Computers in Human Behavior 40 (2014) 133–143

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Measuring flow in gamification: Dispositional Flow Scale-2

Juho Hamari [a,b,*], Jonna Koivisto [a,1]

[a] Game Research Lab, School of Information Sciences, FIN-33014 University of Tampere, Finland
[b] Department of Information and Service Economy, Aalto University School of Business, P.O. Box 21220, 00076 Aalto, Finland

ARTICLE INFO

Article history:

Keywords:
Flow
DFS-2
Gamification
Games for health
Exergames
Persuasive technology

ABSTRACT

This paper measures flow in the context of gamification and investigates the psychometric properties of the Dispositional Flow Scale-2 (DFS-2). We employ data gathered from users of an exercise gamification service ($N = 200$). The results show that the original DFS-2 factorial structure does result in a similar model fit as the original work. However, we also present a factorial respecification that satisfies more recent model fit thresholds. Beyond validating the original DFS-2 instrument in the context of gamification, the psychometric analysis and the respecifications suggest that the components of flow divide into highly correlated conditions of flow (which were also found to be more salient in the context of gamification: *autotelic experience, balance of skill and challenge, control, clear goals,* and *feedback*) and into possible outcomes (*merging action-awareness, concentration, loss of sense of time*, and *loss of self-consciousness*) from achieving flow.

1. Introduction

The flow state has been widely used to describe an optimal experience characterized as a state of being fully focused and engaged in an activity (Csíkszentmihályi, 1975, 1990). Csíkszentmihályi (1990) initially defined flow as an experience, which is likely to occur when the demands of the task and the abilities of the performer are balanced. In other words, the individual performs at the height of their skills, and the task is optimally challenging (Nakamura & Csíkszentmihályi, 2002). Since then, the theoretical constitutes and measurement instruments for flow have become more refined. Current factorial structures for flow measurement have converged to contain nine dimensions (Csíkszentmihályi, 1990; Fu, Su, & Yu, 2009; Jackson & Eklund, 2002), although the constructs might slightly differ qualitatively between models. Flow is regarded as an especially pertinent experience for challenging activities where individual skill and concentration are important, such as in sports (Csíkszentmihályi, 1990; Jackson & Eklund, 2002) and in games (Hsu & Lu, 2004). The Dispositional Flow Scale-2 (Jackson & Eklund, 2002) is currently one of the most used measurement instruments for flow. Thus far, the psychometric properties of the Dispositional Flow Scale-2 (DFS-2) have been investigated in both the context of exercise (Jackson & Eklund, 2002; Kimiecik & Jackson, 2002) as well as in video gaming (Procci, Singer, Levy, & Bowers, 2012; Wang, Liu, & Khoo, 2009). The psychometric properties of DFS-2 have been found to be acceptable in both contexts; albeit more varied in the context of video gaming.

Recently, the contexts of exercise and video gaming have been increasingly converging through introduction of game design into a variety of exercise and health related activities. In general terms, this development is referred to as "gamification" (Deterding, Dixon, Khaled, & Nacke, 2011; Hamari, 2013; Hamari, Huotari, & Tolvanen, 2014; Hamari & Lehdonvirta, 2010; Huotari & Hamari, 2012; McGonigal, 2011), which refers to computers/technology being used to affect people's motivations and behavior through game-like systems. Due to gamification and the general proliferation of video games in contemporary society, people are believed to be increasingly engaging in activities that are more likely to induce the flow state. Therefore, through gamification of common activities, being in a flow state could be increasingly relevant with respect to everyday experiences. Especially, activities requiring perseverance and commitment from the individual, such as exercise, ecological consumption and education have widely been the targets of using computer technologies in changing human behavior (e.g. gamification and persuasive technologies – Hamari, Koivisto & Sarsa, 2014; Hamari, Koivisto & Pakkanen, 2014). Computer-supported gamified services such as Nike+, Zombies, Run!, Fitocracy, and Runkeeper all aim at structuring, supporting and motivating the exercise activities (see Hamari & Koivisto, 2013; Koivisto & Hamari, 2014) through the provision of optimally

* Corresponding author at: Game Research Lab, School of Information Sciences, FIN-33014 University of Tampere, Finland. Tel.: +358 50 318 6861.

E-mail addresses: juho.hamari@uta.fi (J. Hamari), jonna.koivisto@uta.fi (J. Koivisto).

[1] Tel.: +358 45 126 5525.

http://dx.doi.org/10.1016/j.chb.2014.07.048
0747-5632/

difficult challenges and feedback. Therefore, gamified exercise is an interesting avenue for further flow research in addition to the separate contexts of exercise and gaming which both have been dominant veins of flow research. However, currently there are no studies investigating flow in the aforementioned computer-supported gamified exercise and examining the applicability of current measurement instruments in this setting. Furthermore, for the purposes of study of flow in gamification, validation of measurements instruments for furthering this vein of study is required.

This paper seeks to contribute to the vein of literature examining the psychometric properties of flow measurement instruments (Jackson & Eklund, 2002; Procci et al., 2012; Wang et al., 2009). In this paper particularly, we investigate the psychometric properties of the Dispositional Flow Scale 2 (henceforth DFS-2) in the context of the computer-supported gamification. We test model fit indices, validity and reliability of the original DFS-2 model by Jackson and Eklund (2002) and the respecification made by Procci et al. (2012). Furthermore, we present our respecification of the scale in the studied context. Findings of the study regarding the validity of the DFS-2 model in gamification of exercise are discussed.

2. Background

2.1. Flow

In his seminal work, Csíkszentmihályi (1990) studied athletes, artists and professionals in various fields in order to understand the flow experience and its antecedents. Based on his findings, he defined the following conditions for flow: (1) Perceived challenges of the activity match and stretch the capabilities of the individual, thus producing an experience of being fully engaged in the task and acting on the height of his/her skills (Csíkszentmihályi, 1990; Nakamura & Csíkszentmihályi, 2002). (2) The goals of the activity are explicit and reachable, and one receives instant feedback for his/her progress on the activity (Nakamura & Csíkszentmihályi, 2002). To further describe the state of flow, Csíkszentmihályi (1990) defined the nine dimensions of flow which are common to the experience: (1) a balance between the challenge of the task and skills of the individual, (2) a merging of action and awareness, i.e. one performs the activity almost "automatically", (3) clear perceived goals, (4) unambiguous feedback, (5) focusing on the task at hand, (6) a sense of control of the activity, (7) a loss of self-consciousness or a reduced awareness of self, (8) time transformation, i.e. sense of time becomes distorted, and (9) an autotelic, intrinsically rewarding experience, implying that the activity in itself is a reason for performing it instead of any external objectives (Csíkszentmihályi, 1990; Nakamura & Csíkszentmihályi, 2002). These nine dimensions have served as a basis for several measurement instruments (e.g. Fu et al., 2009; Jackson & Eklund, 2002) for capturing the flow state and the conditions for experiencing it.

In contexts related to human behavior and computers, flow has most notably been studied in human–computer interaction (Webster, Trevino, & Ryan, 1994), video games (Hsu & Lu, 2004; Procci et al., 2012; Wang et al., 2009; Weibel, Wissmath, Habegger, Steiner, & Groner, 2008), instant messaging (Lu, Zhou, & Wang, 2009), mobile technologies (Chen, 2006), web sites in general (Skadberg & Kimmel, 2004), and game-based learning (Admiraal, Huizenga, Akkerman, & Dam, 2011; Fu et al., 2009).

