A model for explaining primary school students' learning outcomes when they use multi-user virtual environments

Emmanuel Fokides

Springer

CrossMark

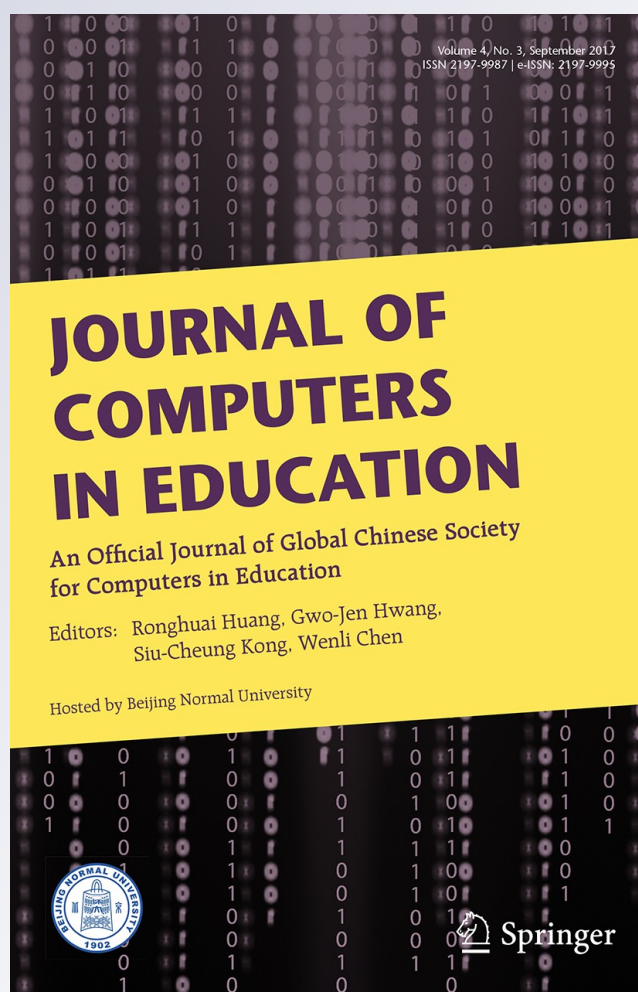# A model for explaining primary school students' learning outcomes when they use multi-user virtual environments

Emmanuel Fokides[1]

Abstract The study examines the subjective factors that affect primary school students' learning outcomes when using multi-user virtual environments (MUVEs). Eight variables were used so as to build a research model that was partially based on the models developed by Salzman et al. and the technology acceptance model. Structural equation modeling was used for parameter estimation and model testing. Self-reported data, as well as the results of three knowledge acquisition tests, were gathered from 396 sixth-grade students who used a MUVE in which the main theme was space exploration. Results analyses revealed a good model fit, and 56% of the variance in the learning outcomes was explained. Out of the 25 research hypotheses, 17 were supported. Perceived ease of use, enjoyment, motivation, and perceived learning effectiveness were the most influential factors. On the basis of these findings, it is suggested that software developers need to focus on the game-like features of their applications and that educators need to foster in-world collaboration between students.

Keywords Model · MUVEs · Primary school students · Structural equation modeling · Subjective constructs

Introduction

Various technological innovations have multiplied and diversified the ways in which students learn and interact with the learning material. One such technology, which has seen rapid development in recent years, is 3D multi-user virtual environments (MUVEs), a sub-genre of virtual reality (VR). In short, VR

✉ Emmanuel Fokides
  fokides@aegean.gr

[1] Department of Primary School Education, University of the Aegean, 1 Dimokratias str., Rhodes 85132, Greece

🗌 Springer

applications are 3D simulations of real or imaginary environments. A much-debated aspect of these applications is presence/immersion, the sense one has that he/she is present in the virtual environment and not in the real world (Rizzo et al. 1998). Depending on the hardware and software used, VR applications can vary from fully immersive (that use sophisticated equipment, such as head-mounted displays and haptic devices), to simple "low-tech" desktop applications (that use just mid-range computers) (Levin 2011). What is more, in MUVEs, multiple users can simultaneously use the same simulation and hence the term "multi-user." Thus, interactions are possible not just between the user and the virtual objects but also with other users present in the virtual world.

Constructivism provides the theoretical framework for the educational uses of VR and MUVEs (Dickey 2005). This learning theory supports the idea that learning is an active process; knowledge is constructed on the basis of what learners already understand and as they make connections between new and old information (Ertmer and Newby 2013). Collaboration between users, peer feedback, visual, and audio stimuli are but a few of VR's features that have an educational interest (Zheng and Newgarden 2011). The above, lead to—probably—the most significant benefits for education, that of incentives for learning and active learning (O'Neil et al. 2005; Mikropoulos and Natsis 2011). Together with learning activities (both in-world and in-classroom), the educational process becomes more effective (Martin et al. 2011).

Indeed, the relevant literature demonstrating the educational benefits when using VR/MUVEs in diverse educational settings is quite extensive (e.g., Merchant et al. 2014). While most studies report—more or less—satisfactory learning outcomes, there is too much speculation on why these results were achieved, with the focus being on the teaching methods that were used, on the educational settings, and on the technological aspects of the applications (Merchant et al. 2014; Mikropoulos and Natsis 2011). The problem with focusing just on the above is that a number of important key factors that affect students' learning experience when using VR/MUVEs applications are left out of the equation. Therefore, we cannot (and do not) have an in-depth understanding of their complex interactions or to the extent of which each factor affects the learning outcomes. For example, the inclusion of psychological factors is not common; the learner's psychological perspective has hardly been studied (Yaman et al. 2008). On the other hand, studies that included such factors do exist. Merchant et al. (2015) examined the undergraduate students' acceptance of MUVEs. Self-efficacy in MUVEs was correlated with course learning in another instance (deNoyelles et al. 2014). Presence had also been the factor of interest in a number of studies (e.g., Hassell et al. 2012). Self-efficacy and perceived system efficacy and their impact on the effectiveness of virtual training systems were examined by Jia et al. (2014). Vrellis et al. (2016) correlated satisfaction when using MUVEs with the learning outcomes. Self-efficacy, presence and usability were used for developing a model that tried to explain the observed learning outcomes in desktop VR by Merchant et al. (2012). Finally, Lee et al. (2010) tried to explain the learning outcomes in VR using a number of VR's features together with presence, motivation, cognitive beliefs, control, and reflective thinking.

From the above-mentioned studies, it can be concluded that there is a wide range of factors used for explaining the learning outcomes in VR/MUVEs. Then again,

these studies had very few in common; different factors were used as key determinants of the learning outcomes or of the learning experience, different types of VR/MUVEs were used, and the learning subjects were also diverse. The only common element that they had was that they were narrowly focused on young adults (mainly university students). Thus, it can be argued that:

- More research is needed toward a more comprehensive inclusion of factors or toward the inclusion of factors not yet tested.
- Results of non-multiuser VR applications are probably not applicable to MUVEs since users' collaboration was not taken into account.
- Younger ages, for example, primary school students, are still understudied.
- The sample sizes, in some cases, were rather small. Also, the learning subjects were too specialized. Therefore, the results are not easily generalized.

Taking into consideration the above ascertainments, it becomes evident that there is the need for a more comprehensive inclusion of factors that affect the learning experience and, subsequently, the learning outcomes when students use MUVEs. These factors can be either subjective/psychological constructs or certain MUVEs' features. In addition, there is the need to have a better understanding of how these factors interplay with each other. These are exactly the objectives of the study at hand. It presents the development and testing of a model for the examination of a number of factors involved in the learning experience when using MUVEs. In addition, and because of the lack of studies in younger ages, the target group was primary school students aged around 12 (sixth-grade). In the coming sections, a brief review of the literature regarding the factors affecting the learning outcomes when using MUVEs is presented, followed by the research rationale and methodology, and the results analyses. Subsequently, results are discussed, and the conclusion completes the work.

