USING FORMATIVE MEASUREMENT MODELS TO EVALUATE THE EDUCATIONAL AND MOTIVATIONAL VALUE OF AN AR-BASED APPLICATION

Costin Pribeanu
National Institute for Research and Development in Informatics, Bucharest, Romania
E-mail: pribeanu@ici.ro

Abstract

With the explosion of new technologies for e-learning there is an increasing demand to assess the educational and motivational value of a new application. Augmented reality (AR) is a promising technology that creates new learning experiences by integrating real objects from the traditional school into a computing environment. A research challenge is to better understand the relationships between various factors of interest for the successful deployment of educational systems in primary and secondary schools. There are several approaches to evaluation that are based on quantitative methods. In recent years there is an increasing interest in taking an alternative perspective to measurement by using formatively measured constructs. This paper will highlight several benefits in using formative measurement models to evaluate the educational and motivational value of an AR-based e-learning application. The evaluation target is a Chemistry learning scenario that has been developed in the European project ARiSE – Augmented Reality in School Environment. Based on our previous work we developed a new evaluation instrument that includes both reflectively and formatively measured constructs to evaluate the ergonomic, educational, and motivational quality of a desktop AR application. The preliminary results from a pilot study show the extent to which specific features of the Chemistry scenario are positively influencing the educational and motivational value.
Key words: formative measurement models, e-learning, augmented reality, educational value, motivation.

Introduction

An important problem in education is how to engage students with appropriate information technologies during the learning process. In this respect, AR-based technologies are creating new opportunities for designers. Desktop AR systems integrate real objects from the traditional school in a computing environment. This facilitates learning by doing and places the learner into the center of the learning process which in turn could significantly increase the motivation to learn. As many authors pointed out, learning by doing is captivating and creates a user experience that is similar to computer games thus being more attractive for the young learner (Brom et al., 2011; Vos et al., 2011).

ARiSE (Augmented Reality in School Environments) was a research project funded by the European Commission under FP6-027039. The project created an Augmented Reality Teaching Platform (ARTP) in order to test the pedagogical effectiveness of using AR technologies in class and the usability of the target platform. A specific objective was to test the extent to which ARTP is enhancing students' motivation to learn.

Previous work in evaluating ARTP were based on qualitative studies (Vilkoniene, 2008, Vilkonis et al., 2008) to assess the educational value, quantitative studies (Balog & Pribeanu,

2009) to assess technology acceptance and a mix of methods to assess usability (Pribeanu et al., 2008). The last two investigations were based on an evaluation instrument aimed at including various factors that are relevant for a technology acceptance model (TAM), such as perceived ease of use, perceived usefulness, and perceived enjoyment (Davis, 1989). These factors were conceptualized as reflectively measured constructs. The estimation results of our TAM model for ARTP revealed a relatively low variance explained for the perceived usefulness and perceived enjoyment which in turn suggested some limitations of the evaluation instrument.

The objective of this paper is twofold. The first objective is to briefly summarize our previous work with formatively measured constructs that appeared to be promising for the evaluation of AR-based interactive systems. This work was done on the ARTP using the existing samples collected during the project. The second objective is to present some preliminary results in evaluating the educational and motivational value of a Chemistry learning scenario developed onto ARTP. The interaction paradigm for this learning scenario is "building with guidance" and is targeted at understanding the periodic table of Chemical elements, the structure of atoms / molecules, and chemical reactions. This work was done using a new evaluation instrument that is based on both formatively and reflectively measured constructs. In this respect we will present a set of causal indicators that are influencing the educational and motivational value of the target scenario.

The rest of this paper is organized as follows. Some methodological aspects regarding measurement models are briefly summarized in the next section. Then, we will present our previous work with formative measurement models. Next, we will present the method and the evaluation results from a pilot study. The focus is on the specification and estimation of three formatively measured constructs that are relevant for the educational and motivational value of the target application. The paper ends with conclusion and future research directions.

