

Meskill, C. and Swan, K. (1993) Assessing CALL Software: A Computerized Database Approach. in CALICO Transactions. Duke University. 104-106.

Assessing CALL Software: A Computerized Database Approach

Carla Meskill and Karen Swan, University at Albany, State University of New York

A recent MacWorld special issue decries inequities of computer usage for language minority students in U.S. public schools. Key to this perceived failure is lack of teacher training in pedagogically effective utilization (MacWorld, September 1992). The schools, like universities, colleges and adult learning programs, however, continue to invest in hardware and Computer Assisted Language Learning (CALL) software. Development and distribution of CALL software is consequently on the rise. The number of packages available to educators has doubled in the past two years alone. However, while software products currently abound, there persists a dearth of empirical data to support the effectiveness of CALL as a tool for teaching and learning foreign and second languages. Moreover, the bulk of studies undertaken to date report mixed results regarding pedagogical effectiveness. A common conclusion is that effectiveness ultimately depends on carefully calculated and implemented teacher strategies; strategies that are in sync with the needs and ethos of a given classroom (Robinson et al, 1985; Johnson, 1991; Perez, 1984). Where does this leave the language instructor who needs to assess the potential match between software products and a specific teaching and learning situation?

Language professionals are frequently called upon not only to judge the potential viability of CALL for their institutions, but to also ask themselves what CALL has to offer them in terms of their own classrooms and teaching. Relevant information has historically come from two sources. The individual can turn to CALL software reviews in professional publications and to colleagues for anecdotal recommendations. Accessing and sorting through these sources is by no means a straightforward task. First, access to relevant publications and the time to read and catalog reviews can be problematic. Second, perspectives from which this kind of information is presented, both in published and anecdotal forms, shape critiques and recommendations. As such, they clearly affect a particular slant on software products, a slant that the teacher in need of information germane to her own needs and experiences and the needs of her students and institution may not share nor find sufficiently helpful. Another form of guidance in this assessment process, software evaluation review sheets, also saw a period of popularity in the early 1980's (Strei, 1983). Such review guides were an attempt to prompt the novice CALL user to look at a standard list of surface features. Typical software review sheets continue to be limited to formats such as checklists concerned with general categories of software characteristics.

The language teacher's task, though clearly aided by these forms of guidance, might be facilitated via a more directed method of inquiry that offers multiple perspectives on software viability.

This presentation outlines procedures employed in a CALL software assessment project that undertook to establish critical review criteria and methods of accessing applications evaluations once they were completed. Software packages, designed for both native speakers and non-native speakers of English, were analyzed for their potential effectiveness along a range of dimensions established within the framework of communicative language teaching. These reviews were entered into a computer database containing a number of attribute fields that graphically represent the degree to which each software package satisfies the CALL design criteria. In database form, attributes of the applications reviewed can be manipulated for comparison, and generalized conclusions regarding the match of design and perceived effectiveness readily reached.

The Project

Fifteen teachers of English as a second language participated as software reviewers in piloting the database project. Although at the time none of the teacher/reviewers was a CALL user, each had read extensively the available literature concerning actual and potential uses of CALL and had discussed at length design effectiveness and ineffectiveness as related to theory and methods of language learning. The fact that reviewers were not experienced CALL users may be perceived on the one hand as limiting the breadth and credibility of their findings. On the other hand, however, the fact that these reviewers were uninitiates in terms of actual practice appeared to have enhanced their capacity for 1) generating review criteria as true desiderata; that is, their perceptions of the role of the computer in language teaching and learning were untainted by prior experiences both positive and negative; and 2) viewing applications via multiple perspectives; that is, for example, from the point of view of the language learner, of the traditional language teacher, or of the teacher who adheres to communicative language teaching principles. They were open minded in approaching a task that by nature provokes subjective judgements. Not having had prior opportunities to formulate strong opinions, their assessments can be considered unique. Future plans for the database project include the capacity for comparison of novice versus full-fledged CALL user evaluations. It will be possible, then, in future versions to compare and contrast reviewer evaluations based on level of experience as CALL users. Comparisons on this dimension will allow database users to view applications from multiple perspectives.

This group of teacher/reviewers, having read and discussed CALL design issues in the context of language teaching and learning practices, participated in generating the review criteria described below.

Review Criteria

The group agreed that information typically included in software reviews such as titles, authors/vendors, distribution, hardware requirements, and population targeted was fundamental to these evaluations. Spaces and prompts for this information consequently heads evaluation guide sheets. Skills intentionally targeted, that

is, those that are transparent in the design of the software application, were also noted in the initial identifier cluster of information.

Indices of program quality in terms of technical quality, content clarity, ease of use, navigation, and the nature of input and feedback were included as partially subjective measures in the second review cluster.

