The State of Children’s Software Evaluation—Yesterday, Today, and in the 21st Century

WARREN BUCKLEITNER
Michigan State University
Children’s Software Revue, 44 Main Street
Flemington, NJ 08822, USA
buckleit@aol.com

Choosing the best books, manipulatives, toys, and software is an important and essential task for any anyone who works with children. As computer use becomes more common in home and classroom learning, the selection of software takes on even more importance. Despite the growing need for software review information, however, the number of software evaluation studies and published papers on the topic has decreased steadily since its peak year in 1984.

This article examines the current state of children’s software evaluation practice and predicts what the future of educational software evaluation might look like in light of the potential of the dynamic nature of educational and edutainment software. Key issues are discussed such as the best way to evaluate the appropriateness of software for children at each age-group and the most efficient means of making this information available to parents and teachers.

The task of evaluating an interactive electronic experience with some degree of reliability and validity presents a unique set of challenges not previously encountered in the study of traditional materials such as books. Because the software experience is an interactive and multidimensional one, the evaluation of children’s software must look at the pedagogy used in its de-
velopment as well as design features (how help is provided, the accommodation of different abilities, the ease with which a child can access the program’s components, etc.).

Contrast this with the evaluation of children’s literature, a task in which entire courses are dedicated in teacher education programs. With children’s books, the basic technology of ink on paper has changed little over the past 300 years. Books are a two-dimensional, fixed, often linear medium, and making judgments about the quantity (how many pages) and the quality of the graphics and text is a relatively straightforward process. If books are two-dimensional, software is three-dimensional, with the addition of the psychology of the user interface and all of the issues related to the timing of events. In addition to decisions about the quantity (how many activities or some measure of replay value) and quality (narration, animation), an evaluator is placed in the position of making judgments about such things as the degree of child control the product provides, and the clarity of the menu design. In some ways, evaluating a software experience can be like evaluating a teacher/child interaction.

Some of questions to bring to the task of software evaluation:

1. What is software? While this may seem like a trivial question, it is essential first ground to cover in any discussion of software evaluation. Since it was first used in a technical computing journal in 1960, the word “software” has come to have many different meanings. The current common usage refers to software as the programming code stored on CD-ROMs that transforms a computer into a set of games, a word processor, or perhaps an Internet browser, and it is this definition that is used for this paper. But the introduction of talking toys with embedded microprocessors and ROM chips has blurred any conventional definition of software. As the Internet continues to develop, the differences between interactive and linear television programming will converge, further expanding the types interactions children have with various types of programming. As a result, it is important to come to the task of software evaluation with an encompassing knowledge of the various kinds interactive electronic experiences that the current technology presents.

2. What is the intended purpose of the software, and where is the software intended to be used? The intended function and context is essential to consider when evaluating software. Is the software designed to teach a specific skill such as letter recognition in a quiet classroom setting? Or is it’s purpose to simulate an on-screen puppy, provide a set of tools for
The State of Children's Software Evaluation

creating multimedia presentations, or to give the child a fun experience while racing Hot Wheels cars? Whatever the intention is, does it do the job well? It is the task of the teacher or evaluator to apply the appropriate evaluation criteria to a software product in order for a resulting review to be useful.

3. What is the developmental level of the intended audience? While all software needs to have an intuitive menu design, software for children needs to be carefully designed by individuals who have a good understanding of the way children think. For example, an opening menu that requires reading will not work in a preschool setting. Likewise the theme of the graphics, style of music, and choice of characters can have an influence on how children react to a software product.

4. How does the software compare with similarly designed, competitive products? In order to be of value, a review must consider the entire population of similar products in order to make conclusions about the value of the product. If a similarly designed product is on the market for half the price, this needs to be mentioned.

5. What theoretical orientation do you bring to the software evaluation process? Every educational perspective has its champions in software. Constructivist thinking has influenced programs like LEGO Mindstorms, while behaviorism’s reinforcement strategies are well applied in programs like JumpStart Kindergarten, where children earn “Smart Stars” for completing activities. In making judgment calls on these products, a reviewer should not only be able to identify the underlying theoretical influence in a product, but must also be aware of his or her own theoretical orientation.