Methodological Aspects

In information systems research a distinction is made between two types of model: structural models and measurement models. The measurement model describes the causal relationships between a construct (latent variable) and its measures (indicators, items, observed variables). The structural model describes the causal relationships between constructs. Before estimating and assigning semantics to the structural model we have to correctly specify the measurement model (Anderson and Gerbing, 1988).

According to the direction of causal relationships, we distinguish between two types of measurement model: reflective and formative. There are distinct characteristics of each measurement model that were discussed in detail by Edwards & Bagozzi (200), Diamantopoulos & Winklhofer (2001), Jarvis et al. (2003), and Diamantopoulos et al. (2008).

In the reflective measurement models, the causal direction is from construct to indicators which are also termed as manifest variables. A change in constructs is reflected in simultaneous changes in all indicators. Therefore items are interchangeable and elimination of one of them doesn't change the construct domain. Measures should be positively correlated and the measurement model should have convergent validity.

In the formative measurement models the causal direction is from indicators to construct. Indicators are not interchangeable since each is capturing a distinct cause. Since the measures are defining the construct, a census of indicators is recommended. There are no assumptions on unidimensionality and correlations between indicators. However, colinearity should be avoided. Indicators don't have an error term and items are intercorrelated.

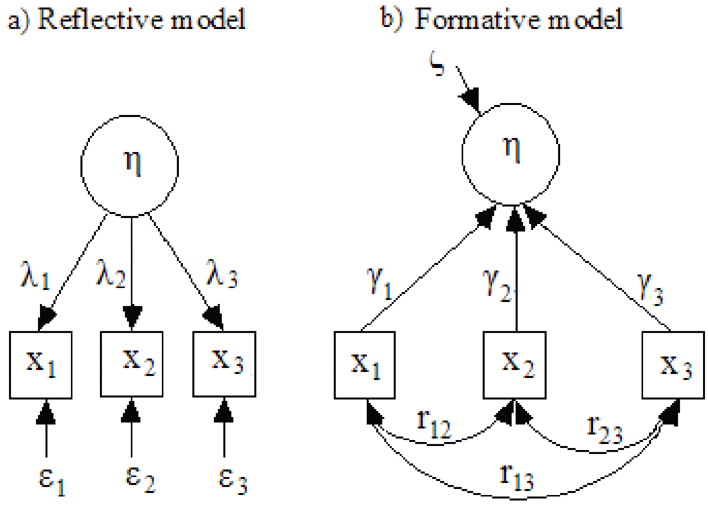

Figure 1: Reflective and formative measurement models.

Boolen (2011) distinguish between causal and composite (formative) indicators. Causal indicators share a common theme (conceptual unity) and may influence one or several latent variables. The error term accounts for indicators not taken into account. Composite indicators are completely determining the latent variable so there is no error term.

A formative measurement model taken in isolation is under identified and cannot be estimated. Most authors recommend achieving identification based on a 2+ rule: specifying effects (outcomes) of the formative constructs on at least two other variables that are reflectively measured. The effect variables could be: two reflective indicators (MIMIC model), two reflective constructs, or a reflective construct and a reflective indicator.

The selection of the outcome variables is just as important as is the selection of indicators (Diamantopoulos, 2011). As Wilcox et al. (2008) pointed out, the selected effect variables are determining the empirical meaning of the formative construct and the set of indicators. According to recent studies, there are several criteria to assess the validity of formative indexes (Diamantopoulos, 2011; Franke et al., 2008) : adequate coverage of the construct's domain (content validity), absence of multicollinearity, significant γ-coefficients, complete mediation of the effect of indicators on the outcome variables, significant influence (ß-coefficients) on the outcome variables, acceptable fit with the data.

Previous Work with Formative Models

In this section we summarize our previous work with formative models. We specified and estimated two formative models measuring the ergonomic quality of the ARTP and a formative model measuring the motivational value.

