Field Test Evaluation of Educational Software: A Description of One Approach

Mariela Tovar
Nicholas Barker

Abstract: Educational evaluators in general have traditionally recognized the need to incorporate data from potential users in designing evaluation studies. In the field of courseware evaluation, however, there has been a great deal of emphasis placed on expert judgment as a source of data for evaluating computer-based educational materials. Although courseware reviews are extremely useful, they are not substitutes for field tests; each provides a different type of information that evaluators may use in order to determine the quality of an instructional product.

This paper reports on the evaluation of a courseware designed to assist the writing of the lower-case alphabet. The main objective of the article is to demonstrate an evaluation design which provided adequate answers to our evaluation questions, allowed us to perform multiple comparisons to support our conclusions, and was also practical enough to be used in a normal classroom situation without disturbing everyday activities. Three criteria for selecting a design are presented followed by a description of the courseware evaluation.

A common theme in the educational computing literature is the concern about the poor quality of courseware available. Consequently, many efforts have been directed toward providing developers and users of educational software with guidelines for evaluation. Many authors and organizations provide reviews and evaluation guidelines for assessing the merits of computer-based materials (e.g., Bitter & Camuse, 1984; Lathrop & Goodson, 1983; MicroSIFT, 1983; National Council of Teachers of Mathematics, 1981; Senn, 1983). Most of these guidelines place a great deal of emphasis on expert judgment and consideration of principles of instructional design, as sources of data for evaluating computer-based materials. This approach overlooks the fact that traditionally, educational evaluators have recognized the need to incorporate data from potential users in designing evaluations studies (Popham, 1974; Scriven, 1973; Stake, 1973).

As educators have become more critical, they have perceived the need for conducting courseware evaluation studies which incorporate student data under actual conditions of use (Bramble & Mason, 1985; King & Roblyer, 1984; Ragsdale, 1982; Steffin, 1983; Steinberg, 1983). Courseware reviews are extremely useful, but they are not substitutes for field tests; each provide a different type of information that evaluators may use in order to determine the quality of an instructional product. This point is illustrated by Weston (1986) in the context of formative evaluation methods. The author points out that "Experts should be called upon to judge those factors which fall within their area of expertise, such as accuracy, completeness, instructional quality..." (p. 8); however, Weston says, they "should not be expected to anticipate problems students may have with the materials..." (p. 8).

Designing and conducting field test evaluations, require some modifications to traditional research designs in order to account for the practical limitations and constraints imposed by most educational settings. A design is basically a plan which dictates how data will be gathered during an investigation. Campbell and Stanley (1963) provide an extensive discussion of experimental and quasi-experimental design and their potential threats to validity. True, experimental designs are often difficult to use in school settings since they require the presence of one or more control groups. Frequently, evaluators are faced with practical constraints which prevent them from implementing such designs (e.g., subjects of availability, difficulty in ramdomization, ethical considerations, disruptions of everyday scheduling, etc.).

We believe that in addition to consideration of threats to internal validity, evaluators selecting designs for field testing should consider designs that:

a) Answer their evaluation questions;
b) Collect the maximum amount of evidence on which to base conclusions of instructional effectiveness; and
c) Take into consideration the practical limitations of the setting in which the evaluation is conducted.

This article reports the evaluation of a courseware designed to assist the writing of the lower-case alphabet.[1] Although both formative and summative evaluation and procedures were carried out, the report will focus on the latter. The main objective of the article is to demonstrate an evaluation design which provided adequate answers to our evaluation question, allowed us to perform multiple comparisons to support out conclusions, and was also flexible and practical enough to be used in a normal classroom situation without disturbing everyday activities. The article will describe the evaluation method and will conclude with a discussion of the evaluation design in terms of the three criteria for selection specified above.

Description of the Courseware
The learning of handwriting skills requires constant repetition on the part of the student, as well as monitoring by the teacher in order to reinforce the correct letter formation. If this monitoring process could be achieved through a student-paced and self-

---

[1]The courseware was designed by the second author.

instructional format, the teacher would be able to use this time in more constructive ways with his or her students.

The main objective in the design of the courseware was to create a tool that could be used as an effective monitor for the learning and remediation of the lower-case alphabet. As such, a system was needed that could accept graphic information supplied by the student (a letter), compare the student input with a template stored in the computer memory, and display the results of this comparison to the student for critical analysis. The system would allow students to repeat their attempts until they could produce a letter that matched the template. It was very important that the system would be self-instructional and easy to operate.