6. Does the software take advantage of the latest technology? Most of the research regarding the evaluation of software was done from 1982 to 1986 (Figure 1), when the Apple IIe with 128K of memory represented the state of the art. Today, the same $2000 that one would have spent to purchase an Apple II buys a computer that is 300 times faster, with 32 MB of memory, and Internet Access. Present day computers have text-to-speech capability, the ability to recognize children’s voices, and the ability to play high quality music and full-motion video. Software size is no longer limited to an unreliable floppy disk—CD-ROMs provide dependable vehicles for delivering large amounts of interactive content at a cost less than a floppy disk.

7. What is the history of the software in question, and what is the current “state of the art” of comparable software? Because microcomputer technology is changing so quickly, it is important that the date that any re-
view was conducted is given, along with the version of software, and the hardware requirements needed to make the program function properly. For programs like Reader Rabbit, there are dozens of different versions, updates, and packaging options. This information must be included in any review in order for it to be useful.

Who Evaluates Educational Software?

The job of evaluating the quality and appropriateness of software products over the years has been taken on by government agencies, universities, private for-profit publications, non-profit consumer groups, educational conventions and trade groups. These reviews are delivered to the public in the form of booklets, periodicals, or more recently on the Internet and television programming. For educators, librarians, and parents who are relying on software evaluations that have been conducted by others, it is essential to consider the review organizations orientation and areas of possible bias. Here is a sampling of different sources for software evaluation along with a discussion of strengths and weaknesses.

A Brief History of Educational Software Evaluation

In the early 1980’s, the romance and novelty of computers software led to a flurry of software evaluation activity (Lathrop & Goodson, 1983; Rucker 1985; Jones & Vaughan (1983); Mead 1983). One way to measure the extent of this activity is to search the ERIC database using the key words “Software and Evaluation.” This technique shows that there was more published work on the topic for the year 1984 than any other year (Figure 1) with a total of 419 papers, studies, and software evaluation related articles. During that year, there were 22 software evaluation services in the US and Canada, according to the publication “Only the Best,” a book which has synthesized the review information of services over the years, and which contains a list of the various review sources over the years. (Neill & Neill, 1985-1992; ASCD, 1992 to present).

It is interesting that the amount of software evaluation activity has decreased significantly in subsequent years, despite the gradual increase in computer use with children and the dramatic increase in the computer’s potential for delivering educational experiences via CD-ROMs and the Internet.
Table 1
Categories of Who Reviews Children's Educational Software

