



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Development and Validation of Mathematics Courseware Usefulness Evaluation Instrument for Teachers

¹N. Sahari, ²A.A. Abdul Ghani, ²H. Selamat and ³A.S. Md. Yunus

¹Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

²Faculty of Computer Sciences and Information Technology, Universiti Putra Malaysia, Malaysia

³Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia

Abstract: The primary purpose of this study was to develop an instrument for evaluating the usefulness of mathematics courseware and to provide psychometric evidence of validity and reliability. Throughout four phases of this study, 696 participants were involved and six types of MCs were used. Each participant was required to evaluate the courseware heuristically and complete the Mathematics Courseware Usefulness Evaluation Instrument (MCUE). Based on the theoretical perspective, a hypothesized model with three factors which were usability, functionality and efficiency, five sub-factors and seven criteria were proposed. This study presents some empirical evidence of whether the evaluation model and its underlying metrics are reliable and valid for determining the usefulness of Mathematics Courseware. A pool of evaluation metrics were collected based on the MC preliminary evaluation survey, related articles on MC reviews and based on several existing evaluation instruments. Through several experiments, we validate an eight-dimension usefulness attribute involving Ease of use, Attractiveness, User control, Concept presentation, reinforcement, assessment, accuracy and learning support material and 56 metrics.

Key words: Usefulness model, usefulness metrics, construct validity, exploratory factor analysis, criterion-related validity

INTRODUCTION

Mathematics Courseware (MC) packages encourage learners to use them effectively if they are designed appropriately. In addition to text and reference books in the markets and with a large number of instructional software titles available, students find it hard to choose the right learning material. Choosing a good courseware has become a main concern in learning with Computer Aided Instruction (CAI) by educational stakeholders. When considering whether or not to use a CAI material, it is recommended that teachers carry out a full evaluation to reveal whether or not the developers have tackled a subject in an appropriate way. However, with the vast development of MC, teachers have problems to select the right one (Herring *et al.*, 2005; Yushau *et al.*, 2004). Evaluation of MC with a valid and reliable instrument must be conducted in order to select the suitable courseware.

There are two types of evaluation; formative and summative evaluation. Formative evaluation is an ongoing process conducted along every step of the development process with the aim of obtaining data to increase the effectiveness of the design and has become an essential part in software development. Summative evaluation, on

the other hand, is an evaluation performed on the finished product in real situations so as to determine that the software is suitable for the job it has been designed to do. Shiratuddin and Landoni (2002) mentioned that when evaluating new under development or modifying existing educational software, formative evaluation is performed. While, summative evaluation is used to select, implement and guide decision whether to continue using the software. The methods that are used to conduct summative evaluations are survey methods, which are interviews and questionnaire. Since interviews may incur higher cost and in many cases the interviewees opinion might affect respondents issues of interest, questionnaire-based surveys are more popular.

There are several questionnaires available in the literature. Most of them measure specific elements of software. For example, Nielsen's attributes of Usability (NAU), Nielsen's Heuristic Evaluation (NHE) (Nielsen, 1994) and Practical Heuristic for Usability Evaluation (PHUE) (Perlman, 1997) measure usability attribute of user interface by heuristic. These instruments measure the global system usability. Some instruments measure the effectiveness of specific functions of a system and how the system supports users in completing their tasks such as Perceived Usefulness and Ease of Use (PUEU) by

Davis and Venkatesh (2004) and Computer System Usability Questionnaire (CSUQ) by Lewis (1995). Those instruments are suitable for Human Computer Interaction (HCI) experts.

