

## **THE RELATIONSHIP BETWEEN THE TEACHERS' EVALUATIONS AND THE STUDENTS' LEARNING OUTCOMES USING MULTIMEDIA DRILLS IN PRIMARY EDUCATION**

**Piret Luik**

Tallinn University, Estonia

### **Abstract**

*The present study highlights the implementation of information and communication technology in education via the use of many educational software programs, which every teacher can use with their students. Some learning materials are produced by individuals who are not aware of the pedagogical principles and do not know how to produce effective educational software. Therefore, the question arises: Are the teachers competent in choosing multimedia learning materials? And if they compose multimedia learning materials by themselves, are they aware of the features of multimedia which are effective or which hinder learning? The study was carried out to seek answers to these questions and find out the relationships between the teachers' evaluation of software used and the learning outcomes by students in primary education. The results are based on a correlation study where 34 multimedia drills and practice materials were used for learning. Each multimedia drill was evaluated by ten primary teachers, who had used these drills with their students and saw how the students managed with each particular drill. The evaluations by the teachers were gathered by means of questionnaires, using a 10-point scale covering 17 aspects of the drills. Each drill was used by 80 students of the 3<sup>rd</sup> form from the same schools. The learning outcomes provided by the students were evaluated by pre- and post-tests.*

**Key words:** primary education, teachers' evaluation, learning outcomes, multimedia drills

### **Introduction**

In Estonian schools, the teacher chooses the learning materials for students. Traditional learning materials (printed text-books and work-books) are composed by competent authors. All traditional learning materials go through an evaluation process by experts, and these materials gain approval from the Ministry of Education. With the implementation of computers in schools, a number of different educational software has rapidly increased in Estonia. Different kinds of multimedia learning materials are available from the Internet as well. These kinds of multimedia learning materials are not usually evaluated by experts.

Some learning materials are produced by the individuals who are not aware of the pedagogical principles and do not know how to produce effective educational software.

There are courses on how to technically compose multimedia learning materials for pre-service teachers in different teacher education curricula, but there are few courses where evaluation procedures for such kind of materials are taught. Therefore, the question arises as to whether teachers are competent in choosing appropriate multimedia learning materials. And if they compose multimedia learning materials by themselves, are they aware of the features of multimedia materials which are effective or which hinder learning? This study was carried out to answer these questions and find out the relationships between the teachers' evaluations and the learning results by students.

## Review of the literature

There are many studies investigating students' ratings of educational software (Diederer, Gruppen, Hartog, & Voragen, 2005; Herring, Notar, & Wilson, 2005) and students' learning outcomes achieved by working with these learning materials (Mikk & Luik, 2003; Jacobson, 2006; Ngu & Rethinasamy, 2006; Luik, 2007). Some handbooks and papers provide suggestions for efficient evaluation of educational software of different types (Boyle, 1997; Phillips, 1997; Hughes, 1998; Higgins, 2000; Alessi & Trollip, 2001). But there are very few studies which deal with the relationships between the teachers' evaluations of the learning materials and the students' learning results achieved by working with these learning materials. Nevertheless, it is an important area, because if the teacher composes or chooses ineffective learning materials believing that they are efficient, the students are unlikely to achieve high learning outcomes.

Rana (2002) has investigated primary teachers' expectations about a Web-Based Learning System. The research questions are the following: *What kind of Web-Based Learning System can be effective according to the opinions of teachers?* and *How do primary teachers rate navigability, presentation, easiness of use and suitability in teaching?* Rana (2002) found out that the teachers preferred websites with hyperlinks and the colourful nature of the website. But the teachers were of the opinion that the text colour should be balanced against the background colour. Also, the teachers liked clear instructions and easy to follow navigation of the websites. As the learning in primary school should be more like a game, the primary school teachers liked the reward system in the websites.

Nurmi and Lehti (2003) have investigated a little over 500 European teachers' opinions on the usefulness of digital learning materials. The results indicated that the teachers needed a variety of different kinds of digital materials. The teachers wished to use traditional types (content plus self-test section) of learning materials besides other types of materials. The teachers considered a possibility to choose, (re)use, combine, modify and implement learning materials as very important.