Factors affecting the learning outcomes when using MUVEs

In order to develop a model for explaining how certain factors affect, and ultimately shape, the learning experience and the learning outcomes in MUVEs, one has to determine which factors to include. Among the first to propose a model of how VR's affordances work with other factors in shaping the learning process and the learning outcomes were Salzman et al. (1999). In short, their model suggests that: (a) VR's features (e.g., multisensory cues and immersion) influence learning, (b) the concept/subject being learned moderates how VR's capabilities influence the learning process, (c) the learner's characteristics (e.g., gender, computer experience) play a role in shaping the learning process and may also interact with VR's features in influencing learning, and (d) VR's affordances, as well as the individuals' characteristics, affect both the interaction experience (e.g., usability) and the learning experience (e.g., presence, motivation), which, in turn, influence learning (Fig. 1).
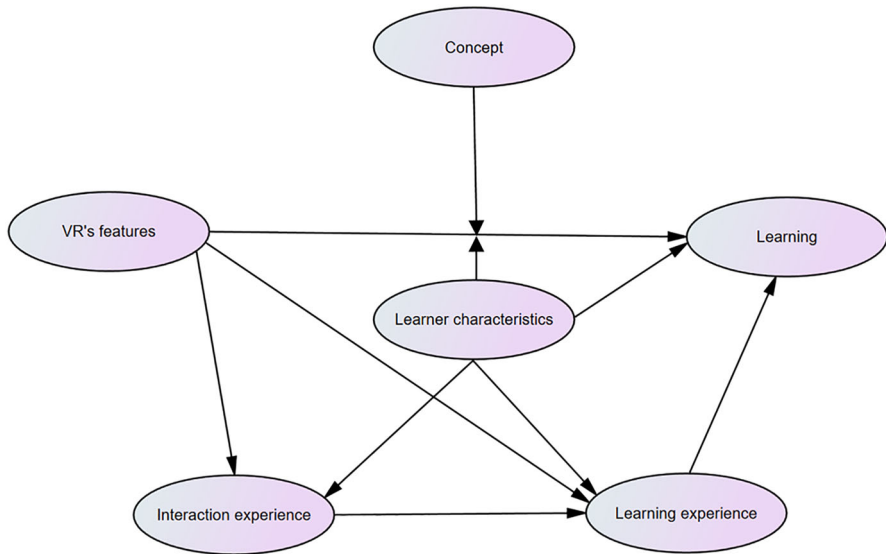
Fig. 1 Salzman et al.'s theoretical model

While this model provided useful insights on what factors can be considered for inclusion in the study's model, one has to keep in mind that it was proposed for immersive VR systems and also collaboration, a key MUVEs' feature, was absent. Therefore, other models relevant to VR and MUVEs, such as the ones presented in the previous section, were also considered. It was found that even if they had common factors (in some cases), these were not viewed in the same way. For example, MUVEs'/VR's features (e.g., realism) were defined either in terms of technology (which, in essence, is an objective view) or in terms of quality and accessibility (which is a subjective view). Dissimilar and, sometimes, contradictory views regarding a factor may lead to confusion and different interpretations of a study's findings. Therefore, it was decided that all the factors included in the proposed model would be subjective ones, as viewed through the eyes of the user and how he/she interprets them.

Widening the literature review even further, models regarding the use of other ICT tools were considered. Among them, the Technology Acceptance Model (TAM) (Davis et al. 1989) is broadly used. The TAM tries to interpret one's intentions to use technological tools and, by extension, the learning outcomes resulting from their use. This is done by modeling the causal relationships between perceived ease of use, perceived usefulness, attitude toward an ICT tool, and behavioral intention to use this tool. Summarizing the literature review, seven factors were identified that were commonly used for explaining the learning processes and outcomes when individuals, regardless of their age, use MUVEs and other ICT tools. These factors can be roughly grouped into three major categories:

MUVEs' features, affordances, and constraints

- *Perceived realism* Although from a technical perspective, the simulation's realism is the result of how detailed the virtual objects are and, in general, how close to reality their behavior is, it can also be viewed as a subjective feature, because individuals perceive it differently. Studies indicated that realism plays an important role when individuals use MUVEs/VR applications acting as a facilitator of the learning experience (Dalgarno and Lee 2010; Harrington 2012; Lee et al. 2010; Wan et al. 2007).
- *Perceived ease of use* Perceived ease of use is included in many studies that utilize the TAM (e.g., Davis et al. 1989). It is also included in models regarding VR (e.g., Lee et al. 2010). In all cases, it has been found to play a major role in one's experience when using diverse ICT tools.
- *Collaboration* One of the fundamental MUVEs' features is that they allow collaboration between users (including social interactions and peer feedback). Studies have shown that it is an important element in the learning processes that take place in MUVEs (Zheng and Newgarden 2011) as well as in other ICT applications (Gomez et al. 2010).

State of mind

- *Enjoyment* The fun and pleasure which one feels in a MUVE can be defined as the degree to which a user considers that its use is an enjoyable experience (Ducoffe 1996). Studies have shown that these positive feelings contribute to knowledge acquisition (Faiola et al. 2013; Harrington 2012; Wrzesien and Raya 2010).
- *Presence* While studies indicated that presence is affecting the learning outcomes (Bulu 2012; Mikropoulos and Natsis 2011; Lee et al. 2010), one might argue that it should be included as a factor in a MUVE only if sophisticated equipment, such as head-mounted displays, are used. Actually, presence has generated a lot of controversies. There are researchers who supported that it solely depends on the equipment used (e.g., Schuemie et al. 2001; North and North 2016). Others supported the idea that it depends on the individual's personality (e.g., Nunez 2004) and that it can be included as a factor even in desktop VR applications, having a significant impact on the learning outcomes (e.g., Hassell et al. 2012; Lee et al. 2010). Since the matter is not resolved, it was decided to include presence as a construct in the present study, even if it was planned not to use equipment that enhances presence in a MUVE.

Learning enabling features

- Perceived usefulness is also a construct in the TAM and refers to the extent to which a person believes that using an ICT tool would enhance his/her productivity and performance. It also affects the course of the learning process (Sun et al. 2008). Perceived usefulness was also considered as a construct in other studies regarding VR's impact on learning (e.g., Lee et al. 2010; Sharda et al. 2004). Since certain aspects of perceived usefulness are not directly related to learning (e.g., productivity), in this study its meaning was narrowed, and it was renamed to perceived learning effectiveness. Thus, perceived learning effectiveness is defined as the extent to which a person believes that a MUVE is a learning enabler, a facilitator of the learning process, compared to other teaching methods.
- Incentives for learning/motivation. Incentives for learning are not a static attribute, but they are inherently volatile and sensitive to the way the content is presented (Linnenbrink and Pintrich 2002). Research has indicated that the 3D presentation of the MUVE, interaction with its objects and the increased control on what the user selects to view, can influence motivation and, as a result, the learning outcomes (Benbunan-Fich and Hiltz 2003; McLellan 2004; Piccoli et al. 2001).

Research model and hypotheses

The proposed research model, shown in Fig. 2, includes the factors presented in the previous section. Since it is an untested model, and since the literature suggests diverse relationships among the factors, all possible relationships were added. The study's hypotheses were based on these relationships:

H1a–e Collaboration significantly and positively influences presence, enjoyment, motivation, perceived learning effectiveness (PLE), and the learning outcomes.

H2a–e Realism significantly and positively influences presence, enjoyment, motivation, PLE, and the learning outcomes.

H3a–e Perceived ease of use (PEU) significantly and positively influences presence, enjoyment, motivation, PLE, and the learning outcomes.

H4a–d Presence significantly and positively influences enjoyment, motivation, PLE, and the learning outcomes.

H5a–c Enjoyment significantly and positively influences motivation, PLE, and the learning outcomes.

