Int. J. Human–Computer Studies 159 (2022) 102746

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

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

Self-report measure of dispositional flow experience in the video game context: Conceptualisation and scale development

Xiaowei Cai *, Javier Cebollada, Mónica Cortiñas

Institute for Advanced Research in Business and Economics (INARBE), Department of Business Management, Public University of Navarre (UPNA), Arrosadia Campus, 31006 Pamplona, Spain

ARTICLE INFO

Keywords:
Flow experience
Flow scale
Video games
Scale development

ABSTRACT

The flow theory has been widely applied to explain video game players' gaming and purchasing behaviour. However, due to the conceptual and empirical flaws of the current measurement instruments, researchers can hardly apply them to measure dispositional flow experience of adult video game players. In this research, we conceptualised flow experience and developed its measurement instrument in the video game context. To achieve these objectives, we conducted five phases with different participants in each of them: conceptualisation of the constructs and item generation ($n = 13$), expert judging ($n = 5$), pre-test ($n = 96$), initial development and validation ($n = 289$), and advanced development and validation ($n = 593$). We applied both qualitative and quantitative analysis to conceptualise and measure flow experience of video game players, including grounded theory and several statistical tools of latent variable modelling. We obtained a scale of 28-items that performs well in the first-order model. Moreover, we tested three hierarchical structure of flow experience: unidimensional model, independent antecedent model, and hierarchical antecedent model. Results show that hierarchical antecedent model is the best structure to represent flow experience. We named our scale *Video Game Dispositional Flow Scale* (VGDFS).

1. Introduction

Flow experience refers to those moments when everything comes together to create a state of absorption and enjoyment in what one is doing (Csikszentmihalyi, 1975, 1990). Researchers have argued that the flow experience is responsible for the positive emotions during video game playing (Hoffman and Novak, 1996; Michailidis et al., 2018; Nah and Hall, 2014). Therefore, flow theory has been widely applied in video game studies. On the one hand, researchers have studied the consequences of flow experience, and found that flow experience was positively related to the intention to play video games (Chang, 2013; Hsu and Lu, 2004; Shin and Shin, 2011; Zhou, 2013), attitude towards playing video games (Ha et al., 2007), satisfaction (Kim and Ko, 2019; Sepehr and Head, 2018), loyalty (Hsiao and Tang, 2016), and purchase intention of in-game goods (Animesh et al., 2017; Hamari and Keronen, 2017; T. Huang et al., 2017). Thus, due to the positive consequences of flow experience, studying flow experience is critical to analyse gaming behaviour for both researchers and practitioners in the video game industry. On the other hand, researchers have investigated the antecedents of flow experience and found that players' character identification (Soutter and Hitchens, 2016) and functional-based in-game goods purchase (Cai et al., 2020) lead to flow experience.

In marketing and consumer research, the questionnaire approach is widely used to measure psychometric properties. The first step in a scale development process is the conceptualisation (DeVellis, 2016; Mackenzie et al., 2011; Netemeyer et al., 2003), as the validity of what is being measured rest largely on the conceptual definition of the domain (Netemeyer et al., 2003). In defining flow experience, Csikszentmihalyi (1975) described six components. Later, other researchers established their measurement models of flow in different contexts. Among the various conceptualisations of flow, Jackson's (1996) nine components structure in the context of sports stands out, because it offers a comprehensive characterization of flow (Moneta, 2012). Moreover, scholars established the conceptualisation of flow in the video game context by incorporating two additional dimensions, immersion and social interaction, and they named it *GameFlow* (Sweetser et al., 2017; Sweetser and Wyeth, 2005). However, as Michailidis et al. (2018) pointed out, immersion and flow do not appear to be conceptually

* Corresponding author.
E-mail address: xiaowei.cai@unavarra.es (X. Cai).

https://doi.org/10.1016/j.ijhcs.2021.102746
Received 21 March 2021; Received in revised form 23 November 2021; Accepted 24 November 2021
Available online 29 November 2021

distinct. Therefore, a problem of construct proliferation[1] (Bergkvist and Eisend, 2021; Podsakoff et al., 2016) may exist in the GameFlow framework.