2.2. DFS-2

Some of the most widely used measurement instruments for flow have been developed by Jackson and colleagues (Jackson & Eklund, 2002; Jackson & Marsh, 1996; Jackson, Martin, & Eklund, 2008). The Flow State Scale (Jackson & Marsh, 1996) was designed to examine the flow experience in a given situation, while the Dispositional Flow Scale (DFS) (Jackson et al., 2008) assesses the tendency of experiencing flow (Jackson & Eklund, 2002). Both of the models build on the nine dimensions of flow by Csíkszentmihályi (1990). Jackson and Eklund (2002) further refined the models and replaced some of the initial items of the DFS. The resulting version was titled DFS-2.

Since then, the DFS-2 has been widely applied especially to the study of various physical activities, education, arts and digital gaming. In the context of physical activity, the scale has been employed for study of flow in, for example, recreational outdoor activities (Whitmore & Borrie, 2005), physical education (González-Cutre, Sicilia, Moreno, & Fernández-Balboa, 2009), among elite athletes (Hodge, Lonsdale, & Jackson, 2009), and various sports (Crust & Swann, 2013; Gucciardi, Gordon, & Dimmock, 2009; Jackson & Eklund, 2002; Jackson et al., 2008; Koehn, Morris, & Watt, 2013; Nicholls, Polman, & Holt, 2005). In other settings, the DFS-2 has been applied for study of flow, for example, among musicians and music students (Fritz & Avsec, 2007; Sinnamon, Moran & O'Connell, 2012) and in studying (Cermakova, Moneta, & Spada, 2010).

Most pertinently to the scope of this paper, the model fit of the DFS-2 has been investigated in the contexts of online games (Wang et al., 2009) and video gaming (Procci et al., 2012). In their study, Wang et al. attempted to validate the adequacy of the instrument in internet gaming. However, the DFS-2 did not provide an appropriate model fit (Wang et al., 2009). Procci et al. (2012) sought to validate the scale in a video gaming setting. The hierarchical model did not yield sufficient model fit in their study either. Therefore, Procci et al. (2012) attempted to respecify the model, but the respecification also failed to meet the set standards for acceptable goodness-of-fit. See Table 1 for the fit indices of the studies by Wang et al. (2009) and Procci et al. (2012).

It should be noted, however, that both Wang et al. (2009) and Procci et al. (2012) had set the thresholds for Comparative Fit Index (henceforth CFI) and Non-Normed Fit Index (henceforth NNFI) at 0.95, which is considered an excellent level for the index (Hu & Bentler, 1999). The thresholds are higher than the one used by Jackson and Eklund (2002) in their validation of the instrument (see Table 1).

Currently, the research in the context of computer-supported gamification is scarce. Both, the context of games and exercise have been regarded as one of the most likely contexts for people to experience flow. Therefore, we believe that also the context of gamified exercise affords an interesting area for studying it. Furthermore, flow is often discussed to be an important psychological target state pursued by gamification efforts (Huotari & Hamari, 2012; McGonigal, 2011). However, as with all psychometric measurement, the validity of the instrument is of crucial importance (Jackson & Eklund, 2002; Procci et al., 2012) in capturing the researched phenomenon. Thus, as the first step in conducting research on flow in gamification, the psychometric properties of the DFS-2 in a gamified setting will be examined.

3. The empirical study

3.1. Data

The data was gathered via an online questionnaire from the users of a service called Fitocracy that gamifies exercise. At its core, Fitocracy enables their users to track exercise. The system does not contain automatic tracking. Instead, the users enter the details of their activities into the service. The service contains a database of exercises ranging from, for example, various physical outdoor and recreational activities to gym activities and sports. Depending

Table 1
Goodness-of-fit indices of the DFS-2 validation studies.

Paper	Context	CFI/NNFI threshold for acceptable fit used in the study	Specification	*N*	χ^2	*df*	CFI	NNFI	RMSEA	CI90 lo – hi/PCLOSE	Fit?*
Jackson & Eklund (2002)	Physical activity*	0.90	First-order	574	1427.219	558	0.912	0.901	0.052	0.049, 0.055	Yes
			Second-order	574	1606.487	585	0.897	0.889	0.055	0.052, 0.058	Yes
Wang et al. (2009)	Online gaming*	0.95	First-order	1578	1925.49	558	0.936	0.927	0.047	0.045, 0.049	No
			Second-order	1578	1522.58	548	0.954	0.947	0.040	0.038, 0.042	Yes
Procci et al. (2012)	Video games*	0.95	Hierarchical (second-order)	314	1351.995	558	0.906	n/a	0.067	0.063, 0.072	No
			Respecified	314	1348.158	539	0.901	n/a	0.069	0.065, 0.074	No

Indices: χ^2 = Chi Square, CFI = Comparative Fit Index, NNFI = Non-Normed Fit Index, RMSEA = Root Mean Square Error of Approximation, PCLOSE = p of close fit.
* As reported in the original work.

on the type of the activity, the user may enter the length, the intensity, the distance or the weights used in each exercise. As a gamifying feature, the service calculates a point value which is based on an amount of points allocated to the given exercise and further adjusted according to the details of the user's activity. For example, lifting heavier weights gives more points than lifting lighter weights, or swimming for 30 min gives more points than swimming for 20 min. Thus, the users gain points for each exercise they track in the service.

Furthermore, users gain level-ups and unlock achievements (on game achievements, see Hamari & Eranti, 2011) based on the points gained. Users can also perform quests by tracking an exercise that corresponds to a given set of conditions, or challenge other users into duels. In addition to the system giving feedback to the users on their progress, the community members may provide feedback to one another and communicate with each other. With features such as group-forming, profile-building, and content sharing the service holds similarities with many social networking

Table 2
Length of experience, age, gender and exercise information of the respondent data.

Gender	Frequency	Percent	Length of experience	Frequency	Percent
Female	102	51	Less than 1 month	24	12
Male	98	49	1–3 months	38	19
Age (mean = 29.5, median = 27.5)			3–6 months	29	14.5
Less than 20	9	4.5	6–9 months	26	13
20–24	51	25.5	9–12 months	33	16.5
25–29	54	27	12–15 months	38	19
30–34	41	20.5	15–18 months	7	3.5
35–39	22	11	More than 18 months	5	2.5
40–44	16	8	Exercise sessions per week (mean = 5.3, median = 5.0)		
45–49	3	1.5	1–4	83	41.5
50 or more	4	2	5–9	106	53.0
			10–14	6	3.0
			15 or more	5	2.5
			Exercise hours per week (mean = 7.2, median = 6.0)		
			1–4	51	25.5
			5–9	99	49.5
			10–14	40	20.0
			15 or more	10	5.0

Table 3
The used thresholds for goodness-of-fit indices.

Indicator	Excellent	Acceptable	Source
Model fit measure			
CFI	>0.95	>0.90	Hair et al. (2010), Hu and Bentler (1999)
NNFI	>0.95	>0.90	
RMSEA	Good < 0.5, passable 0.5–1		
*Validity & Reliability**			
Square root of the AVEs	Square root of the AVE of each construct < all correlations between it and other constructs		Fornell and Larcker (1981)
Inter-correlations*	No inter-correlations are higher than 0.9		Pavlou, Liang, and Xue (2007)
Item loadings	Items are most highly loaded with the intended construct		Chin (1998)
	CR > AVE > MSV > ASV		Hair et al. (2010)
AVE	>0.5		
CR	>0.7		

Indices: CFI = Comparative Fit Index, NNFI = Non-Normed Fit Index, RMSEA = Root Mean Square Error of Approximation, CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
* It should be noted that high inter-correlations between constructs might not in reality pose a problem since all the constructs are regarded as components of flow and therefore are expected to be highly correlated.