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Strengths/Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>City, State and Regional Departments of</td>
<td>California Instructional Technology Clearinghouse, Columbus Public Schools</td>
<td><strong>Strengths</strong>: Reviews are written with a school setting in mind, making these more appropriate sources for teachers. Reviewers have no ties to the industry, which increases the validity. <strong>Weaknesses</strong>: Reviews often lag behind the current market. Because multiple reviewers are used with no evidence of inter-rater reliability training, the reliability of the reviews can be decreased.</td>
</tr>
<tr>
<td>Education, Public School Districts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Software and Hardware Industry</td>
<td>Apple Computer Software Guides</td>
<td><strong>Strengths</strong>: These booklets are distributed free of charge, and can be useful for learning about the software for a particular platform. <strong>Weaknesses</strong>: Reviews are written to favor a particular platform. Reviews may be dated or not comprehensive.</td>
</tr>
<tr>
<td>Software Catalogs</td>
<td>Educational Software Institute (ESI) BrainPlay.com SmarterKids.com Amazon.com</td>
<td><strong>Strengths</strong>: Catalogs can provide useful, up-to-date price and platform information. In some cases, the online versions of these catalogs make it possible to leave comments about software products. <strong>Weaknesses</strong>: There is an underlying sales motive behind the reviews. Software descriptions can be written to emphasize strengths and diminish weaknesses.</td>
</tr>
<tr>
<td>Commercial magazines and Internet-based</td>
<td>Family PC Technology &amp; Learning Media &amp; Methods Curriculum Administrator</td>
<td><strong>Strengths</strong>: Can be more timely and often more accurate, with useful side-by-side comparisons of competitive products. Can be comprehensive. <strong>Weaknesses</strong>: These magazines may be supported by advertising revenue from the software publishers whose products are reviewed. Critical reviews of highly marketed programs can be hard to find. Reviews can be written by a variety of reviewers with little or no inter-rater reliability, and the reviews may not be comprehensive. Money may be taken from publishers in exchange for the use of award seals.</td>
</tr>
<tr>
<td>publications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trade Group Award Programs</td>
<td>The Software Publishers Association Codies (SPA) EdPress</td>
<td><strong>Strengths</strong>: These awards are given by industry insiders who vote by ballot. <strong>Weaknesses</strong>: There are often entry fees required to join, which eliminates the smaller companies and reduces the validity of the awards.</td>
</tr>
<tr>
<td>Universities</td>
<td>Southern Illinois University Arizona State University</td>
<td><strong>Strengths</strong>: can be objective and more analytical <strong>Weaknesses</strong>: Often not comprehensive or reliable</td>
</tr>
<tr>
<td>Independent Experts, “Gurus”</td>
<td>The Oppenheim Toy Portfolio, Dr. Toy</td>
<td><strong>Strengths</strong>: Reviews are written by educators with a strong child development background, or knowledge of how children play. <strong>Weaknesses</strong>: Because these experts also review toys, they may not be as knowledgeable about software related issues. Money may be taken from publishers in exchange for the use of award seals.</td>
</tr>
<tr>
<td>Non-Profit Mission Driven Organizations</td>
<td>Parent’s Choice Foundation EPIE Institute (Educational Products Information Exchange)</td>
<td><strong>Strengths</strong>: Can provide comprehensive, objective coverage of the software market. <strong>Weaknesses</strong>: Can be dependent on external funding so the comprehensiveness can vary from year to year.</td>
</tr>
<tr>
<td>Independent Experts, “Gurus”</td>
<td>The Oppenheim Toy Portfolio, Dr. Toy</td>
<td><strong>Strengths</strong>: Reviews are written by educators with a strong child development background, or knowledge of how children play. <strong>Weaknesses</strong>: Because these experts also review toys, they may not be as knowledgeable about software related issues. Money may be taken from publishers in exchange for the use of award seals.</td>
</tr>
<tr>
<td>Non-Profit Mission Driven Organizations</td>
<td>Parent’s Choice Foundation EPIE Institute (Educational Products Information Exchange)</td>
<td><strong>Strengths</strong>: Can provide comprehensive, objective coverage of the software market. <strong>Weaknesses</strong>: Can be dependent on external funding so the comprehensiveness can vary from year to year.</td>
</tr>
</tbody>
</table>
Figure 1. The number of studies or papers related to educational software evaluation from 1980 to 1997 (ERIC database)

The first and largest systematic software evaluation effort was conducted by the Educational Products Information Exchange, or EPIE. Since 1967, the nonprofit group has been reviewing textbooks, AV materials, and other educational resources, and turned its attention to software full time including software, in the form of a database known as TESS (The Educational Software Selector, Komoski & Haven, 1983-1993). TESS was available from 1983 to 1993 in print form, and on CD-ROM in subsequent years. EPIE was one of the first groups to apply the searchable database as a means of helping educators find specific products.

The 1980s also saw the first standardized evaluation instruments for children’s software, from High/Scope Educational Research Foundation (Buckleitner, 1985) and the Haugland and Shade DAP Scale (1988). These checklist-based software evaluation forms were an attempt at quantifying the factors in software design that have been associated with the effectiveness of software products for young children.

In the 1996, the first review databases became available online, and Amazon.com, the online bookstore, demonstrated the potential for a commercial for merging evaluation information with the purchasing process.
These Internet-related technologies have helped to lay the blueprint for the future of the way software evaluation information may be created and distributed in the 21st century.

**The Future of Software Evaluation**

Given the rate at which microcomputer-related technology has proliferated over the last 20 years, the demand for believable software review information is likely to increase as more educators attempt to utilize fit computer-based activities into their lesson plans.

