Independent Validation of the Video Game Dispositional Flow Scale With League of Legends Players

NICK VON FELTEN, Center for General Psychology and Methodology, University of Basel, Switzerland
FLORIAN BRÜHLMANN, Center for General Psychology and Methodology, University of Basel, Switzerland
SEBASTIAN A. C. PERRIG*, Center for General Psychology and Methodology, University of Basel, Switzerland

Flow is a highly influential concept across many research domains. Because of the close links between enjoyment and flow, it is also a central concept in game research. Accordingly, many measurement instruments in the form of questionnaires have been developed to measure flow in video games. However, limited independent validation work has been carried out to date. In this project, we evaluated the recently developed Video Game Dispositional Flow Scale (VGDFS) in the context of League of Legends. An online survey posted on the online community reddit.com yielded a sample of $N = 140$. Confirmatory factor analysis did not support the originally proposed structure. Subsequent exploratory factor analysis resulted in inconclusive findings. Convergent validity with other questionnaires for the measurement of interest, enjoyment, and flow could be demonstrated in our data, although with unusually high correlations between the VGDFS and another flow measure. For divergent validity, correlations with questionnaires for the measurement of boredom and trait anxiety in the hypothesized direction could be found. Findings suggest that the VGDFS needs further evaluation in different video game contexts while also hinting at measurement problems possibly related to conceptual issues with flow.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; **HCI theory, concepts and models**; • **Applied computing** → **Computer games**.

Additional Key Words and Phrases: flow; gaming; player experience; scale; replication; validation; psychometrics

ACM Reference Format:
Nick von Felten, Florian Brühlmann, and Sebastian A. C. Perrig. 2022. Independent Validation of the Video Game Dispositional Flow Scale With League of Legends Players. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '22 EA), November 2–5, 2022, Bremen, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3505270.3558351

1 INTRODUCTION

For some people, observing someone play a video game feels like watching a zombie. The player seems absorbed into the game and does not react to external cues. While this might seem like a miserable experience from an exterior view, this state, often referred to as *flow*, *"in which an individual is completely absorbed in activity without reflective self-consciousness but with a deep sense of control" [28, pp. 150–153]*, is quite enjoyable to the player [7, 8, 28, 35]. Due to the association of flow with enjoyment, it represents an important construct in game research. [35]. Consequently, several questionnaires were constructed in the long lasting endeavor of measuring flow (see Cai et al. [4] for a review). However, with this rich selection of questionnaires, conceptual problems arise [4]. Although most models share certain properties, there is no consensus on one particular flow model. Most notably the quantity of dimensions, as well as

*Corresponding Author.

Manuscript submitted to ACM

the existence of a higher-order factor, are still debated. Further, some of the dimensions are argued to be prerequisites to flow (i.e., Clear Goals, Challenge-skill Balance, and Unambiguous Feedback) [4, 27]. To clean up some of these doubts Cai et al. [4] conducted a literature review on flow in gaming and found that there is no valid measure for dispositional flow in the leisure game context suitable for adults, only for children and adolescents. However, adults constitute the largest video game population, with 76% of players over the age of 18 [11]. Cai et al. [4] thus constructed and validated the Video Game Dispositional Flow scale (VGDFS). They chose the flow construct to be comprised of one single flow component represented by nine dimensions following the theory by Jackson and Marsh [16]. The dimensions are described in Table 1. Because the authors validated their measure with a general and consequently unspecific video gaming population, the authors requested validation in specific gaming contexts as it remains unclear how the questionnaire performs for players of specific games. For this purpose we chose to independently validate the VGDFS, focussing on a Multiplayer Online Battle Arena (MOBA) game. While flow experiences were found to occur in most games [18], MOBAs were found to offer greater challenge than other genres, which fits the challenge-skill aspect of flow [18]. With a wide variety of champions to choose from, players can have different levels of proficiency with each one. With different combinations of champions and different players each round, MOBAs offer high variation of challenge between games. The same holds for the dimensions of Clear Goals and Unambiguous Feedback. Concerning Clear Goals, the ultimate goal in MOBAs is very clear, the short-term goals can be more ambiguous. Feedback can be immediate with a wrong play resulting in death or loss of structures, clearly related to the Unambiguous Feedback dimension. Therefore, we considered League of Legends (LoL) [31] a suitable game to investigate the psychometric properties of the VGDFS. LoL is not only a very popular MOBA, but with 180 million active players a month [22], also one of the most popular online games in general. Also, players of LoL have been studied numerous times in past research [e.g., 2, 19], further stressing the importance of having scales that work with players of LoL. Moreover, flow was found to be associated with performance [13] which fits LoLs large e-sports scene and general goal for in-game performance [21]. This not only marks flow as an essential construct for studying LoL players but also reinforces the need to investigate the psychometric quality of the VGDFS in this specific context to account for the characteristics and possible differences of LoL to other games.

