Integrating Multi-User Virtual Environments in Modern Classrooms

Yufeng Qian
Northeastern University, USA

IGI Global
DISSEMINATOR OF KNOWLEDGE

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

Chapter 9
Factors Affecting Primary School Students' Learning Experiences When Using MUVEs:
Development and Validation of a Scale

Emmanuel Fokides
University of the Aegean, Greece

Pinelopi Atsikpasi
University of the Aegean, Greece

ABSTRACT

3D multi-user virtual environments (MUVEs) are considered a technological advancement that, in the coming years, will have a significant impact on how students learn. On the other hand, the factors that shape the learning experience in such applications are not well studied. This chapter is an attempt to fill that gap. It reports the development and validation of a scale to measure the factors that come into play when primary school students use MUVEs in formal educational settings. Perceived learning effectiveness, perceived ease of use, presence, motivation, perceived application realism, interactions, enjoyment, as well as collaboration were used to develop a questionnaire that initially included 34 items. A total of 352 sixth-grade primary school students used a MUVE in formal educational settings and the aforementioned questionnaire was administered to them. The exploratory and confirmatory factor analysis revealed the existence of 7 factors and 24 items that were retained in the final version of the scale. The factor structure of the questionnaire is also discussed.

INTRODUCTION

Information and communication technologies have significantly increased the level of independence students have when learning. They have also multiplied and diversified the ways in which students can learn and interact with the learning material. One technology that attracts the interest of educators and researchers is virtual reality (VR). VR is an "umbrella" term and various sub-genres do exist, one of which is 3D multi-user virtual environments (MUVEs). In recent years, there is a growing use of MUVEs in diverse educational settings (e.g., formal and informal learning) and in almost all learning domains (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). This calls for the use of instruments and methods assessing not only the learning outcomes they yield, but also assessing the interactions between a number of key factors that affect students' learning experience when using them. By knowing how certain subjective constructs and MUVEs' features interplay with each other, we can have an in-depth understanding of how MUVEs affect the learning outcomes. This, in turn, will allow us to develop strategies to maximize the impact of the positive elements and, at the same time, to minimize the impact of the negative ones, thus increasing the odds of achieving better results in terms of knowledge acquisition.

There are certain steps in this process, the first one being the development and validation of scales to measure the factors that come into play when students use MUVEs for educational purposes. The study presents the development of a scale to measure exactly this. A number of factors, namely perceived learning effectiveness, perceived ease of use, presence, motivation, perceived application's realism, interactions, enjoyment, as well as collaboration, were included in the scale. The reasoning for selecting these factors, the research methodology, and the results of the exploratory and confirmatory factor analysis of the scale are discussed in the coming sections.

BACKGROUND

VR applications are 3D simulations of real or imaginary environments that "fool" the human senses; users have the feeling of being in a real environment (Hew & Cheung, 2010). Depending on the hardware and software used, VR can vary from fully immersive (that uses sophisticated equipment, such as head-mounted displays and haptic devices for the provision of somatosensory feedback), to simple "low-tech" desktop applications (that use just mid-range computers) (Levin, 2011). Furthermore, in MUVEs, multiple users can simultaneously use the same simulation; thus, they can interact not only with the virtual objects but also with each other. Principles drawn from constructivism provide the theoretical framework for the educational uses of MUVEs (Dickey, 2005). According to this theory, learning is an active process and knowledge is constructed on the basis of what learners already understand and as they make connections between new and old information (Ertmer & Newby, 2013). Social interaction, peer feedback, collaboration between users, visual and audio stimuli are but a few of their features that have an educational interest (Zheng & Newgarden, 2011). These, lead to, probably, the most significant benefits for education, that of incentives for learning and active learning (O'Neil, Wainess, & Baker, 2005).

There are numerous studies, in all levels of education, demonstrating the educational benefits when using VR/MUVEs. For example, in science education, the non-textually mediated presentation of the content, allows students to understand complex concepts (Squire, Barnett, Grant, & Higginbotham, 2004). Moreover, by encompassing both small and large scales, by allowing side-by-side comparisons, and by

presenting the content in multiple perspectives in time as well as in space, issues related to scientific misconceptions can be addressed (Barnett, Yamagata-Lynch, Keating, Barab, & Hay, 2005).

While most of studies related to the educational uses of MUVEs 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 and, in general, on the educational settings (Merchant et al., 2014). Fewer studies have examined which factors, other than the above, are involved and how they interact with each other. More importantly, the inclusion of psychological factors is not common. As Yaman, Nerdel, and Bayhuber (2008) noted, the learner's psychological perspective has hardly been studied in computer simulation-based learning. On the other hand, studies that examined such factors, provide useful insights. For example, Merchant, Keeney-Kennicutt, and Goetz (2015) examined the acceptance of MUVEs when teaching chemistry to undergraduate students. deNoyelles, Hornik, and Johnson (2014) correlated certain aspects of self-efficacy in MUVEs with course learning, having as a target group university students, but this time financial accounting was the learning subject. On another instance, presence was the factor of interest and the target group was, once again, university students (Hassell, Goyal, Limayem, & Boughzala, 2012). Jia, Bhatti, and Nahavandi (2014) found that self-efficacy and perceived system efficacy have an impact on the effectiveness of virtual training systems. Pre-service teachers' satisfaction in MUVEs, in correlation with the learning outcomes, was examined by Vrellis, Avouris, and Mikropoulos (2016). Merchant, Goetz, Keeney-Kennicutt, Kwok, Cifuentes and Davis (2012) used the learners' characteristics (self-efficacy and presence) and usability, in order to develop a model that tried to explain the observed learning outcomes in desktop VR. Finally, on a more systematic inclusion of factors, Lee, Wong, and Fung (2010) also developed a model to explain the learning outcomes in VR. Their model used a number of VR's features together with presence, motivation, cognitive beliefs, control, and reflective thinking, while the target group was, this time, secondary school students and the learning subject was the frog's anatomy.