A look at the past and current methods of software evaluation shows that practices are behind the times, often unreliable, biased by commercial interests, and rarely verified by independent studies. What is needed are more studies such as that done by Escobedo and Evans (1997) where the ratings assigned by the published software methods are compared with actual child selection, or Tammen and Brock (1997), where middle school students are asked to identify issues they feel are important for the evaluation of software programs. In the Escobedo and Evans study, children were videotaped choosing between 13 software programs that had been assigned ratings with the Haughland and Shade (1990) evaluation instrument. Results showed some positive relationships between highly rated programs and children’s preferences, although some of the titles identified by the scale as developmentally inappropriate were preferred by children. Additional findings revealed that children preferred the software that provided the opportunity to interact, and interactions were the defining characteristic that motivated selections.

We should begin asking questions about what software experiences are best at promoting different aspects of children’s learning. For example, do rote learning experiences at the computer actually result in higher achievement test scores? In the future of software evaluation, specific characteristics such as these should be considered. Other trends with potential include:

- The democratization of the review process via the Internet. Amazon.com, the online bookstore, serves as a proven model for how review information can be shared by way of an public shared database. These online electronic forums hold potential for increasing the validity of reviews, because anyone can contribute a differing opinion on a particular software product.
- Integration of reviews with lesson plans. Review information doesn’t do any good unless it is correlated with local, state and national learning objectives, and unless teachers have access to it when they are planning
their lessons. In the 21st century, teachers will have electronic lesson plans, with access to review databases that will help them match a software activity with a particular learning setting.

Objectivity in educational research is not a new concept, nor are the basic ideas of validity and reliability in measurement. As we move into the 21st century, our children deserve rigorous, well constructed evaluation methods applied to the products they use that are subject to public criticism and evaluation. When computer-savvy educators have access to this kind of solid, evaluative information, they can plan and teach more effectively, and tap the power and excitement that computers can bring to the learning process.

Table 2
Online Sources of Educational Software Reviews

**BrainPlay.com**
http://www.brainplay.com
This catalog contains software reviews from several different sources, including the Boston Computer Museum and older reviews from Children’s Software Revue. A database of children’s video reviews is also available on this site.

**California Instructional Technology Clearinghouse**
http://clearinghouse.k12.ca.us
This site contains the Technology in Curriculum (TIC) Evaluations database, with reviews of over 2000+ different educational products.

**Children’s Software Revue**
http://www.childrenssoftware.com
Contains a searchable database of 3400 software reviews, with the ability to add reviews, or publically comment on a review. Content for the print publication and website is written by educators.

**EPIE Institute**
http://www.interhelp.com/epie_tess.htm
EPIE (Educational Products Information Exchange) will soon be providing free public access to the TESS (The Educational Software Selector) database, thanks to a grant from the Department of Education.

**EvaluTech**
http://www.evalutech.sreb.org
EvaluTech is a free, searchable database of curriculum-related print, non-print and technology reviews specifically for Pre-K-12 educators based on the NCPDI print publication (North Carolina Department of Public Instruction, Educational Technology Programs Division).
Family PC
http://www.familypc.com
The web site for the commercial print magazine, with information on late-breaking hardware and software. A good source of family tested reviews and interesting articles.

Only the Best: The Annual Guide to the Highest Rated Software and Multimedia
http://www.ascd.org/services/newotb/webpromo.html
This annual book and CD takes a survey of 20 or so review organizations, and lists just the top rated software.

Superkids Educational Software Review
http://www.superkids.com
This site contains software reviews with ratings (for home use), along with some interesting articles.

Software Publisher’s Association
http://www.spa.org/project/resource.htm
This site provides a listing of educational software review sources. Some are dated, but overall this is useful.

Southern Illinois University School of Education (Edwardsville)
http://leader.soed.siue.edu/Evaluations/EvalProject.html
This online database provides an example of an experimental project to gather and share evaluations of educational software. The evaluations are being completed by teachers in the area, and the results are being stored in a database and shared with anyone who might be interested.

Technology & Learning Magazine
http://www.techlearning.com
Offers a searchable database of reviews from past issues of the magazine.

Worldvillage
http://worldvillage.com
A nice source of illustrated reviews from 40 different US-based reviewers. Reviews cover games, the Internet, education, and more.

Ziff Davis Software Library
http://www.hotfiles.com/educate.html
ZDNET Software Library contains ratings of shareware and software demos that can be downloaded on the spot.

References