Most software is developed for real users. Thus, users have their right to choose software that satisfies their needs. There are several instruments that measure end-user computing satisfaction which are, End-user adoption of animated interface (Serenko *et al.*, 2007), Questionnaire for User Interface satisfaction (QUIS) (Huang *et al.*, 2004), User Interface Rating Form (UIRF) (Cartwright, 2006), Computer User Satisfaction Inventory (CUSI) (Jordan, 2000) and After Scenario Questionnaire (ASQ) (Lewis, 1995). In general those instruments evaluate the users reaction to a specific computer interface, system terminology and information, affect and competence of a computer system. Usability Measurement Inventory (SUMI) is the only available questionnaire for the assessment of usability of software, which has been developed, validated and standardized on European wide basis (Van Veenendaal, 1999).

To measure educational software, the above instruments could be used to measure the usability and accessibility of the software. At the same time it is important to consider the learning effectiveness factor (Thurber *et al.*, 2002; Elisavet and Economides, 2003) in developing and evaluating educational software. Thurber *et al.* (2002) have developed Mathematics Curriculum-based Measurement (M-CBM) to measure mathematics computation, application and reading skills. Research by Elisavet and Economides (2003) produced evaluation instrument for hypermedia courseware. This instrument consists of four main factors, which are, evaluation of the content, organization and presentation of the content, technical support and update process and evaluation of learning. The questionnaire has 111 questions that can be used to evaluate web-based and stand alone educational packages. This instrument is suitable for expert evaluators such as developers, HCI experts and instructional designers.

There is variance among the above instruments in the naming, variety and completeness of their dimension. However there are common elements that are frequently used namely user satisfaction, usability and curriculum-based *per se*. Although evaluation of educational effectiveness is the primary area with which developers of learning software must concern (Furner and Daigle, 2004; Nisanci, 2000), most previous studies, however, have focused on the multimedia or courseware design adequacy. Software for learning might be usable but not educational, or vice versa. The final goal must be for systems to be both usable and educational. Only a few instruments have examined the quality of both educational and HCI adequacy. Moreover, developers usually design their own evaluation instruments in order

to evaluate the courseware that they have developed. Only a few employed systematic procedure to develop and validate an instrument for educational software (Ou-Yang, 1997; Van Schaik and Ling, 2005).

The most important factor to consider in selecting a courseware is the external quality factor. Students, that are, users, tend to use or not use a courseware according to the extent to which they believe the courseware is useful to them. The usefulness attribute means that using the courseware helps them to understand their mathematical concepts better and engages them in learning activities. Since teachers are more aware of their students needs, they are the most reliable group to determine whether a mathematics courseware is suitable for their students learning. Therefore, this study focuses on the development of a reliable and valid evaluation instrument to measure the usefulness of MC by teachers. The instrument provides teachers with external quality elements, which are pedagogical adequacy as well as Human-Computer Interaction (HCI) elements.

MATERIALS AND METHODS

When constructing evaluation tools, there are test validation theories that must be considered. There are types of validity; content, construct, criterion and reliability.

The development of the usefulness instrument was conducted in a four-phase study, as shown in Fig. 1. Throughout this study, six types of high school MCs were used. The MCs were chosen from different