The AR platform consists of 4 independent modules organized around a table on which real objects are placed (Wind et al., 2007). The platform has been registered by Fraunfofer IAIS (Spinnstube®). The real objects are a periodic table and a set of colored balls. The evaluation instrument had 28 questions (on a Likert 1-5 scale) and 2 open questions: free description of most positive and most negative aspects. The items are measuring various factors: ERG (ergonomics), PEOU (perceived ease of use), PU (perceived usefulness), PE (perceived enjoyment), and INT (intention to use).

In order to specify and estimate the formative models we used the data collected in 2008. We analyzed the initial sample for Biology scenario of 139 observations for normality (skewness and kurtosis), univariate and multivariate outliers. Then we transformed the data (square root extraction) and we repeated the analysis and successively removed 9 observations. This results in a working sample with 130 observations that present moderate deviations from normality. In order to cross validate the model on another sample, we used the Chemistry scenario data. We performed the same data analysis procedure on the initial sample and successively removed 11 observations, thus getting a working sample with 128 observations with moderate deviations from normality. The formative models were estimated with AMOS 17.0 for Windows (Arbuckle, 2007).

The ergonomic quality is a key factor influencing both PU and PE. By ergonomic quality we refer to the extent to which a system is too easy to understand, easy to learn how to use, and easy to use. A formative model is useful to measure distinct usability aspects, such as the quality of visual and auditory perception (ERG-P) and the ease of interaction and collaboration (ERG-O). The latent variables are influencing the overall ease of use (PEOU1) and a reflective construct measuring the ease o learning how to operate with ARTP (ease of understanding, ease of learning and ease of remembering how to operate). More details regarding the indicators and effect variables could be found in (Pribeanu, 2011).

We specified and estimated both models on the Biology scenario and cross validated them on the Chemistry scenario. The results are presented in Table 1 (structural models).

Table 1. Summary of estimation results for ERG-P and ERG-O.

ERG-P	Biology		Chemistry	
	(γ/ß)	Sig. (p)	(γ/ß)	Sig. (p)
Indicators				
ERGP1	0.36	< 0.001	0.29	0.010
ERGP2	0.30	0.002	0.31	0.002
ERGP3	0.21	0.010	0.32	0.004
ERGP4	0.29	0.002	0.24	0.047
Effects				
PEOU1	0.63	< 0.001	0.61	< 0.001
PEOL	0.91	< 0.001	0.75	< 0.001
Explained variance				
ERG-P	78%		67%	
PEOL	83%		56%	

ERG-O	Biology		Chemistry	
	(γ/ß)	Sig. (p)	(γ/ß)	Sig. (p)
Indicators				
ERGO1	0.27	0.006	0.24	0.018
ERGO2	0.21	0.030	0.30	0.002
ERGO3	0.30	0.003	0.24	0.016
ERGO4	0.33	0.001	0.38	<0.001
Effects				
PEOU1	0.66	< 0.001	0.49	< 0.001
PEOL	0.87	< 0.001	0.93	< 0.001
Explained variance				
ERG-O	62%		55%	
PEOL	87%		86%	

All γ-coefficients are significant and the latent variables are completely mediating the effect of their indicators on the effect variables. All ß-coefficients are significant and both models show very good fit with the data, according to the cut-off values of quality indices (Hair et al, 2006).

The analysis of γ-coefficients revealed useful insights and makes it possible a comparison. As regarding ERG-P, the accuracy of visual perception has a similar weight in both scenarios. The vocal explanations (ERGP3) have a more important contribution in the Chemistry scenario, where some difficult concepts are explained in the introduction. As regarding ERG-O, selecting a menu item (ERGO2) is more difficult in the Chemistry scenario, since the student has the hands on the colored balls. Correcting the mistakes is more important in the Biology scenario because of the difficulty to correctly select small organs.