H6a–b Motivation significantly and positively influences PLE and the learning outcomes.
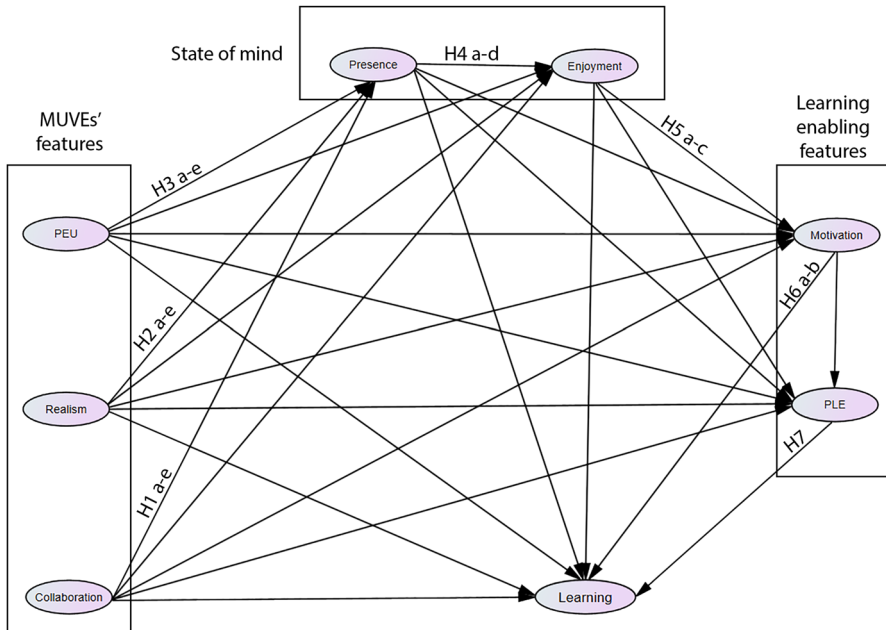
Fig. 2 The proposed model

H7   PLE significantly and positively influences collaboration and the learning outcomes.

In this study, the learning outcomes were used as the dependent variable, while all the other factors were used as independent variables. For the purposes of structural equation modeling, collaboration, realism, and PEU were considered as exogenous variables, while presence, enjoyment, motivation, and PLE were endogenous variables.

Method

To test the proposed model, a project was designed and implemented utilizing a MUVE developed and used in previous research projects. Space exploration is its theme, and it was developed using OpenSimulator (http://opensimulator.org/). The MUVE has two levels and a total of three different areas. On the ground level, the first area presents the history of rockets, from the first rockets to the space shuttle, rocket engines, and a rocket launch pad. The second area presents space suits, moon vehicles, and the Mars Rover (Fig. 3). Scripts allow interactions to take place; the user can launch the Apollo 11, ignite rocket engines, disassemble a multistage rocket, put rockets side by side and compare them. Videos and slide presentations are also included, providing more detailed information on all objects. The second level is placed high in the sky and presents man-made satellites. As in the first level,
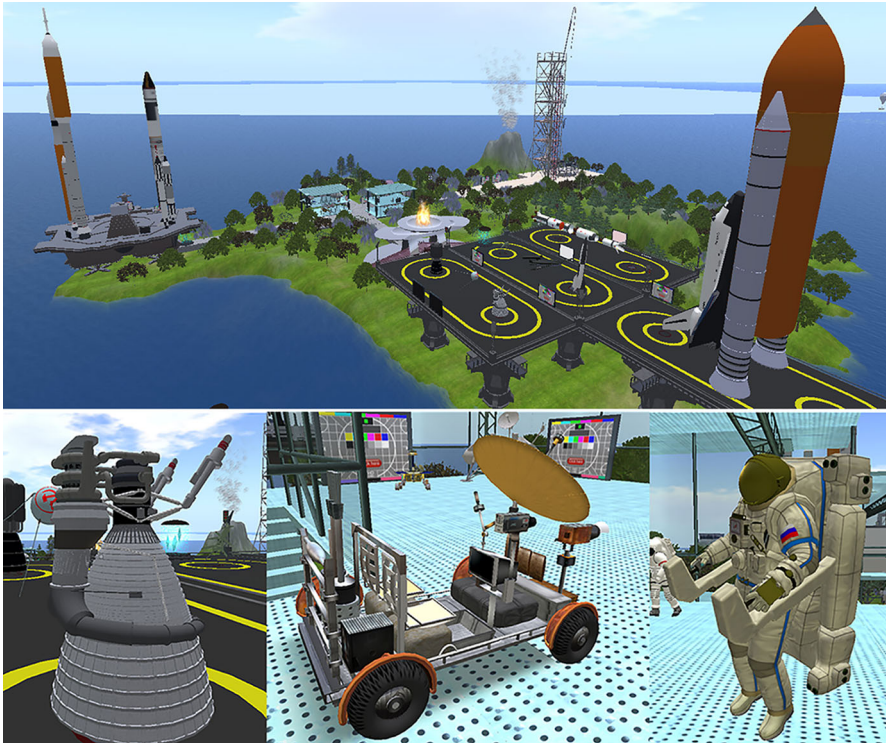
Fig. 3 Screenshots from the application

the users can interact with the objects (e.g., put satellites side by side and compare them or disassemble them and get information on each of their main parts). The learning objectives of the application are not just to provide detailed information, but to allow users to understand how they function and to make comparisons. The application was installed on a computer running as an Opensimulator server in each participating school.

The study's target group was sixth-grade primary school students (12-year-olds). Since the science curriculum at this grade (as well as at all the previous ones) does not include any subjects/lessons related to space exploration, it was assumed that prior knowledge was minimal and totally depended on what students have learned by themselves. A total of 437 students participated in the study, coming from 20 primary schools in Athens, Greece. Students used the MUVE for a total of 6 h (three two-hour sessions). In each session, they could visit one of the three application's areas, explore the environment, study the relevant material in whatever order they liked and for as long as they liked, chat with each other (through instant messaging or through voice, since the MUVE had this feature enabled), take notes, and to collaborate in-world. The classes' teachers were also present in the MUVE, guiding students to the areas of interest and answering their questions. Because of the limited number of computers, each class was divided into two halves, so that each

student used his/her own computer. This limitation resulted in having to repeat each session twice in each school.

Prior to the beginning of the project, students' parents were briefed, and their written consent for their children's participation was obtained. The sixth-grade teachers of the participating schools were also briefed, and they were explicitly asked not to intervene in terms of trying to teach students anything related to the MUVE during their regular teaching. In addition, in one two-hour session, students explored the affordances and constraints of MUVEs using another very small one without any learning material in it. This was deemed as necessary in order to proactively face difficulties when using the main application. Also, at this stage, students selected and edited their avatars. The project lasted from early September till late November 2016 (it was not implemented simultaneously in all schools).

Instruments

Immediately following the end of each session, a short test was administered to students, having 15 questions each (multiple choice, fill-in-the-blanks, right-wrong), all relevant to the material presented in the MUVE. It has to be noted that about half of the questions were about facts and figures (e.g., facts and figures about rockets and satellites), while on the other half the questions were about concepts' comprehension (e.g., how rockets and satellites function, differences, and comparisons).

Also, immediately following the end of the last session, the MUVEs learning factors scale (MLFS) (Fokides and Atsikpasi 2017) was administered to students. The MLFS is a newly developed scale, measuring the factors that were presented in a previous section. Specifically, for measuring the perceived learning effectiveness, four questions were adapted from the corresponding questions in the computer attitude scale (Selwyn 1997) measuring perceived usefulness. This scale is used and validated in a large number of studies examining intentions to use diverse technologies (including MUVEs) in an educational context (e.g., Teo and Noyes 2011; Teo and Lee 2010). The four questions measuring perceived ease of use were also the ones used in the above-mentioned scale. For measuring presence, four questions from Novak et al.'s (2000) presence questionnaire were used. Four items

Table 1 Factors' mean scores

Factor	Min	Max	*M*	SD
Presence (20)	5	19	12.40	3.01
Collaboration (15)	3	14	6.31	2.59
PLE (20)	8	20	15.80	2.91
PEU (20)	8	20	14.96	2.85
Motivation (20)	4	20	13.74	3.33
Realism (20)	5	19	12.54	2.60
Enjoyment (20)	5	20	13.85	3.03
Tests' scores/learning outcomes (45)	10	42	25.43	7.78

from the Intrinsic Motivation Inventory (McAuley et al. 1989; Tamborini et al. 2010) were used for assessing the enjoyment when using a MUVE. Witmer's and Singer's (1998) questionnaire for assessing presence provided four items for measuring perceived realism. For measuring motivation, four questions were adapted from a relevant questionnaire (Martens et al. 2007). Finally, three questions for measuring collaboration were improvised by the MLFS's authors. The scale is presented in the Appendix.

