



D5.2: ITEM ANALYSIS RESULTS



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 732420.

This communication reflects only the author's view. It does not represent the view of the European Commission and the EC is not responsible for any use that may be made of the information it contains.

Task 2.1- Definition of main topics and performance indicators (construction of metrics)

Contract number:	732420
Project acronym:	e-Confidence
Project title:	Confidence in behaviour changes through serious games
DELIVERY DATE	31/08/2018
Author(s):	Barbara Rončević Zubković, Tamara Martinac Dorčić, Svjetlana Kolić-Vehovec, Sanja Smojver-Ažić.
Partners contributed:	P5 – USAL
Date:	27/07/2018
Version:	1.0
Abstract:	This document presents the psychometric characteristics of questionnaires applied in the post-test. Internal consistency and test-retest reliability were examined.
Status:	<p>X PU (Public)</p> <ul style="list-style-type: none"> • PP Restricted to other programme participants (including the Commission Services) • Restricted to a group specified by the consortium (including the Commission Services) (please specify the group) • Confidential, only for members of the consortium (including the Commission Services)



1. DOCUMENT REVISION LOG

VERSION	DATE	DESCRIPTION	AUTHOR
0.1	17/07/2018	The draft contains information about psychometric analysis of the scales	FHSS
0.2	23/07/2018	The suggestions from USAL are accepted or discussed.	
1.0	27/07/2018	Final version	FHSS



2. CONTENT

1. DOCUMENT REVISION LOG	3
2. CONTENT.....	4
3. INTRODUCTION	5
4. FACTOR ANALYSIS.....	6
5. INTERNAL CONSISTENCIES OF THE SCALES.....	9
6. TEST-RETEST RELIABILITIES OF THE SCALES.....	12
7. CONCLUSIONS	14
8. REFERENCES	16



3. INTRODUCTION

The psychometric properties of the questionnaires applied in the pre-test were already examined and showed in D4.2. The results showed that most scales applied for the measurement of the variables have adequate reliability. Here, the psychometric properties of the questionnaires applied in the post-test are presented.

Reliability is essential part of the psychometric properties of a measuring instrument. It refers to consistency of measurements: from time to time (test-retest), from form to form (parallel forms), from item to item (internal consistency), or from one rater to another (rater-to-rater) (Knapp & Mueller, 2010). While in pre-test we analysed internal consistency of the scale, in the post-test we can also analyze test-retest reliability.

Internal consistency reliability (Cronbachs' alpha) is the most commonly used indicator of the reliability. Its rationale is that the individual items of the scale should all be measuring the same construct and be highly intercorrelated (Hair, Black, Babin, & Anderson, 2010).

The purpose of test-retest reliability is to assess measurement consistency of one instrument from one time point to another. The correlation between the two sets of measurements is the most common way to assess test-retest reliability (Knapp & Mueller, 2010). In the social and behavioural sciences, reliability coefficients over .70 are often considered acceptable, while Cronbachs' alpha coefficient may decrease to .60, with minimum item to total correlation of .50 (Hair et al., 2010).

Parallel forms reliability and rater-to-rater reliability are not applicable to eConfidence measurement instruments, as we had only one form of each instrument and all answers were given by students on standardized scales.

Therefore, the internal consistencies and test-retest reliabilities of the scales in the post-test are presented and internal consistencies are compared to pre-test indicators.

For the Game Experience Questionnaire and Game Quality scale, that were not used in the pre-test, but were adapted from already existing questionnaires, a factor analysis was run in order to explore component structure, and check whether the structure of Game Experience Questionnaire is aligned with the one obtained in the original survey.



4. FACTOR ANALYSIS

In analysing component structure of Game Experience Questionnaire we followed the procedure described by authors of the original scale (Poels, deKort, & IJsselsteijn, 2007). We used the concise version of the Game Experience Questionnaire (GEQ; IJsselsteijn, de Kort, & Poels, 2013): In-game version, but extended with one additional item for each component in order to increase reliability. The exploratory factor analysis was conducted on entire sample in order to obtain the structure that could be applicable across different samples (Spanish and English speaking students, students that played different games). Principal axis factoring with oblique (oblimin) rotation was applied. The analysis yielded three factors with eigenvalues greater than 1. The pattern matrix is shown in Table 4.1.