Bos (2003) has carried out research on teacher rankings of several sets of exercises on the basis of the expected effectiveness. She (Bos, 2003) found out that the individual characteristics of the children did not have any influence on the ranking of the exercises.

Kurz and Middleton (2006) studied pre-service math teachers during one course, seeking to find out how they determine the learning and usability afforded by the software as it may possibly relate to students' learning. The authors used heuristic evaluations of educational software, and they concluded that, after a specific software course, the pre-service teachers were able to distinguish between the features and describe how these features support or hinder the learning process of their future students.

The previous research has examined studies on the teachers' expectations and rankings of learning materials, but not how these rankings correlate with the students' learning outcomes. The study objective is to reveal the link between teachers' valuations and students' learning outcomes working with educational software. Basing on the results of Rana (2002) and studies of characteristics of software (Mikk & Luik, 2003; Luik, 2007), the following hypotheses are put forward:

1. There is a negative correlation between teachers' evaluation of aspects, which are inherent only for computer-based learning materials, and students' learning outcomes.
2. There is a positive correlation between the evaluation of plainness of the content and the students' learning outcomes and between the evaluation of the effectiveness of the drill, compared with traditional practice, and students' learning outcomes.

## **Method**

### *Sample*

One-group quasi-experiment was carried out with 37 boys and 43 girls from four Estonian schools. The schools were different. One school was selected from the centre of the county, two schools were from the small towns and one was from the countryside. The students' groups were of mixed ability. High-achieving and low-achieving students were determined basing on the results of an achievement test. There were 18 low-achievers, 31 middle-achievers and 31 high-achieving students in the study. All the students had experience in learning with computers, and only 14% of the 3<sup>rd</sup> form students mentioned that their skills in using a computer were not good before the study.

Because drills are considered to be useful for learning basic math skills, foreign language, vocabulary and spelling (Trotter, 1998; Alessi & Trollip, 2001), the studies in basic maths and English as a foreign language were used in the experiment. In total, 15 drills about maths and 19 drills about English as a foreign language were used in the experiment. It was considered appropriate to test the students from the 3<sup>rd</sup> form (age 9–10), as, in Estonian schools, the multiplication tables are taught in the 3<sup>rd</sup> form, and many drills are used for learning multiplication tables and arithmetic operations. Also, English as a foreign language is taught from the 3<sup>rd</sup> form in Estonian schools, and mostly vocabulary is taught in this form.

Each piece of educational software was evaluated by ten primary teachers from the same schools, who had used these drills with their students so that the teachers saw how the students managed with the particular drill. All the participating primary teachers were female – there are very few male primary teachers in Estonia. The age of the participating primary teachers varied from 27 to 49 years ( $M=37.1$ ;  $SD=6.3$ ). All participating teachers rated their computer skills at least as fair, and all the participating teachers had graduated from at least one course related to the usage of computers in the classroom. All the teachers in the study had used computers in their classroom activities. Eight of participating teachers had used computer applications for preparing multimedia learning materials for their students.

### *Instruments*

The teachers' evaluations were gathered by questionnaires. The teachers evaluated, on a 10-point scale, 17 aspects of the drills: the effectiveness of the drill comparing with the traditional practice, plainness of the content, pleasantness of the drill, interest of the presentation of the learning material, simplicity of manipulating, design, sounds and colours of drill materials, fitness of feedback, child-friendliness, suitability for students' age, attractiveness of the drill materials, suitability for students' computer-skills, appropriateness of pace, playfulness and competition. The reliability of the questionnaire (Cronbach's alpha) was .87.

The learning outcomes of students were evaluated by means of tests, which were composed by experienced teachers who did not participate in this study. The tests were in two versions, and both forms of the test were in a paper-pencil format. A range of questions about basic skills (multiplication table or adding or subtracting, translating words from English to Estonian and from Estonian to English) were included in the tests. The tests were composed by teachers of the particular subject. As the tests differed in the number of items, the percentage of the student's score was calculated. The reliability (Cronbach's alpha) of the tests was .76-.86, and their validity was confirmed by the experts. The experts were two teachers of mathematics, two teachers of English as a second language and four class-teachers. The first expert-teacher of the particular subject and two class-teachers reviewed the tests and made corrections where needed. Then, the second expert-teacher of the particular subject and the other two class-teachers reviewed the tests again.