The hardware used in designing and running the courseware was an Apple II computer and a graphics tablet peripheral called a PowerPad. This tablet does not require a special pen to communicate with the computer; any stylus-like implement can be used. The software that was created for this courseware has two parts. First, there is the "bank" of alphabet letters that are brought to the screen, segment by segment, on the prompting of the student. Second, there is the communication between the computer and the tablet.

The student/program interaction can be described in a number of steps:

a) The student secures a sheet of paper on the graphic tablet by two small clips. This ensures that the paper is positioned exactly so the user's attempt would superimpose the model to be displayed on the screen;

b) The student presses the key of the letter that is to be practiced;

c) A flashing dot on the screen, indicates where the segment starts forming, on a grid-like background of four horizontal lines;

d) The first segment is created dynamically on the screen. It is drawn slowly, as if drawn by an invisible hand;

e) The student copies this segment on the sheet (on which is drawn the same grid pattern as on the screen) resting on the tablet. The pressure of the user's pencil on the tablet transmits the attempt to the computer memory;

f) When the spacebar is pressed again, the flashing dot indicates where the next segment is going to start;

g) Prompted by the student, the next segment appears on the screen and is copied as before on the sheet;

h) When the letter is completed, the student attempt appears superimposed over the model on the screen and the two letters can be examined for discrepancies; and

i) By pressing the key of the letter, the process starts again.

The courseware was designed so as to incorporate the major elements identified by the literature as important for the successful learning of handwriting skills. These elements are: copying, critical analysis, immediate and accurate feedback, and presenting the student with dynamically-formed segments of the letter to copy (Charles, 1971; Gibson, 1972; Lally, 1982; Lally and McCleod, 1981; McCleod and Proctor, 1979; Smith and Murphy, 1982). These elements had not been combined previously into one instructional technique.

METHOD

Main Evaluation Questions
The evaluation of the courseware examined three questions:

Will the use of this courseware in the course of a normal learning routine significantly increase the number of letters written to criterion by learning disabled children?

Will any significant improvement be sustained over a post-intervention period of three weeks?

Is the courseware an appropriate teaching tool for the remediation of handwriting in a classroom environment? That is to say, is the courseware truly self-instructional?

Sample
A total of 16 children participated in this evaluation. They were chosen from grades one and two of the regular programme in an Elementary school in Montreal. All the children had been diagnosed as Moderately Learning Disabled (MLD) according to the criteria of the Department of Education of Quebec (1981). All subjects were unable to form ten or more letters of the alphabet to criterion. The criterion for each letter was established in cooperation with two teachers. In the case of borderline cases, a majority decision determined whether or not a student was included in the evaluation.

Eleven of the subjects were from grade two, and five were from grade one. They were randomly assigned to one of two groups of eight subjects each.

Evaluation Design
Some educational objectives, such as the learning of handwriting skills, are achieved through a gradual and progressive learning process. It is important, when intervening in such a process, to demonstrate that any significant increase in learning is due to the intervention and is not merely a product of simple continuation of the process. Ideally, a successful intervention should significantly accelerate the learning process that would occur without it.

The design chosen for this evaluation was an adaptation of the multiple time-series design as described by Campbell and Stanley (1963). A time series design is the most efficient way to establish the existence of a true effect by observing the stability of the base-line of the pre-intervention learning (the pre-tests) and the stability and strength of the post-intervention learning (the post-tests). The stability of any increase in learning is an important factor. If the test scores increase immediately after the intervention, yet the effect weakens over the complete post-testing period, it would indicate that the increase is due largely to the novelty of the situation (i.e., the presence of computer courseware). To indicate that true learning has taken place, the data would have to indicate long-term stability after the removal of the intervention.

The main evaluation design (Figure 1) was a staggered time-series format. Two groups of eight subjects each were pre-tested for three weeks, given the intervention for two weeks,

and then post-tested for three weeks. However, Group B was administered the first pretest as Group A were experiencing the third one. Thus the time-span for the complete evaluation was ten weeks. The advantage of this design was that the two groups could act as control for each other at two critical points, therefore keeping the time variable constant.

Figure 1.
Main Evaluation Design (Staggered Time-Series).