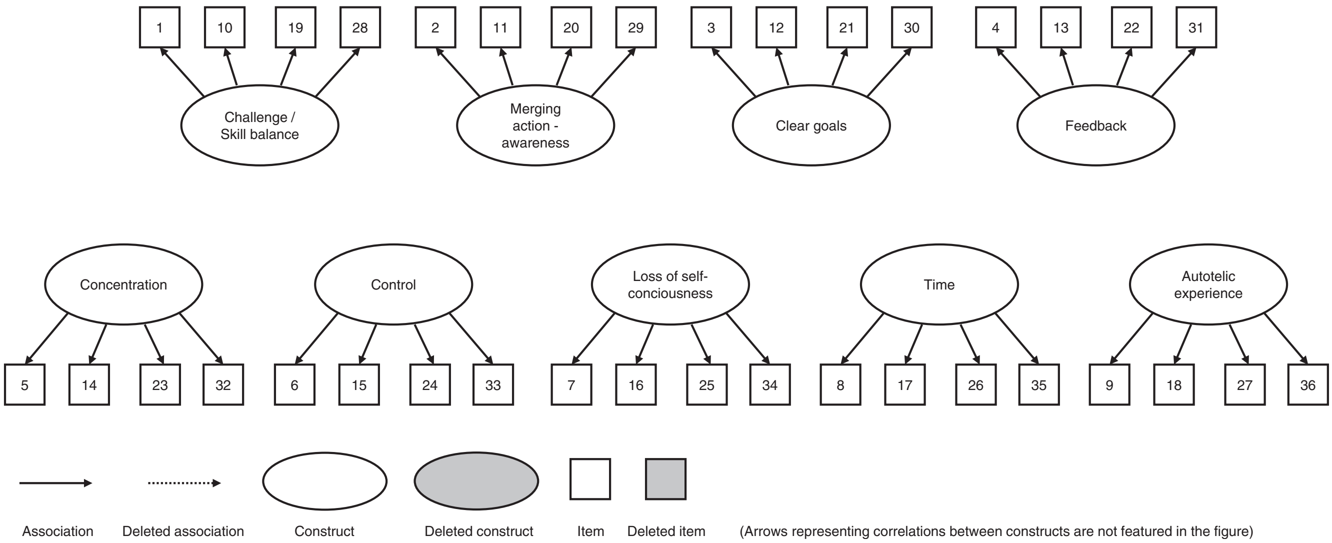

Fig. 1. Original DFS-2 factorial structure.

services (Baker & White, 2010; Boyd & Ellison, 2008; Lin & Lu, 2011; Ellison, Steinfeld & Lampe, 2007; Pfeil, Arjan & Zaphiris, 2009).

The data gathering was executed by posting a description of the study and the survey link to the discussion forum and groups within the service. Only registered users of the service were able

Table 4
Step 1a: correlation table.

	CR	AVE	MSV	ASV	T	CSB	MAA	G	F	C	CTRL	LSC	A
T	0.914	0.726	0.152	0.081	**0.852**								
CSB	0.841	0.573	0.762	0.476	0.266	**0.757**							
MAA	0.845	0.579	0.434	0.285	0.383	0.659	**0.761**						
G	0.847	0.583	0.805	0.515	0.233	0.873	0.597	**0.763**					
F	0.883	0.654	0.767	0.397	0.290	0.674	0.494	0.876	**0.809**				
C	0.902	0.699	0.677	0.404	0.290	0.750	0.621	0.774	0.676	**0.836**			
CTRL	0.886	0.660	0.805	0.499	0.230	0.860	0.584	0.897	0.802	0.823	**0.813**		
LSC	0.956	0.843	0.170	0.102	0.072	0.412	0.374	0.377	0.168	0.362	0.377	**0.918**	
A	0.884	0.659	0.637	0.389	0.390	0.786	0.481	0.798	0.705	0.577	0.767	0.232	**0.812**

The bolded figures represent the square roots of the AVE of the corresponding constructs.
Indices: CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
Constructs: T = Time transformation, CSB = Challenge-Skill Balance, MAA = Merging Action & Awareness, G = clear Goals, F = Feedback, C = Concentration, CTRL = Control, LSC = Loss of Self Consciousness, A = Autotelic experience.

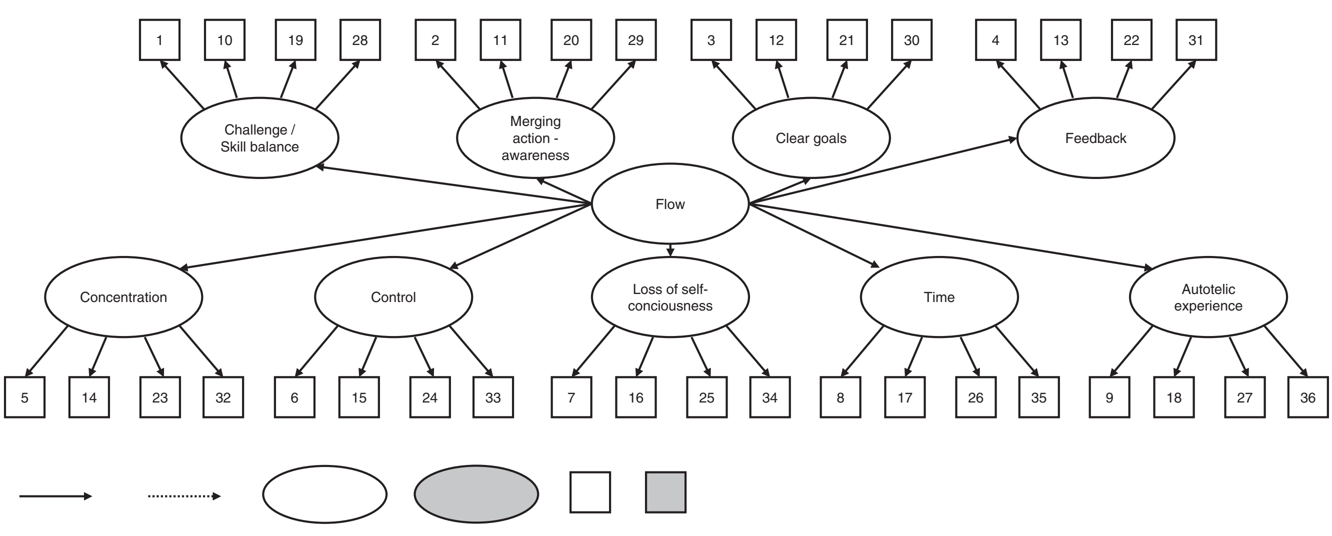

Fig. 2. Original DFS-2 s-order factorial structure.

