EDUCATIONAL SOFTWARE EVALUATION FORM FOR TEACHERS

Yılmaz Kara
University of Atatürk, Turkey
yilmazkaankara@yahoo.com

ABSTRACT

The purpose of the study was to develop an educational software evaluation form to provide an evaluation and selection instrument of educational software that met the requirements of some balance between mechanics, content and pedagogy that is user friendly. The subjects for the study comprised a group of 32 biology teachers working in secondary schools in the central part of the Erzurum province in Turkey. The biology teachers were asked to evaluate three educational software packages for their classes, schools, students, and individual use. The data from biology teachers were analyzed by the SPSS, using the frequency, reliability and factor routines. The body of the form covers 40 evaluation criteria in four major categories: Content, Student Involvement, Ease of Use, and Design, Esthetics.

Keywords: Educational Software Evaluation; Educational Software Selection; Evaluation Criteria.

INTRODUCTION

Instructors have been using computers and their components extremely because of their advantageous aspects in classrooms nowadays. The proliferation of computer assisted learning (CAL) has, in fact, affected the structure and culture of education. As the new language of education, it has facilitated the development of more student-centered approaches through the production of interactive educational software for teaching and learning. Since the introduction of new technologies into schools over the past several years, there have been frequent and continual expressions of concern for the quality of computer software (Komoski, 1984).

Evaluating and selecting the best, the appropriate software is a very important component of success in using educational software in both educational and corporate settings (Lee, Choi & Byun, 1996). Instructors, as well as parents, often have difficulty distinguishing quality instructional software from the trivial and ineffective software now on market. Guidelines and sources of knowledge about evaluating and selecting quality software program are limited (Chang & Osguthorpe, 1987). As more of the instructional software is placed on the market the need for careful review of the material prior to its purchase becomes increasingly necessary. Educational software, like all other educational material, should be evaluated with a thorough and detailed evaluation before it is used in the classroom (Heller, 1991).

Schools use educational software, teachers are increasingly having their students use the programs for a variety of learning activities. Many teachers are finding educational software to be a valuable teaching tool that offers a bonanza of learning activities for students. Teachers must give students materials that are not offensive, outdated, or biased. Teachers need to provide objective data regarding the instructional effectiveness of the educational software before purchase. There is a need for good to know how to evaluate software and not naively rely on advertisements. Evaluators of educational software must have the skills to critically evaluate and decisions for not only about format, but also content and the process of learning (Herring, Notar, and Wilson, 2005).

In a study conducted by the Eisenhower National Clearinghouse at Ohio State University, the researchers examined 200 software packages intended for use in social studies and language arts instruction. Based on the evaluations, most of the software lacked important content. The researchers noted that the software was entertaining, but not instructional (USA Today Magazine, 1998). In an earlier study, the same researchers examined 175 math and science software programs and found the same results (BioScience, 1997). One of the major reasons for poor software in the schools deals with the evaluation of that software by education professionals. Despite the increasing number of software titles in education, the area of educational software evaluation has been increasingly more and more muddled because of a lack of consensus among software evaluators (Hardin & Patrick, 1998; Huber & Guise, 1995).

With all of the software that exists today, it has become increasingly difficult for educators and curriculum experts to keep up with which software programs are the best. Since the 1980's, many different academicians have attempted to determine what to examine when deciding to purchase educational software and what makes software effective. Every major researcher who has created a series of terms for an individual to check off when evaluating software has created a new scheme. These new schemes have just confused the arena of software evaluation. All types of schemes have been created to evaluate software. Most of the evaluation forms concentrated on the technology and mechanics and did not look at pedagogy and content. It was also noted that there is no standard evaluation criteria or common terminology. It was decided to develop an evaluation form to provide an evaluation and selection instrument of educational software that met the requirements of some balance between mechanics, content and pedagogy that is user friendly. The purpose of this study is to develop an educational software evaluation form that engages in-service teachers in the selection of materials for development of instructional activities that can be used in multidisciplinary environments.

METHOD

This study consisted of the development of an evaluation form for educational software programs to provide in-service teachers how to evaluate educational software programs in an effective way for their classes, schools, students, and individual use. However, the form can be used by anyone else with knowledge of educational software. Participants were given the brief written information and steps to evaluate educational software in order to fulfill the evaluation requirements on the evaluation form. A detailed description of the brief written information and steps to evaluate educational software is presented below:

Educational software evaluation

Multimedia, hypermedia training materials and computer based and aided instructions are part of the teaching profession. Teachers have for years been on textbook selection committees. Now teachers will be on software selection committees. People who have had experience in the use of a variety of software and a good software evaluation form are critical to selection committees. This exercise will give you a leg up on the rest of your fellow teachers for this most important role.