### *Procedure of the experiments*

The students were asked to accomplish pre-tests before studying the particular unit to determine their prior knowledge. After that, the students practiced particular skills using the multimedia software. All the students worked independently with 34 different drills. After completing their practicing, the students filled in the post-tests. All the students were asked to study all drills. Separately, the teachers gave rankings for the different aspects of the same 34 drills.

### *Data analysis*

Statistical package SPSS 11.5 for Windows was used for data analysis. The mean test scores of all students and the mean evaluations of the teachers were calculated in the case of each drill. Also, the mean test scores for the boys and girls and the high- and low-achieving students were calculated. The Pearson correlation analysis revealed a significant relationship between different pre-test and post-test scores. Due to the significant correlations, the co-variation analysis was used in order to calculate the mean adjusted post-test scores. In the co-variation analysis, the theme was as a factor, and the scores from the pre-test – as a covariant.

The data was analysed using correlation analysis in order to test the significance of the relationships between the different aspects of evaluations by teachers and learning outcomes by students. The results from both the mean adjusted post-test score and post-test score were used. The mean adjusted post-test score indicated an increase in the learning, because the pre-test score was accounted for in this indicator. Since assessment is taken according to a post-test rather than an increase in the learning in Estonia, the post-test scores were used as well. Also, the correlation coefficients between the teachers' mean evaluations and the learning outcomes by boys and girls and high- and low-achieving students were calculated.

## **Results**

The Pearson correlation coefficients between different aspects of mean evaluations by the primary school teachers and students' mean adjusted post-test scores are provided in Table 1.

Table 1. The Pearson correlation coefficients (r) between different aspects of mean evaluations by the primary teachers and students' mean adjusted post-test scores (MAPTS)

<b>Aspect of the evaluation by the primary teachers</b>	<b>r with the MAPTS all students</b>	<b>r with the MAPTS for boys</b>	<b>r with the MAPTS for girls</b>	<b>r with the MAPTS for high-achieving students</b>	<b>r with the MAPTS for low-achieving students</b>
Effectiveness of the drill comparing with the traditional practice	-.44**	-.40*	-.36*	-.23	.15
Plainness of the content	-.18	-.04	-.20	.30	-.27
Pleasantness of the drill	-.09	-.10	-.07	-.20	.11
Interest of presentation of learning material	-.09	.06	-.16	-.20	.12
Simplicity of manipulating	-.09	.04	-.14	.22	-.05
Design of drill	-.22	-.14	-.23	-.38*	.20
Sounds of drill materials	-.17	-.06	-.20	-.39*	.27
Colours of drills materials	-.25	-.24	-.20	-.41*	.29
Fitness of feedback	-.08	-.33	.08	-.26	.17

*Sequel to Table 1 see on p. 64.*

	<i>Sequel to Table 1.</i>				
Child's friendliness	-.18	-.22	-.11	-.18	.02
Suitability for students' age	-.35*	-.36*	-.25	.09	-.18
Attractiveness of drill	-.20	-.18	-.18	-.41*	.22
Suitability for students' computer-skills	-.19	-.33	-.06	-.25	.21
Appropriateness of pace	-.11	-.22	-.03	-.30	.38*
Appropriateness of playfulness	-.09	-.01	-.11	-.20	.11
Appropriateness of competition	.19	.02	.23	-.15	.05

\* Statistically significant at the 0.05 level

\*\* Statistically significant at the 0.01 level

If the teachers evaluated the effectiveness of the drill higher as compared with the traditional practice, all students', boys' and girls' learning increased (mean adjusted post-test) was lower. If the teachers evaluated the drill as more suitable for students' age, all students' and boys' learning increase was lower. If the high-achieving students received low learning increase, the teachers evaluated the appropriateness of design of the drill, appropriateness of sounds of drill materials, appropriateness of colours of drill materials and appropriateness of attractiveness of drill higher. If the teachers' evaluations about the appropriateness of pace were higher, the low-achieving students' learning increase was higher, too.