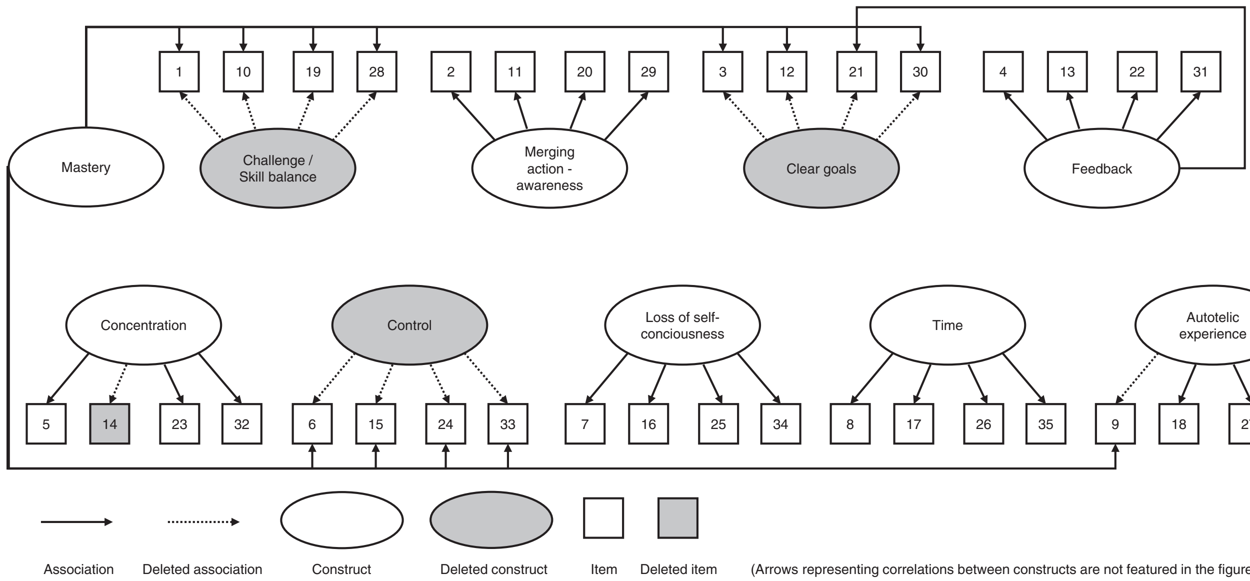

Fig. 3. Factorial structure in Procci et al. (2012) respecification.

Table 5
Step 2: correlation table.

	CR	AVE	MSV	ASV	MAA	Mastery	F	T	A	C	LSC
MAA	0.845	0.579	0.408	0.258	**0.761**						
Mastery	0.932	0.535	0.714	0.447	0.639	**0.731**					
F	0.897	0.636	0.714	0.338	0.491	0.845	**0.797**				
T	0.914	0.726	0.156	0.089	0.382	0.260	0.274	**0.852**			
A	0.881	0.715	0.637	0.313	0.477	0.798	0.704	0.395	**0.846**		
C	0.899	0.750	0.692	0.344	0.620	0.832	0.683	0.286	0.559	**0.866**	
LSC	0.956	0.844	0.161	0.087	0.375	0.401	0.188	0.072	0.227	0.357	**0.919**

The bolded figures represent the square roots of the AVE of the corresponding constructs.
ndices: CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
Constructs: T = Time transformation, MAA = Merging Action & Awareness, F = Feedback, C = Concentration, LSC = Loss of Self Consciousness, A = Autotelic experience, Mastery = see Fig. 3.

to access the survey. The respondents were entered in a prize draw for a $50 Amazon gift certificate. Altogether 200 valid responses were gathered. The sample size satisfies several different criteria for sufficient sample size. According to Hoelter (1983), samples sizes of 200 and above are sufficient for structural equation modeling data analyses. Bentler and Chou (1987) propose a minimum ratio of 5 respondents per 1 construct in the model (for the model of the present paper the threshold would be 180 respondents). Hair, Black, Babin, and Anderson (2010) suggest the same rule of thumb for factor analyses. Loehlin (1998) suggests that 100 respondents are required and 200 would be preferable. Anderson and Gerbing (1984, 1988) recommend a threshold of 150 respondents for models where constructs comprise of three or four indicators. Hair et al. (2010) suggest similar thresholds but suggest that complex models with few indicators per construct (less than three) and low communalities might require larger samples. In this study, all of the constructs comprise from minimum of four or more items and the communalities are high (referring to a more explorative SEM). Based on these sources, the sample size should be sufficient for the analyses carried out herein.

See Table 2 for the demographic information of respondents as well as details regarding length of experience with the service and reported amounts of exercise.

3.2. Measurement

As the Dispositional Flow Scale-2 is dispositional rather than a state scale, it measures the tendency to experience flow in a given setting rather than flow experience in a single given instance of activity (Jackson & Eklund, 2002). Here, we were interested in flow in the computer-supported gamified exercise particularly. Therefore, instructions for answering were provided in the survey accordingly so that it was made sure the respondents understood that all of the statements referred to exercise performed when using the Fitocracy service (instead of exercise in general).

DFS-2 scale consists of constructs based on the nine dimensions of flow defined by Csíkszentmihályi (1990). Thus, the constructs included in DFS-2 are the following: *time (transformation)* (T), *challenge-skill balance* (CSB), *merging of action and awareness* (MAA), *clear goals* (G), *feedback* (F), *concentration* (C), *control* (CTRL), *loss of self-consciousness* (LSC), and *autotelic experience* (A). Each of the nine constructs contains four items. Together the constructs form a 36-item scale for measuring the flow experience. The DFS-2 relies on self-reported data. Therefore, similarly to previous DFS-2 studies, the items were measured on a Likert scale ranging from strong disagreement (1 on the Likert scale) to strong agreement (7 on the Likert scale) with the statement.

3.3. Procedure

After the data collection, the procedure for the analyses consisted of 6 steps each investigating model fit indices as well as validity and reliability of the given factorial structure (see Table 3 for thresholds): (1) Confirmatory factor analysis (henceforth CFA) for both original single and second-order DFS-2 factorial structures by Jackson and Eklund (2002), (2) CFA for the Procci et al. (2012)

Table 6
Explorative factor analysis.

	1	2	3	4	5	6
f01CSB	**0.416**					
f10CSB	**0.562**	0.406				
f19CSB	**0.697**					
f28CSB	**0.562**	0.434				
f03G		0.600				
f12G	**0.492**	0.473				
f21G	**0.550**	0.449				
f30G			**0.669**			
f06CTRL	0.474		**0.488**			
f15CTRL	0.427	**0.510**				
f24CTRL	**0.526**	0.500				
f33CTRL	**0.546**	0.488				
f09A	0.629					
f18A	0.691					
f27A	0.801					
f36A	0.778					
f05C		0.713				
f14C		0.807				
f23C		0.711				
f32C		0.743				
f04F			0.837			
f13F			0.834			
f22F			0.649			
f31F			0.734			
f07LSC				0.910		
f16LSC				0.922		
f25LSC				0.898		
f34LSC				0.918		
f08T					0.877	
f17T					0.887	
f26T					0.779	
f35T					0.907	
f02MAA						0.725
f11MAA						0.713
f20MAA						0.707
f29MAA						0.819
Variance extracted	16.2	14.5	12.1	10.5	9.6	8.7

Bolded: larger than some item in the core construct.
Item codes referring to the original corresponding constructs: T = Time transformation, CSB = Challenge-Skill Balance, MAA = Merging Action & Awareness, G = clear Goals, F = Feedback, C = Concentration, CTRL = Control, LSC = Loss of Self Consciousness, A = Autotelic experience.

respecification for video games, (3) Exploratory factor analysis (henceforth EFA), (4) CFA on the factorial structure suggested by EFA, (5) finding a factorial structure with adequate fit via omitting items, (6) finding an adequate factorial structure via partial second-ordering. The confirmatory factor analyses were carried out in AMOS 21. The exploratory factor analyses as well as data screening were carried out in SPSS 21. We opt for model fit indices used also in other papers investigating DFS-2.

3.3.1. Step 1a: CFA for single-order DFS-2

The test indicated an adequate fit ($\chi^2 = 1044.955$, CFI = 0,918, NNFI = 0.907, RMSEA = 0.066, CI90 = 0.060–0.073) if we adopted the same model fit indices as in the original work. However, when using current thresholds, the model does not have adequate fit (Table 3). This first-order model (Fig. 1) also shows rather high inter-correlation between several constructs (Table 4). Most notably *challenge-skill balance* (CSB), *clear goals* (G), *control* (CTRL) and *autotelic experience* (A) are highly correlated with each other. Furthermore, there seems to be some overlap between *clear goals* (G) and *feedback* (F).