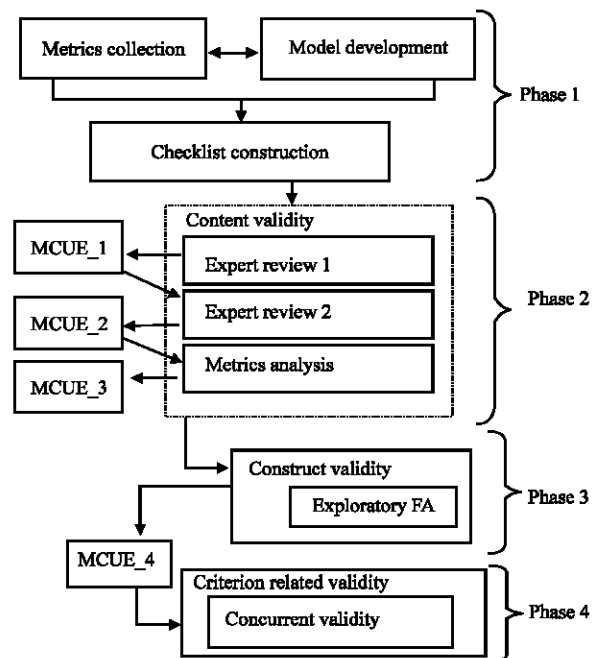


Fig. 1: Methodology of the study

Table 1: No. of participants

Phase	Participants
1	2 mathematics lecturers 3 courseware developers
2	10 expert reviewers (8 mathematics lecturers and 2 assistance director of Curriculum Development, Malaysian Ministry of Education) 35 mathematics teachers
3	620 mathematics teachers
4	26 pre-service teachers

developers from local and other countries. In this study a total of 696 participants were involved (Table 1).

Each participant was required to evaluate the courseware heuristically and to complete the instrument in the form of either checklist or questionnaire. The first phase of the study was to construct a checklist from the mapping of the metrics collection and the developed usefulness model. In the second phase, the checklist was reviewed by experts and teachers for content validity check. Construct validity which was Exploratory Factor Analysis (EFA) was done in the third phase. The main idea of doing this was to explore the metrics in the instrument. Within factor analysis, the orthogonal rotation using varimax was used because results of this method are more likely to be replicated in future studies. Moreover, It is believes that more than one factors or dimensions accounts for the variance in the data (Hair *et al.*, 1998). Finally, in the fourth phase the criterion validity which is concurrent validity was tested.

USEFULNESS EVALUATION MODEL

The first phase of the study applied an in-depth literature review to obtain an evaluation model for usefulness quality attribute of a mathematics courseware that could be used by teachers. The usefulness evaluation model is constructed based on the study of McCall FCM model, Boehm model and ISO 9126 (Ortega *et al.*, 2000) and Nielsen (1993) acceptance model.

The model emphasizes to the related attributes to measure mathematics courseware before acceptance, which is an evaluation before it is used by students. As shown in Fig. 2, usefulness attribute is divided into two sub-attributes, which are utility and usability. Utility refers to teachers perception of MC educational adequacy and measured by functionality and efficiency factors. Functionality factor includes the suitability criterion as a tutorial material and the curriculum accuracy of MC. Suitability criterion is measured by concept presentation, reinforcement and assessment. While, efficiency factor of a MC is measured through the availability of its learning support materials.

Usability, in the context of MC, is related to technical adequacy that engages students in using the courseware, which is ease of use and attractiveness. Ease of use criterion is measured primarily by its friendliness,

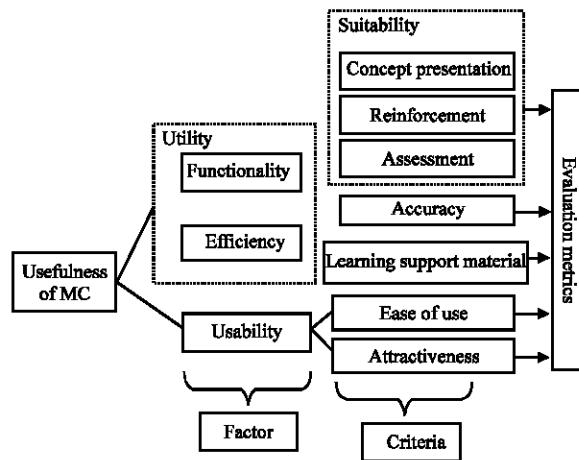


Fig. 2: Usefulness evaluation model

intuitiveness and ease of navigation. Attractiveness refers to MC user interface, readability, appearance and visual design. All usefulness criteria would be measured by evaluation metrics.