The Pearson correlation coefficients between different aspects of mean evaluations by the primary teachers and students' mean post-test scores are provided in Table 2.

Table 2. The Pearson correlation coefficients (*r*) between the different aspects of mean evaluations by the primary teachers and students' post-test scores (PTS)

Aspect of the evaluation by the primary teachers	<i>r</i> with the PTS all students	<i>r</i> with the PTS for boys	<i>r</i> with the PTS for girls	<i>r</i> with the PTS for high-achieving students	<i>r</i> with the PTS for low-achieving students
Effectiveness of the drill comparing with the traditional practice	-.23	-.17	-.24	-.33	-.29
Plainness of the content	.21	.20	.18	.30	.35
Pleasantness of the drill	-.07	-.32	.09	-.20	-.02
Interest of presentation of learning material	-.33	-.40*	-.24	-.20	-.17
Simplicity of manipulating	-.00	-.03	.01	.22	.12
Appropriateness of design of drill	-.44**	-.39*	-.42*	-.38*	-.20
Appropriateness of sounds of drill materials	-.43*	-.32	-.45**	-.39*	-.29

*Sequel to Table 2 see on p. 65.*

*Sequel to Table 2.*

Appropriateness of colours of drill materials	-.49**	-.37*	-.50**	-.41*	-.26
Fitness of feedback	-.05	.04	-.10	-.26	-.11
Child's friendliness	-.06	-.19	.02	-.18	-.01
Suitability for students' age	.15	.02	.21	.09	.26
Appropriateness of attractiveness of drill	-.49**	-.36*	-.51**	-.41*	-.19
Suitability for students' computer-skills	-.33	-.26	-.32	-.23	-.20
Appropriateness of pace	-.20	-.26	-.18	-.31	-.30
Appropriateness of playfulness	-.25	-.16	-.28	-.23	-.13
Appropriateness of competition	.08	-.03	.13	-.17	.04

\* Statistically significant at the 0.05 level

\*\* Statistically significant at the 0.01 level

There were no positive correlations between any students' post-test scores and teachers' evaluations. If the teachers evaluated appropriateness of design of drill, appropriateness of colours of drill materials and appropriateness of attractiveness of drill higher, all students, boys, girls and high-achieving students got fewer points in post-tests. The teachers' evaluations about the appropriateness of sounds of drill materials were negatively related to the post-test scores of all students, girls and high-achieving students. If the teachers evaluated the interest of presentation of learning material higher, the boys' post-test scores were lower.

## Discussion

Despite the fact that primary teachers spend almost all of the school-day with their students and for that reason should know their students and should be able to choose suitable educational software for them (Mei Mei-Yan, Walker, & Huang, 1999), this study indicated the opposite result. All the teachers' evaluations, which were significantly related to the students' post-test scores, were negatively correlated. Besides the significant correlations between the teachers' evaluations and the students' mean adjusted post-test scores, there was only one positive correlation coefficient – if the teachers evaluated appropriateness of pace higher, the low-achieving students' learning increase was higher, too. This meant that if the teachers were of the opinion that the particular multimedia learning materials were good for some aspect of learning, the students gained lower results working with this drill and the opposite when the teachers were of the opinion that the particular learning material was poor for some aspect of learning.

The teachers also evaluated the effectiveness of a particular multimedia drill compared with traditional practice. This evaluation more directly indicated the effectiveness of the multimedia drill in the opinion of teachers. Unfortunately, this evaluation was negatively correlated with the mean adjusted post-test score of all students and for boys and girls. The

teachers were not able to recognise more effective drills. Correlations of the post-test scores with the evaluations by teachers on the effectiveness of a particular drill, compared with the traditional practice, was not significantly different. There was no reason to believe that if the primary school teachers chose multimedia learning materials instead of the traditional practice, the students would achieve higher learning results.