Data analyses

A number of students had to be excluded from the study because they were absent in one or more sessions. Also, questionnaires with unengaged (with no variance) responses were removed. Consequently, the final sample size was 396 cases. The distribution of boys and girls was fairly even (48 and 52%, respectively). Three scores were computed on the basis of the correct answers in the three tests. These scores were added as variables, in addition to the 27 variables of the MLFS, and formed the learning outcomes factor. Scores from items on each subscale were summed to provide individual scores on each construct (Table 1).

Exploratory and confirmatory factor analysis

Since the MLFS is a newly developed scale, it was decided to conduct an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA), in order to re-establish the underlying dimensions between the variables and the latent constructs and to re-test its convergent and discriminant validity. The study's 396 cases satisfied Tabachnick and Fidell's (2007) rule of thumb for at least 300 cases, and are considered very good according to Comrey and Lee (2013). In addition, Cattell's (1978) suggestion for a ratio of 3–6 times as many observations as variables and an absolute minimum of 250 observations was also satisfied. Skewness and kurtosis indices were small and well below the recommended level of |3| and |10|, respectively (Kline 2005). The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy index was .91, the Bartlett's Test of Sphericity was significant ($p < .001$), and the extraction communalities were above the .5 level (Tabachnick and Fidell 2007; Hair et al. 2006). Therefore, the data were well suited for factorial analysis. Principal axis factor analysis (PAF) with oblique rotation was used for assessing the underlying structure for the 30 items. PAF accounts for the covariation among variables, thus, it is suitable for structural equation modeling (SEM) that was to follow (Kline 2005), while for research involving human behaviors oblique rotation is considered to produce more accurate results (Costello and Osborne 2005).

As hypothesized, 8 factors were extracted (MLFS's 7 factors, plus the learning outcomes) using both the Kaiser's (1960) criterion (eigenvalue > 1) and the more recommended scree test (Costello and Osborne 2005) (Fig. 4). No variables were dropped, and all factors had 3 or more strong loadings (Table 2). All items loaded high on their respective factors (>.6) and each factor averaged above the .7 level, as

recommended by Hair et al. (2006). There were no significant cross-loadings between items, and there were no correlations between the factors greater than .7. The total variance explained by the 8 components was 69.70%. The reliability scores of all constructs using Cronbach's alpha was between .85 and .97, and the overall score was .92, exceeding DeVellis's (2003) guidelines (>.70). All these findings suggested that the internal consistency of the constructs and of the overall scale was satisfactory.

The resulting factor structure was inputted into AMOS 24 to perform CFA. The standardized estimates were regarded as very good (ranging from .74 to .99) (Hair et al. 2010). The items explained more than half the amount of variance of the latent variable that they belong since all of their R^2 values were above .50 (ranging from .54 to .97) (Table 3). The average variance extracted (AVE) in all cases was above the .50 level as suggested by Hair et al. (2010). Also, the measurement items were loaded with significant t-values on their theoretical constructs, and all the reflective indicators were significant at the .001 level (Table 3). Therefore, the convergent validity was satisfactory. The presence of discriminant validity was evaluated by comparing the square root of the AVE for any given factor with the correlations between that factor and all other factors. Since the variance shared between a factor and any other factor was less than the variance that the construct shares with its
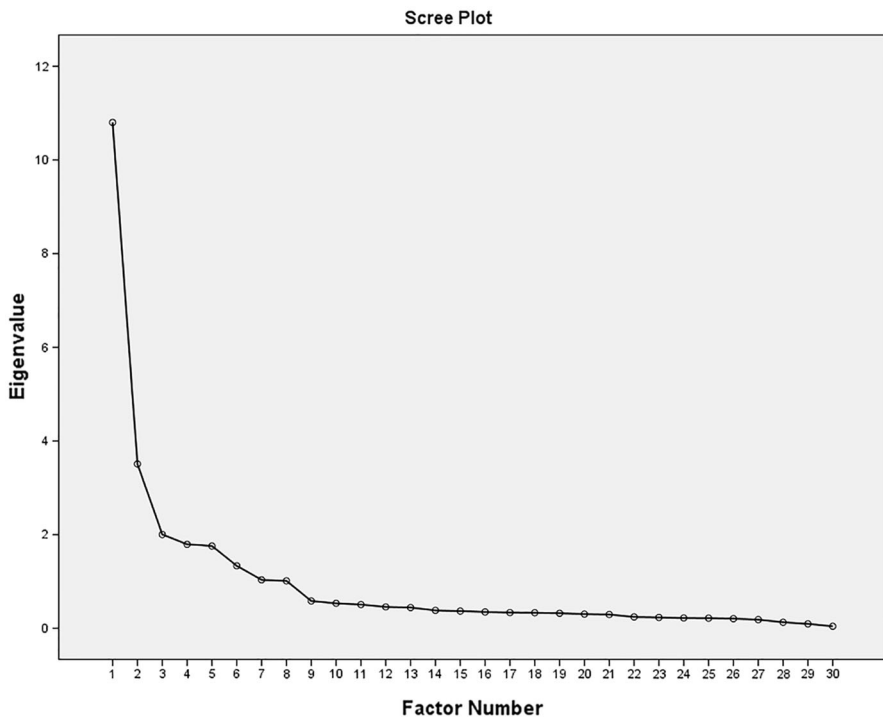

Fig. 4 Scree plot of the eigenvalues

Table 2 Exploratory factor analysis

Item	Factor loadings								Communalities
	Enj	PEU	Real	Pre	PLE	Score	Mot	Col	
Enj4	.88								.76
Enj2	.88								.83
Enj1	.78								.62
Enj3	.77								.60
PEU4		.85							.65
PEU2		.78							.68
PEU1		.76							.65
PEU3		.75							.68
Real2			.86						.69
Real1			.75						.56
Real4			.74						.61
Real3			.68						.53
Pre2				.90					.76
Pre4				.75					.58
Pre3				.75					.57
Pre1				.74					.61
PLE2					.82				.74
PLE1					.81				.71
PLE3					.76				.64
PLE4					.75				.64
Score2						.96			.93
Score1						.96			.97
Score3						.90			.88
Mot1							.83		.68
Mot2							.82		.72
Mot3							.80		.75
Mot4							.75		.70
Col2								.88	.75
Col3								.87	.78
Col1								.77	.66
Eigenvalues	10.80	3.51	2.00	1.79	1.76	1.33	1.03	1.01	
% Variance explained (Total 67.05)	35.09	10.62	5.67	5.05	4.71	3.54	2.68	2.35	
Cronbach's α	.90	.88	.85	.86	.89	.97	.91	.88	
Total = .92									

Extraction method: PAF. Rotation method: Oblique. Values < .30 are omitted for clearance of presentation

Enj enjoyment, *PEU* perceived ease of use, *PLE* perceived learning effectiveness, *Pre* presence, *Real* realism, *Mot* motivation, *Col* collaboration

Scores = tests' scores/learning outcomes

Table 3 Results for the measurement model

Item	SE	t-value	R^2	AVE
Enj4	.88	–	.77	.69
Enj2	.93	26.22	.87	
Enj1	.75	18.37	.57	
Enj3	.75	18.17	.56	
PEU4	.78	–	.61	.65
PEU2	.82	17.21	.68	
PEU1	.80	16.71	.64	
PEU3	.82	17.01	.66	
Real2	.81	–	.65	.59
Real1	.74	14.90	.54	
Real4	.79	16.04	.62	
Real3	.74	14.91	.54	
Pre2	.85	–	.73	.62
Pre4	.75	16.30	.57	
Pre3	.77	16.65	.59	
Pre1	.77	16.82	.60	
PLE2	.86	–	.74	.68
PLE1	.84	20.60	.70	
PLE3	.79	18.89	.63	
PLE4	.79	18.93	.63	
Srores2	.97	–	.93	.93
Scores1	.99	58.93	.97	
Scores3	.94	42.37	.88	
Mot1	.82	–	.66	.71
Mot2	.85	19.631	.72	
Mot3	.86	19.94	.74	
Mot4	.84	19.26	.70	
Col2	.85	–	.72	.72
Col3	.89	20.08	.78	
Col1	.81	18.41	.65	

This value was fixed at 1.00 for model identification purposes

SE standardized estimate, *AVE* average variance extracted

measures, it was concluded that the discriminant validity was satisfactory in all cases (Fornell et al. 1982) (Table 4).

