PER LINGUAM VOL. 6 NO. 1 1990

Evaluating educational courseware: Principles and guidelines

Jannie J. Botha

For quite some time there has been serious concern about the quality of educational courseware. Not only has the educational market been flooded with inferior courseware but good, intelligent material has not been forthcoming. The aim of this paper is to give an overview of the principles and guidelines for evaluating educational courseware so that similar evaluation procedures can be used by teachers. If a sound evaluation procedure is not followed the chances are that programs will be bought solely on the strength of well written advertisements and not after sound evaluation. The result of this approach will be that worthwhile programs will gather dust on shop shelves while inferior programs will be purchased. As soon as the novel effect of computers wears off, instructors will turn their attention to suitable courseware. If this is not possible, they may decide that computer-assisted instruction is a waste of time. In this paper the focus is on different types of evaluation as well as effective evaluation procedures. Examples of programs in the field of second and foreign language teaching are included.

Daar heers al lank ernstige kommer oor die kwaliteit van opvoedkundige programmatuur. Nie alleen is die opvoedkundige mark deur minderwaardige programmatuur oorstroom nie, maar uitstekende, intelligente materiaal het ook nie die lig gesien nie. Die doel van hierdie artikel is om'n oorsig te gee oor die beginsels en riglyne om opvoedkundige programmatuur te evalueer sodat soortgelyke prosedures deur onderwysers gebruik kan word. Indien 'n sterk evalueringsprosedure nie gevolg word nie, is die kanse goed dat programme uitsluitlik as gevolg van goedgeskrewe advertensies gekoop sal word en nie na deeglike evaluering nie. Die gevolg van hierdie benadering sal wees dat verdienstelike programme stof sal vergader op winkelrakke terwyl minderwaardige programme gekoop word. Sodra die nuwigheid van rekenaars verflou, sal instrukteurs na geskikte programmatuur begin kyk. Indien dit nie beskikbaar is nie, mag hulle besluit dat rekenaargesteunde onderrig 'n vermorsing van tyd is. In hierdie artikel word gekyk na verskeie tipes evaluering, asook effektiewe evalueringsprosedures. Voorbeelde van programme oor tweede en vreemdetaalonderrig word ingesluit.

1 Introduction

Part of my research project conducted during 1985 in the U.S.A. included the evaluation of existing computer-based language programs (Botha 1986). It was found that there had been serious concerns about the quality and applicability of existing CAI (in particular CALI) materials.

The main aim of this article is to give an overview of the procedures for reviewing programs so that schools and universities could set up similar evaluation procedures. If a valid evaluation procedure is not followed, the chances are that programs will be bought exclusively because of well-written advertisements and not after thorough evaluations. The end result could be that worthwhile programs gather dust on shelves, while less worthwhile programs are bought and used.

In 1982 the Office of Technology Assessment (OTA) of the U.S.A. Congressional Board established four reasons for the lack of quality courseware in computer-based education (as referred to by Truett 1984 — original source not available):

- 1 Many of the technologies employed in education are still new.
- The production of high quality courseware is expensive.
- 3 Programmers and design experts are in short supply. (This is even more so in the Republic of South Africa.)
- There is sometimes a lack of good copy protection with the result that good programs are pirated. This could have serious consequences for the courseware markets in general.

Referring to 1 above, one should point out that microcomputers have been around since the end of the 1970s. The novel effect could have been a factor in the earlier years of CBE but should become less so towards the end of the present decade.

It is a pity that microcomputers have had to overcome so many hurdles towards the establishment of their position in education in general. Educational innovation has been around since 1950 — the so-called **programmed instruction** at that stage. Truett's viewpoint is that educators are now putting the ideas of the 1960s into practice.

With reference to 2 above, it should be granted that time is money. It was calculated in the U.S.A. that it could cost up to \$12 000 to produce one good quality CAI program (Truett 1984). Brain time is the most important ingredient in the courseware development. It has been found, however, that small developers with much lower overheads, can also produce quality courseware. Larger companies have the advantage that they can buy the services of a large number of good programming personnel. They therefore produce programs at a much faster rate; though, as said before, this still does not always guarantee quality. Despite the high development costs, software companies are still booming in the U.S.A.