The motivational value is also a key factor influencing both the perceived usefulness and the intention to use. By motivational value we refer to the perceived enjoyment in learning with ARTP (intrinsic motivation). The formatively measured latent variable (PE) was estimated by adding two reflectively measured constructs: PU (perceived usefulness) and INT (intention to use). A formative model is useful to measure distinct aspect (facets) of the perceived enjoyment. More details regarding the indicators and effect variables could be found in (Pribeanu, 2012).

We specified and estimated both models on the Biology scenario and cross validated them on the Chemistry scenario. The results are presented in Table 2.

Table 2. Summary of estimation results for PE.

PE	Biology		Chemistry	
	(γ/ß)	Sig. (p)	(γ/ß)	Sig. (p)
Indicators				
PE1	0.29	< 0.001	0.27	< 0.001
PE3	0.24	0.001	0.10	0.108
PE4	0.18	0.045	0.17	0.031
PE5	0.27	0.001	0.48	0.001
PE6	0.19	0.010	0.26	0.002
Effects				
PU	0.88	< 0.001	0.73	< 0.001
INT	0.87	< 0.001	0.90	< 0.001
Explained variance				
PE	87%		96%	
PU	76%		53%	
INT	77%		82%	

The item PE2 had non significant γ-coefficient and was eliminated. All the other γ-coefficients are significant and the latent variables are completely mediating the effect of their indicators on the effect variables. All ß-coefficients are significant and both models show very good fit with the data, according to the cut-off values of quality indices (Hair et al, 2006).

The preference for the Chemistry scenario was obvious in all ARiSE studies and the formative model brings some additional insights.

The Chemistry scenario was perceived as much more exciting (PE5) than the Biology scenario. The pleasure to interact with real objects (PE3) has a lower height in the Chemistry scenario (because balls were not stable so students had difficulties in simulating chemical reactions). Students perceived both scenarios as interesting (PE1) and they liked learning with ARTP (PE6). The model explains more variance in PU for the Biology scenario and more variance for PE and INT in the Chemistry scenario. The influence of the formatively measured latent variable is much higher on INT than on PU in the Chemistry scenario.

Preliminary Results From a Pilot Study

Based on the conclusions drawn from our previous work we started the development of a new evaluation instrument, having both reflectively and formatively measured constructs. In this study we focus on a set of 8 causal indicators pointing to specific features of the ARTP. There are several typical AR capabilities, such as: 3D visualization, animation, vocal interface for learning and guidance, and haptic feedback. There are also some specific features for this scenario: augmentation of the atom structure, building a molecule from atoms, and simulation of chemical reactions. The description of the causal indicators set is given in Table 3.

Table 3. The set of causal indicators.

Item	Description
ARF1	The augmentation helps to understand the chemical structure of an atom
ARF2	Building a molecule from atoms helps to understand Chemistry
ARF3	Simulating a Chemical reaction with ARTP helps to understand it better
ARF4	Interacting with colored balls symbolizing atoms is a good idea
ARF5	Using ARTP helps to understand the periodic table
ARF6	Performing exercises with ARTP is useful to test my Chemistry knowledge
ARF7	Vocal explanations help interacting with ARTP
ARF8	ARTP creates a feeling of control over the learning process

The causal indicators are influencing three formatively measured constructs (ARF-PEF, ARF-PU, and ARF-PE that are in turn mediating the effects of their indicators on several reflectively measured variables: PEF (perceived efficacy), PU (perceived usefulness), and PE (perceived enjoyment). The list of outcome variables is presented in Table 4. These variables are measuring two facets of the educational value (perceived efficacy and overall usefulness for learning) and the motivational value of the target application.

Table 4. The effect variables.