Table 4.1. Results of exploratory factor analysis for Game Experience Questionnaire

Original dimension	Item	Factor		
		1	2	3
flow	I was fully occupied with the game.	.69		
immersion	I felt that I could explore things.	.68		
immersion	I was interested in the games' story.	.68		
challenge	I thought it was hard.	.67		-.39
positive affect	I thought it was fun.	.67	-.41	
challenge	I felt challenged.	.67		
positive affect	I enjoyed.	.66		
challenge	I had to put a lot of effort to it.	.65		
immersion	I felt imaginative.	.62		
negative affect	I felt bored.	-.54	.46	
flow	I lost track of time.	.52		
flow	I forgot everything around me.	.49		
tension	I felt irritable.		.75	
tension	I felt frustrated.		.74	



tension	I felt annoyed.		.66	
negative affect	It gave me a bad mood.		.66	
negative affect	I found it tiresome.		.53	
competence	I felt skilful.			.64
competence	I was good at it.			.54
competence	I felt successful.			.47
positive affect	I felt good.			.41

Note. Factor loadings above .30 are presented

The first factor was comprised of items that are mostly related to positive experiences during game play. It consisted of 12 items that had factor loadings higher than .48. The first factor comprised items that relate to immersion, flow, challenge, and positive affect. Five items loaded high (more than .40) solely on the second factor that related to experiences of tension and negative affect. Two items loaded on both factors but were saturated with the first factor to a higher degree (boredom negatively, and fun positively). The third factor related to feelings of competence, and four items from original competence scale and one item from positive affect scale loaded on that factor.

The obtained structure is not unexpected because positive and negative experiences are usually distinguished. However, the four original scales are grouped together, so obtained factor solution is not very informative. We also run the exploratory factor analysis on separate samples but the obtained structure was somewhat different when taking into account language (Spanish and English speaking) and game (School of Empathy and Go online), but seven factors have not emerged. We also tried to replicate the original structure of Game Experience Questionnaire by determining the number of factors to 7, but the factors did not align with the original structure. However, in order to make the results more informative and comparable to other studies we decided also to explore the internal consistencies of the original scales.

The Cronbachs' alpha coefficients for the three scales that were computed based on obtained factors, as well as for each dimension of the original scale will be show in the following chapter.

We also explored the component structure of Game Quality scale. The exploratory factor analysis was conducted on entire sample. Principal axis factoring with oblique (oblimin) rotation was applied. The analysis yielded two factors with eigenvalues greater than 1. The pattern matrix is shown in Table 4.2.



Table 4.2. Results of exploratory factor analysis for Game Quality Scale

Original dimension	Item	Factor	
		1	2
User interface aesthetics	The characters are appealing.	.88	
Satisfaction	I would recommend this game to my peers.	.82	
User interface aesthetics	The game design is attractive.	.73	
Operability	My character in the game reacted promptly to command.	.51	
Satisfaction	I prefer learning with this game to learning through other ways.	.51	
Learnability	The instructions in the game are clear and easy to understand.		.79
Accessibility	The fonts used in the game were easy to read.		.58
Learnability	The feedback in the game was helpful.	.35	.56
Operability	It happened that this game stopped unexpectedly.		

The first factor consisted of items assessing user interface aesthetics, satisfaction and operability. We name that factor Game satisfaction. The second factor related to learnability and accessibility, and we named that factor Learnability. The item *It happened that this game stopped unexpectedly* was not loaded on either of the obtained factors, even when the item was reversely coded. The internal consistency coefficients are calculated for each factor and for the entire scale, omitting the item related to unexpected stopping of the game, as it reduced reliability from .84 to .79.



5. INTERNAL CONSISTENCIES OF THE SCALES

The internal consistency indices obtained for each subscale and dimension of Game Experience Questionnaire and Game Quality Scale in Spanish and English speaking sample are presented in the Table 5.1.

Table 5.1. Reliability (internal consistency) of the Game Experience Questionnaire and Game Quality Scale

			Cronbach's Alpha	
			Spanish sample	English speaking sample
GAME EXPERIENCE				
Obtained factor structure	Positive experiences (1 st factor)	11 items	.89	.90
	Negative experiences (2 nd factor)	6 items	.77	.83
	Competence (3 rd factor)	4 items	.65	.80
Original factor structure	Immersion	3 items	.78	.84
	Tension	3 items	.83	.77
	Competence	3 items	.56	.71
	Flow	3 items	.79	.70
	Negative Affect	3 items	.64	.66
	Positive Affect	3 items	.80	.92
	Challenge	3 items	.76	.69
GAME QUALITY				
Game satisfaction (1 st factor)		5 items	.81	.85
Learnability (2 nd factor)		3 items	.76	.76
Total		8 items	.83	.85



As can be seen from Table 5.1., subscales of the Game Experience Questionnaire that were constructed based on obtained factor structure show satisfactory reliabilities. The first subscale was named Positive experiences, the second Negative experiences, and the third Competence. The item related to boredom was included in the second subscale because it referred to negative experience so it was more related to content of the second factor. Additionally it reduced the reliability of the first scale that was related to positive experiences.