	Tests / Weeks									
	1	2	3	4	5	6	7	8	9	10
Group A	0	0	0	0$_X$	0$_X$	0	0	0		
Group B			0	0	0	0$_X$	0$_X$	0	0	0

Terrel and Lynyard (1982) used a similar design when evaluating the SpeakSpell spelling aid. They noted that a particular advantage of such a design is that it is simple and flexible enough to be adapted to the organizational patterns of a school with a minimum of disruption. A design of this kind can be imposed on a normal classroom routine without disturbing the working atmosphere significantly. Three designs were derived from the main evaluation design in order to make the following comparisons:

> *Comparison 1.* The results of Group A over the intervention period were compared to the results of Group B over the pretesting period (Figure 2). The design for this comparison was a 2 by 2 mixed factorial with treatment (intervention) as a between-group factor and testing as a within-group factor.

> *Comparison 2.* The results from the intervention period of Group B were compared to the results from Group A over the posttesting period (Figure 3). The design for this comparison was the same as the one used for comparison one. For comparisons 1 and 2, the test for week 3 was used as a covariate. This was used to account for individual differences in handwriting ability before the courseware was introduced.

> *Comparison 3.* The third comparison was between the combined posttests of both groups. The results of the two groups were collapsed over an eight-week period into a one group time-series design (Figure 4).

Figure 2.
Evaluation Design for Comparison 1.

	Tests / Weeks		
	3	4	5
Group A	O	O$_X$	O$_X$
Group B	O	O	O

Figure 3.
Evaluation Design for Comparison 2.

	Tests / Weeks		
	3	6	7
Group A	O	O	O
Group B	O	O$_X$	O$_X$

The independent variable in this evaluation was the introduction of the courseware for a ten-day period in the normal learning routine of the subjects. The dependent variable was the subjects' performance in the series of weekly tests. The subjects' score in each test indicated the number of errors in letter formation (i.e., letters not formed to criterion).

Procedure

One day before introducing the courseware, the subjects participated in a familiarization session with the computer and the graphics tablet. This was intended to minimize any confusion concerning the manipulation of the courseware during the intervention period.

A time for the weekly testing was arranged with the classroom teacher. All subjects were tested for their ability to write each letter to criterion. The subjects were tested as a group, following a standard procedure, and using a standard instrument, a lined sheet of

Figure 4.
Evaluation Design for Comparison 3 (Collapsed Data).

Tests / Weeks

1	2	3	4	5	6	7	8
0	0	0	0_X	0_X	0	0	0

paper. The letter "a" was written on the blackboard and the subjects' were asked to copy it onto their sheets. After a 15-second interval the procedure was repeated for the letter "b", and subsequently for the rest of the alphabet. Each letter of the alphabet was tested and the procedure was consistent throughout. The sheets were then collected and each letter was marked (1 or 0), according to the agreed upon criterion. The total test score for each subject, indicated the number of letters notreaching criterion (out of a total 26).

For the first two weeks Group A followed the testing procedure previously described. At the third test, Group B joined in for the first time. At the fourth, fifth, sixth, seventh and eighth tests both groups were tested together. For tests nine and ten, only Group B was tested

The intervention of the courseware consisted on ten daily sessions for each subject over a two-week period. Group A worked with the program during weeks four and five and Group B during weeks six and seven. A table was set up in the corner of the room to support the hardware and facing away from the class. A schedule was drawn up so that the subjects could spend 20 minutes a day working with the program without disturbing the rest of the class.