There may be a short supply of good programmers and design experts (cf. 3 above) but unqualified individuals seldom produce quality programs. Often programs are produced by programmers who are not educators, or vice versa. Most developers agree now that it takes a team effort to produce quality programs. Such a team should consist of a subject expert, instructional designer, and an experienced programmer.

With regard to the piracy of programs (4 above), it must be pointed out that it could jeopardize the courseware market. This becomes a problem, for example, when programs are exported

to foreign countries where different copyright laws apply and where control over piracy is almost impossible.

2 Types of evaluations

The evaluation of programs entails much more than this term suggests. Evaluations could be done at different stages during the development and marketing process, and also by different interested parties.

2.1 Formative evaluation:

This evaluation takes place while a program is still under development. Recommendations made at this stage can still be implemented in subsequent revisions. Characteristic of this type of evaluation is that it makes provision for repetitive evaluation. Furthermore, the information gained could be used for immediate revision long before the program is marketed.

2.2 Summative evaluation:

The goal of this evaluation is to assess the value of a completed product. This evaluation is conducted before a product is marketed. It has the most value when it is undertaken by the consumer.

2.3 Publisher evaluation:

It was found by Truett, during a survey in 1983, (Truett 1984) that out of 56 publishers nine did not do any field evaluation of their courseware. Although the majority did evaluate the courseware their evaluation methods did not make provision for comparisons on common grounds. Because a program is field tested during development or shortly afterwards, it does not necessarily mean that it is a useful program for a classroom situation. On the other hand, field testing is not a waste of time. Very few publishers doubt the value of direct student/teacher interaction with courseware.

2.4 Published reviews:

It may have become clear from the previous paragraph that the publisher evaluations may not be a guarantee that courseware is of high quality. Reviews in periodicals could provide valuable information in this regard. But such reviews should not be consulted in isolation. Not all reviewers can be considered experts in their respective fields. Published reviews should therefore be seen as an additional source, never as a primary source.

2.5 School evaluation:

The courseware should meet and achieve the objectives laid down by the school authorities (regional, district, or individual schools). In the final instance, only the teacher could determine exactly what these objectives are, and whether the objectives are met. The teacher, from his/her classroom experience, should determine whether the content as well as the instructional techniques are suitable.

Later in this paper a more complete exposition will be given of an evaluation procedure. At this stage it is necessary to point out that a team effort is very important in the evaluation of courseware in schools. Such a team needs not be composed of teachers from within one school. Teachers from the same school district or region could be members of an evaluation team. The main task of an evaluation team should be to identify potential problems.

3 Effective evaluation procedures

There can be no doubt that effective evaluation procedures should form part of any CBE (and CALI) environment. Only in this way can one ensure that high quality programs are bought or developed and that such programs are used.

3.1 Preparations for evaluation of courseware:

One of the first aspects to be considered deals with the composition of a review team. It is self evident that members of a review team should have a commitment, and a unique outlook on CALI, and CBE in general. According to Truett (1984) a review team should include the following members:

- (a) teachers from the specific subject area and/or grade level;
- (b) students at an applicable grade level;
- (c) a media specialist/librarian;
- an expert in microcomputers and their applications to (language) teaching, as well as their capabilities;
- (e) administrators, e.g. a principal of a school, or head of a department, and
- (f) others, e.g. parents or representatives from the PTA, and a school committee.

It must be emphasized that a review team should never exclude a staff member who will be actively involved in implementing a program.

The final composition of a review team will be determined by local circumstances. A basic principle should be that the final team should be as close as possible to the one outlined above. In the metropolitan areas in Southern Africa, schools should make use of staff at institutions

for tertiary education; for example, universities and technicons. There seems to be no reason why well-composed evaluation teams should not be used in the South African situation.

The advantages of a team approach to evaluate are that time is saved, a range of experience is available, some attention is given to previewing requests, and written requests to publishers should get faster responses.

There are also **disadvantages** in the team approach: a committee may spend more time on evaluation than an individual would; the arrangements of meetings and locations may be difficult; accessibility of computers and programs must be assured well in advance; there may be unresolved differences over the final decision to purchase a program.