The abovementioned studies had little in common. Besides having different factors as key determinants of the learning outcomes or of the learning experience, they examined different types of VR/MUVEs, and the learning subjects were also diverse. Thus, it can be argued that:

- While there are many studies examining the relationship between VR and the learning outcomes, they are either focused on the educational settings or on specific VR's features (e.g., presence). Fewer studies encompassed three or more factors that come into play in an educational VR/MUVE. Thus, more research is needed toward a more comprehensive inclusion of factors.
- Taking into account the specific characteristics of MUVEs (e.g., in-world collaboration), there is a good chance that the results of studies that examined non-multiuser VR applications are not applicable to MUVEs.
- The emphasis appears to be on young adults, while younger ages, for example, primary school students, are still understudied.
- The small the sample sizes and the specialized learning subjects do not allow the generalization of the results.

On the basis of the above, the purpose of the study was to develop a scale that will allow the examination of a wide range of factors involved in the learning experience when using MUVEs. Because of the lack of studies in younger ages, it was decided that the target group would be primary school students aged around 12 (sixth-grade).

THEORETICAL AND EMPIRICAL BASIS OF THE INSTRUMENT

In order to develop a scale for measuring the factors that affect, and ultimately shape, the learning experience in MUVEs, one has to determine which factors to include. Salzman, Dede, Loftin, and Chen (1999), were among the first to propose a model of how immersive VR's affordances work with other factors in shaping the learning process and the learning outcomes. They theorized that: (a) VR's features (e.g., multisensory cues and 3D immersion) are likely to influence learning (learning process and outcomes), (b) the concept being learned is likely to moderate how VR's capabilities influence the learning process, (c) the learner's characteristics (e.g., spatial ability, gender, computer experience) should play a role in shaping the learning process and may also interact with VR's features in influencing learning, and (d) it is likely that VR's affordances, as well as individual characteristics, affect both the interaction experience (e.g., usability) and the learning experience (e.g., presence, motivation), which, in turn, influence learning (Figure 1).

Although the model identified by Salzman and her colleagues provided useful insights on what factors can be considered for inclusion on the scale, one has to keep in mind that it was proposed for immersive VR systems and also collaboration, a key MUVEs' feature, is absent. Therefore, other technology-mediated models relevant to VR and MUVEs as well as to ICT tools, in general, were considered. It was found that these models not only use a variety of factors but even if they have common factors, these are not viewed in the same way. For example, in some cases, technology factors are illustrated in terms of technology features while in others they are illustrated in terms of quality and accessibility. However, the literature review revealed eight factors that are 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:

Figure 1. Salzman et al.'s theoretical model
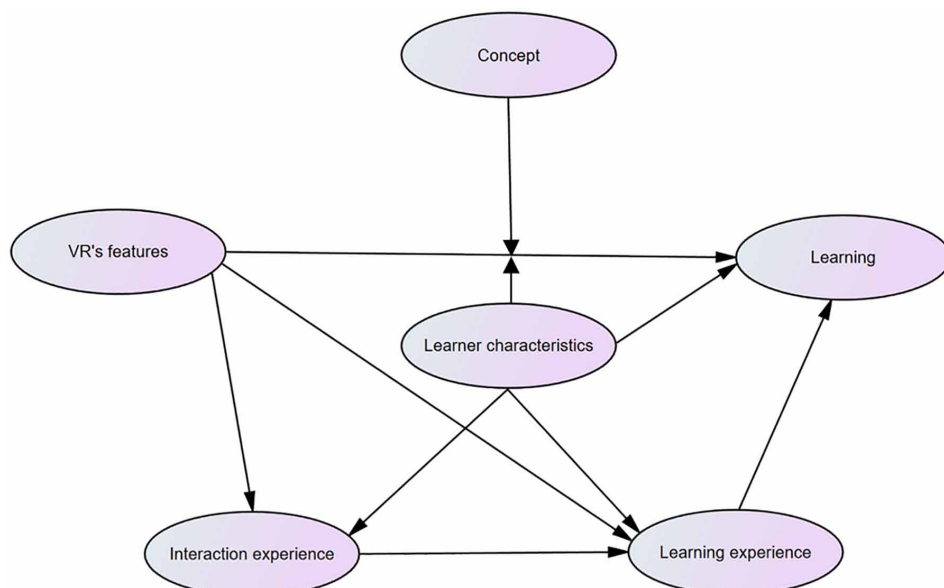
MUVEs' Features, Affordances, and Constraints

- **Perceived Realism:** From a technical perspective, the simulation's realism varies depending on how detailed the virtual objects are and, in general, how close to reality their behavior is. On the other hand, it is a subjective feature, because individuals perceive it differently and also plays an important role in one's experience when using MUVEs or VR applications in general (Dalgarno & Lee, 2010; Lee et al., 2010).
- **Interaction:** In addition to realism, increased interactions with the objects included in the simulation, also add to the user's experience (Dalgarno & Lee, 2010, Lee et al., 2010).
- **Perceived Ease of Use:** Perceived ease of use has been found to play a major role in one's experience when using diverse ICT tools (Davis, Bagozzi, & Warshaw, 1989). As a construct, it is included in a large number of studies that utilize the Technology Acceptance model (Davis et al., 1989) which tries to interpret people's intentions to use technological tools and, by extension, the learning outcomes resulting from their use. It is also included in models regarding VR (Lee et al., 2010; Sun, Tsai, Finger, Chen, & Yeh, 2008).
- **Collaboration:** As it was already mentioned, MUVEs allow social interactions, peer feedback, and collaboration between users, that are all considered as important elements in the learning processes that take place in MUVEs (Zheng & Newgarden, 2011).

