

# Specification and Validation of a Formative Index to Evaluate the Ergonomic Quality of an AR-based Educational Platform

C. Pribeanu

## Costin Pribeanu

National Institute for Research and Development in Informatics - ICI Bucharest  
Romania, 011455 Bucuresti, Bd. Maresal Averescu, 8-10  
E-mail: pribeanu@ici.ro

### Abstract:

The ergonomic quality of educational systems is a key feature influencing both the usefulness and motivation for the learner. Desktop Augmented Reality (AR) systems are featuring specific interaction techniques that may create additional usability issues affecting the perceived ease of use. Measuring key usability aspects and understanding the causal relationships between them is a challenge that requires formative measurement models specification and validation. In this paper we present an evaluation instrument based on two main formative indexes that are capturing specific usability measures for two AR-based applications. The formative indexes are forming a second order formative construct that acts as predictor for both the general ease of use and ease of learning how to operate with the application.

**Keywords:** formative measurement model, formative index, augmented reality, usability, ergonomic quality.

## 1 Introduction

Educational systems based on desktop AR technologies are creating an appealing user experience for the learner by integrating real life objects into computer environments. Touching and holding real objects is increasing the students' motivation to learn and could better support active and collaborative learning [18], [22]. As AR technologies become more wide-spread, there is an increasing interest in their ergonomic quality. Designing for usability is not easy in emerging technologies, like AR systems, which are featuring novel interaction techniques [5], [15].

The ISO standard 9126-1 defined usability as the capability of a software system to be easy to understand, easy to learn how to operate with, easy to operate with, and attractive, when used under specified conditions [19]. By ergonomic quality we refer to the first three usability aspects: ease of understanding, ease of learning how to operate, and ease of operating with a software system. How to measure and improve the usability of interactive systems is a key research topic in HCI. A research challenge is to better understand the relationships between different usability measures as well as between usability and other factors of interest [17].

In a previous work we developed a measurement model that was grounded in the technology acceptance models (TAM) theory [9] in order to explain the causal relations between various factors influencing the intention to use of an AR-based educational platform [3], [4]. Although the structural model was useful to test some typical TAM hypotheses the variance explained was small and several items targeting specific usability aspects were eliminated in order to achieve the unidimensionality required by a reflective measurement model. Moreover, reflective measurement assumes the same antecedents for reflective indicators (as manifest variables) so causal relations are estimated at construct level [8], [24]. These shortcomings suggest looking for an alternative modeling approach.

In this paper we present a measurement model for the evaluation of the ergonomic quality of applications developed onto an AR-based educational platform. The Augmented Reality Teaching Platform (ARTP) was developed in the framework of the ARiSE (Augmented Reality for

School Environments) European project. Two AR applications implementing learning scenarios for Biology and Chemistry were developed and tested onto ARTP.

The measurement model consists in two sets of formative indicators that are measuring two dimensions of the ergonomic quality of desktop AR applications: the quality of visual and auditory perception and the ease to operate and collaborate in a constrained space. The two indexes are forming a second order formative construct. In order to achieve identification requirements we used as outcome variables a reflective construct measuring the perceived ease of learning how to use ARTP and a general reflective item measuring the overall ease of use. The formative measurement model was estimated on the Biology scenario data. Then we cross validated the models on the Chemistry scenario data.

The rest of this paper is organized as follows. In the following section we describe the formative measurement models and discuss some methodological aspects related to the specification, identification and validity. In section 3 we present and discuss the estimation results with the Biology scenario data. In section 4 we present the results of a confirmatory assessment of the formative measurement model using the Chemistry scenario data and we comparatively discuss the results for each scenario. The paper ends with conclusion and future research directions.

## 2 The formative measurement model

### 2.1 Reflective vs. formative measurement models

A measurement model describes the relationships between a construct (latent variable) and its measures (indicators, items) while a structural model describes the relationships between different constructs [12], [13]. The causal relation between a construct and its measures could be from construct to measures (reflective model) or from measures to construct (formative model). There are distinct characteristics of each measurement model that were systematically presented and discussed in detail in [6], [10], [12], [20].

In the reflective measurement model the indicators are manifest variables of the latent variable. A change in the constructs is reflected in simultaneous changes in all indicators. As such, the 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 and discriminate validity.