The procedures to generate a candidate list of metrics from the domain of all possible metrics representing the constructs are discussed below:

USEFULNESS METRICS

A pool of evaluation metrics were collected based on MC preliminary evaluation survey where three developers and two mathematics lecturer who were directly involved in developing and evaluating courseware were asked to list down educational and technical factors and criteria of good mathematics courseware based on open ended questions given. In addition to that, other metrics were collected from related articles on MC reviews (Cook, 2002; Hawkes, 2000; Lawson, 2000) and based on several existing evaluation instruments (Ellisavet and Economides, 2003; Storey *et al.*, 2002; Yushau *et al.*, 2004). The collection of statements metrics were judged and assigned to a tentative checklist. Ultimately, the preliminary survey and literature-based produced a checklist with 85 metrics consisting of three factors, which were functionality, usability and efficiency and five sub-factors which were suitability, accuracy, ease of use, attractiveness and learning support materials. The numbers of metrics to the corresponding criteria are shown in Table 2.

In the next phase, 10 panel members or experts were asked to review this 85-metrics instrument through two rounds of content validity. The checklist was sent to eight experts, who were lecturers from three local universities for the first round of content validity. The experts were

assigned to explore three mathematics courseware in order to identify any metrics they deemed relevant to the dimensions. The experts were encouraged to add or drop metrics and dimensions.

Generally all experts agreed that the metrics were comprehensive and relevant to all criteria and factors. Feedbacks from the panel such as, redundant and not measurable metrics, unclear and confusing sentences were reviewed. Due to recommendations from the panel, 19 metrics were eliminated, 12 metrics were modified, 53 metrics remained and one metric added. Eventually a total of 66 metrics were obtained. The experts supported the factors and criteria for usefulness evaluation. The 66 metrics sentences were revised and formed a set of questionnaire with five-point-Likert-scale with response choices ranging from 1-5 representing

strongly disagree, disagree, quite agree, agree and strongly agree respectively. This instrument is called a MC usefulness evaluation instrument version 1 (MCUE_1).

In the second round content validity check, MCUE_1 was sent out to four experts (two of them were involved in the first review) to review the appropriateness of its format. The panel reviewed on operational issues such as metrics sentence clarity and sequence, as well as relevant metric to its criteria and factor. Based on the panel judgments, two metrics were modified and one metric was added to the accuracy criterion. Consequently the revised questionnaire was based on 67-metrics and was called MCUE_2.

Next MCUE_2 was administered to 35 mathematics teachers. To determine the internal consistency of the instrument, analysis of reliability using Alpha Cronbach indices were used. The reliability of each criterion is; suitability (0.860), accuracy (0.867), ease of use (0.835), attractiveness (0.917) and efficiency (0.959). These values indicate that the reliability of the instrument is good (Kitchenham *et al.*, 2002; Konting, 2004). From Table 3 an examination of the metrics comprising the

Table 2: Criteria and metrics of the checklist

Factor	Sub-factors	Metric
Functionality	Suitability	1-23
	Accuracy	24-38
Usability	Ease of use	39-50
	Attractiveness	51-63
Efficiency	Learning support materials	64-85