Item	Description
Perceived efficacy (PEF)	
PEF1	ARTP would help me to learn with less effort
PEF2	ARTP would help me to understand the lesson better
Perceived usefulness (PU)	
PU1	I find ARTP useful for learning
PU2	After using ARTP my Chemistry knowledge will improve
Perceived enjoyment (PE)	
PE1	I like learning with ARTP
PE2	ARTP motivates me to learn

The main purpose of the pilot study was to test the new evaluation instrument. The target application was the Chemistry scenario. The learning scenario for chemistry has an introduction and three lessons. Each lesson has several exercises. For a detailed description of the learning tasks see Vilkonis et al., 2008). The sample was pretty small (N=71), students 7th grade from 8 schools in Bucharest. After testing, the students were asked to answer a questionnaire by rating the items on a 5-point Likert scale. The data was collected in May-June 2012. More details regarding the experiment could be found in (Iordache et al., 2012).

In order to comply with the requirements for an estimation based on structural equation modeling (SEM) techniques we perform data transformation (variable reflection and square root extraction) in order to reduce the skewness and the number of outliers. With the use of a $p < 0.001$ criterion for Mahalanobis distance no multivariate outliers among the cases were found.

We estimated both a MIMIC model and a structural model in each case, in order to check the stability of causal indicators. We added one reflectively measured item (PU1 in ARF-PEF and ARF-PE and PEF2 in ARF-PU) In order to achieve identification for the structural model. The models are presented in Figure 2, where x_1… x_4 are the causal indicators, y_1-y_3 are the effect indicators, η is the formatively measured construct, and η_1 is the reflectively measured construct. The validity check revealed that only four indicators should be kept in each model (the other four were eliminated, so only four were represented in Figure 2).

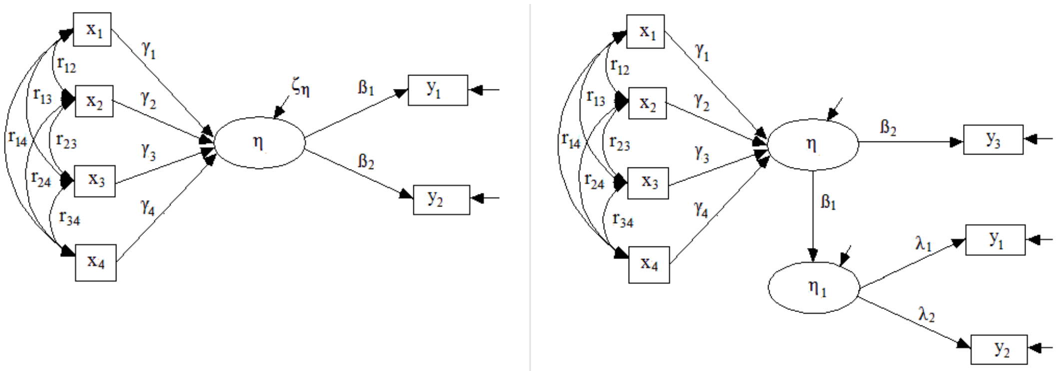

Figure 2: The MIMIC and structural models for estimation.

The evaluation results for the structural models are presented in Table 5. The formatively measured construct η takes the values ARF-PEF, ARF-PU and ARF-PE. The reflectively measured construct η_1 takes the values PEF, PU, and PE. The item y_3 takes the values PU1, PEF2, PU1.

Table 5. Estimation results (structural models).

ARF	ARF-PEF		ARF-PU		ARF-PE	
	(γ/ß)	Sig. (p)	(γ/ß)	Sig. (p)	(γ/ß)	Sig. (p)
Indicators						
ARF1			0.23	0.029		
ARF2					0.23	0.043
ARF3	0.31	0.002				
ARF4					0.21	0.042
ARF5	0.43	<0.001	0.32	0.004	0.38	0.003
ARF6			0.32	0.036		
ARF7	0.21	0.027				
ARF8	0.40	<0.001	0.37	<0.001	0.48	0.002
Effects						
η1	0.97	<0.001	0.98	<0.001	0.99	<0.001
y3	0.71	<0.001	0.71	<0.001	0.77	<0.001
Explained variance						
η	83%		81%		69%	
η_1	95%		95%		98%	

All causal indicators are useful since they are significant in at least one model, although the weight on each latent variable is context specific. The results highlight their relative importance for the educational value (PEF and PE) and motivational value (PE). All γ-coefficients are significant and the latent variables are completely mediating the effect of their indicators on the effect variables. All ß-coefficients are significant and all models show very good fit with the data (with one exception, the MIMIC model for ARF-PU). The set of causal indicators for each latent variable is stable between the MIMIC and structural models. The fit indices for the structural models are presented in Table 6.