Another gap we want to investigate is how the measured constructs are associated with theoretically close constructs measured with other questionnaires. Cai et al. [4] found evidence for convergent and divergent validity between the nine subscales with an average variance extracted approach. However, the correlations with different established measures can give important insights into the relatedness of different constructs [5], which is especially important for the gaming industry, where flow is often used as an inference for enjoyment [35]. Further, divergent validity with other scales can allow inferences about the absence of undesirable traits in many games (e.g., anxiety or boredom). We provide initial evidence for a never before mentioned 4-factor structure to the flow construct. But also found evidence for the proposed 9-factor structure. Further, we identified that the dimensions Sense of Control and Loss of Self-consciousness seem to be problematic in general, while we found Clear Goals to be problematic specifically for LoL. We establish initial evidence for convergence with a different flow scale and an enjoyment scale, as well as preliminary evidence for divergence with an anxiety and a boredom scale. Overall, the present work aims to contribute to developing the flow concept by investigating the VGDFS.

2 METHOD

To achieve our goal, we recruited 171 participants, of which 140 remained after data cleaning, to fill out a series of standardized questionnaires—including the VGDFS—in an online survey.

Table 1. Descriptions of flow dimensions adapted from Cai et al. [4].

Dimension	Description
Action-Awareness merging (AAM)	Actions and awareness are merged, which means that players are absorbed and feel completely immersed in the game world.
Autotelic Experience (AE)	Describes the experience of completely intrinsic motivation, which is caused by flow experiences. This is seen as a determinant of enjoyment.
Challenge-skill Balance (CSB)	Players can enter the experience of Flow, when the challenges of the game and the skills of the player are in an equilibrium, which means the tasks are subjectively not too easy or too hard to achieve.
Clear Goals (CG)	Players know exactly how to proceed. Because the goals of the upcoming tasks are clear to them and they are attentive to relevant in-Game cues.
Concentration (CON)	Players in a flow state, are fully focused on the tasks in the game and their concentration is only directed towards the specific game-tasks.
Loss of Self-consciousness (LSS)	The consciousness outside of the game-world is diminished and the players experience their ego inside of the game world when in a flow state.
Sense of Control (SC)	While experiencing flow, players feel to be in control of the game world.
Transformation of Time (TT)	Describes a warped perception of time in the real world while experiencing a flow state.
Unambiguous Feedback (UF)	In-Game progess (feedback) is accessible immediately and is apparent.

2.1 Measures

2.1.1 Video Game Dispositional Flow Scale (VGDFS). The main questionnaire of interest was the VGDFS by Cai et al. [4], which assesses dispositional flow in video games. Dispositional flow can be understood as the propensity or frequency with which someone can experience flow [15]. The VGDFS consists of 28 items with a 7-point Likert-type response format, ranging from 1 (*Never*) to 7 (*Always*), measuring nine dimensions of flow (see Table 1).

2.1.2 Scales for Convergent Validity. For convergent validity, we chose the Dispositional Flow Scale Short (DFS-S) [17] because the scale was constructed based on the same theory as the VGDFS and was shown to be of good psychometric quality. The DFS-S is measured on 5-point Likert-type scales ranging from 1 (*Never*) to 5 (*Always*). We opted for the short version because the VGDFS is already quite long, and we wanted to minimize the participant burden. Second, we used the Interest/enjoyment part of the Intrinsic Motivation Inventory (IMI) [32]. This 7-point Likert-type scale ranges from 1 (*Not at all true*) to 7 (*Very true*). The Interest/enjoyment subscale is a self-report measure of intrinsic motivation and is frequently used in isolation [32]. We selected the IMI based on the proposed connection between enjoyment and flow and the notion of intrinsic motivation in some dimensions of flow (e.g., *Autotelic Experience*). Following the authors' recommendation, we slightly adapted the wording to fit the LoL context.

2.1.3 Scales for Divergent Validity. To assess divergent validity, we chose the short boredom proneness scale (SBPS) [34] and the short trait anxiety inventory (STAIT-5) [37]. The SBPS is measured by a 7-point Likert-type scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*) and the STAIT-5 uses a 4-point scale ranging from 1 (*Not at all*) to 4 (*Very much so*). Both scales are of comparable psychometric quality to their long versions [29, 34, 37]. Again, we chose the short scales to minimize participant burden. Both scales should correlate negatively with the VGDFS because flow is supposed to occur at a balance between skill and challenge [8]. If skill and challenge are not in balance (the game is overly challenging and player skill is low, or vice versa), this hinders flow and leads to anxiety or boredom.