The criteria presented here for evaluating software is not exhaustive. For example, where software purchasing depends on available funds, a cost-benefit/effectiveness analysis would be useful or may be required. Also, there may be economic and political details, which vary broadly from school to school, have an impact on evaluative findings.

When preparing for an evaluation everyone should be involved in the evaluation and the interpretation of results which will increase the use evaluative results to improve learning. The evaluation form is based on applying variables in the technical, instructional, and organizational levels of software. The evaluation form is set up for the school practitioner whose insights on the processes and outcomes of software use are among the most important of all stakeholders. The researcher will provide the educational software programs for evaluation.

Steps to evaluate multimedia software

-briefly examine manual and other documentation to ensure you have a clear idea of the purposes and capabilities of the software program

-fill out general information of evaluation form.
-insert CD, follow directions and work through program (for exceptionally long programs complete a representational portion)
-work through program again making errors and responses you think your students might make (most effective method here is to have one or some of your actual students work through the program)
-complete multimedia software evaluation form (you may need to do this while program is running to allow you to refer to the computer and program while completing the form).

Three software programs were given to the teachers to evaluate. These software packages where: Vitamin (SEBIT Education and Information Technologies Inc.), Rediscover Science and Math (Edunetics Interactive) and Bioscopia (Ruske & Puhretmaier Edutainment). The form was completed by all biology teachers (total, n=32) working in secondary schools in the central part of the Erzurum province in Turkey. All subjects were university graduates with majors in biological science, and most were experienced biology teachers with around ten years teaching experience. Teachers were given one week to complete the form, after which time they were returned to the researchers for assessment. Teachers were reassured that their questionnaire answers and the results of this research would not be used for any purpose other than for this study. The data were analysed by the SPSS statistical package (SPSS Inc., 1988), using the frequency, correlation, reliability and factor routines.

RESULTS

To develop the Educational Software Evaluation Form (ESEF), a pool of items was collected by adapting items from before studies (Aşkar ve Köksal, 1987; Deniz, 1988; Eseryel, 2002; Gibbs, Craves, and Bernas, 2001; Karaman, 1996; Lee, 1997; Maden, 1996; Nisanci, 2000; Numanoğlu, 1992; Orhan, 1995; Shade, 1996; Şimşek, 1995; Taylor, 1987; Truet and Gillespie, 1984) and writing new items. The author developed 11 additional items for the initial pool of items. These items were included after consulting with some experts in technology education. As a result, the initial pool of items in the scale included a total of 57 items. These 57 items were then presented using a five-point Likert scale (from strongly agree, agree, neutral, disagree to strongly disagree) to a group of Turkish biology teachers for item analysis. All of the items were presented in Turkish. The translation between English and Turkish (if required) in this study was completed by the author, and the experts in English Language validated the translation

.

Table 1: Rotated factor loadings for the four factors of the ESEF

Items	Factor loading
Content	
Does the content meet the objectives of your intended use?	0.689
Is the content appropriate for your intended audience?	0.463
Is the content accurate?	0.685
Is the content current?	0.601
The sited is NOT biased.	0.712
Is there real substance and depth of the content?	0.619
Does the content present subject matter in an interesting, lively, and compelling way?	0.731
Are the educational objectives clearly stated?	0.643
Are specific learning skills addressed by the software content?	0.738
Does the program offer a range of difficulty levels?	0.530
Student Involvement	
Is the interaction with the content quality?	0.397
Does the content promote higher-level thinking?	0.537
Does the lesson personalize instruction appropriately?	0.455
Is the program appealing to a wide audience?	0.699
Do the children return to this program time after time?	0.503
Is the step size appropriate for the kind of learners and the learning task?	0.741
Are the speeches and sounds meaningful to children?	0.856
Are lesson activities, content, and procedures likely to motivate to perform?	0.741
Is the program responsive to a child's actions?	0.687
Is the theme of the program meaningful to children?	0.559
Ease of Use	
Is the software easy to install?	0.681
Is the software easy to learn the first time you use it?	0.479
Is the software easy to use once you've learned to use it?	0.780
Will additional training be required?	0.686
Are tutorials provided (on screen or online)?	0.664
Is there a helpful user manual?	0.374
Is it easy to navigate through the software?	0.701
Does the opening screen give clear directions?	0.607
Can your child use the program independently after first use?	0.689
Would this program operate smoothly in a classroom?	0.555
Design, Esthetics	
Is there a balanced use of graphics, text and sound?	0.515
Are graphics used appropriately?	0.472
Do graphics make sense to the user?	0.761
Is sound used appropriately?	0.769
Do graphics make sense to the user?	0.661
Are buttons varied, obvious and easy to use?	0.781
Are spelling and grammar used correctly?	0.806
Does the software track student progress?	0.582
Are necessary skills to operate the program within the developmental range of your students?	0.542
Does the software track student progress?	0.464