Another aspect, which forms part of the preparation for courseware evaluation, is the acquisition of the programs from publishers. In the U.S.A. this may not be a problem under normal circumstances. In South Africa acquiring programs for evaluation purposes is a real problem, apart from being far from the publishers. One gets the impression that, apart from being far from the publishers, suppliers of hardware are not really interested in stocking courseware, or helping potential buyers to acquire courseware from the U.S.A. and the United Kingdom. There are several ways to alleviate this problem. Many U.S.A. courseware publishers sell demonstration discs at a very reasonable price, or even make demonstration copies available free of charge. As will be pointed out later in this paper, it is rather risky to evaluate only demonstration copies of a program.

A courseware clearinghouse should be established so that interested parties could obtain information about programs from this source. Such a clearinghouse should be able to provide useful information about available programs, published reviews, and addresses of vendors where programs could be obtained. Further, staff in a clearinghouse should also be able to give guidance with regard to the evaluation of programs.

Another way to solve the above-mentioned problem, and not necessarily in contrast to the clearinghouse concept, is for education authorities to negotiate with a local computer dealer, but preferably a publishing house, to acquire courseware for evaluation purposes. An agreement should be reached about the subsequent orders for programs. It is advisable in this case to make arrangements with more than one local supplier so that the prices of programs can be compared.

The latter arrangements will only be possible if well established review teams (with the necessary backing) have been established.

3.2 Guidelines for evaluation of courseware:

Before specific guidelines are discussed, it will be helpful to look at some general principles for evaluation. In general courseware should be judged on three important characteristics:

- 3.2.1 Content: The content should have educational value, and it must be consistent with the subject area. It should also meet the goals of the curriculum. It should also be free from errors, as well as racial, sexual or ethnic bias. Racial and ethnic bias are completely unacceptable in second language learning.
- 3.2.2 **Instructional techniques:** The purpose of a program should be well defined, and it must be capable of achieving this purpose. The presentation must be clear and logical. Further, it must adhere to the valid principles of instructional design, must be at the appropriate level of difficulty, must be motivational, stimulating and creative.
- 3.2.3 **Technology:** Are support materials necessary and are they available? Is the information on the screen and in the printed manuals clear? Will students be able to use courseware independently? The design should also employ the capabilities of the computer(s) for which the lessons were designed.

Any evaluation or review team should be able to identify inferior courseware without losing too much time. Bramble (1985:173) listed aspects which could be regarded as a checklist for eliminating poor courseware.

1 Audible response to student errors: When a student makes an error the computer should not announce it to all the other students in the room by emitting a sound. Also, when a student makes as error, instant feedback is very impotant. It is not acceptable to present the next question without giving feedback to a student. Responses in the target language are preferable.

If a student answers incorrectly, he will probably want another chance. If a second attempt is also wrong, then feedback should be given to lead him to discover the right answer.

- 2 Rewarding of failure: A student should not find it more fun to make a mistake than to provide the correct answer. (Hangman has been described as a popular and enduring game but it also shows a flaw in instructional design. Incorrect answers in Hangman are rewarded and therefore reinforced. The game also provides neutral feedback to correct answers a character is inserted at the right place, and the player is not "hanged".) (Cf. Botha 1986:31.)
- 3 Any sound that cannot be controlled: Sound can be effective for motivation, or reward. Constant repetition of the same sounds can become boring. In a program such as **Grammar Examiner** it is possible to turn off the sound.
- 4 Technical problems: A program should run smoothly and without errors.
- 5 Uncontrolled advancing of screen display: If a student cannot control the rate of the screen

display in accordance with his needs, then this seriously defeats individualization.

6 Inadequate technical instructions: Instructions should be clear and appropriate for a student's reading level. A student should not have to guess or determine by trial and error how to operate a program. An area where programs are particularly weak is in the loading instructions. It was found during evaluations that programs had been adapted for IBM PC's from Apple microcomputers, for example, but that the loading instructions had not been adapted accordingly or adequately. This is a source of frustration for students and teachers alike.

A student would like to ensure that he is in the right program after it has started. Therefore the program title should verify this. A student will also feel insecure if he does not know how the program works (Tuttle 1983:35-39). For example, if a student sees a question, how much of the answer must be entered before it will be accepted by the program? This and other information should be stated clearly in the documentation or in the program itself.