State of Mind

- **Enjoyment:** The fun and, in general, the 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 the positive feelings, such as fun and enjoyment, contribute to knowledge acquisition (Harrington, 2012).
- **Presence:** Presence is defined as the sense one has in a VR application that he/she is present there and not in the real world (Rizzo, Wiederhold, & Buckwalter, 1998). Studies indicated that as a factor is affecting the learning outcomes (Bulu, 2012; Lee et al., 2010). One might argue that presence 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 debate that dates back at the first VR applications. There are researchers who supported that it solely depends on the equipment used (e.g., North & North, 2016; Schuemie, Van Der Straaten, Krijin, & Van Der Mast, 2001). Others supported the idea that it heavily depends on the individual's personality (e.g., Nunez, 2004), arguing that even reading a book can generate the feeling of presence. Studies have indicated that presence is indeed a factor in desktop VR applications and that it significantly influences the learning outcomes (e.g., 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 Technology Acceptance Model and refers to the extent to which a person believes that using an ICT tool would enhance his/her productivity and performance and affects the course of the learning process (Hong & Tam, 2006). Perceived usefulness is also considered as a construct in other studies regarding VR's impact on learning (e.g., Lee et al., 2010; Sharda et al., 2004; Sun et al., 2008). In this study, perceived usefulness was considered as perceived learning effectiveness, 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. Modern cognitive theories do not consider incentives for learning as a static attribute but as inherently volatile and sensitive to the way the content is presented (Linnenbrink & Pintrich, 2002). Researchers believe 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 (McLellan, 2004).

Having selected the factors, the next step was to develop the scale itself. For that matter, a number of questionnaires were considered that satisfied the following inclusion or exclusion criteria: (a) to have been tested and validated in studies concerning VR or MUVEs, (b) in multiple item factors (factors that were examined using multiple questions) the questions that had high loadings were selected, while (c) questions with intermediate or low loadings were excluded, and (d) questions coming from single variable factors (factors that were examined using only one question) were considered for inclusion only if their loadings were exceptionally high.

Specifically, five questions regarding perceived learning effectiveness 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 & Lee, 2010; Teo & Noyes, 2011). The five questions measuring perceived ease of use were also the ones used in the above-mentioned scale. For measuring presence, five questions from Novak's, Hoffman's and Yung's (2000) presence questionnaire were used. The Intrinsic Motivation Inventory is a multidimensional measurement for assessing participants' subjective experience of enjoyment related to a given activity (McAuley, Duncan, & Tammen, 1989; Tamborini, Bowman, Eden, Grizzard, & Organ, 2010). Four items from this scale were used for assessing the enjoyment when using a MUVE. Witmer's and Singer's (1998) questionnaire for assessing presence, provided a total of seven questions for measuring perceived realism and interaction. For measuring motivation, four questions were adapted from two relevant questionnaires (Martens, Bastiaens, & Kisrcher, 2007; McAuley et al., 1989). Finally, four questions for measuring collaboration were improvised, since no questionnaires that satisfied the inclusion criteria were found.

The pool of questions was translated into Greek by three groups. Each group consisted of one psychologist and one computer science professional with experience in MUVEs, both experts having proficiency in the English language. The resulting three different versions were back-translated into English and then viewed by another group of five experts. A unified version was obtained through a consensus meeting held with these experts aiming to assess the semantic adaptation and, thus, the initial version of the MUVEs learning factors scale (MLFS) was formulated, having a total of 34 questions.

METHOD

To confirm the factorial structure of MLFS and for validating it, a pilot project was designed and implemented. For that matter, it was decided to use a MUVE that was developed and used in previous research projects. Space exploration and the related facts and concepts are the themes of this MUVE, which was developed using OpenSimulator (http://opensimulator.org/). There are two levels in the MUVE. On the ground level, the technology behind space exploration is presented. A wealth of highly detailed 3D models is included; from the first rockets to the space shuttle, space suits, rocket engines, a rocket launch pad, moon vehicles, the Mars Rover, to name but a few (Figure 2). Scripts allow interactions to take place; the user can launch the Apollo 11, ignite rocket engines, disassemble a multistage rocket and see its stages, put rockets side by side and compare them, teleport from one place to another. 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. Gravity is set to zero so that users can float in space. As in the first level, the users can interact with objects (e.g., put satellites side by side and compare them or disassemble them and get information on each of their main parts).