In the formative measurement model the measures are defining the conceptual meaning of the 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 [6]. There are no assumptions on unidimensionality and correlations between indicators. However, collinearity should be avoided. Indicators don't have an error term and items are intercorrelated. Although there is an error terms at construct level this is not a measuring error but a disturbance accounting for other causes not specified by the model [11]. The nomological net of formative indicators could differ as this is a distinct feature of the formative measurement [20].

A formative measurement model taken in isolation is under identified and cannot be estimated. Jarvis et al. and Diamantopoulos et al. 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 [12], [20]. The outcome variables could be: two reflective indicators (MIMIC model), two reflective constructs, or a reflective construct and a reflective variable. The selection of the outcome variables is just as important as is the selection of indicators [11], [14]. According to Wilcox et al., the selected effect variables are determining the empirical meaning of the formative construct and the set of indicators [26].

The proper specification of the measurement model is a precondition before analyzing and

assigning a meaning to the structural model [1]. According to Jarvis et al., there are many studies in literature that are based on inappropriate specification of the measurement models [20]. In recent years, there is an ongoing debate regarding the formative versus reflective specification of various constructs and the appropriateness of measurement scales that are frequently used in different domains [12].

Taking the appropriate measurement perspective is not a simple issue. As pointed out by Jarvis and colleagues, based on the analysis of 178 papers published in four top journals in marketing research, there are about 29% cases (reported at 1192 constructs) of misspecification [20]. Moreover, the authors themselves experienced difficulties in classifying 14% of constructs featuring both reflective and formative characteristics. Wilcox and colleagues argued that a construct is not inherently formative or reflective so the researcher has a choice to take a perspective or another [26]. In this respect, the specification of alternative models is useful since is providing with more insights into the field of study.

## 2.2 Experiment, samples and data analysis

ARTP is a "seated" AR environment: users are looking to a see-through screen where virtual images are superimposed over the perceived image of a real object placed on the table [27]. Two AR-based applications were developed onto this platform (see Figure 1).

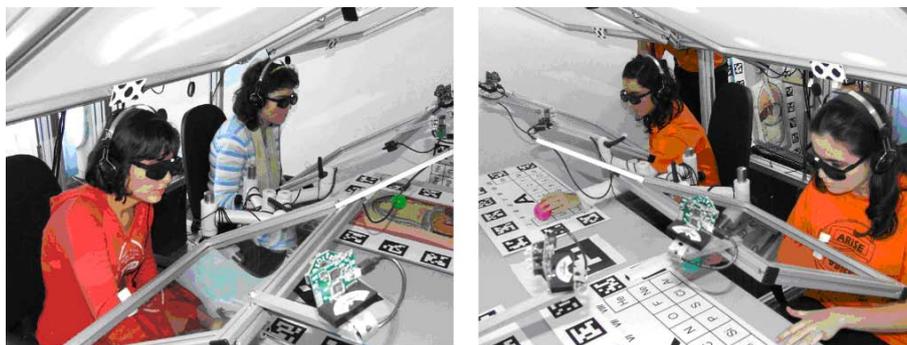


Figure 1: Students testing the ARTP learning scenarios: Biology (left) and Chemistry (right)

The first application implemented a Biology learning scenario for secondary schools. The implemented paradigm was "3D process visualization of hidden processes" and was targeted at enhancing the students' understanding and motivation to learn the human digestive system. The real object is a flat torso of the human body. A pointing device having a colored ball on the end of a stick and a remote controller Wii Nintendo as handler has been used as interaction tool that serves for three types of interaction: pointing on a real object, selection of a virtual object and selection of a menu item.

The second application implemented a Chemistry scenario. The implemented paradigm was "building with guidance" and was targeted at enhancing the students' understanding and motivation to learn the periodic table of Chemical elements, the structure of atoms / molecules, and the chemical reactions. The real objects were the periodic table of chemical elements and four sets of colored balls symbolizing atoms. The remote controller Wii Nintendo has only been used as interaction tool for confirming a selection.

The test was conducted in 2008, on the ICI's platform which is equipped with 4 ARTP modules. A total number of 139 students (13-14 years old), from which 65 boys and 74 girls tested the platform. All were 8th grade students enrolled in 3 general schools in Bucharest. None of them was familiar with the AR technology. The students came in groups of 7-8, accompanied

by a teacher. Each student tested the platform twice: once for the Biology scenario and second time for the Chemistry scenario. Each scenario consists of a demo lesson and a number of exercises.