7 Errors of any kind: Mistakes in factual content, grammar, spelling, etc. are not acceptable at any stage.

8 *Insults, sarcasm, and derogatory remarks:* It is unforgivable if students are attacked or belittled by remarks made in the program. For example, a program should not call a student names every time an error is made.

9 Poor documentation: It was found that many evaluated programs did not specify any instructional objectives, and these had to be inferred from the documentation. Suggestions for adapting the program for other activities, or grade levels, should also be clear. It should not be necessary for the teacher to read through the entire documentation (sometimes not more than one page) to find information about additional activities, for example.

10 Refusal to supply a backup copy: If a publisher placed special codes in the program to prevent it from being copied, then he should supply a backup copy free of charge for emergency use. There should also be a replacement policy for damaged or worn discs.

If the above-mentioned specifications are not met, then a review team may find that time was wasted by considering a program.

3.3 Process and criteria of evaluation

It should be clear by now that courseware evaluation is not a haphazard process performed by individuals in their own offices.

The EVALUATION PROCESS should be systematic for this could lead to a wiser expenditure of already dwindling educational budgets (Truett 1984).

Programs could be used in three ways in the classroom: as adjunct materials, as a complete course in the curriculum, or as a management system. The last mentioned is not purely for

teaching but could play a very important role in determining student progress through the program. A management system therefore is not used directly for instructional purposes. Teachers should determine ahead of time how a program will be used.

The process of evaluation should begin with determing the instructional objectives. This approach provides an opportunity to focus on the entire activity. It is recommended that a review team should be interested in how a program could be incorporated into the curriculum.

The next step in the evaluation process is to *identify possible programs*. This information could be obtained from software directories, company catalogues, consortium newsletters (cf. CALICO), and journals. A review team should assemble a list of titles for dissemination. In this respect a clearinghouse could play a valuable role.

After a list of programs has been compiled, it is necessary to *examine* some of the *programs* more *closely*. This examination may form the basis for ordering preview copies. Any description in a catalogue or other source is virtually worthless if it does not contain enough information to answer the following six basic questions:

- 3.3.1 Will the program suit the objectives?
- 3.3.2 Is it designed for the level of the target audience?
- 3.3.3 Will it work on the equipment available in the institution?
- 3.3.4 How much does it cost, and what is included for the price?
- 3.3.5 By whom is it published or distributed?
- 3.3.6 How can it be ordered?

It was mentioned earlier that many programs did not include explicit *instructional objectives* (sometimes for obvious reasons!). The objectives of a program could be regarded as valuable guide posts. Truett states that there is no more important information in the program description than the objectives.

Information about the *student level* is also very important. Often such a wide range is specified that teachers should rightly wonder whether such a specification serves any purpose, e.g. a typical American specification (K 12) would translate into South African terms as Sub A/Grade 1 - Std. 10. Any teacher will confirm that such a range level could hardly be covered by only one program, or even a series of programs.

If a clear specification of the hardware requirements is not given, the review team runs a risk that the program cannot be used on a school's equipment. Think in this regard of the compatibility of different graphics cards. For example, a true Hercules graphics card is required by Missing Links. Special attention should be paid to the DOS (Disc Operating

System) required. As is the case with many other peripheral devices, there is very little standardization among the different versions of DOS. With more and more compatibles or hybrid systems coming on the market — usually at ridiculously low prices — more non-standard versions of DOS will also enter the market. Furthermore, the stated memory requirements should be checked against the maximum memory of the computers in an institution.

Usually the *name* and the *address* of a company should give an indication of the quality to be expected from a program. A good program will seldom emerge from a company which has a name for inferior programs.

A final point in the process of evaluation is to request a preview copy — usually with the option of returning it within 30 days. If a company offers a demonstration copy instead, a team may be running a risk. It is not advisable to buy a program after only a part (usually the best part) has been evaluated. A demonstration copy will hardly ever reveal errors in a program.

Once the evaluation has been fully comprehended, a review team should take note of the criteria to be applied in the evaluation procedure.

The CRITERIA FOR EVALUATION is almost the most important part of the entire evaluation procedure.

The criteria discussed below should give evaluators a framework for judging courseware. At the end of the discussion an evaluation form is included in which all the discussed criteria have been taken into account.