The various conceptualisations of flow experience have given rise to distinct flow measurement scales. Drawing on the original study of flow (Csikszentmihalyi, 1975), researchers developed the first self-report measure, the flow questionnaire (FQ) (Csikszentmihalyi and Csikszentmihalyi, 1992). Moreover, following Jackson's (1996) conceptualisation of flow in the sports context, she and her colleagues conducted a great number of studies to develop and refine the flow scale, including the development of the flow state scale (FSS) (Jackson and Marsh, 1996), the establishment of the nomological network of flow (Jackson et al., 1998), the development of the refined version of FSS, Flow State Scale-2 (FSS-2), the dispositional version, Dispositional Flow Scale-2 (DFS-2) (Jackson and Eklund, 2002), and the development of the short versions of FSS-2 and DFS-2 (Jackson et al., 2008). While the dispositional scale measures the frequency of the flow experience across time domains, the state scale measures the extent of the flow experienced in an activity.

In the context of video games, researchers built upon the Gameflow measurement model (Sweetser et al., 2017; Sweetser and Wyeth, 2005) to develop *EGameFlow*, a scale to measure learners' enjoyment of e-learning games (Fu et al., 2009). Additionally, researchers refined items used in the previous conceptual and empirical studies to develop a flow scale for the video game context (Fang et al., 2013). Researchers have also developed several flow alike scales to measure the immersion and engagement in the context of video games, in which the flow experience is one of the key dimensions. These scales are the immersion questionnaire (IQ) (Jennett et al., 2008), the game engagement questionnaire (GEQ) (Brockmyer et al., 2009), the user engagement scale (UES) (Wiebe et al., 2014), and the consumer video engagement scale (CVES) (Abbasi et al., 2017, 2019).

However, under some circumstances, existing measurement tools are either inappropriate or does not exist at all (DeVellis, 2016). Inducement of the flow experience is context-dependent (Fang et al., 2013). Previous evidence suggests that the flow experienced in the video game environment is different from that experienced during physical activity, possibly because video gaming is less physically demanding than sport (Wang et al., 2009). Furthermore, previous attempts to measure flow experience in video game context (Fang et al., 2013) combined items meant to measure dispositional flow with items for measuring state flow. Meanwhile, although DFS-2 has been verified as an adequate instrument to measure dispositional flow in some information systems, such as gamification setting (Hamari and Koivisto, 2014), its applicability in the video game context has been questioned due to the mixed findings (Procci et al., 2012; Wang et al., 2009). While the DFS-2 was valid for measuring dispositional flow in adolescent players (Wang et al., 2009), researchers failed to replicate the findings among adult players (Procci et al., 2012). Similarly, although the EGameFlow scale was validated for the e-learning games context, researchers found that the original 31-item scale underperformed in the entertainment game setting (Chen et al., 2018). Meanwhile, IQ, GEQ, UES, and CVES contain only some of the dimensions of the flow experience. Dropping any component implies the loss of the corresponding definitional aspects of flow experience (Engeser and Schiepe-Tiska, 2012).

This research has two objectives. The first objective is to conceptualise the flow experience in the video game context while avoiding the construct proliferation problem. The second objective is to develop a dispositional flow scale applicable to adult video game players.

Through reviewing the literature of flow experience and interviewing video game players, we adapted the conceptualisation of the nine dimensions of flow experience (Jackson, 1996) from the sports context to the video game context. Additionally, we proposed that flow experience could be operationalised through three different models: an unidimensional model, an independent antecedent model, and a hierarchical antecedent model. Through reviewing literature and applying grounded theory (Charmaz, 2014), we generated an initial pool of 64 items. Later, we followed a structured scale development procedure (Churchill, 1979; DeVellis, 2016; Hinkin, 1995; Mackenzie et al., 2011; Netemeyer et al., 2003) to trim, modify, and validate the initial item pool through different quantitative phases, including pre-test, initial scale validation, and advanced scale validation, until we had a refined scale with 28 items. Finally, we named the scale that we created in this research as *video game dispositional flow scale* (VGDFS).

This study has both theoretical and practical implications. On the theoretical side, we clarified the dimensions of the flow experience in the video game context, as well as introducing the most supported operationalisation of flow: the hierarchical antecedent model. On the practical side, we developed a 28-items dispositional flow scale. Compared with previous research (Procci et al., 2012), we found that the VGDFS outperformed the DFS-2 for adult players in the video game context. The adult players in the entire gamer population are especially important because the majority of video game players (67%) are adults between 20 and 50 (Bosman, 2020). Moreover, unlike copyrighted scales like DFS-2, we responded to the initiative of Hays et al. (2018) by removing potential financial barriers to the use of psychometric measures.

2. Literature review on flow theory, flow scales, and their application in video game studies

Csikszentmihalyi (1975) conceptualised the flow as an experience that occurs when the challenge of the task and the skill of the performer are balanced. A person in such a state can make full use of whatever skills are required and receives clear feedback on his/her action (Csikszentmihalyi, 1975). At the same time, there is no time to get bored or to worry about what may or may not happen (Csikszentmihalyi, 1975).