Table 3: Metrics reliability

Metric	Corrected metric-total correlation	Alfa Cronbach if metric deleted	Metric	Corrected metric-total correlation	Alfa Cronbach if metric deleted
Suitability (0.860)			Ease of Use (0.835)		
a1	0.579	0.850	c32	0.647	0.806
a2	0.380	0.856	c33	0.452	0.829
a3	0.314	0.859	c34	0.509	0.823
a4	0.122	0.865	c35	0.757	0.786
a5	0.489	0.852	c36	0.795	0.785
a6	0.493	0.852	c37	0.434	0.832
a7	0.561	0.849	c38	0.290	0.846
a8	0.611	0.847	c39	0.629	0.807
a9	0.668	0.847	Attractiveness (0.917)		
a10	0.636	0.846	d40	0.696	0.911
a11	0.521	0.850	d41	0.691	0.910
a12	0.140	0.862	d42	0.500	0.916
a13	0.579	0.848	d43	0.714	0.909
a14	0.628	0.846	d47	0.735	0.909
a15	0.193	0.868	d48	0.577	0.914
a16	0.213	0.862	d49	0.657	0.911
a17	0.602	0.846	d50	0.595	0.913
a18	0.617	0.846	d51	0.631	0.912
a19	0.494	0.852	d52	0.682	0.910
a20	0.312	0.858	d53	0.520	0.919
Accuracy (0.867)			Learning support materials (0.891)		
b21	0.610	0.854	e54	0.708	0.877
b22	0.699	0.845	e55	0.599	0.884
b23	0.715	0.845	e56	0.703	0.877
b24	0.613	0.852	e57	0.689	0.878
b25	0.576	0.855	e58	0.578	0.888
b26	0.556	0.858	e59	0.756	0.876
b27	0.535	0.857	e60	0.712	0.877
b28	0.497	0.860	e61	0.526	0.885
b29	0.519	0.859	e62	0.499	0.886
b30	0.491	0.869	e63	0.349	0.891
b31	0.661	0.854	e64	0.479	0.887
d44	0.611	0.912	e65	0.525	0.885
d45	0.795	0.907	e66	0.642	0.881
d46	0.704	0.910	e67	0.338	0.892

attribute of usefulness scale indicates that metrics a4, a12, a15 and a16 have the lowest corrected metric-total correlations. This indicates that those metrics should be reviewed or eliminated (Ebel and Frisbie, 1986). If these four metrics were removed from the scale, the alpha if item deleted column shows that overall reliability would increase slightly. Therefore deletion of these metrics considered appropriate. From the analysis, metric a4 was modified and three others were eliminated and the improved instrument MCUE_3 with 64 metrics was formed. MCUE_3 was then tested for construct validity.

CONSTRUCT VALIDITY

Exploratory Factor Analysis (EFA) was employed in order to determine and validate the construct. MCUE_3 was administered to 626 mathematics teachers in five regions. In this study, Peninsular Malaysia is divided into four regions (Southern, Western, Northern and Eastern) and East Malaysia is considered one region. One state was chosen randomly from each region. Schools in each region were selected randomly and all mathematics teachers in the schools involved in the testing. They were required to explore a MC and completed the MCUE_3 questionnaire.

Based on their responses, factor analysis was carried out to determine the dimensionality of the construct. In the quest for a stable factor structure, an iterative procedure that began with submitting the items to a factor analysis procedure with varimax rotation was employed. Selection number of factors to retain was based on metric loading and eigenvalue. Since our goal is to examine the most important loadings in interpreting the factor solution, we decided to use a cut-off point of 0.40 for item loadings and eigenvalue greater than 1.00 as suggested by Hair *et al.* (1998).

Prior to conducting the exploratory factor analysis, two indicators were examined in order to determine

whether the sample was appropriate for such an analysis. The Kaiser-Meyer-Olkin measure of sampling adequacy index was 0.962 which was very good and Barlett's test of sphericity was significant, $\chi^2(N = 620) = 2016$ and $p < 0.0001$ indicating that sample and correlation matrix were adequate for analysis (Field, 2005). At the end of the factor analysis procedure, a total of nine metrics deleted and 56 metrics remained. The factor analysis revealed 10 factors with eigen values of more than one (23.343, 3.114, 2.964, 2.638, 1.695, 1.297, 1.268, 1.192, 1.128 and 1.053). The procedure resulted in a 56 metrics instrument that accounted for 62.018% of the variance.