After testing, the students were asked to answer a usability questionnaire by rating the items on a 5-point Likert scale (1-strongly disagree, 2-disagree, 3-neutral, 4-agree, and 5-strongly agree). The questionnaire has 28 closed items and 2 open questions, asking users to describe the most 3 positive and most 3 negative aspects. The first 24 closed items are targeting various dimensions of the ARTP such as ergonomics and usability (items 1-14), perceived utility (items 15-17), perceived enjoyment (items 18-21) and intention to use (items 22-24). The last four items were to assess how the students overall perceived the platform as being easy to use, useful for learning, enjoyable to learn with, and exciting.

In order to estimate the new measurement model we used the Biology scenario data. We analyzed the initial sample of 139 observations for normality (skewness and kurtosis), univariate and multivariate outliers. We transformed the data (square root extraction) and we repeated the analysis and successively removed 9 observations. The final sample has 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. The final sample has 128 observations with moderate deviations from normality.

### 2.3 Model specification and identification

According to our knowledge, there are few approaches to formative index construction for the usability and / or ease of use [21]. Although the perceived ease of use and perceived usability are frequently used in information systems research, in almost all studies they are specified as reflectively measured constructs. As such, their indicators have a limited contribution (as manifest variables) to explain the effect of usability problems.

Since the objective of this study is to analyze the relationships between different aspects related to the ergonomic quality of the ARTP, 15 items in the usability questionnaire are of interest, from which 11 are formative measures and 4 are reflective measures. The 15 items (presented in Annex 1) are grouped into four constructs and a single item measure:

- The quality of visual and auditory perception (ERG-P): clear observation and superposition, easy to read the information on the screen, and easy to understand the vocal explanations.
- The ease of interaction and collaboration and collaboration (ERG-O): comfortable work place, easy to select a menu item with the remote control, easy to correct errors, and easy to collaborate with colleagues.
- The ease of adjusting the devices and accessories (ERG-A), i.e. the see-through screen, stereo glasses and head phones.
- The ease of learning (PEOL): easy to understand, easy to learn and easy to remember how to use ARTP.
- The general item measuring the overall ease of use (PEOU1).

The first three constructs are composite indexes measuring distinct usability aspects that are specific to an AR-based learning application. As such, the indicators are not interchangeable and elimination of any of them will alter the conceptual domain of the construct. For example, if we analyze the three items measuring the quality of the visual perception, each is targeting a

different usability aspect. The clarity of observation through the screen is a hardware issue while the clarity of superposition between the augmentation and the real object is a software issue. Reading the information on the screen relates to augmentation, messages to the user and menu items.

Note that apart from the specific AR devices and accessories there are also several usability aspects which are specific to a given application. For example, in the Biology scenario the user selects an organ by pointing on flat torso of the digestive system which is a real object shared by to students staying face-to-face. In the Chemistry scenario, the students create a molecule by bringing together several colored balls symbolizing atoms. In this respect, the interaction with the remote control, the correction of mistakes (selection errors) and the collaboration between students depend on the real objects registered with the application. Therefore a formative model is an appropriate measurement perspective.

The ergonomic quality of ARTP is a multidimensional construct conceptualized as a composite of formative indexes. Each dimension is a formative index measuring a set of specific usability aspects. Each index is assumed to have a significant positive influence on two general usability aspects: the perceived ease of learning how to use ARTP (the construct PEOL) and on the overall ease of use (the general item PEOU1).

## 2.4 Validity of the formative indexes

According to recent studies, there are several criteria to assess the validity of formative indexes [6], [10], [12], [14]: adequate coverage of the construct's domain, absence of multicollinearity, indicator validity, significant  $\gamma$ -coefficients, complete mediation of effects, significant influence ( $\beta$ -coefficients) on outcome variables, and acceptable fit with the data.

Although a census of indicators is ideal to cover the scope of a formative index, this is rarely possible. In our model, each index is addressing a distinct aspect of the ergonomic quality of ARTP. Since formative indicators are also capturing critical usability aspects as indicated in previous studies (e.g. [23]) the coverage of the domain is acceptable.

The collinearity of formative indicators was analyzed with the VIF (Variation Inflation Factor) statistic for each index. VIF values were in the range 1.183-1.946 for the Biology scenario, respectively 1.085-1.715 for the Chemistry scenario below the 3.3 cut-off value [12].