2.2 Participants

In total, 171 participants recruited from the r/leagueoflegends subreddit and the r/leagueoflegends discord channel completed the survey. Data cleaning followed recommendations by Brühlmann et al. [3]. First, we excluded all participants below the age of 18 ($n = 9$), given that this was a participation requirement explicitly mentioned in the consent form. Next, we excluded all participants who indicated that we should not use their data in the SRSI UseME item ($n = 9$) [25].

Further, participants who gave wrong answers to both an instructed response item and a bogus item were excluded (n = 13). After data cleaning, the final sample consisted of N = 140. Concerning gender, 122 participants identified as men and 13 as women. One participant identified as non-binary, two preferred not to disclose, and two chose to self-describe. The primarily male sample was foreseeable because the subreddit is known to be predominantly male [12]. Participants' age ranged from 18 to 48 (M = 24 years, SD = 4.35). Sixty-three (45%) participants worked full-time, 22 (15.71%) part-time and 55 (39.29%) were unemployed. Most participants played the game on Microsoft Windows (n = 138), and two on MAC OS. Most players played up to twelve hours per week (n = 37) and up to three hours per session (n = 47). Demographics were adopted from Cai et al. [4] to ensure comparability. The sample was mostly comparable in terms of employment and time spent playing (total hours and hours per session), but not in gender and age, with our sample being more male-dominated and most participants under 30 (91.43%). We further requested the in-game ranking. The most common rank was platinum (n = 35), followed by silver and diamond (both n = 25). All ranks, from unranked to challenger, were present at least once.

2.3 Procedure

Participants completed an online survey created with EFS Unipark. After providing informed consent, they gave demographic information. Next, they were instructed to reflect on when they felt most "in the zone" while playing LoL. This scenario task was adapted from Jackson et al. [17]. Afterward, participants filled in the VGDFS, followed by the convergent and divergent questionnaires. Finally, they answered the SRSI UseME question [25] and could give feedback or ask questions. A printout of the online survey, including exact item wording for all questionnaires used, can be found on the OSF repository https://osf.io/sfr39/.

3 RESULTS

3.1 Confirmatory Factor Analysis

We first performed confirmatory factor analyses (CFA) to test the model fit of the VGDFS. For these analyses, we chose the proposed models by Cai et al. [4]; first, the original model following the theory by Jackson and Marsh [16] and then the model with the best model fit in the study by Cai et al. [4], which was the hierarchical antecedent model with the prerequisite *Clear Goals*, *Unambiguous Feedback*, and *Challenge-skill Balance*. Mardia [23]'s test indicated a violation of the multivariate normality assumption. Therefore, we chose to use the robust maximum-likelihood estimator. Results suggested that the proposed models do not fit the data, neither the unidimensional model [$\chi^2(341) = 594.53$, $p < .001$, $\chi^2/df = 1.747$, $CFI = .86$, $SRMR = .13$, $RMSEA = .07$, $GFI = .77$, $AGFI = .72$,$IFI = .86$], nor the hierarchical antecedent model [$\chi^2(314) = 561.97$, $p < .001$, $\chi^2/df = 1.64$, $CFI = .85$, $SRMR = .13$, $RMSEA = .08$, $GFI = .77$, $AGFI = .72$, $IFI = .85$]. While RMSEA was below the cutoff used by Cai et al. [4], it was still not optimal according to other cutoffs [14, 33]. In the original study by Cai et al. [4] the standardized loadings from *Flow* to *Transformation of Time* and *Loss of Self-consciousness* were quite low, which was replicated in our research in both models (Unidimensional model: $Flow \sim TT = .138$ and $Flow \sim LSS = .256$; Hierarchical antecedent model: $Flow \sim TT = .158$ and $Flow \sim LSS = .286$). Further, Cai et al. [4] excluded items with standardized residuals greater than |4.0|, low loadings, and misspecification according to modification indices. Looking at the standardized residuals, we found that in both models all the items in *Loss of Self-consciousness* and *Transformation of Time* shared an unacceptable degree of error according to the authors' standards. The loading for item LSS2 was also considerably lower than for the other items of *Loss of Self-consciousness*, both in the unidimensional (LSS2: 0.67; LSS1: 0.95; LSS3: 0.87) and hierarchical antecedent model (LSS2: 0.67; LSS1: 0.95;

LSS3: 0.87). Further, modification indices pointed at a misspecification between the latent variables *Transformation of Time* and *Loss of Self-consciousness*. We thus concluded that further investigation of the VGDFS was required.