The analysis of internal consistencies of the original dimensions of Game Experience Questionnaire showed that in most cases Cronbach's alpha values are satisfactory, especially considering that scales consist of only 3 items. One exception is competence scale in Spanish sample with Cronbach alpha .56. Obtained reliabilities suggest that the scales could be used as proposed by original factor structure.

The Cronbach's alpha values of the Game Quality subscales that were computed according to the results of the factor analysis also show satisfactory internal consistency on both samples. The scale can also be used as a whole (consisting of eight items), as it shows good internal consistency on both samples.

The internal consistency indices obtained for the variables from Theory of planned behaviour (TPB) and personal variables (Assertiveness, Social skills, Empathy, and Friendship) Spanish and English speaking sample are presented in the Table 5.2.

Table 5.2. Reliability (internal consistency) of the scales assessing TPB and personal variables

		Cronbach's Alpha	
		Spanish sample	English speaking sample
INSTRUMENTS FOR MEASURING VARIABLES ADAPTED FROM THE TPB			
TPB in bullying situations			
Attitudes toward bullying	6 items	.57	.53
	4 items	.61	.61
Attitudes toward protecting the victim	6 items	.86	.82
	5 items	.91	.88
Control	3 items	.27	.39
Self-efficacy	3 items	.84	.76



Subjective norms	3 items	.63	.64
Behavioural intention	3 items	.75	.84
TPB in safe use of the internet			
Attitude	4 items	.80	.89
Control	3 items	.75	.83
Self-efficacy	3 items	.78	.85
Subjective norms	3 items	.72	.81
Behavioural intention	3 items	.80	.83
PERSONAL VARIABLES			
Assertiveness	8 items	.42	.48
Empathy	6 items	.76	.90
Social Skills	8 items	.81	.93
Friendship	3 items	.48	.77

As can be seen from Table 5.2., in most cases Cronbach's alpha values are greater than .70, so we can conclude that the most of the scales have satisfactory reliabilities, which means that the items in the scale measure the same construct. Similar coefficients are obtained in the pre-test (see D4.2 *Pre-test data*).

However, some scales have lower internal consistency: Attitudes toward bullying, Control and Subjective norms related to protecting the victim, Assertiveness, and Friendship show values lower than .70. Similar results were obtained in the pre-test. When two items are omitted from the scale Attitudes toward bullying the internal consistency increases to .61. The internal consistency of Subjective norms related to protecting the victim is higher than .60 so it can be considered acceptable. Since the internal consistency of the Control related to protecting the victim is very low (.27 and .39) so we cannot treat it as a scale, but one item will be used in subsequent analysis. The results that we have obtained for the Assertiveness scale (CABS) are similar to those obtained in the study by Mesa et al. (2013) who also obtained low levels of reliability. However, that scale is analyzed separately for each type of reaction (assertive, aggressive, and passive) in subsequent analyses.



6. TEST-RETEST RELIABILITIES OF THE SCALES

The test-retest reliabilities of the scales are shown in Table 6.1. Spearman correlations are calculated because most of the variables are not normally distributed. Test-retest reliabilities are calculated only on control groups because it is not appropriate to calculate test-retest on the whole sample when design of the study includes experimental intervention. It is expected that only participants that are not involved in intervention treatment would give the similar answers on two occasions. Here, the results for knowledge variables are also shown. It is expected that those scales do not have high internal consistency because they assess different aspects of knowledge, but the results on these scales should remain on the same level in control group in both measurement points, i.e. they should have high test-retest reliability.

Table 6.1. Reliability (test-retest) of the scales

	Test-retest correlations	
	Spanish sample	English speaking sample
INSTRUMENTS FOR MEASURING VARIABLES ADAPTED FROM THE TPB		
Behaviour in bullying situations		
Bullying behaviour	.32**	.50**
Victimization	.52**	.70**
Knowledge in bullying situations		
Recognizing bullying situations	.61**	.22
Recognizing appropriate reactions	.33*	.56**
TPB in bullying situations		
Attitudes toward bullying	.22	.26
Attitudes toward protecting the victim	.19	.50**
Control	-.16	.51**
Self-efficacy	.36**	.59**
Subjective norms	.42**	.56**