The documentation which accompanies a program should be evaluated. The minimum information should include directions for the user, a validation statement of the program's effectiveness, a teacher's manual containing information about vocabulary, preparation and follow-up activities, and a description of the assessment instrument. The documentation which accompanies many programs can be improved. As mentioned earlier in this paper, in the case of adapted programs, documentation for a new operating environment must be adequate.

The review team should also decide whether the documentation is complete. The following question with regard to the documentation should also be asked: Are the objectives clearly stated? If they have to be inferred, or are implied, serious questions should be asked about the quality of the program. Furthermore: are any student materials provided? Are teacher's materials present? Is the technical documentation adequate, complete? Is there a flowchart showing the general logic of the program? If a flowchart is not present, the general logic should be very clear from the rest of the documentation. Is the documentation free of errors? Is there enough information for a novice to start the program? Is there a glossary in which error messages from within the program are defined? It can be very frustrating to encounter an error message without being able to find out what it means. Is the courseware in a well designed container which will withstand rough student handling?

A good investigation should be made of the **educational value** and the **instructional objectives**. It should be clear to the team that the program appears to achieve the stated educational objectives. There should also be an indication that the program has positive affective outcomes and does not put the student off. The learning which takes place in the program should be generalizable to real-life situations. This is very important from a communicative language teaching viewpoint. It should also be determined whether the graphics, colour, and sound used in the program are really necessary, i.e. from an educational point of view. It should not be too difficult to determine the mode of instruction (e.g. drill-and-practice, simulation, etc.) for this should already be clear from the documentation. The team should determine in this regard whether the mode of instruction is appropriate for the material presented and the level of presentation.

A more difficult question deals with the instructional potential of the microcomputer, and whether this potential has been used to its fullest extent. A relevant question in this regard concerns the use of the memory: if a program was written for microcomputers with only 64k, while many microcomputers today already employ a minimum of 640k, then a program could run slower due to frequent disc accesses. This in turn could be a source of frustration among students.

It almost goes without saying that the content of a program should receive due consideration. Special reference should be made to the accuracy and appropriateness of the content. A variety of questions should be asked about the content: Does the content appeal to the target audience? This could be determined after a trial run with the actual audience. Is the content factual and accurate, or has it become dated? Is it free of racial, ethnic, and sexual stereotypes? Is it free of grammatical and spelling errors? It goes without saying that this is extremely important in language programs. Is the content well-organized, in a logical sequence, from concrete to abstract, from simple to complex? It is important to determine whether a teacher can modify the content. This could be a great help because often teachers are bound to reject programs which do not fit exactly into a curriculum.

Another criterion in evaluation concerns user-interaction. It is very important to involve students in a program. If this does not happen, the computer is used as a very expensive pageturning device. Interaction also concerns the level of difficulty allowed by the program. Students will not be interested in using a program if the level of difficulty is too low or too high. A further concern should be the feedback students receive during the program. Apart from the feedback in the target language, the feedback should also indicate whether a student may try more than once. Students might also like to know whether they may review material before taking a test for example. It is further important to know if a program makes use of different types of questions, e.g. open-ended, multiple choice, sentence type answers, and even short paragraphs.

If a program can be used with more than one student at a time, it fosters cooperation, and it could be considered for small group work. Group work is considered as an important ingredient of communicative language teaching.

A very important consideration is how much typing is expected of a student. This question should be considered in the light of a computer literacy project in progress at a school. If such a project is present, the amount of typing required will not be a determining factor. If no such project is being offered or will not be part of the school activities in the near future, then typing requirements should be judged from a different perspective.

The review team should also pay close attention to the technical consideration. These considerations relate to the programming techniques used, and the ease of use of the program. If a program has been well constructed, it will be easy to start. The program should be user-proof. If error messages are displayed (and this is preferable) then their meaning should either be clear from the message itself, or additional information should be given in the documentation. It should also be clear from the program itself whether the sound can be turned off.

A general impression could be obtained by looking at the screen. Is it uncluttered? Are the graphics understandable? Will lack of colour (e.g. monochrome microcomputer) render the program ineffective?