The general item PEOU1 is an overall measure of the ergonomic quality of ARTP which qualify it for using as criterion validity. An analysis using Pearson's rho indicated that there are significant positive linear relationships between PEOU1 and the formative indicators of ERG-P and ERG-O but no significant correlations with the formative indicators of ERG-A. Nevertheless, in both samples ERG-A indicators are positively correlated with the formative item ERGO1. This suggests that ERG-A is not a distinct dimension of the ergonomic quality of ARTP but only an antecedent of a formative indicator measuring the comfort with the workplace. A regression analysis on the Biology data sample showed that ERGA1 and ERGA2 are two antecedents of ERGO1 (standardized coefficients  $\beta_{ERGA1} = 0.191$ , sig=0.046 and  $\beta_{ERGA2} = 0.185$ , sig=0.039). The regression analysis on the Chemistry data sample confirmed this finding ( $\beta_{ERGA1} = 0.156$ , sig=0.083 and  $\beta_{ERGA2} = 0.255$ , sig=0.005).

In order to estimate the formative indexes we used a MIMIC model and a structural model presented in Figure 2. The models were estimated using AMOS 17.0 [2]. Each index has n formative indicators, more specifically n=4 for ERG-P and ERG-O, and n=3 for ERG.

There are four outcome variables in the MIMIC model. Three of these reflective indicators are further grouped in the structural model that features 2 outcome variables: the general item PEOU1 (overall ease of use) and the reflective construct PEOL (ease of learning how to operate). All outcome variables are closely related to the focal construct as they measure general aspects

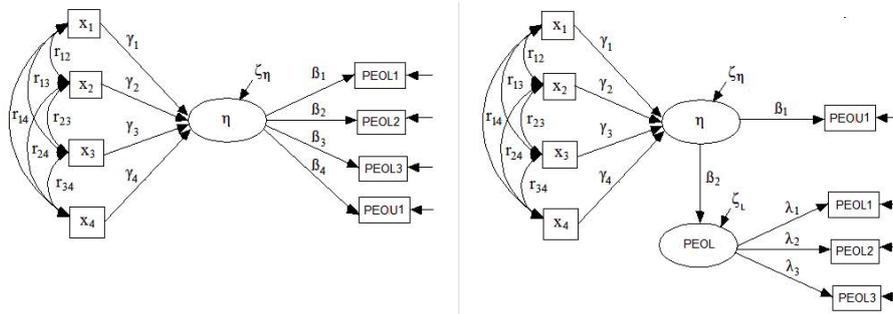


Figure 2: Estimation of formative indexes with MIMIC (left) and structural models (right)

of the perceived ergonomic quality.

There are three general hypotheses assessed with these models:

1. There is a significant contribution of the formative indicators to the composite index ( $x_i \rightarrow \eta$ ,  $i=1\dots n$ ).
2. There is a significant positive influence of the composite index on the perceived ease of learning how to use ARTP ( $\eta \rightarrow \text{PEOL1}$ ,  $\eta \rightarrow \text{PEOL2}$ ,  $\eta \rightarrow \text{PEOL3}$  in the MIMIC model, respectively  $\eta \rightarrow \text{PEOL}$  in the structural model).
3. There is a significant positive influence of the composite index on the overall ease of use ( $\eta \rightarrow \text{PEOU1}$ ).

Since the structural model includes a reflectively measured construct, the internal consistency and convergent validity should be assessed. The scale reliability and unidimensionality were analyzed with SPSS 16.0 and Amos 17.0. The consistency of scale (Cronbach's alpha) was 0.701 for the Biology scenario and 0.704 for the Chemistry scenario which is acceptable. Convergent validity was assessed by examining the standardized factor loadings, composite reliability, and average variance extracted for PEOL in each scenario [16]. Almost all factor loadings are over the minimum recommended level of 0.60. The composite reliability was 0.711 for the Biology scenario and 0.704 for the Chemistry scenario, above the minimum recommended value of 0.70 in each scenario. The average variance extracted was 0.456 for the Biology scenario and 0.438 for the Chemistry scenario. Overall, PEOL construct has an acceptable convergent validity.

### 3 Estimation results on the Biology scenario data

#### 3.1 First order formative indexes

The results of MIMIC and structural model estimations for ERG-P and ERG-O are presented in Table 1. All  $\gamma$ -coefficients are significant at  $p < 0.05$  level thus supporting the first hypothesis. There are small differences between the magnitudes of  $\gamma$ -coefficients in the two models. The variance of the error term associated with the formative index is small in each model, so the formative index is sound and each formative item has a distinct contribution to the explained variance [11].