Accompanying these initial sets of information is the reviewers' prose description of the application. It was determined by the group that this description should be brief and objective and thereby serve as the base into which further subject commentary would hook. This portion of the evaluation guide, then, represented an opportunity to simply describe what the particular piece of software looks like and what it ostensibly does. Opportunity for critical analysis was reserved for later sections.

In developing and refining the evaluation criteria, the group concluded that to investigate possibilities for use of the software from various perspectives, they would need to consider uses of the software potentially beyond or contrary to the transparent skills level for the population targeted. That is, rather than accept only the level(s) indicated directly or indirectly by the package, they would look further to subjectively and creatively determine alternative and expanded uses for additional groups of learners.

Expanding on proscribed level was also of necessity carried over into expansion of proscribed audiences as several of the packages reviewed were language arts programs designed for native speakers of English. Once having determined that a native speaker package was adaptable for ESL, additional specifications regarding adaptation, appropriate second language skills level(s) were indicated according to software external tasks envisioned.

Interpretative criteria for the reviews included: content complexity (this is in terms of the level of cognitive processing in conjunction with linguistic processing required by the content represented in the software); motivation (how engaging students would find the content and genre of activities available); "learning mileage" (how much mileage students and their teacher might get from the package both in terms of time and quality of language practice); and open-endedness on the part of the student (as defined by Young, 1988) and open-endedness in terms of the teacher who might wish to modify the program (Underwood, 1984).

In keeping with the goal of multiple perspectives on the software evaluated, evaluation guides included prompts to reviewers to analyze the viability and value of the package for:

- One student (single user)
- Two students (a pair of users)
- Groupwork (large group or whole class)

It is under these categories that reviewers were directed to expand on specific pedagogical approaches, both transparent and potential,

for each of the three configurations. Reviewers found this section key to their thinking about using CALL as a catalyst for meaningful and creative interaction among students and the teacher. They considered ways to emphasize oral/aural practice of specific types of useful discourse around the package under review, then, within the context of each of these three configurations. Notes regarding strengths and weaknesses of the software design for single users, pairs and groups of users were also made in the context of configuration possibilities.

Finally, whether or not the kinds of language learning provided by the context of and activities in each package were superior by virtue of being on-line was to be indicated. That is, reviewers were prompted to determine whether the computer was an asset or whether the kinds of practice offered by the software could be done just as well off-line.

To prepare a database with optimal access for comparative purposes, visual tools in the way of icons were developed for several of the review criteria. Reviewers were to indicate the degree of quality of these measures by sketching in or numerically indicating the number of icons that correspond to their assessment. The scale extends from .5 icons (very poor) to 5 icons (excellent).

In field testing the evaluation scheme for the database, teacher/reviewers worked in pairs. As they were new to CALL, this served to enhance the review process in that comparison of perceptions came to mediate richer subjective assessments.

The Database

Reviewer evaluations were entered into a graphical database written in ToolBook (Asymetrics). The CALL software database, in its pilot form, consists reviewer responses to the criteria outlined above. Currently database users can access information concerning twenty-two software applications. They can also, as mentioned earlier, cross-reference applications by specified identifiers and compare evaluative information. Users are able to access reviews of application by specific platform or by potential uses as a whole-class presentation device, for example. In the same manner, users can contrast the transparent skills targeted with subjective reviewers' comments regarding uses beyond what a specific package prescribes. The architecture of the database allows users to add their own reviews and their own notes to any field.

Future Enhancements

Future plans include the capacity to access a list of software users to whom further questions can be directed, evaluations of software tools (wordprocessing and spreadsheets, for example) and telecommunications applications. Information regarding distributor review policies will also be included and frequently updated. Users will eventually be able to access relevant book, journal and newsletter titles where they can find additional information regarding CALL software.

REFERENCES

- Johnson, Donna (1991) Second language and content learning with computers: Research in the role of social factors. in Dunkle (ed) Computer-Assisted Language Learning and Testing: Research Issues and Practice. New York: Newbury House.
- Perez, Bertha (1984) Selecting Computer Software for Limited English Speakers. Intercultural Development Research Association Newsletter. September. p 1-8.
- Robinson, Gail (et al) Computer assisted instruction in foreign language education: A comparison of the effectiveness of different methodolgies and different forms of error corrections. EDRS:ED 262 626.
- Strei, Gerry, (1983) Format for the evaluation of courseware used in computer-assisted language instruction (CALI). CALICO Journal, 1, 2, p 43-46.
- Underwood, John (1984) Linguistics, Computers and the Language Teacher. Rowley, MA: Newbury House.
- Young, Richard (1988) Computer-Assisted Language Learning Conversations: Negotiating an Outcome. CALICO Journal, March, 65-83.