The evaluation was conducted in a regular grade two classroom comprising 25 students. The grade one subjects entered the room just for their turn at the computer. Initially, the appearance of the hardware caused some disturbance in the class. This was compounded by the news that only certain class members were to use it. As a result, arrangement were made so that the rest of the class would get some time at the computer on the completion of the evaluation.
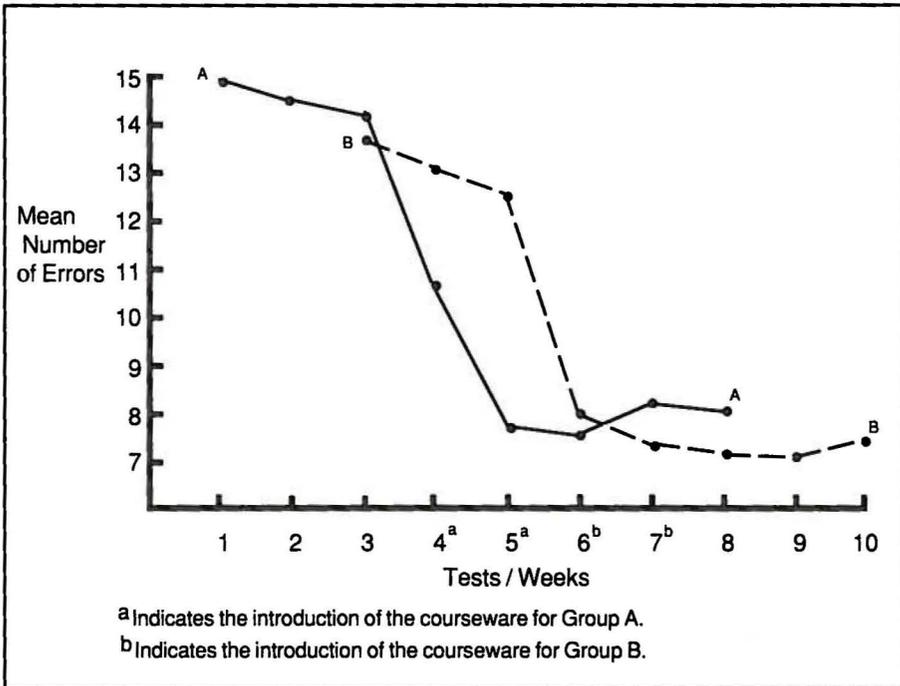
At the start of each session, the subject would ask the teacher which letters were to be practiced. A limit of five letters per session was set. At the end of each session the subjects showed the sheets to the teacher for review and comments. The subjects were then allowed to keep the used sheets.

For the first two days, the evaluator stayed with each subject at the computer, explaining the program and observing reactions. After these two sessions, the evaluator withdrew from the computer and observed the subjects from a distance, appearing to be busy with other work. The reactions of the classroom teacher and the general atmosphere in the classroom were also monitored. It was thought that this informal data could not only augment the results of the statistical analysis, but also provide information concerning the interaction between the courseware and elementary teachers and children.

RESULTS

The means and standard deviations for the number of errors recorded (letters not reaching criterion) are shown in Table 1 (See next page). The means are expressed graphically in Figure 5.

Figure 5.
Number of Errors in Letter Formation for Groups A and B.

a Indicates the introduction of the courseware for Group A.
b Indicates the introduction of the courseware for Group B.

The analysis of covariance for the first comparison (tests 4 and 5, see Figure 2) indicated a significant interaction between the treatment and testing variables, $F (1,12) = 7.17, p = .02$. This would indicate that the mean number of errors in letter formation for Group A decreased significantly with the introduction of the courseware compared to Group B results, where only a slow decrease of errors occurred during the same period. Post-hoc comparisons using the Tukey procedure confirmed the above; a significant difference was found between the means for Tests 4 and 5 for Group A ($p < .05$), but not for Group B.

The scores for Group A and Group B for the second comparison (Tests 6 and 7, see Figure 4) were also analyzed through an analysis of covariance. This allowed a comparison of Group A during the post-intervention period with Group B during the intervention. Results showed a non-significant interaction between the treatment and testing variables. Main effects were also non-significant. Introducing the courseware in Group B had the effect of eliminating the previous differences between the two groups. Also the scores of Group A, were maintained during the post-intervention.

The means and standard deviations for the number of errors recorded in the third comparison (See Figure 4 for the collapsed data) are shown in Table 2 (see next page). A

TABLE 1
Means and Standard Deviation for Number of Errors in Letter Formation:
Groups A and B

	Tests									
	1	2	3	4	5	6	7	8	9	10
GROUP A										
M	14.9	14.3	14.1	10.4	7.9	7.7	8.4	8.2		
SD	4.0	4.7	4.3	4.4	5.1	5.9	5.4	5.5		
GROUP B										
M			13.4	12.7	12.2	8.0	7.3	7.1	7.0	7.4
SD			1.8	2.1	2.0	1.9	2.6	2.6	2.7	2.6

TABLE 2
Cell Means and Standard Deviations for Number of Errors
in Letter Formation: Comparison 3

	Tests							
	1	2	3	4	5	6	7	8
M	14.1	13.5	13.1	9.2	7.6	7.4	7.7	7.2
SD	3.3	3.6	3.4	3.5	3.9	4.4	4.2	4.4

One-way repeated measures analysis of variance was performed. The analysis indicated a significant difference between the means of the weekly testing sessions, $F(7,91) = 43.80$, $p < .001$. Post-hoc comparisons (Tukey procedure) revealed a significant difference ($p < .05$) between the mean number of errors for weeks 3 and 4 (introduction of the courseware).