Fit indices are acceptable, over the recommended values [16]:  $\chi^2=1.115$ ,  $DF=13$ ,  $\chi^2/DF=1.624$ ,  $GFI=0.962$ ,  $CFI=0.974$ ,  $SRMR=0.036$  (ERG-P, structural model), and  $\chi^2=22.963$ ,  $DF=13$ ,  $\chi^2/DF=1.766$ ,  $GFI=0.960$ ,  $CFI=0.958$ ,  $SRMR=0.042$  (ERG-O, structural model).

In both models all  $\beta$ -coefficients are significant ( $p < 0.001$ ), which supports the last two hypotheses. The influence of formative indexes is stronger on the perceived ease of learning how to

Table 1: Estimation results for ERG-P and ERGO - Biology scenario

ERG-P	MIMIC model		Structural model		ERG-O	MIMIC model		Structural model	
	$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)		$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)
Contribution					Contribution				
ERGP1	.33	<.001	.36	<.001	ERGO1	.22	0.018	.27	0.006
ERGP2	.31	0.001	.30	0.002	ERGO2	.22	0.017	.21	0.030
ERGP3	.20	0.010	.21	0.010	ERGO3	.29	0.003	.30	0.003
ERGP4	.27	0.002	.29	0.002	ERGO4	.33	<.001	.33	0.001
Effect variables					Effect variables				
PEOU1	.63	<.001	.63	<.001	PEOU1	.62	<.001	.66	<.001
PEOL			.91	<.001	PEOL			.87	<.001
PEOL1	.64	<.001			PEOL1	.63	<.001		
PEOL2	.73	<.001			PEOL2	.71	<.001		
PEOL3	.61	<.001			PEOL3	.63	<.001		
Variance expl.					Variance expl.				
ERG-O	71%		78%		ERG-O	54%		62%	
PEOL			83%		PEOL			75%	

use ARTP than on the general ease of use. This means that once the user understands and learns how to use the system he finds it easy to use. The highest contributions to ERG-P have the first two items (clarity of observation through the see-through screen and accuracy of superposition). The most important contribution to ERG-O has the last item related to the ease of collaboration with colleagues. The ease of correcting the mistakes proved also to be an important measure for the Biology scenario.

### 3.2 Second order formative index

ERG-P and ERG-O are two distinct dimensions of the ergonomic quality of ARTP that are forming a second order formative construct (ERG). We used the scores of the first order constructs (the predicted values of the multiple regression) as formative indicators in the second order construct. Similar approaches are described in [7], [8]. The estimation results are presented in Table 2.

Table 2: Estimation results for second order construct - Biology scenario

ERG	MIMIC model		Structural model	
	$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)
Contribution				
ERG-P	.65	<.001	.68	<.001
ERG-O	.27	0.006	.30	0.004
Effect variables				
PEOU1	.63	<.001	.63	<.001
PEOL			.90	<.001
PEOL1	.64	<.001		
PEOL2	.71	<.001		
PEOL3	.61	<.001		
Variance expl.				
ERG	75%		84%	
PEOL			81%	

The  $\gamma$ -coefficients are significant in each model. The contribution of the first dimension is much higher showing that the quality of visual perception is a critical requirement for the desktop AR systems. The analysis of modification indices showed that the formative index is completely mediating the effects of its items.

Both  $\beta$ -coefficients are significant ( $p < 0.001$ ), which supports the last two hypotheses. The variance of the error term associated with the formative index is 0.009 (medium effect). The magnitude of the error term is suggesting some other aspects not covered by the indicators.

Fit indices are acceptable, over the recommended values:  $\chi^2=11.759$ ,  $DF=7$ ,  $\chi^2/DF=1.679$ ,  $GFI=0.972$ ,  $CFI=0.984$ ,  $SRMR=0.032$  (structural model).

## 4 Cross validation of the formative indexes on the Chemistry scenario

### 4.1 First order formative indexes

The results of estimation are presented in Table 3. Almost all  $\gamma$ -coefficients are significant at  $p<0.05$  level thus supporting the first hypothesis. There is only one exception: ERGP4 in the MIMIC model, where the  $\gamma$ -coefficient is significant at  $p<0.10$  level. There are relatively small differences between the contributions of each item in each model.