3.3.2. Step 1b: CFA for second-order DFS-2

The test for the second-order model (Fig. 2) indicated fairly similar indices as the original test by Jackson and Eklund (2002), albeit too low to satisfy the recent thresholds (Table 3): $\chi^2 = 1136.599$, CFI = 0.906, NNFI = 0.899, RMSEA = 0.069, CI90 0.063–0.075.

3.3.3. Step 2: CFA for DFS-2 respecification by Procci et al. (2012)

We then tested the model fit of the Procci et al. (2012) respecification of DFS-2 (Fig. 3). The test showed the respecification to provide a poorer fit than the original model: $\chi^2 = 1242.882$, CFI = 0.878, NNFI = 0.865, RMSEA = 0.081, CI90 0.075–0.087. In the Procci et al. (2012) respecification, *challenge-skill balance* (CSB), *clear goals* (G), and *control* (CTRL) have been combined into a single factor titled '*mastery*'. However, in our data *autotelic experience* (A), *feedback* (F) and *concentration* (C) were also associated with these constructs (Table 5) within this factorial structure.

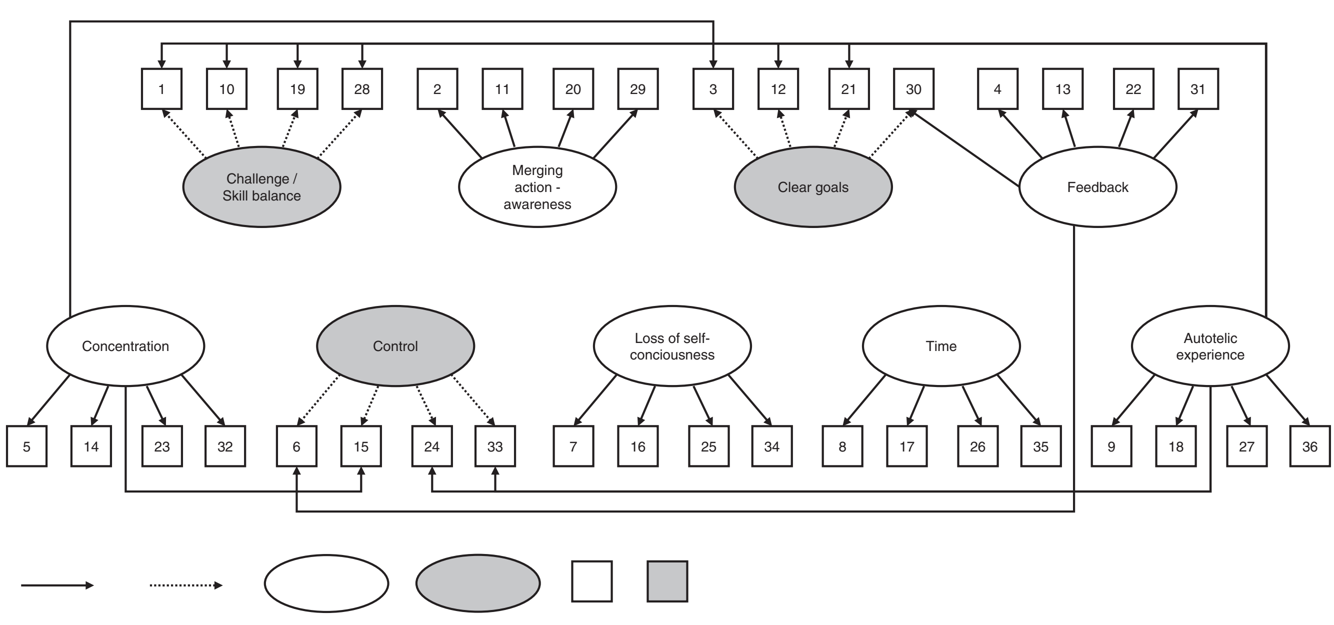

Fig. 4. CFA based on EFA.

Table 7
Step 4: correlation table.

	CR	AVE	MSV	ASV	T	A	C	F	LSC	MAA
T	0.907	0.710	0.158	0.077	**0.843**					
A	0.935	0.548	0.731	0.412	0.287	**0.740**				
C	0.900	0.604	0.731	0.382	0.267	0.855	**0.777**			
F	0.906	0.617	0.689	0.325	0.258	0.830	0.738	**0.785**		
LSC	0.959	0.853	0.154	0.098	0.074	0.375	0.392	0.222	**0.924**	
MAA	0.834	0.560	0.416	0.279	0.398	0.645	0.638	0.526	0.373	**0.749**

The bolded figures represent the square roots of the AVE of the corresponding constructs.
Indices: CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
Constructs: T = Time transformation, MAA = Merging Action & Awareness, F = Feedback, C = Concentration, LSC = Loss of Self Consciousness, A = Autotelic experience.

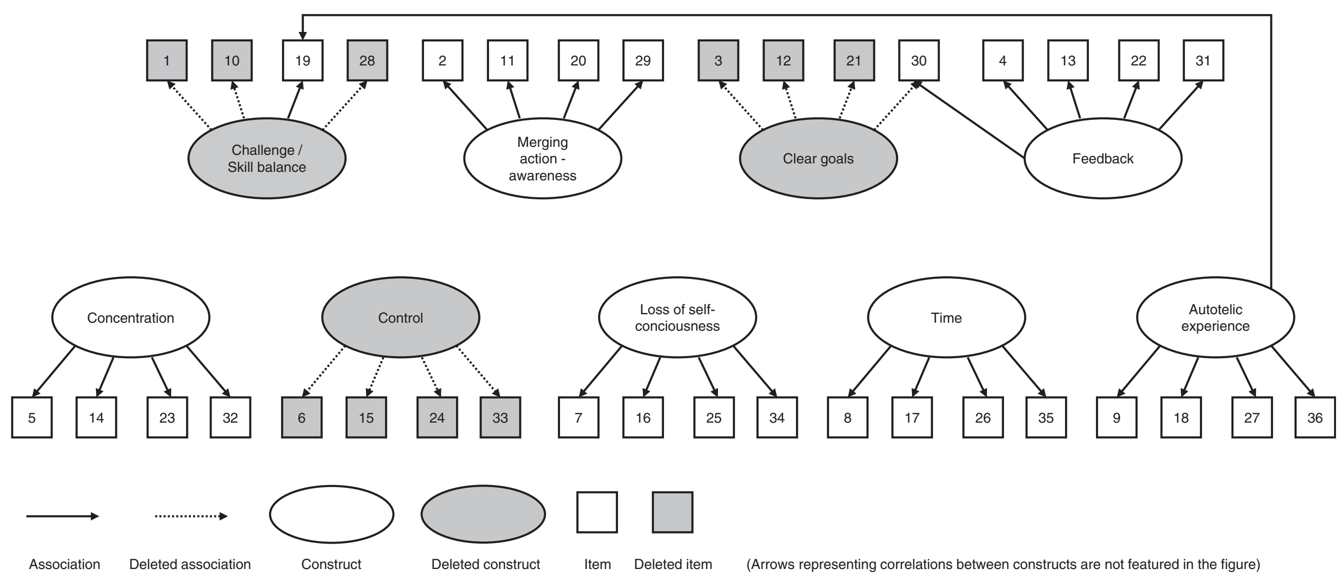

Fig. 5. Respecification via omitting items.

Table 8
Step 5: correlation table.

	CR	AVE	MSV	ASV	T	A	C	F	LSC	MAA
T	0.907	0.711	0.160	0.090	**0.843**					
A	0.895	0.635	0.508	0.263	0.362	**0.797**				
C	0.901	0.697	0.466	0.287	0.283	0.595	**0.835**			
F	0.902	0.649	0.508	0.274	0.270	0.713	0.683	**0.806**		
LSC	0.959	0.854	0.139	0.077	0.074	0.252	0.365	0.206	**0.924**	
MAA	0.833	0.558	0.403	0.248	0.400	0.506	0.635	0.529	0.373	**0.747**

The bolded figures represent the square roots of the AVE of the corresponding constructs.
Indices: CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
Constructs: T = Time transformation, MAA = Merging Action & Awareness, F = Feedback, C = Concentration, LSC = Loss of Self Consciousness, A = Autotelic experience.