In essence, flow theory is a set of intercorrelated constructs, with their definitions and propositions, that systematically describe the relationships among the constructs to explain and predict a set of measurable outcomes (Moneta, 2012). Thus, "*flow cannot be reduced to a single component, and all attempts to take one component of flow as the definitional aspect of flow will consequently disregard essential parts* (Engeser and Schiepe-Tiska, 2012, p. 4)". Csikszentmihalyi (1975) proposed six components of flow experience based on interviews, including *merging of action and awareness, centring of attention, loss of self-consciousness, the feeling of control, coherent and noncontradictory demands*, and *autotelic nature*. In the sports context, Jackson (1996) proposed nine dimensions of flow experience based on the original flow theory (Csikszentmihalyi, 1975, 1990): *clear goals, unambiguous feedback, challenge-skill balance, concentration, action-awareness merging, sense of control, loss of self-consciousness, transformation of time*, and *autotelic experience*. In the video game context, Sweetser and Wyeth (2005) proposed an eight-dimension flow framework (GameFlow), in which *challenge* and *skill* were separated into two distinct dimensions, while *immersion* and *social Interaction* were incorporated as two new dimensions. Previous empirical evidence warranted that flow experience could be represented by a single higher order component only (Beard and Hoy, 2010; Engeser and Rheinberg, 2008; Hamari and Koivisto, 2014; Jackson and Eklund, 2002; Jackson and Marsh, 1996). However, this conclusion is premature (Engeser and Schiepe-Tiska, 2012). Some researchers argued that clear goals, unambiguous feedback, and challenge-skill balance are actually the antecedents of flow experience rather than flow per se (Jackson, 2012; Keller and Landhäußer, 2012; Moneta, 2012; Nakamura and Csikszentmihalyi, 2012). These antecedents were conceptualised to affect flow experience either individually (Jackson, 2012; Nakamura and Csikszentmihalyi, 2012) or jointly (Keller and Landhäußer, 2012; Moneta, 2012).

[1] Construct proliferation problem exists when constructs appears under different names but the conceptual domains overlap (Podsakoff et al., 2016).

Apart from the antecedents of flow, researchers have discussed its consequences (Finneran and Zhang, 2003). One of the most important consequences of flow experience is the enjoyment (Csikszentmihalyi, 1975; Engeser and Schiepe-Tiska, 2012; Landhäußer and Keller, 2012; Sherry, 2004). This consequence is especially important in the video game context. Researchers found that, while enjoyment increased the willingness to play video games, it reduced the willingness to purchase in-game goods (Hamari, 2015). As flow experience and enjoyment are conceptually distinct (Landhäußer and Keller, 2012), researchers also developed a scale to measure enjoyment in the video game context (Fang et al., 2010).

Flow theory has important applications in video game research (Animesh et al., 2017; Chang, 2013; Ha et al., 2007; Hsiao and Tang, 2016; Hsu and Lu, 2004; T. Huang et al., 2017; Kim and Ko, 2019; Sepehr and Head, 2018; Shin and Shin, 2011; Soutter and Hitchens, 2016; Zhou, 2013). Nevertheless, many social science researchers encounter measurement issues (DeVellis, 2016). Sometimes, existing measurement tools are either inappropriate or do not exist at all (DeVellis, 2016). One common response is to rely on existing measurement instruments, although they may be unsuitable or inappropriate (DeVellis, 2016). Another common response is to assume that some newly compiled scales that "look" good and use them directly in the research (DeVellis, 2016). "*In point of fact, most of our measures are only measures because someone says that they are, not because they have been shown to satisfy standard measurement criteria* (Jacoby, 1978, p. 91)." Researchers, therefore, need to realise that by making casual use of a certain measuring instrument, they run the risk of obtaining inaccurate data and results (DeVellis, 2016).

Nevertheless, the development and validation of a new scale is a time-consuming and potentially costly endeavour (Netemeyer et al., 2003). Therefore, if a reliable scale already exists, the value of a new measure may be reduced compared to the costs of developing it (Netemeyer et al., 2003). A new scale should capture the targeted factor either more accurately or more efficiently than existing scales to have incremental validity (Clack and Watson, 1995). Thus, a literature review on flow scales allows us to understand whether we really need a new flow scale in the video game context.