Common method variance (CMV) is the variance caused by the measurement method and not by the constructs that the measures represent (Podsakoff et al. 2003). CMV was a concern since the study was based on perceptual measures at one point in time. It is commonly overlooked—or mentioned as a limitation—in research papers, but if it is present, measurement errors do occur, and the legitimacy of the conclusions is questionable. For that matter, two tests were conducted. The

Table 4 Convergent and discriminant validity

Factor	CR	AVE	Enj	PEU	Real	Pre	PLE	Scores	Mot	Col
Enj	.90	.69	(.83)							
PEU	.88	.65	.44	(.81)						
Real	.85	.59	.43	.49	(.77)					
Pre	.87	.62	.11	.13	.15	(.79)				
PLE	.89	.68	.55	.61	.57	.24	(.82)			
Scores	.97	.93	.57	.58	.43	.18	.67	(.96)		
Mot	.91	.71	.56	.73	.56	.13	.69	.63	(.84)	
Col	.88	.72	.01	.11	.15	.39	.15	.21	.05	(.85)

Diagonal in parentheses: square root of AVE extracted from observed variables. Off-diagonal: correlations between constructs

CR critical ratio, *AVE* average variance extracted

first was Harman's single-factor analysis (Podsakoff and Organ 1986). For the second, a common latent factor (CLF) was added, and the standardized regression weights were compared before and after the addition of the CLF (Gaskin 2013). There was no evidence of CMV in any factor since (a) Harman's single-factor analysis was <50% (36.01%), and (b) the standardized regression weights were not very different when adding the CLF (difference < .1, while Gaskin suggests difference < .2).

All the fit indices appeared to be good, with the exception of χ^2 (Table 5), which has the tendency to indicate significant differences in sample sizes above 200 cases (Hair et al. 2006). Therefore, this anomaly was assumed to be applicable in the present study ($N = 396$).

Structural equation modeling

For testing the fit between the initial research model (Fig. 2) and the obtained data, SEM was to be performed, using AMOS 24. The data were not imputed and a full SEM was conducted. Before conducting the analysis, it was checked whether there were outliers and influential cases and whether the multivariate assumptions were met. A Cook's distance analysis was done, and it was found that no case exhibited abnormal Cook's distance (all cases < 0.025). Linearity was checked by conducting a curve estimation for all the relationships in the model. In some cases, linearity was slightly lower than the strongest relationship between variables but still significant. Thus, it was concluded that the relationships were sufficiently linear to be tested using a covariance-based structural equation modeling algorithm such as the one used in AMOS. Multicollinearity was checked using the variance inflation factor (VIF). The highest value of VIF that was observed was 2.31, well below the recommended maximum of 3 (O'Brien 2007). Thus, the multivariate assumptions were met.

The results of the SEM analysis of the direct effects in the initial model are shown in Table 6. The shaded rows identify the effects that were not statistically

Table 5 Fit indices of the research model

	Result	Recommendation	Reference
χ^2	χ^2 (377, $N = 396$) = 606.38, $p < .001$	ns at $p < .05$	Schumacker and Lomax (2010)
χ^2/df	1.61	1–3	Kline (2005)
SRMR	.030	<.05	Klem (2000) and McDonald and Ho (2002)
TLI	.97	≥.95	Hu and Bentler (1999)
NFI	.93	≥.90	Bentler and Bonett (1980)
RMSEA	.039	<.05	McDonald and Ho (2002)
CFI	.97	≥.95	Hu and Bentler (1999)

Note. ns: not significant

significant and their path coefficients were also small. The model fit, although already satisfactory, may be improved if the non-statistically significant effects are removed from the model, resulting in a simpler and more robust final model. For that matter, all the direct effects were made optional in the model, forming a hierarchy of $2^{25} = 33\,554\,432$ models which was analyzed using the Specification Search Facility available in AMOS 24. The model with the smallest value for BCC_0 was selected as the final model ($BCC_0 = 0.00$), as suggested by Burnham and Anderson (1998). The fit statistics for the final model were satisfactory [χ^2 (385, $N = 396$) = 613.274, $p < .001$, $\chi^2/df = 1.59$, SRMR = .033, TLI = .97, NFI = .93, RMSEA = .039, CFI = .97]. A summary of the hypotheses testing results is shown in Table 7, while Fig. 5 presents the final model.

Mediation

Mediation models are used for providing a more accurate explanation for the causal effect the independent has on the dependent variable. The bootstrapping technique described by Hayes (2013) was used, and it was found that the direct effects of: (a) PEU on the learning outcomes were also mediated through enjoyment, motivation, and PLE, (b) enjoyment on the learning outcomes were also mediated through motivation and PLE, (c) motivation on the learning outcomes were also mediated through PLE. Even though realism did not have a direct effect on the learning outcomes, it had significant indirect effects through enjoyment, motivation, and PLE. Presence also had significant indirect effects, through PLE and collaboration (Table 8).

Multigroup analysis

Gender effects were tested using χ^2 difference tests (Floh and Treiblmaier 2006). The two models were freely estimated, except constraining one path at a time to be equal across groups. It was found that the χ^2 test was not significant in any case, indicating that the effects were not different for boys and girls.

Table 6 Direct effects in the proposed model

Path	Path coefficient (β)	t-value	p
Pre ← Real	.07	1.06	.290
Pre ← Col	.37	6.54	<.001
Pre ← PEU	.05	.78	.433
Enj ← PEU	.30	4.91	<.001
Enj ← Real	.29	4.74	<.001
Enj ← Col	-.09	−1.57	.117
Enj ← Pre	.06	1.10	.270
Mot ← PEU	.53	9.59	<.001
Mot ← Real	.20	3.92	<.001
Mot ← Enj	.24	5.32	<.001
Mot ← Col	−.05	−1.22	.223
Mot ← Pre	.03	0.61	.541
PLE ← PEU	.17	2.52	.012
PLE ← Mot	.34	4.43	<.001
PLE ← Real	.20	3.67	<.001
PLE ← Pre	.11	2.42	.015
PLE ← Enj	.18	3.64	<.001
PLE ← Col	.04	0.83	.409
Scores ← PLE	.34	5.39	<.001
Scores ← Mot	.17	2.29	.022
Scores ← PEU	.16	2.53	.011
Scores ← Enj	.24	5.05	<.001
Scores ← Col	.14	3.36	<.001
Scores ← Real	−.06	−1.16	.245
Scores ← Pre	−.01	−0.32	.752

Post hoc statistical power

The post hoc power analysis for unsupported direct effects hypothesis was conducted using the method described by Soper (2016). For the 7 predictors of the learning outcomes, an observed R^2 of .56 for this factor, a sample size of 396, and a probability level of .05, the observed statistical power was 1.00. This means that the proposed model had an excellent power to detect the significant effects; therefore, it is certain that the non-significant effects that were observed were truly not significant.