As already mentioned, the study's target group was sixth-grade primary school students (12-year-olds). A total of 385 students participated in the study, coming from 18 primary schools in Athens, Greece. It has to be noted that these schools were selected out of a larger number of schools that responded af-

Figure 2. Screenshots from the application
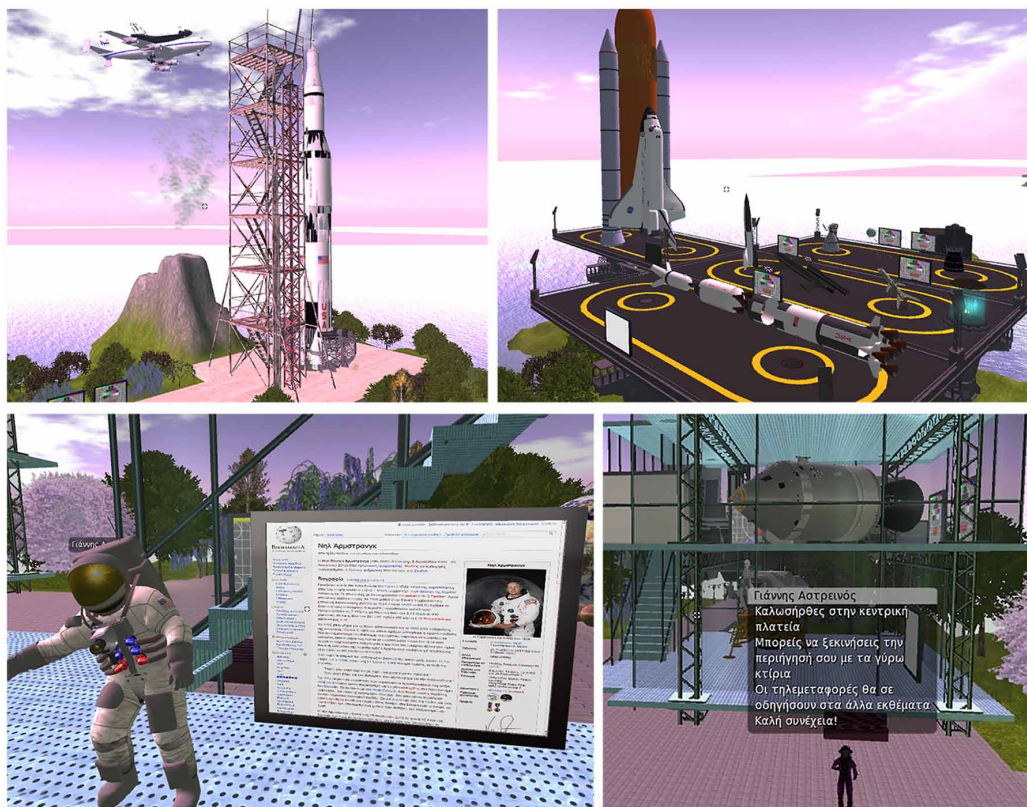
firmatively to the email invitation that was sent to them. Prior to the beginning of the project, students' parents were briefed about the project, its methodology, and objectives. Their written consent for their children's participation was obtained. Written consent was also obtained from the schools' headmasters. 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 or to try to assist them in any way other than providing technical assistance when needed.

Students used the MUVE for a total of eight hours (four two-hour sessions). They were free to explore the environment, talk to each other, and to collaborate in-world. Since the study's objective was to record the students' learning experience in a MUVE, no in-classroom teaching intervention took place. Immediately following the end of the fourth session, MLFS was administered to students. The scale was presented as a list of the 34 items, alternately displayed so that no two items from the same construct appear adjacently, alongside a 5-point Likert scale (worded "Strongly Agree", "Agree", "Neutral", "Disagree" and "Strongly Disagree"). Students were advised that the scale was a survey, not a test and that there were no correct or incorrect answers. They were asked to indicate whether they agree or disagree with each statement and to answer as honestly as possible. Scores were obtained by allocating numerical values to responses: "Strongly Agree" is scored 5, "Agree" is scored 4; "Neutral" is scored 3; "Disagree" is scored 2 and "Strongly Disagree" is scored 1.

The project lasted for about four months, from early February to early June 2016 (it was not implemented simultaneously in all schools).

RESULTS ANALYSIS

Prior to conducting any statistical analysis, all questionnaires were checked for missing data and unengaged responses (cases with no variance in their responses). The number of valid questionnaires left after the initial screening was 352. Since the MLFS was based on translated and adapted versions of questions from multiple sources and a number of questions had to be improvised, an Exploratory Factor Analysis (EFA) had to be conducted in order to establish the underlying dimensions between the variables and the latent constructs. The initial sample was randomly split into sub-samples with 250 cases each, multiple times and, for each, EFA was conducted. The sub-samples' sizes satisfied Cattell's (1978) rule for a sample to variable ratio of 3-6:1 and an absolute minimum of 250 observations. For assessing the underlying structure of the 34 items in the initial version of MLFS, principal axis factor analysis (PAF) with oblique rotation was used. That is because, PAF accounts for the covariation among variables (Kline, 2005) and oblique rotation is considered to produce more accurate results for research involving human behaviors (Costello & Osborne, 2005).

In all cases, the EFA suggested 7 independent and distinct underlying constructs (one less than it was hypothesized). Also, in all cases, some items (4 to 7) had to be dropped, either because they loaded very low on their respective factor or because they were highly correlated with more than one factor. In the coming paragraph, the most strict solution is reported.