3.2 Reliability and Item-analysis

Next, we analyzed the reliability of all constructs using coefficients alpha (α) [6] and omega (ω) [24] following Dunn et al. [9]. We used the bias-corrected and accelerated bootstrap hierarchical ω as recommended by Kelley and Pornprasertmanit [20]. Most constructs showed acceptable internal consistency (values above .70). Only *Sense of Control* seemed problematic, α = .64, 95% CI [.52,.73], ω = .64, 95% CI [.54,.75]. Concerning item difficulties, most items were rather easy (M = 63), though there was no value above 82.98 or below 26.67. We considered all items with variances below 1.00 potentially problematic. Our analysis revealed that the item CON2 (var = 0.97), item AE2 (var = 0.98), and especially item CON1 (var = 0.75) had low variance. There was no need for concern when looking at discriminatory power, with all values above 0.51. The inter-correlation matrix showed no alarming results, though the construct *Sense of Control* had a rather low mean-correlation within its construct (r = 0.38). Refer to the OSF repository for details.

3.3 Exploratory Factor Analysis

After the item analysis, we conducted exploratory factor analysis (EFA). We chose promax-rotation (an oblique rotation) and the principal axis factoring method because of the non-normality of the data and because we expected some factors to be correlated. We did not exclude any items based on the item analysis because less than two indicators per construct can be problematic [26] and we wanted to keep the original set of items. Bartlett's test for sphericity was significant ($\chi^2(378, N = 140) = 2114.8$, $p < 0.001$) and the Kaiser-Meyer-Olkin measure was adequate (KMO = 0.83, all above 0.69). Parallel analysis and inspection of a scree plot suggested a 4-factor solution. The standardized loadings > 0.20 of the 4-factor solution are shown in Table 2. The results suggested that items for *Action-Awareness merging* and *Autotelic Experience* loaded together onto factor four, and items for *Clear Goals* and *Concentration* loaded onto factor one. Items for *Sense of Control* also loaded on the same factor as items for *Clear Goals* and *Concentration*, except for the item SC2, which had low loadings on all factors (max = 0.37). *Transformation of Time* loaded on the same factor as items for *Loss of Self-consciousness*. The items in *Challenge-skill Balance* all loaded onto one factor. All *Unambiguous Feedback* loadings were relatively low (< 0.50), and the items UF1 and UF3 had cross-loadings onto factor one. Further, half of the items had communality values < .50, which was used as a cutoff by Cai et al. [4]. However, no communality was below .30 except for the items AAM2 and SC2.

Because no theory has yet depicted flow as a four-dimensional construct, we also tested a 9-factor solution as proposed by Cai et al. [4] and Jackson and Marsh [16]. We again used a promax rotation and principal axis factoring. The loadings > 0.20 for the 9-factor solution can be seen in Table 3. Interestingly, *Action-Awareness merging*, *Autotelic Experience*, *Challenge-skill Balance*, *Transformation of Time*, *Concentration*, and *Unambiguous Feedback* followed the theoretical assumptions and loaded onto different factors without high cross-loadings. In *Clear Goals*, only the item CG2 loaded highly, while items CG1 and CG3 did not load highly on any factor (max = .44) and had cross-loadings. Further, item SC1 loaded weakly onto the same factor as items for *Challenge-skill Balance*. The other items in *Sense of Control* otherwise loaded onto the same factor. *Unambiguous Feedback* followed theoretical assumptions and had no cross-loadings, although UF2 loaded considerably lower than the other items (0.49). The items LSS1 and LSS3 had cross-loadings on the factor where *Transformation of Time* loaded, except for item LSS2, which only loaded onto one factor. Again, some communalities were below .50, but none below .30. Overall the EFAs suggest that the VGDFS has items that follow the theoretical assumptions well, while others do not. Some of the problematic items were already