Table 3: Estimation results for ERG-P and ERG-O - Chemistry scenario

ERG-P	MIMIC model		Structural model		ERG-O	MIMIC model		Structural model	
	$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)		$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)
Contribution					Contribution				
ERGP1	.23	0.016	.29	0.010	ERGO1	.25	0.009	.24	0.018
ERGP2	.28	0.009	.31	0.012	ERGO2	.27	0.003	.30	0.002
ERGP3	.24	0.010	.32	0.004	ERGO3	.22	0.021	.24	0.016
ERGP4	.20	0.053	.24	0.047	ERGO4	.36	<.001	.38	<.001
Effect variables					Effect variables				
PEOU1	.55	<.001	.61	<.001	PEOU1	.48	<.001	.49	<.001
PEOL			.75	<.001	PEOL			.93	<.001
PEOL1	.70	<.001			PEOL1	.71	<.001		
PEOL2	.67	<.001			PEOL2	.70	<.001		
PEOL3	.52	<.001			PEOL3	.55	<.001		
Variance expl.					Variance expl.				
ERG-O	47%		67%		ERG-O	49%		55%	
PEOL			56%		PEOL			86%	

In both models  $\beta$ -coefficients are significant ( $p<0.001$ ), which supports the last two hypotheses. The variance of the error term associated with the formative index is 0.022 (0.008) for ERG-P and 0.021 (0.016) for ERG-O. Since the magnitude of the error term is small and all indicator coefficients are significant, the formative index is sound and each formative item has a distinct contribution to the explained variance.

Fit indices are acceptable, over the recommended values [16]:  $\chi^2=15.154$ ,  $DF=13$ ,  $\chi^2/DF=1.624$ ,  $GFI=0.973$ ,  $CFI=0.990$ ,  $SRMR=0.038$  (ERG-P, structural model), and  $\chi^2=22.392$ ,  $DF=13$ ,  $\chi^2/DF=1.722$ ,  $GFI=0.958$ ,  $CFI=0.947$ ,  $SRMR=0.048$  (ERG-O, structural model).

The influence of formative indexes is stronger on the perceived ease of learning how to use ARTP than on the general ease of use. The highest contributions to ERG-P have the items ERGP2 (accuracy of superposition) and ERGP3 (understanding the vocal explanation). The contribution of ERGP3 shows the importance of vocal explanations for students. The most important contribution to ERG-O has the last item related to the ease of collaboration with colleagues. The ease of selecting a menu item proved also to be an important measure for the Chemistry scenario.

### 4.2 Second order formative index

The results of structural model estimation are presented in Table 4. Both  $\gamma$ -coefficients are significant. The contribution of each dimension is similar for the Chemistry scenario. The analysis of modification indices showed that the index is completely mediating the effects of its items.

Table 4: Estimation results for second order construct - Chemistry scenario

ERG	MIMIC model		Structural model	
	$\gamma/\beta$	sig.(p)	$\gamma/\beta$	sig.(p)
Contribution				
ERG-P	.45	<.001	.55	<.001
ERG-O	.47	<.001	.49	0.004
Effect variables				
PEOU1	.53	<.001	.55	<.001
PEOL			.83	<.001
PEOL1	.70	<.001		
PEOL2	.68	<.001		
PEOL3	.53	<.001		
Variance expl.				
ERG	63%		80%	
PEOL			69%	

Both  $\beta$ -coefficients are significant ( $p < 0.001$ ), which supports the hypotheses. The variance of the error term associated with the formative index is 0.016 (0.222) which means a medium to large effect. The magnitude of the error term is suggesting some other aspects not covered by the indicators.

Fit indices are acceptable, over the recommended values:  $\chi^2 = 14.932$ ,  $DF = 7$ ,  $\chi^2/DF = 2.133$ ,  $GFI = 0.964$ ,  $CFI = 0.960$ ,  $SRMR = 0.049$  (structural model).

### 4.3 Comparison of results and discussion

The estimation of formative indexes on the Chemistry scenario data cross validated the measurement model and enables a comparison between the two implemented scenarios. The variances explained by the structural models for the formative indexes are higher for the Biology scenario than for the Chemistry scenario.

The variance explained by the model for the second order construct is slightly higher for the Biology scenario. The contribution of ERG-P to the super ordinate index is higher than the contribution of ERG-O in both scenarios but the relative importance is much higher for the Biology scenario. The variance explained by the model for the outcome variable PEOL is also higher for the Biology scenario (81% vs. 69%).

As regarding the ERG-P index, the comparison reveals that understanding of vocal explanations (ERGP3) is the most important item for the Chemistry scenario and the less important for the Biology scenario. This is explained by the fact that the Chemistry demo lesson and exercises were more difficult for students so a clear understanding of the lesson and how to perform the exercises was critical. The accuracy of superposition between the projection and the real object (ERGP2) has a higher importance for Biology.