Next, metrics that belong to each factor employed were examined. Factors 1, 2 and 10 were related to usability criteria, which were named Ease of use (10 metrics), Attractiveness (8 metrics) and User control (3 metrics), respectively. Factors 4, 5, 7, 8 and 9 were related to mathematics teaching and learning strategy. Factors 4, 5 and 7 were named Concept Presentation (7 metrics), Reinforcement (5 metrics) and Assessment (5 metrics), respectively. Factors 8 and 9 were combined and called curriculum accuracy (8 metrics). Metrics from Factor 3 and 6 were combined and formed Learning Support Materials factor (10 metrics). After adding user control criteria to usability factor, as a result there were eight usefulness dimensions not seven as proposed earlier. Another round of reliability tests resulted in alpha values of eight criteria, which were Ease of use (0.909), Attractiveness (0.913), user control (0.801), concept presentation (0.871), reinforcement (0.868), assessment (0.788), accuracy (0.837) and learning support materials (0.866). All factors were highly reliable (Konting, 2004). Table 4 presents the factor loading of the metrics and their corresponding criteria and factors prior to EFA. MCUE_4 with 56 metrics was formed.

Table 4: Metrics and factor loadings

Metrics and loadings	Values	Metrics and loadings	Values
Ease of use		Reinforcement	
1. Icons are meaningful	0.761	29. Multiple examples are provided	0.691
2. Command buttons are clear	0.741	30. Problems enrich critical thinking skill	0.672
3. Active buttons are clearly highlighted	0.733	31. Problems enrich creative thinking skill	0.643
4. Exit or main menu is clearly provided	0.718	32. Multiple problem solving techniques are illustrated	0.624
5. User interface layout help navigation easier	0.689	33. Problems with random parameters are provided	0.572
6. Button with tool tips provided	0.606	Assessment	
7. Computer provides system status to user	0.536	34. Answers can be checked by the system	0.583
8. User controls the sequence of presentation	0.587	35. Self-assessment can be accessed any time	0.563
9. Good interactive control such as scroll bar, edit box and animation button is used	0.482	36. System generates a test report	0.501
10. Help button is available in each screen	0.542	37. Practice problems provide step-by-step worked out solution	0.458
Attractiveness		38. Formative assessment meets learning objectives	0.466
11. Screen background is attractive	0.798	Learning resource materials	
12. Graphics enhance the instructional effect	0.785	39. Calculator is provided	0.821
13. Videos enhance the instructional effect	0.741	40. Graph for mathematical equation is illustrated	0.769
14. Animated illustration is good	0.733	41. Additional reference materials are included	0.751
		42. Visualization power is utilized effectively	0.628

Table 4: Continued

Metrics and loadings	Values	Metrics and loadings	Values
15. Combination of color is appropriate	0.713	43. Instructional games activities are provided	0.617
16. Texts are readable	0.672	44. Feedbacks are provided in appropriate manner	0.537
17. Background music is good	0.615	45. Calculating power of the computer enhances the enjoyment and understanding of concepts	0.531
18. Screen area is used effectively	0.544	46. Experiment and exploration of mathematical idea are interesting	0.493
User control		47. Instructional activities for student with multiple intelligences are provided	0.462
19. User can skip video or animation	0.700	48. Learning tips are provided	0.450
20. Screen layout is consistent	0.568	Mathematics curriculum accuracy	
21. Shortcut is provided for expert user	0.584	49. Mathematics curriculum for secondary school is met	0.770
Concept presentation		50. Mathematics terms are defined correctly	0.745
22. Interactive example is provided for hands-on exploration of a mathematical concept	0.620	51. Good values are embedded indirectly	0.703
23. Activities provide understanding and are meaningful toward formula, theorem and concept	0.583	52. Induction set is presented creatively and simulative to attract students	0.637
24. Examples with visualization are provided		53. Concept is introduced with an induction set	0.629
25. Mathematics concepts are presented with appropriate graphics	0.565	54. Mathematics terms are used correctly	0.575
26. Concepts meet mathematics' topics	0.553	55. Good values of Malaysian society are illustrated	0.575
27. The depth of concept is sufficient	0.541	56. Mathematics concept is presented together with its application	0.523
28. Mathematics experiment and exploration provide information about definition, theorem and the sense of self-discovery	0.503		