In the video game context, researchers attempted to create a flow scale by refining the previous conceptual and empirical results of flow research (Fang et al., 2013). However, the scope for its application is limited by two problems. On the one hand, when researchers (Fang et al., 2013) created the initial item pool, items to measure both dispositional and state flow were included. Therefore, after refinement, there are both items to measure dispositional flow and those to measure state flow in the final scale, which makes it difficult for users to capture either dispositional or state flow. On the other hand, two pairs of two-indicator measures are found in the scale of Fang et al. (2013), which violates the three-indicator rule. Scales that contains two-indicator measurement can lead to some potential problems when estimating the model, such as failure of model identification and Heywood cases (Hair et al., 2013).

There have also been several attempts to measure flow experience capturing some of its components. We name these scales as "flow alike scales". These scales are immersion questionnaire (IQ) (Jennett et al., 2008), game engagement questionnaire (GEQ) (Brockmyer et al., 2009), user engagement scale (UES) (Wiebe et al., 2014), and consumer video engagement scale (Abbasi et al., 2017, 2019). Researchers later refined UES and introduced a shortened version of this scale, although not for the context of video games (O'Brien et al., 2018).

However, measuring all components of flow experience is becoming more prevalent (Engeser and Schiepe-Tiska, 2012). This approach has its advantage because dropping any component implies the loss of the corresponding definitional aspect (Engeser and Schiepe-Tiska, 2012). Based on the original six-components operationalisation of flow (Csikszentmihalyi, 1975), Csikszentmihalyi and Csikszentmihalyi (1988) developed the first measurement instrument of flow experience. Later, Jackson and her colleagues developed a set of scales to measure flow experience in the sports context. These include Long, Short, and Core flow scales, which serve for distinct research purposes (Jackson et al., 2011; Jackson et al., 2008; Jackson and Eklund, 2002; Jackson and Marsh, 1996; Martin and Jackson, 2008). The mentioned scales also have the trait and state versions, thus bringing the total to six different scales. Like other psychological concepts, flow experience also exhibits trait-state distinctions (Jackson et al., 2011). "*It is proposed that flow is a specific Psychological state amenable to state-based assessments, and also that people differ in their propensity to experience flow on a regular basis* (Jackson et al., 1998, p. 360)". While dispositional/trait scales of flow measure the general tendency to experience flow, state scales of flow measure the particular incidence of flow characteristics during a specific event (Jackson et al., 2011).

The flow scales developed by Jackson and her colleagues (Jackson et al., 2011, 2008; Jackson and Eklund, 2002; Jackson and Marsh, 1996) were intended for use in the sports setting. Therefore, other researchers questioned the applicability of these scales outside the sport context, especially in the video game context, because of the relative lack of physical movement involved (Procci et al., 2012). amongst the flow scales developed by Jackson and her colleagues (Jackson et al., 2011, 2008; Jackson and Eklund, 2002; Jackson and Marsh, 1996), FSS-2 and DFS-2 have been assessed in the video game context. The results of the state flow measure, however, were somewhat frustrating: when using FSS-2, video games that were not likely to induce flow still offered participants the opportunity for high levels of sense of control and merging of action awareness (Klarkowski et al., 2015). Regarding the dispositional flow measure, mixing findings have been found. Wang et al. (2009) recruited 1578 secondary school students, and they found that DFS-2 was a valid measurement instrument for assessing dispositional flow experience among child and adolescent video game players. Nevertheless, Procci et al. (2012) replicated the research by recruiting 762 undergraduate students, and they found that DFS-2 was not an adequate measurement instrument for adult video game players. Procci et al., considered that participants' age might affect the validity of DFS-2 in the research. Thus, there is no research in the current literature that empirically demonstrates a valid dispositional measure of flow in the video game context for adult players.

Scholars also have developed flow measures in the video game context. Based on the conceptualisation of the GameFlow framework (Sweetser and Wyeth, 2005), Fu et al. (2009) developed a self-report flow measure in the e-learning game setting, and they named it as EGameFlow. However, although the validity of the EGameFlow scale was verified in the e-learning game setting, researchers assessed its performance in the entertainment game setting and found it to yield low model fit (CFI = 0.892, TLI = 0.879, RMSEA = 0.087) (Chen et al., 2018). Therefore, the applicability of the EGameFlow scale may be limited to the e-learning game setting.

We summarised the above-mentioned dispositional flow scales in Table 1.

A review of the literature revealed two gaps in the research, as viewed both from the conceptual and the operational perspective. On the one hand, there is a lack of conceptual clarity of flow experience in the video game literature, where several frameworks of flow experience co-exist in the literature. On the other hand, we did not find a valid scale to measure dispositional flow experience for adult video game players in the literature. The following sections describe our conceptualisation of the flow experience and our approach for measuring its dispositional aspect in the video game context.