Behavioural intention	.45**	.65**
Knowledge about safe use of internet		
Recognizing appropriate behaviours	.60**	.60**
TPB related to safe use of internet		
Attitude	.29*	.62**
Control	.26	.63**
Self-efficacy	.42**	.60**
Subjective norms	.21	.62**
Behavioural intention	.31*	.65**
PERSONAL VARIABLES		
Assertiveness	.46**	.32*
Aggressiveness	.60**	.56**
Passiveness	.31*	.54**
Empathy	.43**	.44**
Social Skills	.37**	.43**
Friendship	.70**	.47**

Obtained correlations are mostly significant, but do not exceed .70, which indicates that test-retest reliability of the applied scales could not be considered as acceptable. Furthermore, some coefficients are quite different in two samples (Spanish vs. English speaking). It can be concluded that results on measures are not very stable in time. The time between measurement points was quite long (5-7 months) and different uncontrollable factors (e.g. students' social experiences) could have affected the change in the observed variables even though the participants have not been included into experimental intervention. Furthermore, taking into account that the questionnaire was long and that children were already tired at the end of the school year, it is possible that children were not interested in answering the same questionnaire again and responded without carefully reading the items.



7. CONCLUSIONS

In this document the psychometric properties related to reliability of the questionnaires applied in the post-test are presented. Internal consistencies and test-retest reliabilities of the scales in the post-test are calculated. A factor analysis was also run in order to explore component structure of Game Experience Questionnaire and Game Quality scale that were not used in the pre-test.

The exploratory factor analysis was conducted on entire sample in order to obtain the structure that could be applicable across different samples. Principal axis factoring with oblique (oblimin) rotation was applied. The analysis for Game Experience Questionnaire yielded three factors with eigenvalues greater than 1. The first factor was comprised of items that are mostly related to positive experiences during game play, the second factor related negative experiences: tension and negative affect, and the third factor related to feelings of competence. The scales constructed based on obtained factors have satisfactory reliabilities. However, we also calculated the internal consistencies of the original dimensions. They have satisfactory internal consistencies; one exception is competence scale in Spanish sample with Cronbach alpha .56. Obtained reliabilities suggest that the scales could be used as proposed by original factor structure in order to discriminate different experiences and get more informative results.

The exploratory factor analysis for the Game Quality scale yielded two factors with eigenvalues greater than 1. The first factor consisted of items assessing user interface aesthetics, satisfaction and operability, and we named it Game satisfaction. The second factor related to learnability and accessibility and we name it Learnability. The item *It happened that this game stopped unexpectedly* was not loaded on either of the obtained factors. The Cronbach's alpha values of the Game Quality subscales show satisfactory internal consistency on both samples. The scale can also be used as a whole (omitting one item), as it shows good internal consistency on both samples.

Most of the scales that have already been applied in the pre-test have satisfactory reliabilities. However, some scales have lower internal consistency: Attitudes toward bullying, Control and Subjective norms related to protecting the victim, Assertiveness, and Friendship show values lower than .70. Similar coefficients are obtained in the pre-test (see D4.2 *Pre-test data*). When two items are omitted from the scale Attitudes toward bullying the internal consistency increases to .61. The internal consistency of Subjective norms related to protecting the victim is higher than .60 so it can be considered acceptable. Since the internal consistency of the Control related to protecting the victim is very low and cannot be treated as a scale, only one item will be used in



subsequent analysis. The Assertiveness scale (CABS) is analyzed separately for each type of reaction (assertive, aggressive, and passive) in subsequent analyses.

Test-retest reliabilities are calculated only on control groups because it is not appropriate to calculate test-retest on the whole sample when design of the study includes experimental intervention. Obtained correlations are mostly significant, but do not exceed .70, which indicates that test-retest reliability of the applied scales could not be considered as acceptable. It can be concluded that results on measures are not very stable in time. The time between measurement points was quite long and different uncontrollable factors (e.g. students' social experiences) could have affected the change in the observed variables even though the participants have not been included into experimental intervention.



8. REFERENCES

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2010). *Multivariate data analysis*. (7th edition). Pearson Prentice Hall.

IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). *The Game Experience Questionnaire*. Eindhoven: Technische Universiteit Eindhoven.

Knapp, T. R., & Mueller, R. O. (2010). Reliability and validity of instruments. In G. R. Hancock & R. O. Mueller (Eds.) *The reviewer's guide to quantitative methods in the social sciences* (p.337-342). New York: Routledge.

Poels, K., de Kort, Y. A. W., & IJsselsteijn, W. A. (2012). Identification and categorization of digital game experiences: A qualitative study integrating theoretical insights and player perspectives. *Westminster Papers in Communication and Culture*, 9(1), 107-129.