The data were well suited for factorial analysis, since the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy index was .89, the Bartlett's Test of Sphericity was significant ($p < .001$), and the extraction communalities were well above the .5 level (Hair, Black, Babin, Anderson, & Tatham, 2006; Tabachnick & Fidell, 2007) (Table 1). Seven factors were extracted using both the Kaiser's (1960) criterion (eigenvalue > 1) and the more recommended scree test (Costello & Osborne, 2005) (Table 1, Figure

Table 1. Principal axis factor analysis of the retained items

Item	Factor Loadings							Communalities
	Enj	PEU	PLE	Pre	Real	Mot	Col	
Enj2	.93							.87
Enj4	.87							.77
Enj1	.76							.63
Enj3	.76							.59
PEU4		.85						.66
PEU2		.79						.67
PEU1		.77						.63
PEU3		.74						.67
PLE2			.86					.74
PLE1			.82					.69
PLE3			.73					.63
PLE4			.73					.62
Pre2				.92				.80
Pre3				.79				.61
Pre4				.74				.57
Pre1				.74				.64
Real2					.86			.70
Real1					.76			.57
Real4					.73			.59
Real3					.71			.56
Mot1						.85		.69
Mot3						.79		.75
Mot2						.77		.69
Mot4						.70		.68
Col2							.87	.70
Col3							.84	.74
Col1							.74	.63
Eigenvalues	9.20	3.66	1.88	1.79	1.53	1.29	1.02	
% variance explained (Total 67.05)	32.90	12.33	5.80	5.33	4.50	3.57	2.62	
Cronbach's α Total = .91	.90	.88	.89	.87	.85	.90	.86	

Notes. 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

3). These constructs were named as follows: enjoyment (Enj), perceived ease of use (PEU), perceived learning effectiveness (PLE), presence (Pre), realism (Real), motivation (Mot), and collaboration (Col). Interaction was merged with realism, as it will be further elaborated in the coming section. Seven items had to be dropped, all the retained 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-
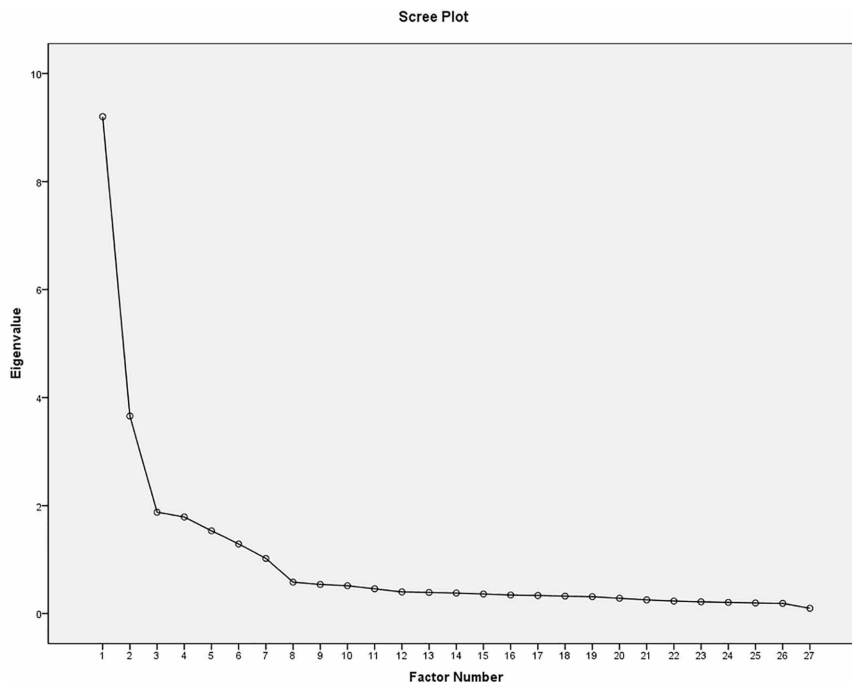
loadings between the retained items and there were no correlations between the factors greater than .7. The total variance explained by the 7 components was 67.05%. The reliability scores of all constructs using Cronbach's alpha was between .85 and .90 and the overall score was .91 (Table 1), suggesting that the internal consistency of the constructs and of the overall scale was satisfactory (deVellis's, 2003).

The final version of the MLFS with the 27 retained items and construct representations is shown in the Appendix.

The resulting factor structure was inputted into AMOS 24 to perform Confirmatory Factor Analysis (CFA). This time, the whole sample was used. Table 2 and Figure 4 show the results of the CFA. The standardized estimates ranged from .73 to .95 and were regarded as very good (Hair, Black, Babin, & Anderson, 2010). All of the R^2 values were above .50, suggesting that items explained more than half the amount of variance of the latent variable that they belong. All the fit indices appeared to be good, with the exception of χ^2 (Table 3). It has to be noted that χ^2 is too sensitive when the sample size exceeds 200 cases. If so, there is a great tendency for it to indicate significant differences (Hair et al., 2006). Therefore, this anomaly was assumed to be applicable in the present study ($N = 352$).

For assessing convergent validity, the average variance extracted (AVE) was measured and it was also checked whether the measurement items were loaded with significant *t*-values on their theoretical constructs. The AVE in all cases was above the .50 level as suggested by Hair et al. (2010). In addition, all the reflective indicators were significant at the .001 level (Table 2). 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 measures, as Fornell, Tellis, and Zinkhan (1982) suggested, it was concluded that the discriminant validity was satisfactory in all cases (Table 4).