Discussion

High $R^2 s$ and structural paths that are significant and substantial (close to .20 and ideally above .30) are required for a model to have a meaningful predictive power (Chin 1988). On the basis of the results, a number of paths had values below this

Table 7 Hypothesis testing results

Hypotheses	Path	Path coefficient (β)	t-value	p
H1a	Col → Pre	.37	6.60	<.001
H1b	Col → Enj	Not supported		
H1c	Col → Mot	Not supported		
H1d	Col → PLE	Not supported		
H1e	Col → Learning	.13	3.47	<.001
H2a	Real → Pre	.10	1.81	.066
H2b	Real → Enj	.29	4.69	<.001
H2c	Real → Mot	.19	3.84	<.001
H2d	Real → PLE	.20	3.71	<.001
H2e	Real → Learning	Not supported		
H3a	PEU → Pre	Not supported		
H3b	PEU → Enj	.30	4.91	<.001
H3c	PEU → Mot	.52	9.57	<.001
H3d	PEU → PLE	.17	2.61	.009
H3e	PEU → Learning	.15	2.50	.012
H4a	Pre → Enj	Not supported		
H4b	Pre → Mot	Not supported		
H4c	Pre → PLE	.13	3.03	.002
H4d	Pre → Learning	Not supported		
H5a	Enj → Mot	.25	5.47	<.001
H5b	Enj → PLE	.18	3.63	<.001
H5c	Enj → Learning	.23	4.95	<.001
H6a	Mot → PLE	.33	4.41	<.001
H6b	Mot → Learning	.15	2.12	.034
H7	PLE → Learning	.32	5.40	<.001

threshold. Then again, even small interaction terms that are statistically significant are important to a model (Chin et al. 2003). Having that in mind, out of the initial 25 hypotheses, 17 were supported by the data. Besides the—more than satisfactory—R^2 of .56 for the dependent variable (learning outcomes), it was found that collaboration, perceived ease of use, enjoyment, motivation, and perceived learning effectiveness were its most significant determinants, since the paths linking them to the learning outcomes were significant ($\beta = .13$, $\beta = .15$, $\beta = .23$, $\beta = .15$, and $\beta = .32$, respectively). Thus, it can be concluded that the model adequately represents the relationships among the factors and possesses the power to explain the learning outcomes when students use MUVEs.

In addition, perceived learning effectiveness ($R^2 = .58$) was influenced by realism, perceived ease of use, presence, enjoyment, and motivation ($\beta = .20$, $\beta = .17$, $\beta = .13$, $\beta = .18$, and $\beta = .33$, respectively), while motivation ($R^2 = .63$) was significantly influenced by realism, perceived ease of use, and enjoyment ($\beta = .19$, $\beta = .52$, and $\beta = .25$, respectively). In contrast, presence ($R^2 = .16$) and
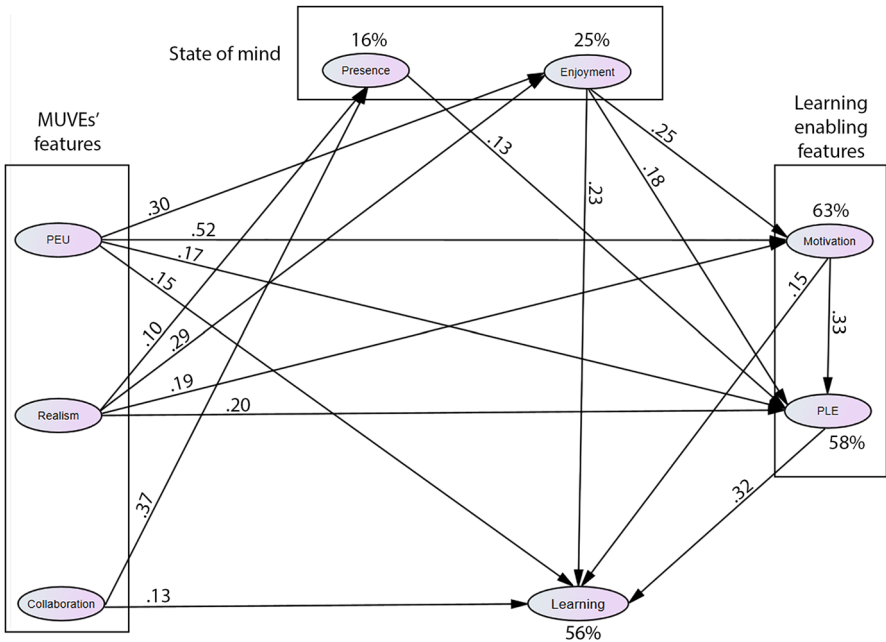
Fig. 5 The final model (non-significant paths were omitted for clearance of presentation)

enjoyment ($R^2 = .25$) were not adequately explained. Nevertheless, presence was influenced by collaboration ($\beta = .37$) and -far less- by realism ($\beta = .10$), while enjoyment was influenced by realism ($\beta = .29$) and perceived ease of use ($\beta = .30$).

It has to be noted that the correlation of the findings of the present study with that of previous research is quite difficult and venturous, for the reason that the proposed model is significantly different from others. Also, it is quite difficult to associate the findings of studies where a single factor was examined (e.g., only presence) with the findings of a model that takes into account multiple factors and their complex relationships, such as the one presented in this study. Nevertheless, it seems that the most influential factor in shaping the learning outcomes was perceived learning effectiveness, followed by enjoyment. Motivation, perceived ease of use, and collaboration were also influential but to a lesser degree. Besides having a strong direct effect ($\beta = .32$), perceived learning effectiveness also acted as a strong mediator of the effects of all the other constructs (except collaboration) (see Table 8). This finding is in contrast to the findings of other studies, where no direct effects were noted (e.g., Lee et al. 2010). Target group differences, different models, and differences in the type of technology being evaluated are plausible explanations for the discrepancy in the findings.

Motivation had a relatively low direct impact on the learning outcomes ($\beta = .15$), but a strong indirect one through perceived learning effectiveness ($\beta = .33$). It also had a strong influence on the aforementioned factor ($\beta = .33$). The role of motivation in shaping the learning outcomes is noted in research

Table 8 Mediation results

Hypothesis	Estimate	p	Mediation effect	Direct effect
Real → Enj → Learning	.29	.001	Yes	No
Real → Mot → Learning	.12	.034	Yes	No
Real → PLE → Learning	.28	.001	Yes	No
PEU → Enj → Learning	.29	.001	Yes	Yes
PEU → Mot → Learning	.32	.038	Yes	Yes
PEU → PLE → Learning	.23	.008	Yes	Yes
Pre → PLE → Learning	.14	.001	Yes	No
Enj → Mot → Learning	.12	.025	Yes	Yes
Enj → PLE → Learning	.18	.002	Yes	Yes
Mot → PLE → Learning	.33	<.001	Yes	Yes

regarding VR (McLellan 2004) as well as other ICT applications (Gomez et al. 2010; Piccoli et al. 2001). Coming to enjoyment, even if it was not adequately explained ($R^2 = .25$), it seems that it is an important construct. It has a significant direct impact on the learning outcomes ($\beta = .23$), has a quite significant impact on both the learning enabling constructs (motivation and perceived learning effectiveness), and mediates the indirect effects of realism and perceived ease of use to the learning outcomes. While other studies related to VR applications have noted its importance (Faiola et al. 2013; Harrington 2012; Wrzesien and Raya 2010), it seems that enjoyments' impact is far more important than it was previously considered.

On the other hand, the role of presence was minimal. It was the least explained construct ($R^2 = .16$), it had a relatively weak direct impact only on perceived learning effectiveness ($\beta = .13$) and an equally weak indirect effect on the learning outcomes (through perceived learning effectiveness, $\beta = .14$). To a certain extent, the feeble role of presence was expected, since, in this study, no specialized equipment was used that would have allowed increased levels of presence. While others have reported significant effects on the learning outcomes in desktop VR applications (e.g., Hassell et al. 2012; Lee et al. 2010), this is neither strongly supported by the findings of the present study nor rejected. Another interesting finding was that the MUVE's realism had a rather small impact on presence ($\beta = .10$), while collaboration had a very strong one ($\beta = .37$). Literature suggests that there are three types of presence: place presence, social presence, and co-presence (Bulu 2012). On the basis of the study's findings, it can be argued that MUVEs foster co-presence (the psychological interaction of the individuals) more than social presence (which is related to the quality of the medium and users' perception of the medium) (Schroeder 2002).