Table 6. Fit indices (structural models).

Model	χ^2	df	χ2/df	GFI	CFI	srmr
ARF-PEF	50.662	7	00.809	00.978	10.000	00.037
ARF-PU	10.103	7	10.443	00.999	10.000	00.007
ARF-PE	10.330	7	10.476	00.966	00.960	00.044

The explained variance is 83% in the formatively measured construct (ARF-PEF) and 95% in the perceived efficacy (PEF). However, the influence on the perceived effectiveness (PEF2) is higher than on the perceived effiency. Most important for this educational facet is the understanding of the periodic table (ARF5) and the feeling of control over the learning process (ARF8). Next, the explained variance is 81% in ARF-PU and 95% in PU, so the formatively measured construct is a very good predictor of PU. Most important is the item ARF8 (feeling of control over the learning process). Finally, the explained variance is 69% in ARF-PE and 98% in PE, so the formatively measured construct is a very good predictor of PE. Most important are the items ARF5 and ARF8.

The results show the relative importance of each specific AR feature for each factor of interest. From far, there are two indicators that are most important for the educational and motivational value of the Chemistry scenario: ARF5 and ARF8.

The set of causal indicators (ARF) depends on the target application. They are specific both to the target discipline and the target platform / application. Therefore, we believe that such a set should be carefully conceptualized for each e-learning system. Regarding the target discipline, the causal indicators should refer to the specific learning goals. Regarding the target platform, the causal indicators are related to the specific interaction techniques. In what concerns the target application, it is important to capture the specific way of implementing the interaction techniques in order to achieve the educational goals. The outcome of the evaluation should provide the designers with a basis to understand which features should be given higher importance and which are less relevant.

Conclusions

Reflective measurement perspective is interested how a latent variable is perceived. The focus of validity is at construct level. Formative measurement perspective is interested how a latent variable is actually measured. The focus of validity is at indicator level. In this paper we presented several formative measurement models based on both previous and recent work. We argue that formative measurement is an equally useful perspective that is able to bring additional insights for the evaluation of the educational and motivational value of an AR-based learning application. Besides this, the estimation of formative models is less demanding as regarding the number of observations, since there are fewer variables to estimate. As such, it could be used during the development of interactive systems in order to provide the developers with useful hints.

The study results revealed a set of causal indicators that act as predictors for the educational and motivational value of the Chemistry application. The main strengths of this work is the conceptualization and estimation of a relatively large set of causal indicators that relate specific ARTP features to the main educational goals of the target application. Overall, the evaluation results are consistent with previous results from both qualitative and quantitative studies. The outcome of using formative measurement models is a detailed analysis of the contribution of each predictor for the various facets of the educational value. While the contribution of the ARTP to a better understanding of the periodic table seems to be a cornerstone for learning Chemistry, the feeling of control over the learning process shows the advantages of a learner-centered approach.

There are several inherent limitations of this work since the study is using a relatively small pilot sample (N=71). The number of observations is at limit even for estimating a simple formative model. In this respect, the study results are exploratory. Also, the methodology for specification and estimation of formative measurement models is not mature yet. In the next future we will focus on the refinement of the questionnaire in order to proceed to data collection for a larger sample.

Acknowledgement

We gratefully acknowledge the support of the ARiSE (FP6-027039) project, funded by the European Commission.

References

Anderson, J. C., Gerbing, D. W. (1988). Structural Equation Modelling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin, 103* (3), 411-423.