From the discussion of the above-mentioned criteria it should be clear that the program evaluation, even by a good review team, is a very complex activity. It is therefore advisable that a review or evaluation form is designed so that some standardization could be brought into the evaluation process. The evaluation form included here was designed for the program evaluation which formed part of this research project. It was found that a completed form of this type contained enough information to facilitate sound, balanced judgements about programs.

4 Conclusion

It should be clear that educational courseware should be thoroughly evaluated. Developers should take note that good courseware requires forethought. Educators, including language teachers, should be reminded that the microcomputer is a powerful educational tool. It is up to teachers in different subject areas to ensure that low quality courseware is not unleashed on the market (Burns 1981:15-16. 49). Burns maintains that courseware could be a real Pandora's chip!

MICROCOMPUTER COURSEWARE EVALUATION FORM 1

1 GENERAL INFORMATION Evaluator: Date: Title of program: Title of package: Publisher/Vendor: Author(s): Target audience: Copyright date: Price: Specific topic: 2 TECHNICAL INFORMATION Computer: Hardware requirements: Data storage: Disc? Memory required? For which other computers? Start with menu? Can program be listed? Can program be modified? Copy protected? Teacher's guide provided? Backup copy provided? **3 EDUCATIONAL INFORMATION** Overview of content:

Objectives:

¹ Adapted from Truett, C. et al. *Choosing Educational Software*. A Buyer's Guide. Libraries Unlimited, Inc. Littleton, Colorado 1984, pp. 70-73.

Prerequisites:		
Worksheets required?	Worksheets supplied?	
Instructional purpose (e.g. remediation):		
Type of program (single or series)?		
Instructional technique (check as applicable):		
Drill and practice	Problem solving Tutorial	
Exploration Game	Simulation Informational	
Management/Administration	Test	
Suitable for: Individual instruction	Small groups	
	Larger groups	
Provision for different levels?	How many?	
Other comments:		
4 RATINGS		
State the rating that reflects evaluation of program. Type/write YES, NO or N/A. Comment if NO was chosen.		
1 Is the content accurate?		
Comment:		
2 Does the content have educational value?	?	
Comment:		
3 Is the content complete?		
Comment:		
4 Is the content free of bias and stereotype	?	
Comment:		

5	Is the content properly sequenced?
	Comment:
6	Is the feedback appropriate?
	Comment:
7	Is the level of difficulty appropriate for the target audience?
	Comment:
8	Does the program accomplish stated or implied objectives?
	Comment:
9	Does the program accept a wide range of possible responses?
	Comment:
10	Does the teacher's guide give relevant information?
	Comment:
1	1 Is the program easy to load into the computer?
	Comment:
12	2 Is the text free of punctuation and spelling errors?
	Comment:
13	3 Are the text and directions easy to read?
	Comment:
14	4 Is the user given adequate time to read and respond?
	Comment:
1:	5 Do colour/graphics/sound enhance the program?
	Comment:

16 Can the user control the rate of presentation?		
Comment:		
17 Can the user control the sequence of presentation?		
Comment:		
18 Does the program branch to correct the level of difficulty contingent upon the user's responses?		
Comment:		
19 Are the supplementary worksheets adequate?		
Comment:		
20 Is a program packed in secure container?		
Comment:		
21 Are the user's correct and incorrect responses recorded?		
Comment:		
22 Can the teacher retrieve the user's final score?		
Comment:		
23 Is a pretest and post-test included?		
Comment:		
24 Is a classroom management system included?		
Comment:		
5 SUMMARY		
Major characteristics (strengths) of the program:		
Major weaknesses of the program:		
Any other comments:		

Bibliography

- BOTHA, J.J. 1986. Using microcomputers in second and foreign language instruction. Stellenbosch.
- BURNS, H. 1981. Pandora's chip: concerns about quality CAI. PIPELINE 6 (2), 15 ff.
- EDUCATIONAL SOFTWARE SELECTOR, The. 1986-87. New York: EPIE and Teachers College Press.
- TOPPER, L. 1972. A checklist for instructional software design. *Educational Technology*. (May), 22-24.
- TRUETT, C. et al. 1984. Choosing educational software. A buyer's guide. Littleton: Libraries Unlimited.
- TUTTLE, H.G. 1983. Programming/evaluating second language CAI. Foreign Language Annals 15(3), 35-39.