A principal components factor analysis with oblique rotation was performed to clarify the structure of ESEF. The purpose of the analysis was to determine whether the factors of ESEF could be meaningfully differentiated from one another. An item was retained only when it loaded greater than $\pm$ 0.30 on the relevant factor and less than $\pm$ 0.30 on non-relevant factor. Thus, the initial 57 items were reduced to 40 items. Only four factors were retained in the final version of the scale and they accounted for 52.1% of variance. The eigenvalues of the four factors from principle component analysis were larger than one: 10.078, 4.738, 3.505, and 2.553. These variables were labeled according to the major items designing each component, they were: content, student involvement, ease of use, design and esthetic. Content subscale accounted for most of the variance (25%) with the factor loading ranging from 0.463 to 0.738. Student involvement component accounted for 11% of the variance with loadings ranging from 0.397 to 0.856. Ease of use accounted for 8% of the variance with loadings from 0.374 to 0.780. Finally, items that loaded on design and esthetic accounted for 6% of the variance with loadings from 0.464 to 0.806. Table 1 presents the factor structure for the ESEF.

A Cronbach's coefficient α was calculated for each of the four sub-scales and the overall scale as a whole. Cronbach's alpha

of the total instrument was 0.91. The alpha coefficient was 0.88 for the content subscale, 0.87 for the student involvement subscale, 0.85 for the ease of use subscale, and 0.84 for the design and esthetic subscale. As Table 2 shows, the α coefficients for all sub-scales were significantly high; suggesting that the internal consistency of the constructs and overall scale is satisfactory.

Table 2 further presents students' average total scores and standard deviations on the four subscales. Teachers scored highest on the ease of use subscale (an average of 3.91 per item, i.e. 39.114/10) and followed by the design and esthetic subscale (an average of 3.82 per item), the student involvement subscale (an average of 3.85 per item), and the content subscale (an average of 3.12 per item). These results imply that teachers, in general, tended to appreciate the ease of use and design, esthetic.

Table 2: Means, standard deviations and reliability coefficients for the ESEF and subscales.

	Items	Mean	SD	Cronbach α
Content	10	30.125	8.748	0.88
Student involvement	10	34.854	8.082	0.87
Ease of use	10	39.114	6.680	0.85
Design and esthetic	10	38.250	6.690	0.84
ESEF	40	142.343	21.864	0.91

Table 3 presents the Pearson correlation coefficients between each of the four sub-scales and the whole scale. The correlations between the sub-scales range between 0.304 and 0.585. The correlations between the sub-scales and the whole scale range between 0.61 and 0.81. This result indicated that there was a high correlation between ESEF and subscales.

Table 3: Inter-correlation matrix for the four sub-scale and overall ESEF.

	Content	Student involvement	Ease of use	Design and esthetic
Content	-			
Student involvement	0.585*	-		
Ease of use	0.346*	0.312*	-	
Design and esthetic	0.304*	0.309*	0.367*	-
ESEF	0.815*	0.760*	0.672*	0.614*

* $p<0.01$

CONCLUSION

Within the theoretical framework laid out by before studies, a scale to evaluate the educational software programs has been developed. The scale consists of four factor analytically distinct sub-scales with high internal consistency, stability and validity. It is hoped that the scale will be of use to educators and researchers. The final scale should be presented as a list of the 40 items, alternately displayed so that no two items from the same construct appear adjacently, alongside a 5-point Likert scale (worded "Strongly Agree", "Agree", "Neutral", "Disagree" and "Strongly Disagree"). Scores from items on each subscale (ranging from 10 to 50) can be summed to provide content, student involvement, ease of use, and design, esthetic scores. The whole scores can also be collectively summed to provide a total score representing the overall evaluation (ranging from 40 to 200). As a normative guide to interpretation, the scores obtained with a sample of 32 teachers gave cut-off scores: at the 25th percentile of 126; at the 50th percentile of 144; and at the 75th percentile of 158. Thus a score below the 25th percentile (126) can be interpreted as "this piece of software is not tested and not beneficial", whereas a score above the 75th percentile (158) can be interpreted as "this would be a good software decision".

The focus of this presentation is on discussing the issues surrounding the evaluation of educational software and on building a comprehensive critter for evaluating and selecting educational software for effective instruction. These criteria will provide a useful framework to help educators and/or trainers select quality educational software for their instructional purposes. However there is no widely agreed upon standards or criteria for educational software evaluation. It is a paramount need to develop not just minimal standards that will act as guidelines to help developers, evaluators, and consumers determine what quality educational software programs are.