Figure 3. Scree plot of the eigenvalues

Table 2. Results for the measurement model

Item	SE	*t*-value	R^2	AVE
Enj2	.95	-	.90	.70
Enj4	.90	27.78	.81	
Enj1	.76	18.99	.57	
Enj3	.73	17.97	.54	
PEU4	.84	-	.61	.64
PEU2	.81	15.59	.67	
PEU1	.86	16.07	.64	
PEU3	.83	15.92	.66	
PLE2	.87	-	.74	.67
PLE1	.79	16.76	.69	
PLE3	.75	18.84	.62	
PLE4	.79	16.83	.63	
Pre2	.78	-	.76	.64
Pre3	.80	16.98	.62	
Pre4	.82	15.90	.56	
Pre1	.81	17.17	.63	
Real2	.83	-	.68	.60
Real1	.73	14.25	.53	
Real4	.78	15.27	.60	
Real3	.75	14.75	.57	
Mot1	.83	-	.65	.69
Mot3	.79	17.86	.73	
Mot2	.86	19.44	.70	
Mot4	.79	18.67	.69	
Col2	.81	-	.66	.68
Col3	.87	16.57	.75	
Col1	.79	15.55	.62	

Notes. - This value was fixed at 1.00 for model identification purposes. SE: standardized estimate. AVE: average variance extracted.

Finally, scores from items on each subscale were summed to provide individual scores on each construct. The 27 individual scores were also collectively summed to provide a total score representing the individuals' overall learning experience in MUVEs (ranging from 0 to 135) (Table 5). As a normative guide to interpretation, the scores obtained with a sample of 352 sixth-grade primary school students gave cut-off scores: at the 25th percentile of 80; at the 50th percentile of 88; and at the 75th percentile of 97 (with an overall range of scores from 47 to 122). Thus, a score below the 25th percentile (80) can be interpreted as a relatively negative experience, whereas a score above the 75th percentile (97) can be interpreted as a relatively positive learning experience when using a MUVE (Table 5).
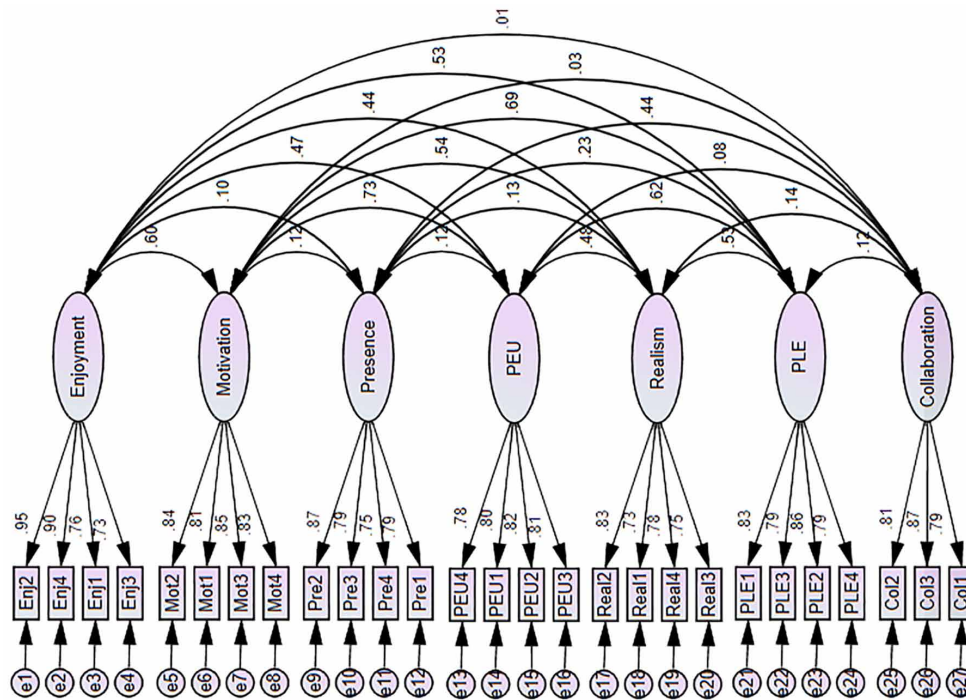
Figure 4. Results of the CFA

Table 3. Fit indices of the research model

	Result	**Recommendation**	**Reference**
χ^2	$\chi^2 (303, N = 352) = 547.95, p < .001$	ns at $p < .05$	Schumacker & Lomax, 2010
χ^2/df	1.81	1 - 3	Kline, 2005
SRMR	.035	< .05	Klem, 2000; McDonald & Ho, 2002
TLI	.953	≥ .95	Hu & Bentler, 1999
NFI	.914	≥ .90	Bentler & Bonett, 1980
RMSEA	.048	<.05	McDonald & Ho, 2002
CFI	.96	≥ .95	Hu & Bentler, 1999

Note. ns: not significant

SOLUTIONS AND RECOMMENDATIONS

Out of the initial 34 items and the inclusion of 8 factors in the original scale, 27 items were retained belonging to 7 factors in the final version. It has to be emphasized that the wording was kept as simple as possible since the questionnaire is addressed to young students. This necessity together with the fact that almost all questions were originally in English (with the exception of the questions measuring perceived collaboration, that were improvised), might have indeed the question "Real2: When interacting with the virtual objects, these interactions seemed like real," although it was supposed to measure interaction, it

Table 4. Convergent and discriminant validity

Factor	CR	AVE	Enj	Mot	Pre	PEU	Real	PLE	Col
Enj	.90	.70	(.84)						
Mot	.90	.69	.60	(.83)					
Pre	.88	.64	.10	.12	(.80)				
PEU	.88	.64	.47	.73	.12	(.80)			
Real	.86	.60	.44	.55	.13	.48	(.77)		
PLE	.89	.67	.53	.69	.23	.62	.53	(.82)	
Col	.86	.68	.005	.03	.44	.08	.14	.12	(.82)