The study's findings support previous research that underlined the importance of perceived ease of use (e.g., Lee et al. 2010). It had direct ($\beta = .15$) and indirect effects on the learning outcomes and had an impact on all the other factors (except presence). It has to be noted that it had an exceptionally strong impact on motivation ($\beta = .52$) and a very strong impact on enjoyment ($\beta = .30$). The application's

perceived realism was also an influential factor, having a strong indirect effect on the learning outcomes (see Table 8) and direct ones to all the other study's constructs, with the strongest being on enjoyment ($\beta = .29$). This finding also confirms previous research noting the importance of an application's realism (Dalgarno and Lee 2010; Harrington 2012; Lee et al. 2010; Wan et al. 2007). From the above, it can be concluded that the MUVEs' realism and ease of use are the most influential factors in shaping the learners' experience. Alas, the same does not hold true for collaboration. Besides having the lowest mean score among the study's factors (see Table 1), it had only a weak impact on the learning outcomes ($\beta = .13$) and a rather strong one on presence ($\beta = .37$). This finding is a cause for concern since collaboration is a key MUVEs feature and also an important factor in shaping the learning outcomes in a MUVE (Mikropoulos and Natsis 2011; Zheng and Newgarden 2011) and in other ICT applications (Gomez et al. 2010). It seems that students were displeased, probably because, to their view, collaboration did not work well and expressed their dissatisfaction to the relevant questions. The short duration of the project (6 h in each school) is a plausible explanation; almost certainly, more time was needed so as students to get aquatinted with the software, master its use, and, among other things, start collaborating efficiently.

As a final note, the model proved to be invariant across genders, meaning that girls and boys are expected to have similar learning outcomes when using MUVEs, and the factors' relationships are also the same. There is no common consensus regarding the role of gender in relation to VR/MUVEs. For example, in a study, it was found that girls outperform boys in the learning outcomes (Nelson and Ketelhut 2008), while in another no gender differences were reported (Wrzesien and Raya 2010).

The study's findings have certain implications for practice. Software developers have to focus on the features of their applications, so as to increase their game-like characteristics because these can act as facilitators to both enjoyment and motivation, as others suggested (Kozlov and Johansen 2010; Wu et al. 2013). Also, the application's realism has to be taken into account, since it affects many of the model's constructs. On the other hand, this has to be done with caution, because increased realism, in essence, means more complex applications and may result in incompatibilities with the computers schools are equipped with. Educators need to pay more attention to the educational settings, fostering collaboration between students when they use MUVEs. This can be done by allowing more time for the use of MUVEs so that students would be able to master their use.

There are limitations to this study that have to be taken into consideration. Despite being meticulous in methodology, one can never be certain about the accuracy or honesty of the participants' answers. The data were collected from primary school students in Greece and the sample size, although adequate for statistical analysis, could have been larger. Therefore, the study's results cannot easily be generalized to other samples. The study's learning subject was space exploration; it is unknown what the results might have been if another subject was chosen. Finally, time restrictions, imposed by the schools, did not allow the project to last for more than 6 h (in each school). If students had more time at their disposal, their views might have been different.

Further validations are required for establishing the model's applicability. Presence and enjoyment require an in-depth examination, since, in this study, these two factors were not adequately explained. Additional factors can be considered that contribute in shaping the learning experience and the learning outcomes in MUVEs. Also, future studies can examine whether the model is invariant in different levels of education, or if it can be used in different types of VR applications, as well as in other types of emerging technologies (e.g., augmented reality) and, thus, increase its usefulness to the researchers.

Conclusion

A model has been developed within the theoretical framework laid out by Salzman and her colleagues and also the Technology Acceptance Model, for explaining the learning experience and the learning outcomes when students aged around 12 use MUVEs. Educators can use the model in order to devise suitable strategies for curricular integration. The model can also be used as a post-course outcome measure of the effectiveness of MUVEs. Researchers can use the model as a comparative measure of the learning experience when using MUVEs in all levels of education, as well as revealing the extent of inequalities between students according to gender, race, and efficacy in using computers. In conclusion, the study contributes to the growing body of research on the educational impact of MUVEs, and it is hoped that the model will be of use to educators and researchers alike.

Appendix

See Table 9.

Table 9 The MLFS scale

Factor	Item	
Perceived learning effectiveness	PEL1	I feel that MUVEs can ease the way I learn
	PEL2	MUVEs are a much easier way to learn compared to the usual teaching
	PEL3	Why use a MUVE? There are easier ways to learn what I want to learn*
	PEL4	MUVEs can make learning more interesting
Perceived ease of use	PEU1	Learning to use the MUVE was easy for me
	PEU2	I found the MUVE easy to use
	PEU3	Whenever I used the MUVE, I needed help because it was not easy for me to use it*
	PEU4	It was easy for me to become skillful at using the MUVE
Presence	Pr1	I forgot about my immediate surroundings when I used the MUVE
	Pr2	When I used the MUVE, I often forgot where I am
	Pr3	When I used the MUVE, the virtual world was more real than the real world

Table 9 continued

Factor	Item	
	Pr4	When I used the MUVE, I felt that my body was in the room, but my mind was inside the world created by the MUVE
Enjoyment	En1	My experience in the MUVE was quite enjoyable
	En2	I would describe my experience in the MUVE as very interesting
	En3	The experience in the MUVE was fun
	En4	I enjoyed experiencing the virtual world in the MUVE very much
Motivation	Mo1	When using the MUVE, I had the impulse to learn more about space exploration**
	Mo2	I tried to explore all the MUVE because everything was so interesting
	Mo3	I wasn't interested in learning using this type of computer program*
	Mo4	This type of computer program did not hold my attention*
Collaboration	Co1	I was displeased because it was impossible to collaborate with others in a MUVE; everyone had a mind of his own*
	Co2	It was interesting that in the MUVE I was doing things together with my fellow students
	Co3	With my fellow students, we were able to jointly decide where to go and what to do in the MUVE
Realism	Real1	The visual display quality of the MUVE distracted me from doing other things
	Real2	When interacting with the virtual objects, these interactions seemed like real
	Real3	There were times when the virtual objects became more real and present for me compared to the real ones
	Real4	The virtual objects seemed like the real objects to me

* Item for which scoring was reversed

** Replace with the learning subject of the MUVE

References

Benbunan-Fich, R., & Hiltz, S. R. (2003). Mediators of the effectiveness of online courses. *IEEE Transactions on Professional Communication, 46*(4), 298–312. doi:10.1109/TPC.2003.819639.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588. doi:10.1037/0033-2909.88.3.588.

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education, 58*(1), 154–161. doi:10.1016/j.compedu.2011.08.024.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer. doi:10.1007/978-1-4757-2917-7.

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum. doi:10.1007/978-1-4684-2262-7.

Chin, W. W. (1988). Issues and opinion on structural equation modeling. *MIS Quarterly, 22*, vii–xvi.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research, 14*(2), 189–217. doi:10.1287/isre.14.2.189.16018.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. New York: Psychology Press.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1–9.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10–32. doi:10.1111/j.1467-8535.2009.01038.x.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003. doi:10.1287/mnsc.35.8.982.

deNoyelles, A., Hornik, S., & Johnson, R. D. (2014). Exploring the dimensions of self-efficacy in virtual world learning: Environment, task, and content. *Journal of Online Learning and Teaching, 10*(2), 255.

deVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Newbury Park, CA: SAGE Publications.

Dickey, M. D. (2005). Brave new (interactive) worlds: A review of the design affordances and constraints of two 3D virtual worlds as interactive learning environments. *Interactive Learning Environments, 13*(1–2), 121–137. doi:10.1287/mnsc.35.8.982.

Ducoffe, R. H. (1996). Advertising value and advertising on the web. *Journal of Advertising Research, 36*(5), 21–35.

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 26*(2), 43–71. doi:10.1002/piq.21143.

Faiola, A., Newlon, C., Pfaff, M., & Smyslova, O. (2013). Correlating the effects of flow and telepresence in virtual worlds: Enhancing our understanding of user behavior in game-based learning. *Computers in Human Behavior, 29*(3), 1113–1121. doi:10.1016/j.chb.2012.10.003.

Floh, A., & Treiblmaier, H. (2006). What keeps the e-banking customer loyal? A multigroup analysis of the moderating role of consumer characteristics on e-loyalty in the financial service industry. *Journal of Electronic Commerce Research, 7*(2), 97–110. doi:10.2139/ssrn.2585491.

Fokides, E., & Atsikpasi, P. (2017). *Factors affecting primary school students' learning experience when using MUVEs. Development and validation of a scale*. Manuscript submitted for publication

Fornell, C., Tellis, G. J., & Zinkhan, G. M. (1982). Validity assessment: A structural equations approach using partial least squares. In B. J. Walker et al. (Eds.), *An assessment of marketing thought & practice* (pp. 405–409). Chicago: American Marketing Association.