Table 5: Correlation of MCUE_4 and IECES

Variable	Pearson correlation	Sig (2-tailed)
IECES Score versus MCUE_5Score	0.742	0.000

CRITERION-RELATED VALIDITY

There are two types of criterion-related validity; predictive and concurrent validity. The first refers to how strongly the scores on the test are related to other behaviors, while the later interested in the relationship between two tests that purport to measure the same domain. In this study concurrent validity is applied. MCUE_4 were compared to a similar instrument called Instrument for Evaluating Chinese Educational Software (IECES) (Ou-Yang, 1997). IECES is a validated instrument with 10 criteria and can be used to measure the external quality of any courseware. Both instruments were used by 26 pre-service mathematics teachers for evaluating a mathematics courseware. Table 5 shows that there are positive correlation of both scores. It supported the criterion validity of MCUE_4 and thus completed validity tests.

DISCUSSION

A key goal of this usefulness instrument development studies is to establish usefulness dimensions through quantitative research involved content validity, reliability test, construct validity and criterion validity. The focus is entirely on the construction of an instrument which is able to yield valid and reliable scores of usefulness dimensions.

In the constructs and metrics selection phase seven criteria or dimensions were identified which were concept presentation, reinforcement, assessment, accuracy, ease of use, attractiveness and learning support material.

Evidence with regards to content validity indicated that the metrics in each dimension can appropriately represent the seven criteria of usefulness measurement as defined in this study. Nevertheless after conducting exploratory factor analysis and after examining the metrics that belong to each criterion, eight criteria seem adequate. After adding user control criterion in the usability factor, the instrument then yield eight dimensions with good reliability.

Finally the criterion validity test was conducted and the result showed that this instrument is comparable positively to a similar existing instrument. The findings of the study lead to the conclusion that MCUE which was developed based on the comprehensive review of literature and the content expert validation processes, was a reliable and valid instrument for measuring the usefulness of mathematics courseware.

CONCLUSION

In conclusion, from a raw collection of 85 metrics the study successfully identified 56 valid and reliable metrics of usefulness attribute of quality. Through a thorough identification process involving four phases of study, the metrics identified offered a valid and reliable measurement of key factors that can be served as usefulness indicators for the evaluation of MC by teachers. Hopefully, with this instrument, teachers might be able to choose the right courseware for their students. Besides using as a summative evaluation tool, by examining the factor loadings, developers could use MCUE as guidelines to ensure that mathematics courseware developed hereafter meet students and teachers needs. In addition to that, this instrument can be used by developers in refining existing courseware.