3.3.4. Step 3: Exploratory Factor Analysis (EFA)

Although the original model met the thresholds of the original work of Jackson and Eklund (2002), we wanted to proceed further with the tests. We moved to run an exploratory factor analysis in order to investigate, which kind of factorial structure the data will converge in. We used principal component analysis with varimax rotation for the exploratory factor analysis.

The resulting factorial structure explained 71.7% of the variance (Table 6). As already suggested by the previous inter-correlation analyses, and as was also partly confirmed by Procci et al. (2012), the items of *challenge-skill balance* (CSB), *clear goals* (G), and *control* (CTRL) seem to mostly load on the same factor. However, in this data also *autotelic experience* (A) is strongly associated with these constructs. All other constructs, however, behave well loading highly only with their expected factors.

3.3.5. Step 4: Confirmatory Factor Analysis (CFA) for the respecified model

Based on the EFA, the constructs *challenge-skill balance, clear goals* and *control* did not form their own factors. Instead, all of the items of these constructs loaded mostly with *autotelic experience*, one item with *concentration* and two items with *feedback* (see Table 6). We then ran a CFA based on the results of the EFA (Fig. 4). The model fit was inadequate: χ^2 = 1208.778, CFI = 0.893, NNFI = 0.882, RMSEA = 0.075, CI90 0.069–0.081. Overall, it seems that *challenge/skill, clear goals* and *control* form a rather uniform construct with *autotelic experience.* However, in CFA *concentration* and *feedback* also seemed to correlate with this dimension to the extent that it would not pass the discriminant validity thresholds (Table 7).

Table 9
EFA after omitted items.

	1	2	3	4	5	6
f04F	0.871					
f13F	0.856					
f31F	0.740					
f22F	0.697					
f30G	0.680					
f16LSC		0.927				
f34LSC		0.927				
f07LSC		0.912				
f25LSC		0.907				
f27A			0.810			
f36A			0.802			
f18A			0.720			
f09A			0.657			
f19CSB			0.638			
f35T				0.904		
f08T				0.892		
f17T				0.885		
f26T				0.774		
f14C					0.786	
f32C					0.784	
f23C					0.751	
f05C					0.729	
f29MAA						0.825
f02MAA						0.755
f11MAA						0.722
f20MAA						0.708
Variance extracted	14.3	14.2	13.1	12.9	11.9	10.9

Item codes referring to the original corresponding constructs: T = Time transformation, CSB = Challenge-Skill Balance, MAA = Merging Action & Awareness, G = clear Goals, F = Feedback, C = Concentration, CTRL = Control, LSC = Loss of Self Consciousness, A = Autotelic experience.

3.3.6. Step 5: Adequately fitting model via omitting items

Following the results of the EFA (Step 3), we omitted all items that had a poorer loading than the items of the core construct of each factor. Resulting in a factorial structure depicted in Fig. 5. In other words, omitting these items effectively omits the three highly inter-correlating constructs almost entirely with the exception of items csb01 and g30. This solution represents the cleanest factorial structure based on the EFA.

With this remaining factorial structure both convergent and discriminant validities were acceptable (Table 8) and model fit adequate: $\chi^2 = 461.259$, CFI = 0.955, NNFI = 0.948, RMSEA = 0.067, CI90 0.048–0.066. Further, the EFA also showed that the remaining model still converged into the same exact factorial solution (Table 9).

3.3.7. Step 6: Adequately fitting model without omitting items

In step 5, we could find a factorial structure that had an adequate model fit even when using the tighter thresholds. However, we wanted to find a factorial structure that would have decent model fit with the entire instrument.

As noted, CSB, G, F, and A are strongly correlated. Qualitatively, all these constructs can be considered as pertaining to experience of mastery, i.e. having clear goals and feedback as an indicator of challenge-skill balance, which further associated autotelic experiences. These aspects of flow are often regarded as conditions of flow (Csíkszentmihályi, 1990; Nakamura & Csíkszentmihályi, 2002), whereas, for example, loosing track of time or self-consciousness as well as merging action-awareness can be considered more as outcomes from reaching flow. Therefore, we modeled a higher-order construct for the four highly correlated constructs, and named it *mastery* (or conditions of flow) similarly to Procci et al. (2012). This structure (Fig. 6) has an adequate model fit if

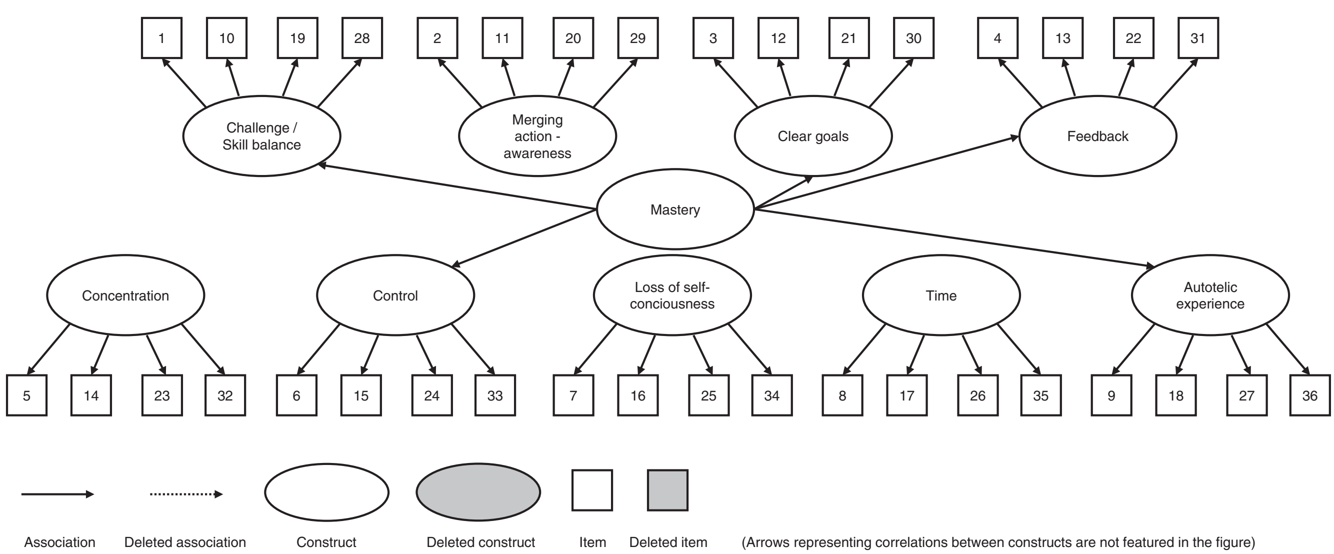

Fig. 6. Respecification via second-order construct.

Table 10
Step 6: correlation table.

	CR	AVE	MSV	ASV	T	MAA	C	LSC	MASTERY
T	0.913	0.726	0.147	0.080	**0.852**				
MAA	0.845	0.579	0.394	0.267	0.383	**0.761**			
C	0.902	0.698	0.666	0.317	0.289	0.621	**0.835**		
LSC	0.956	0.844	0.141	0.103	0.072	0.375	0.362	**0.919**	
MASTERY	0.953	0.804	0.666	0.320	0.289	0.628	0.816	0.368	**0.897**

The bolded figures represent the square roots of the AVE of the corresponding constructs.
Indices: CR = Composite Reliability, AVE = Average Variance Extracted, MSV = Maximum Shared Variance, ASV = Average Shared Variance.
Constructs: T = Time transformation, MAA = Merging Action & Awareness, C = Concentration, LSC = Loss of Self Consciousness, Mastery = see Fig. 6.