As regarding ERG-O, the comparison reveals that the ease of collaboration with colleagues (ERGO4) is the most important item for both scenarios. Selecting a menu item (ERGO2) was easy for the Biology scenario (lowest  $\gamma$ -coefficient) and difficult the Chemistry scenario. This is explained by the fact that the students had to use both hands to manipulate the colored balls (symbolizing atoms) so handling also the remote control became more difficult. Correcting the mistakes (ERGO3) was more difficult for the Biology scenario because of frequent selection errors when students tried to select a small organ.

In both scenarios, ERG-A had a significant positive influence on the formative indicator ERGO1, showing that the ease to adjust the see-through screen and stereo glasses is influencing the comfort on the work place.

## 5 Conclusion and future work

The main contribution of this study is a measurement model for the perceived ease of use of the ARTP featuring a second order formative index with two dimensions: the quality of visual and auditory perception and the ease of interaction and collaboration. These indexes are antecedents of a reflective construct measuring the perceived ease of learning how to use ARTP. The latter could be then integrated in structural models that are based solely on reflective scales.

There are several strengths and limitations of this study. An outcome of this research is the integration of almost all items related to the perceived ease of use that were eliminated in a previous work [4] for unidimensionality and convergent validity reasons. The new measurement model includes 12 of 15 items related to the ergonomic quality. As such, it provides a wider perspective on the ergonomic quality and enables the analysis of specific usability aspects. Second, the estimation of a formative measurement model provides a more detailed information (at indicator level) shedding light on usability aspects that are critical for ARTP and a given learning scenario. Third, the formative indexes were specified and validated with a structural model that addressed all general aspects related to the perceived ergonomic quality: ease of understanding, ease of use and ease of operating with a software system. Since all variables are strongly related to the focal construct the structural model is well supporting an external validity. Up to now, there is no similar model developed for the ergonomic quality of a software system. Fourth, the model was estimated and cross validated on two different samples which enables a comparison between scenarios and makes it possible to further integrate and discuss in more detail the answers at open questions (qualitative data).

As regarding the limitations, the sample used in this study was collected from only 6 classes (3 Romanian schools), having a limited representativeness. Second, both samples are small, at limit for SEM (Structural Modeling Equation) requirements. Third, the convergent validity of the relatively measured construct is at limit (acceptable for an exploratory study). Fourth, the breadth of formative indicators is inherently limited since the evaluation questionnaire was indented to capture the main usability aspects. Fifth, there are inherent limitations since the methodology regarding formative indexes estimation and validation is not mature yet. The usability questionnaire used to collect the data was conceptualized in 2007 while the main recommendations for formative indexes development have been published only in 2008.

Based on this work we intend to develop a new evaluation questionnaire having both formative and reflective items. The questionnaire will be used for the evaluation of a new version of the Chemistry application which is currently under development.

## Acknowledgements

This work was supported by the research projects TEHSIN 503/2009 and ARiSE FP6-027039.

## Bibliography

- [1] Anderson, J.C., Gerbing, D.W. Structural Equation Modelling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin* 103(3), 411-423, 1988.
- [2] Arbuckle, J.L. *AMOS 16.0 User's Guide*. Amos Development Corporation, 2007.
- [3] Balog, A., Pribeanu, C. Developing a measurement model for the evaluation of AR-based educational systems. *Studies in Informatics and Control* 18(2), 137-148, 2009.
- [4] Balog, A., Pribeanu, C. 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, 2010.
- [5] Bach, C., Scapin, D., Obstacles and perspectives for Evaluating mixed Reality Systems Usability. *Proceedings of IUI-CADUI Conference 2004*, 72-79, 2004.
- [6] Bollen, K., Lennox, R. Conventional wisdom on measurement: a structural perspective. *Psychological Bulletin* 110(2), 305-314, 1991.
- [7] Bruhn, M., Georgi, D., Hadwich, K. Customer equity management as formative second order construct. *Journal of Business Research* 61, 1292-1301, 2008.
- [8] Cadogan, J., Souchon, A., Procter, D. The quality of market-oriented behaviors: Formative index construction. *Journal of Business Research* 61, 1263-1277, 2008.
- [9] Davis, F.D. Perceived usefulness, perceived easy of use, and user acceptance of information technology. *MIS Quaterly* 13, 319-340, 1989.
- [10] Diamantopoulos, A., Winklhofer, H. Index construction with formative indicators : an alternative to scale development. *Journal of Marketing Research* 28, 269-277, 2001.
- [11] Diamantopoulos, A. The error term in formative measurement models : interpretation and modeling implications. *Journal of Modeling in Management* 1(1), 7-17, 2006.
- [12] Diamantopoulos, A., Riefler, P., Roth, K. Advancing formative measurement models. *Journal of Business Research* 61, 1203-1218, 2008
- [13] Edwards, J., Bagozzi, R. On the nature and direction of relationship between constructs and measures. *Psychological Methods* 5(2), 155-174, 2000.
- [14] Franke, G., Preacher, K., Rigdon, E. Proportional structural effects of formative indicators. *Journal of Business Research* 61, 1229-1237, 2008.
- [15] Gabbard, J., Swann, E. Usability engineering for augmented reality: Employing user-based studies to inform design. *IEEE Transactions on Visualization and Computer Graphics* 14(3), 513-525, 2008.
- [16] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L. *Multivariate Data Analysis*, Prentice Hall, 2006.
- [17] Hornbaek, K. Current practice in measuring usability: Challenges to usability studies and research. *Int. J. Human Computer Studies*. 64, 79-102, 2006.