Gaskin, J. (2013). SEM series part 5a: Confirmatory factor analysis. Retrieved from https://www.youtube.com/watch?v=MCYmyzRZnIY

Gomez, E. A., Wu, D., & Passerini, K. (2010). Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes. *Computers & Education, 55*(1), 378–390. doi:10.1016/j.compedu.2010.02.003.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: International version* (7th ed.). New Jersey: Pearson.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

Harrington, M. C. (2012). The virtual trillium trail and the empirical effects of freedom and fidelity on discovery-based learning. *Virtual Reality, 16*(2), 105–120. doi:10.1007/s10055-011-0189-7.

Hassell, M. D., Goyal, S., Limayem, M., & Boughzala, I. (2012). Effects of presence, copresence, and flow on learning outcomes in 3d learning spaces. *Administrative Issues Journal: Education, Practice, and Research, 2*(1), 62–73. doi:10.5929/2011.2.1.4.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/10705519909540118.

Jia, D., Bhatti, A., & Nahavandi, S. (2014). The impact of self-efficacy and perceived system efficacy on effectiveness of virtual training systems. *Behaviour & Information Technology, 33*(1), 16–35. doi:10.1080/0144929X.2012.681067.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151. doi:10.1177/001316446002000116.

Klem, L. (2000). Structural equation modeling. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics* (Vol. II). Washington, DC: American Psychological Association.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Kozlov, M. D., & Johansen, M. K. (2010). Real behavior in virtual environments: Psychology experiments in a simple virtual-reality paradigm using video games. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 711–714. doi:10.1089/cyber.2009.0310.

Lee, E. A. L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*(4), 1424–1442. doi:10.1016/j.compedu.2010.06.006.

Levin, M. F. (2011). Can virtual reality offer enriched environments for rehabilitation? *Expert Review of Neurotherapeutics, 11*(2), 153–155. doi:10.1016/j.compedu.2010.06.006.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313–327.

Martens, R., Bastiaens, T., & Kirschner, P. A. (2007). New learning design in distance education: The impact on student perception and motivation. *Distance Education, 28*(1), 81–93. doi:10.1080/01587910701305327.

Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., & Peire, J. (2011). New technology trends in education: Seven years of forecasts and convergence. *Computers & Education, 57*(3), 1893–1906. doi:10.1016/j.compedu.2011.04.003.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60*(1), 48–58. doi:10.1080/02701367.1989.10607413.

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64–82. doi:10.1037/1082-989X.7.1.64.

McLellan, H. (2004). Virtual Realites. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology*. Mahwah, NJ: Erlbaum Associates.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70,* 29–40. doi:10.1016/j.compedu.2013.07.033.

Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Kwok, O. M., Cifuentes, L., & Davis, T. J. (2012). The learner characteristics, features of desktop 3D virtual reality environments, and college chemistry instruction: A structural equation modeling analysis. *Computers & Education, 59*(2), 551–568. doi:10.1016/j.compedu.2012.02.004.

Merchant, Z., Keeney-Kennicutt, W., & Goetz, E. (2015). Predicting Undergraduate Students' Acceptance of Second Life for Teaching Chemistry. *Journal of Online Learning & Teaching, 11*(2)

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*(3), 769–780. doi:10.1016/j.compedu.2010.10.020.

Nelson, B. C., & Ketelhut, D. J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *International Journal of Computer Supported Collaborative Learning, 3*(4), 413–427. doi:10.1007/s11412-008-9049-1.

North, M. M., & North, S. M. (2016). A comparative study of sense of presence of traditional virtual reality and immersive environments. *Australasian Journal of Information Systems*. doi:10.3127/ajis.v20i0.1168.

Novak, T. P., Hoffman, D. L., & Yung, Y.-F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science, 19*(1), 22–42. doi:10.1287/mksc.19.1.22.15184.

Nunez, D. (2004, November). How is presence in non-immersive, non-realistic virtual environments possible? *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (pp. 83–86). ACM. doi: 10.1145/1029949.1029964

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*(5), 673–690. doi:10.1007/s11135-006-9018-6.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474. doi:10.1080/09585170500384529.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Quarterly, 25*(4), 401–426. doi:10.2307/3250989.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879. doi:10.1037/0021-9010.88.5.879.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531–544. doi:10.1177/014920638601200408.

Rizzo, A. A., Wiederhold, M. D., & Buckwalter, J. G. (1998). Basic issues in the use of virtual environments for mental health applications. In G. Riva, B. K. Widerhold, & E. Molinari (Eds.), *Virtual environments in clinical psychology and neuroscience* (pp. 22–42). Amsterdam: IOS Press.

Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence: Teleoperators and Virtual Environments, 8*(3), 293–316. doi:10.1162/105474699566242.

Schroeder, R. (2002). Copresence and interaction in virtual environments: An overview of the range of issues. In F. Ribeiro (Ed.), *Proceedings of the 5th International Workshop on Presence*, (pp. 274–295).

Schuemie, M. J., Van Der Straaten, P., Krijin, M., & Van Der Mast, C. A. P. G. (2001). Research on presence in virtual reality: A survey. *Cyberpsychology & Behaviour, 4*, 183–210. doi:10.1089/109493101300117884.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner' guide to structural equation modeling* (3rd ed.). New York: Routledge.

Selwyn, N. (1997). Students' attitudes toward computers: Validation of a computer attitude scale for 16–19 education. *Computers & Education, 28*, 35–41. doi:10.1016/S0360-1315(96)00035-8.

Sharda, R., Romano, N. C., Jr., Lucca, J. A., Weiser, M., Scheets, G., Chung, J. M., et al. (2004). Foundation for the study of computer-supported collaborative learning requiring immersive presence. *Journal of Management Information Systems, 20*(4), 31–63.

Soper, D. S. (2016). *Post-hoc statistical power calculator for multiple regression* [Software]. http://www.danielsoper.com/statcalc

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*(4), 1183–1202. doi:10.1016/j.compedu.2006.11.007.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson Education Inc.

Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M., & Organ, A. (2010). Defining media enjoyment as the satisfaction of intrinsic needs. *Journal of Communication, 60*(4), 758e777. doi:10.1111/j.1460-2466.2010.01513.x.

Teo, T., & Lee, C. B. (2010). Explaining the intention to use technology among student teachers: An application of the theory of planned behaviour (TPB). *Campus-Wide Information Systems, 27*(2), 60–67. doi:10.1108/10650741011033035.

Teo, T., & Noyes, J. (2011). An assessment of the influence of attitude and perceived enjoyment on the intention to use technology among pre-service teachers: A structural equation modelling approach. *Computers & Education, 57*(2), 1645–1653. doi:10.1016/j.compedu.2011.03.002.

Vrellis, I., Avouris, N., & Mikropoulos, T. A. (2016). Learning outcome, presence and satisfaction from a science activity in Second Life. *Australasian Journal of Educational Technology, 32*(1), 59–77. doi:10.14742/ajet.2164.

Wan, Z., Fang, Y., & Neufeld, D. J. (2007). The role of information technology in technology-mediated learning: A review of the past for the future. *Journal of Information Systems Education, 18*(2), 183–192.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225–240. doi:10.1162/105474698565686.

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55*(1), 178–187. doi:10.1016/j.compedu.2010.01.003.

Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education, 62*, 41–49. doi:10.1016/j.compedu.2012.10.024.

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education, 51,* 1784–1794. doi:10.1016/j.compedu.2008.05.009.

Zheng, D., & Newgarden, K. (2011). Rethinking language learning: Virtual worlds as a catalyst for change. *International Journal of Learning and Media, 3*(2), 13–36. doi:10.1162/ijlm_a_00067.

Emmanuel Fokides is a lecturer in the Department of Primary School Education, University of the Aegean, Greece. His courses focus on the educational uses of Virtual Reality, digital storytelling, Augmented Reality, and Serious Games. Since 1994, he is involved in a number of research projects regarding distance and lifelong learning and the educational uses of Virtual and Augmented Reality. His work is published in several conference proceedings, international volumes, and journals.