Table 2. Factor loadings > 0.20 for the 4-Factor EFA

Dimension	Item	PA1	PA2	PA3	PA4	Communality	Uniqueness	Complexity
Clear Goals	CG1	0.61	-0.21			0.51	0.49	1.4
	CG2	0.52	-0.24	0.22		0.43	0.57	1.8
	CG3	0.57				0.47	0.53	1.4
Unambiguous Feedback	UF1	0.32		0.37		0.36	0.64	2.0
	UF2			0.43		0.31	0.69	1.4
	UF3	0.38	0.22	0.36		0.39	0.61	3.0
Challenge-skill Balance	CSB1			0.78		0.61	0.39	1.1
	CSB2			0.85		0.63	0.37	1.0
	CSB3			0.82		0.64	0.36	1.1
	CSB4	0.24		0.58		0.49	0.51	1.4
Concentration	CON1	0.70				0.54	0.46	1.3
	CON2	0.70		-0.28	0.27	0.56	0.44	1.6
	CON3	0.79		-0.26		0.52	0.48	1.3
Action-Awareness merging	AAM1				0.57	0.42	0.58	1.1
	AAM2				0.43	0.28	0.72	1.5
	AAM3				0.70	0.52	0.48	1.1
Sense of Control	SC1	0.57				0.40	0.60	1.2
	SC2	0.37				0.26	0.74	1.4
	SC3	0.63				0.43	0.57	1.0
Loss of Self-consciousness	LSS1		0.88			0.74	0.26	1.1
	LSS2	0.26	0.60			0.48	0.52	1.4
	LSS3		0.82			0.71	0.29	1.2
Transformation of Time	TT1		0.70			0.56	0.44	1.3
	TT2	-0.28	0.80	0.21		0.70	0.30	1.4
	TT3		0.84			0.69	0.31	1.1
Autotelic Experience	AE1		-0.23		0.56	0.33	0.67	1.4
	AE2				0.64	0.46	0.54	1.1
	AE3			0.24	0.53	0.38	0.62	1.4

Table 3. Factor loadings > 0.20 for the 9-Factor EFA

Dimension	Item	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9	Communality	Uniqueness	Complexity
Clear Goals	CG1			0.24		-0.22				0.39	0.58	0.42	3.4
	CG2								1.17		0.92	0.08	1.2
	CG3	0.22		0.44					0.30		0.55	0.45	2.8
Unambiguous Feedback	UF1						0.74			0.22	0.68	0.32	1.3
	UF2						0.49				0.44	0.56	1.8
	UF3						0.86				0.73	0.27	1.2
Challenge-skill Balance	CSB1	0.92									0.70	0.30	1.2
	CSB2	0.95									0.72	0.28	1.1
	CSB3	0.80									0.65	0.35	1.2
	CSB4	0.57		0.24							0.57	0.43	1.9
Concentration	CON1			0.83							0.68	0.32	1.1
	CON2			0.82							0.66	0.34	1.2
	CON3			0.70							0.55	0.45	1.5
Action-Awareness merging	AAM1				0.74						0.61	0.39	1.2
	AAM2				0.66						0.44	0.56	1.2
	AAM3				0.94						0.84	0.16	1.1
Sense of Control	SC1	0.21							0.33	0.22	0.49	0.51	4.4
	SC2									0.60	0.41	0.59	1.3
	SC3									0.55	0.55	0.45	1.6
Loss of Self-consciousness	LSS1		0.33			0.76					0.88	0.12	1.5
	LSS2		0.23			0.50					0.53	0.47	1.8
	LSS3		0.33			0.64					0.77	0.23	1.6
Transformation of Time	TT1		0.75								0.63	0.37	1.0
	TT2		0.97								0.90	0.10	1.0
	TT3		0.79								0.74	0.26	1.1
Autotelic Experience	AE1							0.85	0.26		0.70	0.30	1.4
	AE2							0.62			0.58	0.42	1.4
	AE3	0.22						0.54			0.48	0.52	1.6

conspicuous in the original validation study, which hints at a general problem with the measurement of *Sense of Control* and *Loss of Self-consciousness*. Further, some items in *Clear Goals* seem to be problematic in the specific gaming context. Interestingly, the four components structure suggested by parallel analysis has not been observed in previous studies, either indicating a problem with the questionnaire or opening an avenue for refining the theoretical structure of flow.

3.4 Convergent and Divergent Validity

To assess convergent validity, all scales hypothesized to be convergent (DFS-S, IMI) were fit into one model to calculate measurement-error-free correlations between the latent variables [10]. The VGDFS correlated according to theory with $r = .54$ with the IMI. Between VGDFS and DFS-S, we found an astonishingly high correlation of $r = .97$. The same

procedure was conducted with the VGDFS and the divergent questionnaires. As expected, the correlations between the VGDFS and the two questionnaires were negative, although not very high. The VGDFS correlated with $r = -.19$ for the STAIT-5 and $r = -.17$ for the SBPS . Furthermore, the correlations between the two convergent and divergent scales were also in the expected direction. The STAIT-5 and SBPS correlated at $r = .58$, and the IMI and DFS-S at $r = .41$.