The Tukey procedure also indicated no significant differences between the scores of the three pretests (the baseline data) and no significant differences among the three posttests. This information is represented graphically in Figure 6.

Figure 6.
Number of Errors in Letter Formation: Data Collapsed Over Eight Weeks.
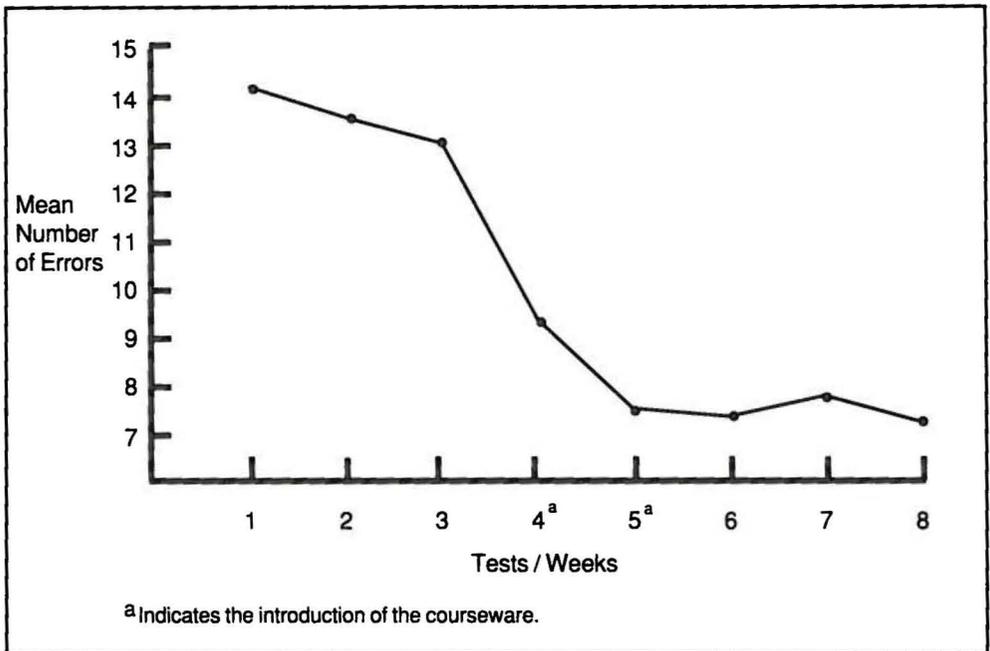
DISCUSSION

The statistical analysis of the data described in the previous section indicates that the intervention of this courseware did increase the number of letters learned to criterion by the students.

The baseline data for both Groups A and B indicate that the learning of this hand-writing skill without using the courseware was progressing at a gradual and statistically insignificant rate. The analyses indicate that a significant increase in learning took place when the courseware was introduced.

The Tukey procedure indicated that there was no significant decrease in learning over the post-test period. When the courseware was withdrawn and the normal learning routine re-established, the level of learning achieved by using the courseware was sustained. This demonstrated that true learning did take place and that the increase was not due solely to the introduction of a new and exciting medium of instruction. It could be argued that this level

of error-free learning could be eventually reached even if the courseware did not intervene, only that the process would take longer. A further long-term study using a traditional control group design would test this hypothesis. Regardless of the results of such a study, the early intervention of the courseware can be justified in two ways. First, this accelerated learning can only increase the motivation and self-confidence of students already struggling with learning disabilities. Secondly, this learning comes about in a self-instructional format, thus relieving the teacher of a time-consuming and repetitive task.

Although the main objective of this article was to demonstrate the evaluation design, we would like to discuss briefly the classroom observations[2]. The students used the courseware in a truly self-instructional manner only after they had grasped the relationship between what happened on the screen and what they had created on their sheet. Once this was established, critical analysis of what had occurred took place automatically. Any discrepancies in letter formation were pin-pointed and the desire to self-correct was expressed, and the process was repeated.