Notes. CR: Critical ratio. AVE: Average Variance Extracted. Diagonal in parentheses: square root of AVE extracted from observed variables. Off-diagonal: correlations between constructs

Table 5. Factors' totals and percentiles

Factor	min	max	M	SD	Percentiles		
					25	50	75
Presence (20)	5	19	12.32	3.04	10	12	14
Collaboration (15)	3	14	6.20	2.47	4	6	8
PLE (20)	8	20	15.68	2.90	14	16	18
PEU (20)	8	20	14.77	2.83	13	15	17
Motivation (20)	4	20	13.59	3.26	11	14	16
Realism (20)	5	19	12.41	2.61	10	13	14
Enjoyment (20)	5	20	13.83	3.05	12	14	16
Total (135)	47	122	88.80	12.89	80	88	97

Note. Maximum score for each factor is reported in parenthesis

proved to be the question with the second strongest loading in perceived realism. The other two questions that were also supposed to measure interaction also loaded to realism, but their loadings were rather small ($< .50$) and it was decided to drop them. As a result, interaction is not represented as a construct in the final scale. Though other studies examined these two factors separately (e.g., Dalgarno & Lee, 2010; Lee et al., 2010), the findings of the present study suggest that it can be merged with realism. A plausible explanation is that since no specialized equipment was used, that allowed increased interactions, these were viewed as part of the application's realistic representation of the virtual environment.

In general, Hair et al.'s (2006) recommendation for high items' loading ($> .60$) and high factors' average ($> .7$) was strictly followed, leading to the exclusion of 7 items in total. Even so, in the final scale, there are no factors having less than 3 items as suggested by Raubenheimer (2004). Actually, only one factor (collaboration) is measured using 3 items, while all the other factors are measured with 4. Yet, the total variance explained by the 27 items was 67.05%, which is more than satisfactory (Hair et al., 2010). Therefore, it can be argued that MLFS is a quite balanced scale since no factor is overrepresented. Moreover, MLFS's reliability and internal consistency, as a whole and per construct, was well above the .70 threshold (85 to .91) which is considered "acceptable" in most social science research situations

(deVellis's, 2003). The same holds true for its convergent and discriminant validity since no problems were noted during the CFA. Thus, it can be concluded that MLFS seems to be a quite robust scale and short, in terms of how many items it has, and, thus, it can be administered in 5 to 10 minutes.

Collaboration had the lowest mean score ($M = 6.20$, $max = 15$), well below the mid-point. Apparently, collaboration did not work well when students used the MUVE and expressed their dissatisfaction to the relevant questions. This finding has to be further studied since other scholars indicated that it is an important factor in shaping the learning outcomes in a MUVE (Zheng & Newgarden, 2011). The short duration of the project in each school (8 hours) is a plausible explanation. Almost certainly, more time was needed so as students to get aquatinted with the software, master its use, and start collaborating efficiently.

Presence and the application's realism had the second and third lowest mean scores respectively. Since no specialized equipment was used, that could have enhanced the sense of presence, this finding was expected. While the results confirm that it is a factor in a MUVE, as others had noted (e.g., Lee et al., 2010), the relatively low mean score suggests that its role might not be that important compared to other factors. The same applies for perceived realism. It was confirmed that it is a factor, as previous literature suggested (Dalgarno & Lee, 2010; Lee et al., 2010), but of a lesser importance. MUVEs require well-equipped computers (in terms of hardware) in order to run smoothly. Alas, about the half of the participating schools had mid- to low-range computers. Accordingly, the settings regarding the quality of the simulation were set to low, and this had a negative impact on students' perception of the application's realism.

The mean scores of incentives for learning and enjoyment were well above the mid-point. The presence of enjoyment as a factor in the scale confirmed Harrington's (2012) views that fun and enjoyment contribute to the overall learning experience when using MUVEs. The same applies for motivation. It seems that students are motivated to learn when using VR/MUVEs applications as others had previously noted (e.g., McLellan, 2004).

Finally, the factors with the highest mean scores were perceived ease of use and perceived learning effectiveness. Given that students are "digital natives" (Prensky, 2001), it was expected that they would not face significant problems when using the application and this was reflected in their answers to the relevant questions. As for perceived learning effectiveness, it has to be reminded that it was a modified version of perceived usefulness. It seems that students formed the view that MUVEs are useful (as facilitators of their learning), confirming previous research on the importance of usefulness in one's experience when using MUVEs (e.g., Lee et al., 2010; Sharda et al., 2004).

Research has demonstrated that VR and MUVEs applications can produce satisfactory results when they are used for educational purposes. While this holds true, there is no common consensus on why individuals learn in these environments. Though studies scrutinizing the impact of certain factors on the learning outcomes do exist, each analyzed either one or a different set of factors and each used different instruments for validating these factors. On the other hand, in this study, by reviewing the relevant literature, seven subjective factors were located and a scale for measuring them was developed and tested. The scale's considerable internal consistency, stability, and validity are indicators that, indeed, these factors are involved in one's learning experience when using MUVEs. Thus, the study's contribution to research is that it proposes an instrument for measuring multiple factors at the same time. Even more, the scale is flexible; it can be used in a variety of VR/MUVEs application, by excluding certain factors, without altering the validity of the instrument. That is because the questionnaire's items load quite strongly on their respective factors and with minimal cross-correlations. For example, the scale can be easily used

in a single-user VR application, by excluding the items measuring collaboration, since in this type of applications collaboration among users cannot take place.