REFERENCES

- Cartwright, W.E, 2006. Exploring games and gameplay as a means of accessing and using geographical information. *Hum. IT*, 8: 28-67.
- Cook, J., 2000. Review of MathAid Precalculus v15.33. *MSOR Connections*, 12: 41-44.
- Davis, F.D. and V. Venkatesh, 2004. Toward preprototype user acceptance testing of new information systems: Implications for software project management. *IEEE Trans. Eng. Manage.*, 51: 31-46.
- Ebel, R.L. and D.A. Frisbie, 1991. *Essentials of Educational Measurement*. 5th Edn., Prentice Hall, New Jersey, ISBN-10: 0132846136.
- Ellisavet, G. and A. Economides, 2003. An evaluation instrument for hypermedia courseware. *Educ. Technol. Soc.*, 6: 31-44.
- Furner, J.M. and D. Daigle, 2004. The educational software/websites effectiveness survey. *Int. J. Instr. Media*, 31: 61-77.
- Hair, J.F. Jr., R.E. Anderson, R.L. Tatham and W.C. Black, 1998. *Multivariate Data Analysis*. 5th Edn., Prentice Hall, Englewood Cliffs, N.J., ISBN: 0-130-41146-9.
- Hawkes, T., 2000. Review of algebra interactive. *MSOR Connections*, 1: 27-30.
- Herring, D.F., C.E. Notar and J.D. Wilson, 2005. Multimedia software evaluation for teachers. *Education*, 126: 100-110.
- Huang, J.H., C. Yang, B.H. Jin and H. Chiu, 2004. Measuring satisfaction with business-to-employee systems. *Comput. Hum. Behav.*, 20: 17-35.
- Jordan, P.W., 2000. *Designing Pleasurable Products: An Introduction to the Human Factors*. 1st Edn., CRC Press, London, ISBN: 978-0-415-29887-2.
- Kitchenham, B.A., S.L. Pfleeger, L.M. Pickard, P.W. Jones, D.C. Hoaglin, El Emam and K. Rosenberg, 2002. Preliminary guidelines for empirical research in software engineering. *IEEE Trans. Software Eng.*, 28: 721-734.
- Konting, M.M., 2004. *Kaedah Penyelidikan Pendidikan*, Kuala Lumpur: Dewan Bahasa Dan Pustaka. 1st Edn., Kuala Lumpur, Malaysia, ISBN: 9836214046.
- Lawson, D., 2000. Ride the rollercoaster with the mathwise calculus cluster. *MSOR Connections*, 1: 25-30.
- Lewis, J.R., 1995. IBM computer usability satisfaction questionnaire. Psychometric evaluation and Instructions for use. *Int. J. Hum. Comput. Interaction*, 7: 57-78.
- Nielsen, J., 1993. *Usability Engineering*. 1st Edn., Academic Press Inc., Boston, Ma.
- Nielsen, J., 1994. Heuristic Evaluation. In: *Usability Inspection Methods*, Nielsen, J. and R.J. Mack (Eds.). John Wiley and Sons, New York, ISBN: 0-471-01877-5 pp: 25-61.
- Nisanci, M., 2000. Instructional software evaluation Criteria Used by the Teachers: Implications from theory to practice. *Society for Information Technology Teacher Edu. International Conference: Proceedings of SITE 2000*, Feb. 8-12, San Diego, California, pp: 6-6.
- Ortega, M., M. Pérez and T. Rojas, 2000. Construction of a systemic quality model for evaluating a software product. *Software Q. J.*, 11: 219-242.
- Ou-Yang, Y., 1997. The development and validation of the Instrument for Evaluating Chinese Software (IECES). Ph.D Thesis, Ohio State University, Columbus.
- Perlman, G., 1997. Practical usability evaluation. *Conference on Human Factors in Computing Systems CHI '97 Extended Abstracts on Human Factors in Computing Systems: Looking to the Future*, March 22-27, ACM New York, USA., pp: 168-169.
- Serenko, A., N. Bontis and B. Detlor, 2007. End-user adoption of animated interface agents in everyday work applications. *Behav. Inform. Technol.*, 26: 119-132.
- Shiratudin, N. and M. Landoni, 2002. Evaluation of content activities in children's educational software. *Eval. Program Plann.*, 25: 175-182.
- Storey, M.A., B. Phillips, M. Maczewski and M. Wang, 2000. Evaluating the usability of web-based learning tools. *Educ. Tech. Soc.*, 51: 91-100.
- Thurber, R.S., M.R. Shinn and K. Smolkowski, 2002. What is measured in mathematics tests, Construct validity of curriculum-based mathematics measures. *Sch. Psychol. Rev.*, 31: 498-513.
- Van Schaik, P. and J. Ling, 2005. Five psychometric scales for online measurement of the quality of human-computer interaction in Web Sites. *Int. J. Hum. Comput. Interact.*, 19: 309-322.
- Van Veenendaal, E., 1999. Questionnaire based usability testing. *International Conference on Product Focused Software Process Improvement Proceeding*. Nov. 1998, Oulu, Finland, pp: 118-134.
- Yushau, B., M.A. Bokhari and D.C.J. Wessels, 2004. Computer aided learning of mathematics: Software evaluation. *Math. Comput. Educ.*, 38: 165-182.