Two days after the introduction of the courseware, five children (all from grade one) were still demonstrating difficulty in using the self-instructional format. With these children, the relationship between the superimposed attempt on the screen and the letter formed on the sheet was stressed. After four days, only two children were still experiencing difficulty. These children tended to be passive in front of the computer and made the least improvement in the testing situations. The fact that almost all of the grade one students experienced some initial difficulty in manipulating the program and the grade two students did not, indicates perhaps that a certain level of maturity is required to comprehend this interactive process. As well, the problem could be attributed to a lack of confidence among children in the few months of their first full school year. The program itself could be improved to include more visual or musical cues to prompt the student when specific input is required.

The self-instructional aspect of the program appeared to aid the concentration level of the students during the self-correcting process. Generally, small-step corrections were made to a letter not formed to criterion, but because of the ease and speed of the feedback process, criterion could be quickly reached by many such small steps once the self-instructional aspect had been mastered. This constant repetition and correction not only allowed the student to discover the dynamic process of letter formation but also allowed for drill-and-practice of the skill which was self-directed and self-paced.

The courseware was designed to achieve positive results over a short period of time, as this can be a frustrating and exacting skill for these children to learn. The attitude towards discrepancies signaled that this success came early. Initially, discrepancies in letter formation as revealed by the superimposed feedback were viewed as errors. A common response to the feedback was "Oh, I got it wrong there." However, as the self-correctional process began to accelerate, these *errors* were seen more as stepping stones to reach the perfect letter, and the feedback then drew such responses as "It needs a bit more this way" and "It's got to be rounder here."

These manifestations of confidence and motivation can be attributed partly to the ability of the courseware to deliver non-judgmental feedback. The program channelled a

---

[2] A more extensive discussion of the observational data has been submitted for publication by the second author to the *Journal of Learning Disabilities*.

student's motivation towards a goal and then allowed him or her endless attempts to reach it, all the while demonstrating clearly where corrections should be made.

The main purpose of this article was to demonstrate the evaluation design. Consequently, it is worthwhile at this point to discuss how the selection of the design employed in this evaluation considered the three criteria specified at the beginning of the article. To recapitulate, we proposed that evaluators should consider designs that: Answer their evaluation question, collect the maximum amount of evidence on which to base conclusions of instructional effectiveness, and take into consideration the practical limitations of the setting in which the evaluation is conducted.

Given our evaluation questions, we needed a design that could be implemented in the course of a normal learning routine. This was required not only for practical purposes but also because the intended use of the courseware was not to replace the teacher, but as a tool that could be used to supplement the learning activities that normally occur in the classroom. We also required a design that would be sensitive to the normal learning process that occurs with the regular teaching activities. It was also considered important to select a design that controlled for the novelty of the presence of the computer in the classroom. By taking repeated measures before and after the treatment, we were able to observe the stability of the pre-intervention learning (the pretests), and the stability and strength of the post-intervention learning (the posttests). This design was more sensitive to our questions than a pretest-posttest control group design, or a posttest-only control group design would have been.

Probably, the strongest aspect of the design relates to the second criteria; the possibility for multiple comparisons on which to base conclusions of instructional effectiveness. A non-instruction control group was not necessary since groups act as control for each other at different points in time. Group B served as a control for Group A in the first comparison. Group A served as a control for Group B in the second comparison. The results showed not only that introducing the courseware in Group A produced significant learning differences, but also that introducing the courseware in B had the effect of eliminating the previous differences between the two groups. This added further support to our conclusions about the instructional effectiveness of the courseware. Furthermore, the design allowed us to collapse the tests from the two groups in order to observe the learning trends during the pre-intervention, the treatment, and the post-intervention period (comparison 3). The previous comparisons, as well as the observational data provided us with a great deal of evidence to support our conclusions.

In terms of practical constraints, the design allowed us to work with a minimum of equipment (only one computer was necessary in order to implement the evaluation). More importantly, it could be implemented with minimum disruption of the classroom. One of the objectives of this evaluation was to assess the impact of the courseware in conditions as close as possible to the normal classroom routine. The evaluation design dictated who used the courseware and for how long, but the teacher had to decide which letters were to be practiced and to continue with the normal routine of the class while each subject took their turn at the computer. This teacher had never experienced a computer in the classroom before. After some initial confusion, the teacher and the evaluation subjects were soon into a smooth running of the schedule. Over the four weeks that the courseware was in operation, the initial skepticism of the teacher was replaced by enthusiasm.