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. Time restrictions, imposed by the schools, did not allow the project to last for more than 8 hours (at each school). If students had more time at their disposal, their views might have been different. Finally, all of the questionnaire's items measured subjective factors (e.g., perceived realism, perceived ease of use, enjoyment, etc.). Thus, perceived learning effectiveness, which, in essence, is the perception of how much one thinks that he/she had learned when using MUVEs, does not verify/guarantee that he/she has actually learned something. Consequently, the scale, without the inclusion of items measuring knowledge acquisition, cannot be used for measuring the objective learning outcomes.

FUTURE RESEARCH DIRECTIONS

Further validations are required for establishing the scale's validity and applicability. Perceived realism and perceived interactions require an in-depth examination, since, in this study, these two factors were merged. Additional factors can be considered that contribute in shaping the learning experience in MUVEs. Also, future studies can examine whether the scale can be used in different levels of education, 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. The full potential of the scale can be realized if, in future studies, the subjective factors that it measures are compared with data that measure the actual learning outcomes (e.g., through knowledge acquisition tests). By doing so, it would be possible to make multiple comparisons, examine how the factors interact with each other and with the learning outcomes, and draw useful conclusions. For example, it would be possible examine if and how subjective factors influence the learning outcomes or if (and to what extent) perceived learning differs from actual learning.

CONCLUSION

Within the theoretical framework laid out by Salzman and her colleagues, a scale has been developed for measuring a number of factors involved in the learning experience of students (aged around 12) when they use MUVEs. The scale consists of seven factor analytically distinct subscales with high internal consistency, stability, and validity. Those teaching in primary schools can use the scale to devise suitable strategies for curricular integration. The scale can also be used as a post-course outcome measure of the effectiveness of MUVEs. Researchers can use the scale as a comparative measure of the learning experience when using MUVEs in primary 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 scale will be of use to educators and researchers.

REFERENCES

Barnett, M., Yamagata-Lynch, L., Keating, T., Barab, S. A., & Hay, K. E. (2005). Using virtual reality computer models to support student understanding of astronomical concepts. *Journal of Computers in Mathematics and Science Teaching*, *24*(4), 333–356.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. 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

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

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 / MERLOT*, *10*(2), 255.

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

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:10.1080/10494820500173714

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

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

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

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, *41*(1), 33–55. doi:10.1111/j.1467-8535.2008.00900.x

Hong, S. J., & Tam, K. Y. (2006). Understanding the adoption of multipurpose information appliances: The case of mobile data services. *Information Systems Research*, *17*(2), 162–179. doi:10.1287/isre.1060.0088

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*(1), 141–151. doi:10.1177/001316446002000116

Klem, L. (2000). Structural equation modeling. In L. Grimm & P. Yarnold (Eds.), *Reading and understanding multivariate statistics* (Vol. 2). Washington, DC: American Psychological Association.

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

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:10.1586/ern.10.201 PMID:21306202

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

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 PMID:2489825

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 PMID:11928891

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 and Teaching / MERLOT*, *11*(2).

North, M. M., & North, S. M. (2016). A comparative study of sense of presence of traditional virtual reality and immersive environments. *AJIS. Australasian Journal of Information Systems*, *20*. 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? In *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'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *Curriculum Journal*, *16*(4), 455–474. doi:10.1080/09585170500384529

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1–6. doi:10.1108/10748120110424816

Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, *30*(4), 59–64. doi:10.4102/sajip.v30i4.168

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 (Cambridge, Mass.)*, *8*(3), 293–316. doi:10.1162/105474699566242

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 & Behavior*, *4*(2), 183–210. doi:10.1089/109493101300117884 PMID:11710246

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*(1), 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., & Sleezer, C. M. (2004). Foundation for the study of computer-supported collaborative learning requiring immersive presence. *Journal of Management Information Systems*, *20*(4), 31–63. doi:10.1080/07421222.2004.11045780

Squire, K., Barnett, M., Grant, J. M., & Higginbotham, T. (2004, June). Electromagnetism supercharged!: Learning physics with digital simulation games. In *Proceedings of the 6th International Conference on Learning Sciences* (pp. 513-520). International Society of the Learning Sciences.

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, MA: Pearson Education.

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), 758–777. 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

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

Yaman, M., Nerdel, C., & Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education*, *51*(4), 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

KEY TERMS AND DEFINITIONS

Multiuser Virtual Environments (MUVEs): 3D virtual environments where multiple users can simultaneously use the same simulation.

Perceived Ease of Use: The subjective view that an application or an ICT tool can be handled/used with no problems.

Perceived Realism: The view one has regarding how close to reality a VR application is.

Perceived Usefulness: The extent to which a person believes that using an ICT tool would enhance his/her productivity and performance.

Presence: The sense one has in a virtual reality application that he/she is present there and not in the real world.

Scale: Concept, device, or procedure used in arranging, measuring, or quantifying events, objects, or phenomena.

Virtual Reality: A technology that allows a user to interact with a computer-simulated environment, whether this environment is a simulation of the real world or an imaginary world.

APPENDIX

Table 6. 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
	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

Table 6. Continued

Factor	Item	
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

Notes. * = Item for which scoring was reversed; ** = Replace with the learning subject of the MUVE