The only positive correlation coefficient was between the evaluation of the appropriateness of pace and mean adjusted post-test score of low-achieving students. There were no statistically significant correlations between the teachers' evaluation of appropriateness of pace and mean adjusted post-test score in the case of other groups of students (all students, boys, girls and high-achieving students). Perhaps the teachers considered mostly the pace when selecting materials, which might be suitable for low-achieving students, because in class teachers usually take into account the learning pace of the low-achieving students.

Most of the negative correlation coefficients were in the case of the aspects describing the appearance and attractiveness of the drill materials. If the teachers evaluated the appropriateness of attractiveness of drill materials more highly, the students achieved lower results working with this drill. The appropriateness of design, sound and colours were overrated by the teachers and were ineffective for the students as well. As the teachers' evaluations for the appropriateness of design, sound and colours were strongly correlated with the evaluation of the appropriateness of attractiveness of the drill materials (correlation coefficients accordingly  $r=.92$ ,  $r=.65$  and  $r=.89$ , all  $p<.01$ ), it might be appropriate to conclude that the teachers evaluated these three aspects according to the attractiveness. The teachers evaluated more colourful drills with multimedia effects and sounds and did not discern the risks of the attractiveness for students.

The evaluations of the appropriateness of design, sound, colours and attractiveness of drill were negatively correlated with the post-test scores in the case of all students, boys, girls and high-achieving students, but there were no statistically significant correlations with the post-test scores for low-achieving students. But in the case of high-achieving students, the teachers' evaluations of the appropriateness of design, sound, colours and attractiveness of drill were negatively correlated both with the post-test scores and with the mean adjusted post-test scores. The reason might be that attractive characteristics motivated the low-achieving students and therefore they achieved higher results with more attractive drills, but the same characteristics hindered the learning for the high-achieving students (Luik, 2009). Also, Van den Bergh and Vrana (1998) declared that increasing fluency might increase boredom of repetition.

Also, the evaluations related to the interest of presentation of learning material were negatively correlated with the boys' post-test scores. Maybe as all the primary teachers were women, they did not know which presentation of learning material was interesting for boys and would increase the boys' learning outcomes.

The teachers' evaluation of the suitability of multimedia learning material for the students' age was negatively correlated with all students and boys' mean adjusted post-test scores. So if the teachers evaluated the drill as more suitable for the students' age, the students achieved lower learning outcomes.

An interesting result was that the teachers' evaluations about the plainness of the content, simplicity of manipulating, fitness of feedback and suitability for the students' computer-skills were not related to any students' test scores. So there was no reason to believe that if the primary school teachers evaluated the multimedia learning material higher, the students would get higher results. Or if the teacher considered the particular multimedia material low, the students would not get good results working with this material. Except the plainness of the content, all other mentioned aspects were characteristics for only the computer-based learning materials. Therefore it might be concluded that the teachers did not recognise the effective characteristics of computer-based possibilities and the students' computer skills.

## **Conclusion**

There were negative correlations between the 'appropriateness of presented sound' and the students' learning outcome, but there were not statistically significant correlations between the evaluation of aspects, which were inherent only for the computer-based learning materials, like learner control and feedback. Also, the evaluation of the plainness of the content was not related to the students' learning outcome. But the most important finding was that when the teachers evaluated the drill as more effective than traditional practice, the students achieved lower learning results. So the present study revealed that the teachers were not able to recognise efficient multimedia learning materials.

Vrasidas and McIsaac (2001) wrote that teacher education programmes play a crucial role for technology-based teaching and learning across the various disciplines. Today's teachers should be able, besides utilising other teaching skills, to utilise instructional technology, particularly computer-based technologies. The teachers who participated in the study were educated in the field of computer literacy, but they were not able to discern effective multimedia learning materials. Noting this, there was no reason to suppose, that they were able to compose efficient multimedia learning materials (PowerPoint presentations, educational web-sites) themselves as well. Therefore courses for pre-service teachers on evaluating and designing efficient multimedia learning materials were needed in teacher education. Because information and communication technologies (ICT) reform is in progress and new applications appear every year, teaching of ICT skills is needed to insure in-service teachers' sustainable development.