4 DISCUSSION

We found that neither the original nor the antecedent-hierarchical model for the VGDFS fit our data in the CFA. We further identified items CON2 (*... I remain concentrated.*), AE2 (*... the gaming session makes me feel great.*), and especially CON1 (*... I focus on the game.*) as problematic items due to low item variance. While statistical, a 4-factor solution was suggested, we also conducted an EFA with nine factors to acknowledge theoretical considerations. Interestingly, items mostly loaded according to theory in the 9-factor EFA except for *Clear Goals*, *Sense of Control*, and *Loss of Self-consciousness*. Lastly, we found evidence for convergent and divergent validity. Overall, our attempt at confirmation was unsuccessful. We identified problematic items and constructs in different steps of the analysis.

In the 4-factor EFA, we found a different picture from the existing theories. *Challenge-skill Balance* and *Unambiguous Feedback* loaded onto the same factor in our data, which corresponds to the proposed closer association between these constructs in the hierarchical antecedent model by Cai et al. [4], albeit without *Clear Goals*.

Several issues emerged with items for *Sense of Control*. The items for *Sense of Control* had low mean-correlations ($r = 0.38$) and low values of internal consistency ($\alpha = .64$, $\omega = .64$). Concerning the EFAs, SC2 (*... I feel a sense of control in the game.*) loaded poorly in the 4-factor EFA, and had low communality in both EFAs, while loadings for item SC1 (*... I can perceive the natural control of the game.*) were unsatisfactory in the 9-factor solution. Not only low loadings but also cross-loadings were a problem in the 9-factor EFA. *Clear Goals* only had one strong loading (CG2), while *Sense of Control* had two (SC2, SC3). Also, the two constructs seem to share some characteristics, as item CG1 (*... I know how to proceed during the gaming session.*) loaded more onto the *Sense of Control*-dominated factor while item SC1 (*... I can perceive the natural control of the game.*) loaded more onto the *Clear Goals*-dominated factor, which again highlights the problematic nature of *Sense of Control*. Because Cai et al. [4] also dropped *Sense of Control* items in their last validation step due to communality issues, we argue that the construct of *Sense of Control* warrants further investigation. Perhaps, it should be investigated if there needs to be more physical movement to achieve a sense of control, which is the argument made by Procci and Bowers [30].

The dimensions *Loss of Self-consciousness* and *Transformation of Time* fell together in our 4-factor solution. Whereas in the 9-factor solution, there were cross-loadings of items LSS1 (*... I forget about things in the real world.*) and LSS3 (*... I forget about what is occurring in the real world.*) on the same factor as the items of *Transformation of Time*. The connection between these constructs was problematic in both the original work and ours, reflected in low CFA loadings. We argue that it should be further investigated if the addition of *Transformation of Time* by Jackson and Marsh [16] compared to the model by Csikszentmihalyi [8] holds in specific gaming contexts.

Taken together, our analyses show that some items of the VGDFS have remained problematic since its construction and that several additional items have unsatisfactory psychometric properties in the specific context of LoL. The EFAs revealed that some constructs might be closely related or even the same. For *Challenge-Skill-Balance* and *Unambiguous Feedback* this closeness was expected, while we could also show an unanticipated relatedness between the dimensions of *Loss of Self-consciousness* and *Transformation of Time* as well as *Clear Goals* and *Sense of Control*. We further highlight several issues with specific items of the VGDFS that that need to be addressed. While all the identified items need further investigation, most concerning were items of the dimension *Sense of Control*. Given that game researchers need

to ensure that their measurement has a solid theoretical foundation [1], such theoretical considerations need to be clarified when applying the VGDFS. To provide some clarification, we plan on expanding the validation of the VGDFS onto other MOBAs and genres. Another interesting finding is the extremely high correlation (r = .97) between the DFS-S and the VGDFS, suggesting that it is questionable whether the 28-item VGDFS is distinct enough to justify its use over the shorter 9-item DFS-S. Lastly, our study found evidence for divergent validity of the VGDFS to the STAIT-5 and SBPS questionnaires. Yet, because the negative correlations were relatively low, we can only provide initial evidence that flow, as measured with the VGDFS, lies at the intersection of skill and balance between anxiety and boredom.

5 LIMITATIONS AND NEXT STEPS

An important limitation is that our recruitment method could have led to a sampling bias, with only highly motivated people participating in or completing the survey. While this could be positive for data quality, it could be possible that specific individuals systematically refrained from participating in the survey. This is also reflected in the in-game ranks, which were relatively high. Indicating that more competitive players were interested in this survey than typical LoL players, although our sample included all ranks from unranked to challenger. In the present work, we did not account for player competitiveness, which we will do in our future work. Because MOBA games are competitive games in general [18] it could be interesting to see if differences in player competitiveness influence dimensions of flow and if this is true for different MOBAs. Therefore, we plan to continue with an initial comparison between LoL and Dota 2 players [36]. Such a comparison will clarify the problematic nature of some dimensions that seem specific to MOBAs and allow for studying the influence of competitiveness. Further, a comparison of similar games could uncover if the genre of the game or specific game mechanics are decisive factors for flow. For instance, in LoL, players can only kill enemy minions and not allied minions, as in Dota 2, which could influence goal clarity. Finally, future research could also compare the VGDFS to the DFS-S and the longer DFS-2 by Jackson et al. [17] to further test the differences between flow outside and inside of a gaming context.