Arbuckle, J. L. (2007). *AMOS 16.0 User's Guide*. Amos Development Corporation.

Balog, A., Pribeanu, C. (2010). The Role of Perceived Enjoyment in the Students' Acceptance of an Augmented Reality Teaching Platform: a Structural Equation Modelling Approach. *Studies in Informatics and Control, 19* (3), 319-330.

Boolen, K. (2011). Evaluating effect, composite and causal indicators in structural equation models. *MIS Quaterly, 35* (2), 359-372.

Brom, C., Preuss, M., Klement, D. (2011). Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. *Computers & Education, 57*, 1971-1988.

Diamantopoulos, A., Riefler, P., Roth, K. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203-1218.

Davis, F. D (1989). Perceived usefulness, perceived easy of use, and user acceptance of information technology. *MIS Quaterly, 13*, 319-340.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quaterley, 35* (2), 335-358.

Edwards, J., Bagozzi, R. (2000) On the nature and direction of of relationship between constructs and measures. *Psychological Methods, 5* (2), 155-174.

Franke, G., Preacher, K., Rigdon, E. (2008) Proportional structural effects of formative indicators. *Journal of Business Research, 61*, 1229-1237.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. (2006). *Multivariate Data Analysis*. 6th Ed., Prentice Hall, 2006.

Iordache, D. D., Pribeanu, C., Balog, A. (2012). Influence of Specific AR Capabilities on the Learning Effectiveness and Efficiency. *Studies in Informatics and Control*, *21* (3), In press.

Jarvis, C. B., Mackezie, S., Podsakoff, M. (2003). A critical review of construct indicators and measurement models misspecification in marketing and consumer research. *Journal of Consumer Research, 30*, 199-218.

Pribeanu, C. (2012). Un model formativ de măsurare a valorii motivaţionale a unei aplicaţii educaţionale bazate pe realitate îmbogăţită. *Revista Romana de Interactiune Om-Calculator, 5* (2), 13-18.

Pribeanu, C. (2011). Specification and validation of a second order formative index to evaluate the ergonomic quality of an AR-based educational platform. *International Journal of Computers, Communication and Control, 7* (4), 720-731.

Pribeanu, C., Balog, A., Iordache, D. D. (2008). Formative user-centered usability evaluation of an augmented reality educational system. *Proceedings of ICSOFT 2008* – The third International Conference on Software and Data Technologies, Porto 5-8 July, INSTICC. 65-72.

Vilkoniene, M. (2008). Influence of Augmented Reality technology upon pupils' knowledge about Chemical bonds and compounds: the results of the experiment. *Proceedings of ICTNSE 2008*, Šiauliai, 116-123.

Vilkonis, R., Lamanauskas, V., Palepsaitiene, R. (2008). The scenario of learning module "Introductory course of Chemistry" for a new learning platform based on augmented reality technology (Prototype 2). *Proceedings of ICTNSE 2008*, Šiauliai, 123-142.

Vos, N, Meijden, H., Denessen, E. (2011). Constructing versus playing an educational game on student motivation and deep learning strategy used. *Computers & Education, 56*, 126-137.

Wilcox, J., Howell, R., Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research, 61*, 1219-1228.

Wind, J., Riege, K., Bogen, M. (2007) Spinnstube®: A Seated Augmented Reality Display System, *In Proceedings of Virtual Environments 2007,* IPT-EGVE 2007 – EG/ACM Symposium Proceedings, 17-23.

Advised by Laima Railienė, University of Siauliai, Lithuania

Received: *October 20, 2012*   Accepted: *December 05, 2012*

Costin Pribeanu	PhD., Senior Researcher, National Institute for Research and Development in Informatics, 8-10, Maresal Al. Averescu, 011455, Bucharest, Romania. E-mail: pribeanu@ici.ro Website: http://rochi.ici.ro/pribeanu/index.htm