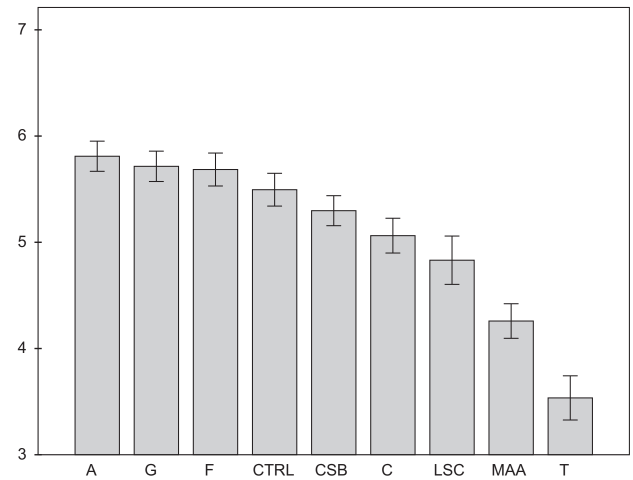

Fig. 7. Means of the responses on the components of flow.

we use the same threshold as Jackson and Eklund (2002) originally: $\chi^2 = 1120.066$, CFI = 0,907, NNFI = 0.900, RMSEA = 0.069, CI90 0.063–0.075. Furthermore, the mastery second-order construct had discriminant validity (Table 10).

We further analyzed the means and deviations of the responses regarding the (original) components of flow in order to see whether the ratings of the different components would provide support for the notion of ‘conditions’ and ‘outcomes’ as well as to generally gauge which flow experiences were most prominent. The data shows that, indeed, in the gamification context, the experiences most commonly linked to gamification in popular discussions were reported to occur more: *autotelic experience* (self-purposefulness/intrinsically motivated experience are often referred to in the discussions on gamification), *having clear goals* and *feedback.* On the contrary, for example *time transformation* as well as *merging action and awareness* were clearly reported to occur less (See Table 12 and Fig. 7).

4. Discussion and conclusions

In this paper we investigated the psychometric properties of the DFS-2 flow measurement instrument in the context of technology-supported gamified physical exercise (Table 11). The results indicate that the original factorial structure by Jackson and Eklund (2002) does provide an adequate fit when same lower thresholds for goodness-of-fit indices are used. However, we wanted to opt for currently favored higher thresholds (Hu & Bentler, 1999; Pfeil et al., 2009). We then moved onto testing a respecification of the model by Procci et al. (2012) of which the factorial structure resulted in a poorer fit. In order to find a suitable factorial structure, as the third step, we conducted an explorative factor analysis. In the fourth step, we conducted a confirmatory factor analysis for a model that was respecified in accordance to the EFA. This method did not yield significantly better results, since it seemed that a few constructs formed combined factors. In the fifth step, we took the factors created by EFA, but omitted items that loaded below 0.6. This model reached the higher thresholds of model fit as well as good convergent and discriminant validity. However, we still wanted to find a factorial structure that reached model fit and validity without omitting any items. Therefore, in the sixth step, we created a second-order construct that contained the constructs with high inter-correlations (CSB, A, G, F, CTRL) as reflective indicators. This model reached good validity and the lower thresholds for model fit.

In their DFS-2 model, Jackson and Eklund (2002) did not consider the causality or relationships between the dimensions of flow. However, theorizations regarding the flow dimensions have considered the *challenge-skill-balance*, *clear goals*, *control*, and *feedback* as conditions required for reaching flow (Csíkszentmihályi, 1990; Nakamura & Csíkszentmihályi, 2002), whereas *loss of self-consciousness*, *time*, *concentration*, and *merging action-awareness* have been considered outcomes from reaching flow. The final respecification created in this study also suggests that the constructs divide into two different categories; the conditional aspects and outcome experiences from reaching the flow state, which are in line with some of the previous considerations (see also Table 12 and Fig. 7 for differences in means). The final respecification presented in this paper is thus also theoretically feasible in addition to presenting good model fit and validity indices.

Furthermore, in our study *autotelic experience* seemed to correlate strongly with other conditions of flow which might pose a deviation from earlier theorizations. This finding would suggest that, at least in the context of computer-supported gamified exercise, the autotelic experience, that is, finding the activity intrinsically motivating, is also a condition for reaching the flow state rather than being an outcome from reaching flow. In a video gam-

Table 11
Goodness-of-fit indices for different factorial structures studied in this paper.

Specification	*N*	χ^2	*df*	CFI	NNFI	RMSEA	CI90 lo – hi/PCLOSE
Step 1a: Original first-order model by Jackson and Eklund (2002)	200	1044.955	558	0,918	0.907	0.066	0.060, 0.073
Step 1b: Original second-order model by Jackson and Eklund (2002)	200	1136.599	585	0.906	0.899	0.069	0.063, 0.075
Step 2: Respecification by Procci et al. (2012)	200	1242.882	539	0.878	0.865	0.081	0.075, 0.087
Step 3: EFA	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Step 4: Respecification based on EFA	200	1208.778	572	0.893	0.882	0.075	0.069, 0.081
Step 5: Respecification with omitted items	200	461.259	279	0.955	0.948	0.057	0.048, 0.066
Step 6: Respecification without omitted items	200	1120.066	577	0,907	0.900	0.069	0.063, 0.075

Indices: χ^2 = Chi Square, CFI = Comparative Fit Index, NNFI = Non-Normed Fit Index, RMSEA = Root Mean Square Error of Approximation, PCLOSE = p of close fit.

Table 12
Means and standard deviation of the components of flow (in descending order).

	A	G	F	CTRL	CSB	C	LSC	MAA	T
Means	5.810	5.715	5.685	5.495	5.298	5.063	4.831	4.259	3.535
SD	1.018	1.025	1.112	1.106	1.013	1.174	1.631	1.168	1.492

T = Time transformation, CSB = Challenge-Skill Balance, MAA = Merging Action & Awareness, G = clear Goals, F = Feedback, C = Concentration, CTRL = Control, LSC = Loss of Self Consciousness, A = Autotelic experience.

ing environment, Procci et al. (2012) also found that *challenge-skill balance, clear goals and control* have a strong inter-correlation. However, *autotelic experience* did not load as strongly with this factor in their study. This could suggest that in gamified exercise, *autotelic experience* is potentially more closely associated with the optimal challenge-skill experience than in pure gaming. Where as in games, the autotelic experience might also be caused by other factors that are missing from the common gamification implementation. Other appeals of video games (see e.g. Hamari & Tuunanen, 2014; Yee, 2006 on gaming motivations), such as audiovisual immersion and stimuli that are commonly missing in the gamification context but often are an elemental aspect of video games. In other words, video game players can potentially reach an autotelic experience even without actual goal-oriented play/mastery which most gamification implementations seem strive for. The data further shows that indeed, in the gamification context, those experience most commonly linked to gamification in popular discussions had been rated to occur more (*autotelic experience* (self-purposefulness/intrinsically motivated experience are often referred in the discussion on gamification), *having clear goals* and *feedback.* Whereas for example *time transformation* as well as *merging action and awareness,* which are also commonly connected with audiovisual immersion, were clearly rated to occur less (See Table 12 and Fig. 7).