6 CONCLUSION

We present an initial step of an independent validation for the VGDFS in the context of LoL. We could not confirm the proposed factor structure using CFA. Although a larger sample is needed to corroborate our findings, we could still identify possibly problematic items and constructs in the VGDFS. Also, our data suggest a 4-factor solution in EFA, which has not been mentioned in the literature before. Furthermore, we found evidence for the theorized 9-factor solution using a second EFA. In addition, we show evidence for convergent validity between the VGDFS and DFS-S and the IMI. However, differences between the VGDFS and DFS-S remain questionable as correlations were unusually high. Likewise, we found weak negative correlations between the VGDFS and STAIT-5, as well as VGDFS and SBPS, which supports the proposed inverse relationships of flow with anxiety and boredom. Overall, our work highlights the need to investigate further the flow construct and its measurement in different gaming contexts, specifically with the VGDFS.

7 DATA AVAILABILITY

All data, analysis code, and materials are available on https://osf.io/sfr39/.

ACKNOWLEDGMENTS

We want to thank Nathalie Herles for their contribution in earlier stages of this project.

REFERENCES

[1] Lena Fanya Aeschbach, Sebastian A. C. Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. Transparency in Measurement Reporting: A Systematic Literature Review of CHI PLAY. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 233 (oct 2021), 21 pages. https://doi.org/10.1145/3474660

[2] Florian Brühlmann, Philipp Baumgartner, Günter Wallner, Simone Kriglstein, and Elisa D. Mekler. 2020. Motivational Profiling of League of Legends Players. *Frontiers in Psychology* 11, 1307 (2020), 18. https://doi.org/10.3389/fpsyg.2020.01307

[3] Florian Brühlmann, Serge Petralito, Lena Aeschbach, and Klaus Opwis. 2020. The Quality of Data Collected Online: An Investigation of Careless Responding in a Crowdsourced Sample. *Methods in Psychology* 2 (04 2020), 100022. https://doi.org/10.1016/j.metip.2020.100022

[4] Xiaowei Cai, Javier Cebollada, and Mónica Cortiñas. 2022. Self-report measure of dispositional flow experience in the video game context: Conceptualisation and scale development. *International Journal of Human-Computer Studies* 159 (2022), 102746. https://doi.org/10.1016/j.ijhcs.2021.102746

[5] Kevin D Carlson and Andrew O Herdman. 2012. Understanding the impact of convergent validity on research results. *Organizational Research Methods* 15, 1 (2012), 17–32. https://doi.org/10.1177/1094428110392383

[6] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 3 (1951), 297–334. https://doi.org/10.1007/BF02310555

[7] Mihaly Csikszentmihalyi. 1990. *Flow: the psychology of optimal experience* (1st ed.). Harper & Row, New York.

[8] Mihaly Csikszentmihalyi. 2000. *Beyond boredom and anxiety: experiencing flow in work and play* (25. anniversary ed.). Jossey-Bass, San Francisco.

[9] Thomas J. Dunn, Thom Baguley, and Vivienne Brunsden. 2014. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology* 105, 3 (2014), 399–412. https://doi.org/10.1111/bjop.12046

[10] Michael Eid, Mario Gollwitzer, and Manfred Schmitt. 2017. *Statistik und Forschungsmethoden* (5th ed.). Beltz, Weinheim ; Basel.

[11] Entertainment Software Association (ESA). 2022. 2022 Essential Facts About the Video Game Industry. Game [PC]. https://www.theesa.com/resource/2022-essential-facts-about-the-video-game-industry/

[12] ITB Esports. 2021. r/League Demographic Information: Breakdown. https://www.docdroid.net/1EZIHWK/rleague-demographics-report-2021-pdf

[13] David J. Harris, Kate L. Allen, Samuel J. Vine, and Mark R. Wilson. 2021. A systematic review and meta-analysis of the relationship between flow states and performance. *International Review of Sport and Exercise Psychology* 0, 0 (2021), 1–29. https://doi.org/10.1080/1750984X.2021.1929402 arXiv:https://doi.org/10.1080/1750984X.2021.1929402

[14] Li-tze Hu and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6, 1 (1999), 1–55. https://doi.org/10.1080/10705519909540118