- [18] Huang, H.M., Rauch, U., Liaw, S.S. Investigating learners' attitude towards virtual reality learning environments: based on a constructivist approach. *Computers & Education* 55, 1171-1182, 2010.
- [19] ISO 9126-1:2001 *Software Engineering - Software product quality*. Part 1: Quality Model
- [20] Jarvis, C.B., Mackenzie, S., Podsakoff, M. A critical review of construct indicators and measurement models misspecification in marketing and consumer research. *Journal of Consumer Research* 30, 199-218, 2003.
- [21] Konradt, U., Christophersen, T., Schaefer-Kuelz, U. Predicting user satisfaction, strain and system usage of employee self-services. *Int. J. of Human-Computer Studies* 64, 1141-1153, 2006.
- [22] Krauss, M., Riege, K., Winter, M., Pemberton, L. Remote Hands-On Experience: Distributed Collaboration with Augmented Reality. *Proceedings EC-TEL 2009*, LNCS 5794, Springer, 226-239, 2009
- [23] Pribeanu, C., Balog, A., Iordache, D.D. Measuring the usability of augmented reality e-learning systems: a user-centered evaluation approach. Chapter 14: *Software and Data Technologies, CCIS 47*, Corderiro, H., Shiskov B, Ranchordas A, Helfert M (Eds.), Springer, 175-186, 2009.
- [24] Ruiz, D.M., Gremler, D., Washburn, J., Carrion, G.C. Service value revisited: specifying a high order formative measure. *Journal of Business Research* 61, 1278-1291, 2008.
- [25] Tabachnick, B. G., Fidell, L. S. . *Using Multivariate Statistics*, 5th ed. Boston: Allyn and Bacon, 2007.
- [26] Wilcox, J., Howell, R., Breivik, E. Questions about formative measurement. *Journal of Business Research* 61, 1219-1228, 2008.
- [27] Wind, J., Riege, K., Bogen M. SpinnstubeŽ: A Seated Augmented Reality Display System, *Virtual Environments: Proc. IPT-EGVE - EG/ACM Symposium*, 17-23, 2007.

### Annex 1 Constructs and items

ERG	MM	Items	Variables
Quality of visual and auditory perception (ERG-P)	F	ERGP1	Observing through the screen is clear
	F	ERGP2	The superposition between projection and the real object is clear
	F	ERGP3	Understanding the vocal explanations is easy
	F	ERGP4	Reading the information on the screen is easy
Ease of interaction and collaboration (ERG-O)	F	ERGO1	The work place is comfortable
	F	ERGO2	Selecting a menu item is easy
	F	ERGO3	Correcting the mistakes is easy
	F	ERGO4	Collaborating with colleagues is easy
Ease of adjusting devices (ERG-A)	F	ERGA1	Adjusting the "see-through" screen is easy
	F	ERGA2	Adjusting the stereo glasses is easy
	F	ERGA3	Adjusting the head phones is easy
Perceived ease of learning to operate (PEOL)	R	PEOL1	Understanding how to operate with ARTP is easy
	R	PEOL2	Learning how to operate with ARTP is easy
	R	PEOL3	Remembering how to operate with ARTP is easy
*** General item	R	PEOU1	Overall, I find the system easy to use

Note: MM(Measurement Model): F (Formative) / R (Reflective)