Obviously, the design selected for this evaluation is not an "all purpose design." Each

evaluation considers unique questions, and is conducted in different settings. The design selected should provide the most credible information in the situation at hand. When the ideal choice is not possible, the next best option can be tried. Discussions of alternative designs for evaluation are available in the literature (e.g., King & Roblyer, 1984; Terrel & Lynyard, 1982; Wagner, 1984). The important thing is not to shy away from conducting field tests because the *perfect design* is not feasible. As King and Roblyer point out:

> If we are to take advantage of the valuable information available from classroom computer projects, alternative designs to study and evaluate the effectiveness of computer-based activities must be employed (p. 23).

In this article, we have tried to illustrate how we solved a particular evaluation problem. We have proposed three general criteria that guided the selection of our evaluation design. We believe that those criteria will be useful to evaluators planning field tests.

REFERENCES

Bitter, G., & Camuse, R. (1984). *Using a microcomputer in the classroom.* Reston, VA: Reston Publishers Co, Prentice Hall.

Bramble, W. J., & Mason, E. J., (1985). *Computers in schools.* New York: McGraw-Hill.

Campbell, D. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton & Miffin.

Charles, J. (1971). The effect of training, self-recording and group contingencies on immediate grade pupils. *Dissertation Abstracts, 32,* 5504.

Gibson, E. J. (1972). Untutored acquisition of writing skills. *Dissertation Abstracts, 32,* 5504.

King, F. J., & Roblyer, M. D. (1984). Alternative designs for evaluating computer-based instruction. *Journal of Instructional Development, 7* (3), 23-29.

Lally, M. (1982). Computer-assisted handwriting instruction and visual/kinaesthetic feedback processes. *Applied Research into Mental Retardation, 27,* 77-82.

Lally, M., & McCleod, I. The effectiveness of computer controlled feedback in handwriting instruction. In R. Lewis & E. D. Tagg (Eds.), *Computers in education.* Amsterdam: North Holland, 1981.

Lathrop, A., & Goodson, B. (1983). *Courseware in the classroom: Selecting, organizing and using educational software.* Menlo Park, CA: Addison-Wesley.

Lewis, R. (1981). Education, computers and microelectronics. *Technological Horizons in Education, 8,* 47-53.

McCleod, I., & Proctor, P. (1979). A dynamic approach to teaching handwriting skills. *Visible Language, 13,* 29-42.

MicroSIFT, Computer Technology Program, NorthWest Regional Educational Laboratory (NWREL) (1983). *Evaluator's guide for microcomputer-based instructional packages* (rev. ed.). Eugene, OR: International Council for Computers in Education (ICCE).

Ministere de l'education du Quebec (1981). *Instructions regarding the organisation of educational services for children with difficulties in adaptation and learning.* (Document No. AE-15-81-01).

National Council of Teachers of Mathematics (1981). *Guidelines for Evaluating Computerized Instructional Materials.* Prepared by the Instructional Affairs Committee of NCTM.

Popham, J. W. (1975). *Educational evaluation.* Englewood Cliffs, NJ: Prentice-Hall.

Ragsdale, R. G. (1982). *Evaluation of microcomputer courseware.* Toronto Ont.: OISE.

Scriven, M. (1973). The methodology of evaluation. In B. Worthen & J. Sanders (Eds.), *Educational evaluation: Theory & practice* (pp. 60-104). Belmont, CA: Charles Jones Publications.

Smith, S., & Murphy, L. (1982). Psycho-motor principles at work in basic skills. *Impact of Science on Society, 32,* 104-109.

Senn, P. R. (1983, May). Six checklists to prepare your classroom for technology. *Social education.*

Stake, R. E. (1973). The countenance of educational evaluation. In B. Worthen & J. Sanders (Eds.), *Educational evaluation: Theory and practice* (pp. 106-125). Belmont, CA: Charles Jones Publications.

Steffin, S. A. (1983). A suggested model for establishing the validity of computer-assisted instructional materials. *Educational Technology, 23* (1), 20-22.

Steinberg, E. R. (1983). Reviewing the instructional effectiveness of computer courseware. *Educational Technology, 23* (1), 17-19.

Terrel, C., & Lynyard, D. (1982). Evaluation of electronic learning aids. *International Journal of Man-Machine Studies, 17,* 59-67.

Wagner, N. (1984). Instructional product evaluation using the staged innovation design. *Journal of Instructional Development, 7* (2), 25-27.

Weston, C. B. (1986). Formative evaluation of instructional materials: An overview of approaches. *Canadian Journal of Educational Communication, 15* (1), 5-17.