[15] Susan A. Jackson, Stephen K. Ford, Jay C. Kimiecik, and Herbert W. Marsh. 1998. Psychological Correlates of Flow in Sport. *Journal of Sport and Exercise Psychology* 20, 4 (1998), 358–378. https://doi.org/10.1123/jsep.20.4.358

[16] Susan A. Jackson and Herbert W. Marsh. 1996. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *Journal of Sport and Exercise Psychology* 18, 1 (1996), 17–35. https://doi.org/10.1123/jsep.18.1.17

[17] Susan A. Jackson, A.J. Martin, and R.C. Eklund. 2008. Long and Short Measures of Flow: Examining Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts. *Journal of Sport and Exercise Psychology* 30 (2008), 561–587. https://doi.org/10.1123/jsep.30.5.561

[18] Daniel Johnson, Lennart E. Nacke, and Peta Wyeth. 2015. All about That Base: Differing Player Experiences in Video Game Genres and the Unique Case of MOBA Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2265–2274. https://doi.org/10.1145/2702123.2702447

[19] Dominik Kayser, Sebastian Andrea Caesar Perrig, and Florian Brühlmann. 2021. Measuring Players' Experience of Need Satisfaction in Digital Games: An Analysis of the Factor Structure of the UPEQ. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Austria) *(CHI PLAY '21)*. Association for Computing Machinery, New York, NY, USA, 158–162. https://doi.org/10.1145/3450337.3483499

[20] Ken Kelley and Sunthud Pornprasertmanit. 2016. Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods* 21, 1 (2016), 69–92. https://doi.org/10.1037/a0040086

[21] Yubo Kou, Xinning Gui, and Yong Ming Kow. 2016. Ranking Practices and Distinction in League of Legends. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 4–9. https://doi.org/10.1145/2967934.2968078

[22] MultiMedia LLC. 2022. *League of Legends player count 2022: how many people play League of Legends?* techacake.com. https://techacake.com/league-of-legends-player-count/

[23] Kantilal Vardichand Mardia. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 3 (1970), 519–530. https://doi.org/10.1093/biomet/57.3.519

[24] Roderick P McDonald. 1999. *Test theory: A unified treatment.* Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA.

[25] Adam W. Meade and S. Bartholomew Craig. 2012. Identifying careless responses in survey data. *Psychological Methods* 17, 3 (2012), 437–455. https://doi.org/10.1037/a0028085

[26] Daniel J. Mundfrom, Dale G. Shaw, and Tian Lu Ke. 2005. Minimum Sample Size Recommendations for Conducting Factor Analyses. *International Journal of Testing* 5, 2 (2005), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

[27] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. *The Concept of Flow.* Springer Netherlands, Dordrecht, 239–263. https://doi.org/10.1007/978-94-017-9088-8_16

[28] Corinna Peifer and Stefan Engeser (Eds.). 2021. *Advances in Flow Research*. Springer International Publishing, Gewerbestrasse 11, 6330 Cham, Switzerland. 453 pages. https://doi.org/10.1007/978-3-030-53468-4

[29] Jiaxi Peng, Wei Guo, Luming Zhao, Xiaochen Han, and Shengjun Wu. 2020. Short Boredom Proneness Scale: Adaptation and validation of a Chinese version with college students. *Social Behavior and Personality: an international journal* 48, 2 (2020), 1–8. https://doi.org/10.2224/sbp.8968

[30] Katelyn Procci and Clint Bowers. 2011. An Examination of Flow and Immersion in Games. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55, 1 (2011), 2183–2187. https://doi.org/10.1177/1071181311551455

[31] Riot Games. 2009. League of Legends. Game [PC]. Riot Games, Los Angeles, CA, USA.

[32] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 1 (2000), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

[33] James H. Steiger. 2007. Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences* 42, 5 (2007), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

[34] Andriy A. Struk, Jonathan S. A. Carriere, J. Allan Cheyne, and James Danckert. 2017. A Short Boredom Proneness Scale: Development and Psychometric Properties. *Assessment* 24, 3 (2017), 346–359. https://doi.org/10.1177/1073191115609996

[35] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment* 3, 3 (2005), 3. https://doi.org/10.1145/1077246.1077253

[36] Valve Corporation. 2013. Dota 2. Game [PC]. Valve Corporation, Bellevue, WA, USA.

[37] Andras N. Zsido, Szidalisz A. Teleki, Krisztina Csokasi, Sandor Rozsa, and Szabolcs A. Bandi. 2020. Development of the short version of the spielberger state—trait anxiety inventory. *Psychiatry Research* 291 (2020), 113223. https://doi.org/10.1016/j.psychres.2020.113223