REFERENCES

Aşkar, P., Köksal, M. (1987). Bilgisayar destekli öğretimde kullanılan yazılım paketlerinin geliştirlmesinde ve değerlendirilmesinde dikkat edilecek noktalar. Eğitim ve Bilim, cilt 12, sayı 66, 20-23.

BioScience. (1997). Educational software gets failing grade. Bioscience, 47, 720.

Chang, L.L. & Osguthorpe, R.T. (1987). An evaluation system for educational software: A self instructional approach. Educational Technology, 27(6), 15-19.

Deniz, L. (1988). Bilgisayar yazılımlarının değerlendirilmesi: Eğitsel yazılımlar. Yayınlamamış yüksek lisans tezi. A.Ü Sosyal bilimler enstitüsü, Ankara.

Eseryel, D. (2002). A Framework for Evaluation & Selection of E-Learning Solutions. In: E-Learn 2002 World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education. Proceedings (7th, Montreal, Quebec, Canada, October 15-19, 2002); see IR 022 093. Support provided by the College Center of the Finger Lakes (CCFL). (ERIC Document Production Service No. ED479454)

Gibbs, W., Craves, P R., & Bernas, R. S. (2001). Evaluation of guidelines for multimedia courseware. Journal of Research on Technology in Education, 54(1), 2-17.

Hardin, L., Patrick, T. B. (1998). Content review of medical educational software assessments. Medical Teacher, 20, 207-212.

Heller, R.S. (1991). Evaluating software: A review of the options. Computers & Education, 17(4), 285-291.

Herring, D.F., Notar, C.E., Wilson, J.D. (2005). Multimedia software evaluation form for teachers. Education, Fall, Vol. 126 Issue 1, p100-111.

Huber, J.T., & Guise, N.B. (1995). Educational software evaluation process. Journal of the American Medical Informatics Association, 2, 259-96.

Karaman, M.K. (1996). Bilgisayar destekli öğretimde kullanılan yazılımnların taşıması gereken eğitsel özellikler ve bazı eğitim yazılımlarının değerlendirilmesi. Yayınlamamış yüksek lisans tezi, Afyon Kocatepe üniversitesi, Sosyal bilimler enstitüsü, Afyon.

Komoski. P.K. (1984). Educational computing: The burden of insuring quality. Phi Delta Kappan, 66, 244-248.

Lee, J.R. (1997). Selecting and evaluating CD-ROMs: A public school media specialist's perspective. Journal of Educational Media & Library Sciences, 34, (4), 368-79.

Lee, S.H., Choi, W., Byun, H. (1996). Criteria for evaluating and selecting multimedia software for instruction. Proceedings of selecting research and development presentations at the 1996 National Convention of the Association for Educational Communications and Technology, 18th Indianapolis, IN.

Maden, C. (1996). Bilgisayar destekli öğretim için hazırlanan eğitim yazılımlarında bulunması gereken öğretim tasarımları ölçütleri ile var olan eğitim yazılımlarının bu ölçütler açısından incelenmesi. Yayınlamamış yüksek lisans tezi. Hacettepe üniversitesi, Sosyal bilimler enstitüsü, Ankara.

Nisanci, M. (2000). Instructional Software Evaluation Criteria Used by the Teachers: Implications from Theory to Practice. In: Society for Information Technology & Teacher Education International Conference: Proceedings of SITE 2000 (11th, San Diego, California, February 8-12, 2000). Volumes 13; see IR 020 112. (ERIC Document Production Service No. ED444534)

Numanoğlu, M. (1992). Milli eğitim bakanlığı bilgisyara destekli eğitim projesi bnilgisayara destekli öğretim yazılımlarında bulunması gereken eğitsel özellikler. Yayınlamamış doktora tezi, Ankara üniversiştesi, sosyal bilimler enstitüsü, Ankara.

Orhan, F. (1995). Bilgisayar ders yazılımlarının değerlendirilmesi için bir model önerisi. Yayınlamamış doktora tezi, Hacettepe üniversiştesi, sosyal bilimler enstitüsü, Ankara.

Shade, D.D. (1996). Software evaluation. Young Children, 5J(6), 17-21.

Şimşek, N. (1995). Yazılımın tasarım standartlarının bilgisayar ortamında öğrenmeye etkisi. Yayınlamamış doktora tezi, Ankara üniversiştesi, sosyal bilimler enstitüsü, Ankara.

Taylor, R. (1987). Selecting effective courseware: Three fundamental instructional factors. Contemporary Educational Psychology, 12(3), 231-43.

Truet, C., Gillespie, L. (1984). Chosing Educational Software: A Buyer’s Guide. Colarado: Libraries Unlimited Inc.

USA Today Magazine. (1998, April). Educational software gets low marks. USA Today Magazine, 126 (2635), 15.