## **References:**

- Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning. Methods and development* (3<sup>rd</sup> ed.). Boston: Ally and Bacon.
- Boyle, T. (1997). *Design for multimedia learning*. Essex. England: Pearson Education Limited.
- Bos, M. (2003). Experienced teachers' expectations about the potential effectiveness of spelling exercises. *Annals of Dyslexia*, 53, 104–127.

- Diederens, J., Gruppen, H., Hartog, R., & Voragen, A. G. J. (2005). Design and evaluation of digital learning material to support acquisition of quantitative problem-solving skills within food chemistry. *Journal of Science Education and Technology*, 14(5/6), 495–507.
- Herring, D. F., Notar, C. E., & Wilson, J. D. (2005). Multimedia software evaluation form for teachers. *Education*, 126(1), 100–111.
- Higgins, K. (2000). Evaluating educational software for special education. *Intervention in School & Clinic*, 36(2), 109–115.
- Hughes, I. E. (1998). “Horses for courses” – categories of computer-based learning program and their uses in pharmacology courses. *Information Services & Use*, 18(1/2), 35–44.
- Jacobson, E. (2006). Computer homework effectiveness in developmental mathematics. *Journal of Developmental Education*, 29(3), 2–8.
- Kurz, T. L., & Middleton, J. A. (2006). Using a functional approach to change pre-service teachers’ understanding of mathematics software. *Journal of Research on Technology in Education*, 39(1), 45–65.
- Luik, P. (2007). Characteristics of drills related to development of skills. *Journal of Computer Assisted Learning*, 23(1), 56–68.
- Luik, P. (2009). Effective characteristics of learning multimedia. In L. T. W. Hin & R. Subramaniam (Eds.), *Handbook of research on new media literacy at the K-12 level: Issues and challenges* (pp. 167–188). New York, London: IGI Global.
- Mei-Yan, L., Walker, D. F., & Huang, J. (1999). Do they look at educational multimedia differently than we do? A study of software evaluation in Taiwan and the United States. *International Journal of Instructional Media*, 26(1), 31–42.
- Mikk, J., & Luik, P. (2003). Characteristics of multimedia textbook that affect post-test scores. *Journal of Computer Assisted Learning*, 19(4), 528–537.
- Ngu, B. H., & Rethinasamy, S. (2006). Evaluating a CALL software on the learning of English prepositions. *Computers & Education*, 47(1), 41–55.
- Nurmi, S., & Lehti, S. (2003). European teachers’ expectations and opinions about learning objects. *Paper presented in the Symposium “Designing Virtual Learning Materials” at the EARLI 2003 Conference*. August 29, 2003; Padova, Italy.
- Phillips, R. (1997). *The developer’s handbook to interactive multimedia. A practical guide for educational applications*. London: Kogan Page
- Rana, P. K. (2002). Evaluation of Web-Based Learning Systems used in the primary education sector. Unpublished Master’s thesis, Brunel University, London, United Kingdom. Retrieved May 15, 2004, from [http://disc.brunel.ac.uk:8080/modres\\_uploads/CS5007B/ci01pk.pdf](http://disc.brunel.ac.uk:8080/modres_uploads/CS5007B/ci01pk.pdf)
- Trotter, A. (1998). *Teaching the basics. Education week on the web*. Retrieved July 20, 2004, from <http://www.edweek.org/sreports/tc98/cs/cs1.htm>
- Van den Bergh, O., & Vrana, S. R. (1998). Repetition and boredom in a perceptual fluency/attributional model of affective judgements. *Cognition and Emotion*, 12(4), 533–553.

Vrasidas, C., & McIsaac, M. S. (2001). Integrating technology in teaching and teacher education: Implications for policy and curriculum reform. *Educational Media International*, 38(2/3), 127–132.

**Correspondence:**

Dr Piret Luik, Faculty of Social Sciences and Education, Institute of Education, University of Tartu, Salme 1a, Tartu 50103, Estonia. Tel: + 372 737 6162; Fax: +372 737 6190. Email: piret.luik@ut.ee