133–142.

Wade, S. E. A synthesis of the research for improving reading in the social studies. *Review of Educational Research,* 1983, *53*(4), 461–499.

Willson, V. L., & Putnam, R. R. A meta-analysis of pretest sensitization effects in experimental designs. *American Educational Research Journal,* 1982, *19*(2), 249–258.

Wise, K. C., & Okey, J. R. A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching,* 1983, *20*(5), 419–435.

Bisherige Untersuchungen auf dem Gebiet der Merkfähigkeitsfragen sind von dem engen Aufgabengebiet der untersuchten Lernergebnisse, vom Vorzug für den Laborkontext anstatt naturalistischer Verhältnisse und von der kurzen Dauer der untersuchten Merkfähigkeitszeitspannen begrenzt worden. Diese Arbeit berichtet Einzelheiten über die Merkfähigkeit der Leistung bei Problemlösungen nach einem Monat, vier Monaten, und sechs Monaten, in drei verschiedenen Untersuchungen, die in Schulumgebungen ausgeführt wurden, die eine Anzahl von Lehrstufen und Fächern umschlossen.

Investigaciones hasta la fecha acerca de problemas de retención estaban limitadas por la gama restringida de los resultados de aprendizaje investigados, la preferencia por contextos de laboratorio más bien que ambientes más naturales, y la corta duración de los periodos de retención estudiados. Este articulo informa acerca de datos referentes a la retención del desempeño en la solución de problemas después de uno, cuatro y seis meses, en tres estudios separados llevados a cabo en ambientes escolares con varios estudiantes en diversas clases.

John A. Ross is an associate professor and Florence J. Maynes a research associate at the Trent Valley Centre of The Ontario Institute for Studies in Education, P.O. Box 719, 150 O'Carroll Avenue, Peterborough, Ontario K9J 7A1.