Beyond the contributions related to the psychometric properties of the DFS-2 scale, this study also suggests that flow should rather be seen as divided between the collection of conditions for reaching flow state and the psychological outcomes that follow from reaching the flow state instead of merely seeing all of the nine dimensions as reflective indicators of the entirety of flow. Previous studies might not have been able to make a distinction along these lines, since, naturally, both the conditions and outcomes are highly correlated as is to be expected. Therefore, we suggest that further studies could conduct experiments which would specifically focus on finding evidence pertaining to causal relationships between the components of the general flow.

Disclosure statement

No competing financial interests exist.

Acknowledgements

This research has been partially supported by individual study grants for both authors from the Finnish Cultural Foundation as well as carried out as part of research projects (40311/12, 40134/13, 40111/14) funded by the Finnish Funding Agency for Technology and Innovation (TEKES).

References

Admiraal, W., Huizenga, J., Akkerman, S., & Dam, G. T. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior, 27*(3), 1185–1194.
Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*(2), 155–173.
Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423.
Baker, R. K., & White, K. M. (2010). Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective. *Computers in Human Behavior, 26*(6), 1591–1597.
Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117.
Boyd, D. M., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230.
Cermakova, L., Moneta, G. B., & Spada, M. M. (2010). Dispositional flow as a mediator of the relationships between attentional control and approaches to studying during academic examination preparation. *Educational Psychology: An International Journal of Experimental Educational Psychology, 30*(5), 495–511.
Chen, H. (2006). Flow on the net–detecting Web users' positive affects and their flow states. *Computers in Human Behavior, 22*(2), 221–233.
Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). London: Lawrence Erlbaum Associates.
Crust, L., & Swann, C. (2013). The relationship between mental toughness and dispositional flow. *European Journal of Sport Science, 13*(2), 215–220.
Csíkszentmihályi, M. (1975). *Beyond boredom and anxiety: Experiencing flow in work and play*. San Francisco: Jossey-Bass.
Csíkszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row.
Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In Proceedings of the 15th international Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, September 28–30, pp. 9–15.
Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook 'friends:' Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168.
Fornell, C., & Larcker, D. (1981). Structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.
Fritz, B. S., & Avsec, A. (2007). The experience of flow and subjective well-being of music students. *Horizons of Psychology, 16*(2), 5–17.
Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101–112.
González-Cutre, D., Sicilia, Á., Moreno, J. A., & Fernández-Balboa, J. M. (2009). Dispositional flow in physical education: Relationships with motivational climate, social goals, and perceived competence. *Journal of Teaching in Physical Education, 28*(4), 422–440.
Gucciardi, D. F., Gordon, S., & Dimmock, J. A. (2009). Evaluation of a mental toughness training program for youth-aged Australian footballers: I. A quantitative analysis. *Journal of Applied Sport Psychology, 21*(3), 307–323.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Englewood Cliffs, NJ: Prentice Hall.
Hamari, J. (2013). Transforming Homo economicus into Homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236–245.
Hamari, J., & Eranti, V. (2011). Framework for designing and evaluating game achievements. In Proceedings of the DiGRA 2011 conference: Think design play, Hilversum, The Netherlands, September 14–17, 2011.
Hamari, J., & Koivisto, J. (2013). Social motivations to use gamification: An empirical study of gamifying exercise. In Proceedings of the 21st European conference on information systems, Utrecht, The Netherlands, June 5–8, 2013.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In Proceedings of the 47th Hawaii international conference on system sciences, Hawaii, USA, January 6–9, 2014.
Hamari, J., Huotari, K., & Tolvanen, J. (2014). Gamification and economics. In S. P. Walz & S. Deterding (Eds.), *The Gameful World: Approaches, issues, applications*. Cambridge, MA: MIT Press.
Hamari, J., Koivisto, J., & Pakkanen, T. (2014). Do persuasive technologies persuade? – A review of empirical studies. In A. Spagnolli, L. Chittaro, & L. Gamberini (Eds.), *Persuasive technology, LNCS 8462* (pp. 118–136). Switzerland: Springer International Publishing.
Hamari, J., & Lehdonvirta, V. (2010). Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science & Applied Management, 5*(1), 14–29.
Hamari, J., & Tuunanen, J. (2014). Player types: A meta-synthesis. *Transactions of the Digital Games Research Association, 1*(2), 29–53.
Hodge, K., Lonsdale, C., & Jackson, S. A. (2009). Athlete engagement in elite sport: An exploratory investigation of antecedents and consequences. *The Sport Psychologist, 23*(2), 186–202.
Hoelter, J. W. (1983). The analysis of covariance structures goodness-of-fit indices. *Sociological Methods & Research, 11*(3), 325–344.
Hsu, C.-L., & Lu, H.-P. (2004). Why do people play on-line games? An extended TAM with social influences and flow experience. *Information & Management, 41*(7), 853–868.
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), 1–55.
Huotari, K., & Hamari, J. (2012). Defining gamification – A service marketing perspective. In Proceedings of the 16th international academic MindTrek conference, Tampere, Finland, 3–5 October 2012, pp. 17–22.
Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The flow state scale-2 and dispositional flow scale-2. *Journal of Sport & Exercise Psychology, 24*(2), 133–150.
Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport & Exercise Psychology, 18*(1), 17–35.
Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport and Exercise Psychology, 30*(5), 561–587.
Kimiecik, J. C., & Jackson, S. A. (2002). Optimal experience in sport: A flow perspective. In T. S. Horn (Ed.), *Advances in sport psychology* (pp. 501–527). Champaign, IL, US: Human Kinetics.

Koehn, S., Morris, T., & Watt, A. P. (2013). Correlates of dispositional and state flow in tennis competition. *Journal of Applied Sport Psychology, 25*(3), 354–369.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179–188.

Lin, K.-Y., & Lu, H.-P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior, 27*(3), 1152–1161.

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Lawrence Erlbaum Associates Publishers.

Lu, Y., Zhou, T., & Wang, B. (2009). Exploring Chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. *Computers in Human Behavior, 25*(1), 29–39.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. London: Jonathan Cape.

Nakamura, J., & Csíkszentmihályi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). New York: Oxford University Press.

Nicholls, R., Polman, R. C. J., & Holt, N. L. (2005). The effects of individualized imagery interventions on golf performance and flow states. *Athletic Insight: The Online Journal of Sport Psychology, 7*(1), 43–66.

Pavlou, P. A., Liang, H. G., & Xue, Y. J. (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly, 31*(1), 105–136.

Pfeil, U., Arjan, R., & Zaphiris, P. (2009). Age differences in online social networking: A study of user profiles and the social capital divide among teenagers and older users in MySpace. *Computers in Human Behavior, 25*(3), 643–654.

Procci, K., Singer, A. R., Levy, K. R., & Bowers, C. (2012). Measuring the flow experience of gamers: An evaluation of the DFS-2. *Computers in Human Behavior, 28*(6), 2306–2312.

Sinnamon, S., Moran, A., & O'Connell, M. (2012). Flow among musicians: Measuring peak experiences of student performers. *Journal of Research in Music Education, 60*(1), 6–25.

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a Web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior, 20*(3), 403–422.

Wang, C. K. J., Liu, W. C., & Khoo, A. (2009). The psychometric properties of dispositional flow scale-2 in internet gaming. *Current Psychology, 28*(3), 194–201.

Webster, J., Trevino, L. K., & Ryan, L. (1994). The dimensionality and correlates of flow in human–computer interactions. *Computers in Human Behavior, 9*(4), 411–426.

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior, 24*(5), 2274–2291.

Whitmore, J. G., Borrie, W. T. (2005). Exploring the usefulness of the dispositional flow scale for outdoor recreation activities. In The Proceedings of the 2005 Northeastern recreation research symposium, Bolton Landing, New York, USA, April 10–12, 2005, pp. 371–378.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior, 9*(6), 772–775.