3. General methodology

To develop a reliable and valid scale to measure the dispositional flow for the context of video games, we followed the rigorous steps of the scale development process (Churchill, 1979; DeVellis, 2016; Hinkin, 1995; Mackenzie et al., 2011; Netemeyer et al., 2003). The entire scale

Table 1
Summary of flow scales and flow alike scales.

Scale type	Name of scale	Author(Year)	Theoretical foundation	Psychometric type	Number of items	Number of points	Structure	Context of development	Validation in video game context	Performance in video game context
Flow scales	LONG Dispositional flow scale (DFS-2)	Jackson and Eklund (2002)	Flow theory	Dispositional	36 items	5 point	Second-order	Physical activity	Yes	Performed well amongst adolescents but underperformed amongst adults
	SHORT Dispositional flow scale (S DFS)	Jackson et al. (2008)	Flow theory	Dispositional	9 items	5 point	First-order	Physical activity	No	Unknown
	CORE Dispositionnal flow scale (C DFS)	Martin and Jackson (2008)	Flow theory	Dispositional	10 items	5 point	First-order	Physical activity	No	Unknown
	LONG Flow state scale (FSS-2)	Jackson and Eklund (2002)	Flow theory	State	36 items	5 point	Second-order	Physical activity	Yes	Questionable
	SHORT Flow state scale (S FSS)	Jackson et al. (2008)	Flow theory	State	9 items	5 point	First-order	Physical activity	No	Unknown
	CORE Flow state scale (C FSS)	Martin and Jackson (2008)	Flow theory	State	10 items	5 point	First-order	Physical activity	No	Unknown
Flow alike scales	EGameFlow	Fu et al. (2009)	GameFlow	Dispositional	56 items	7 point	First-order	Serious video game	Yes	Performed well in the serious game setting but underperformed in the entertainment game setting
	Game engagement questionnaire (GEQ)	Brockmyer et al. (2009)	Flow theory, immersion, presence, absorption, dissociation	Dispositional	19 items	3 point	First-order	Violent video game	Yes	Performed well in the scale development study
	Consumer video engagement scale	Abbasi et al. (2017)	Engagement	Dispositional	29 items	7 point	Second-order	General video game	Yes	Performed well in the scale development and validation studies
	Immersion questionnaire	Jennett et al. (2008)	Flow theory, cognitive absorption, presence, immersion	State	31 items	5 point	First-order	First person shooter game	Yes	Performed well in the scale development study
	User engagement scale (UES)	Wiebe et al. (2014)	Engagement	State	28 items	5 point	First-order	Online video game	Yes	Performed well in the scale development study
	Refined user engagement scale (R UES)	O'Brien et al. (2018)	Engagement	State	30 items	5 point	First-order	E-shopping	No	Unknown
	Short user engagement scale (S UES)	O'Brien et al. (2018)	Engagement	State	12 items	5 point	First-order	E-shopping	No	Unknown

development process serves to ensure several types of validity, including face validity, content validity, convergent validity, discriminant validity, and nomological validity (DeVellis, 2016; Hinkin, 1995; Mackenzie et al., 2011; Netemeyer et al., 2003). There are five phases in our research, which serve to control the mentioned types of validity. While phases 1 and 2 control the face and content validity, phases 3, 4, and 5 control the convergent, discriminant, and nomological validity. Phase 5 also checks for measurement invariance across different gender and age groups among video game players. Table 2 summarises the research process.

In this research, we conducted several phases of data collection, including both online interviews and online surveys. In almost all phases (except the expert judging phase), our target population are video game players from the United States. We selected the samples from the United States video game market because it was the largest gaming market in terms of global revenue, totalling $36.9 billion in 2009 (Newzoo, 2019a). Driven by growth in console game revenues, it overtook China for the number one position (Newzoo, 2019a).

To approach the target population, we use a leading online panel, Prolific,[2] which serves as the sampling frame in our research. The use of an online panel saves both time and money while few disadvantages were observed (Casler et al., 2013). Among the online panels available, we selected Prolific because of several merits. On the one hand, Prolific has no reported functional shortcomings (Palan and Schitter, 2018). On the other hand, empirical results show that the participants on Prolific are more honest and less exposed to common research tasks than participants on other online panels, such as Mturk (Peer et al., 2017). Nevertheless, due to the policy restrictions of Prolific,[3] researchers are not able to screen the participants inside the questionnaire. If researchers wish to acquire a customised population that the default screening function does not provide, they need to conduct a two-steps sampling process[4] to precisely locate the target population. The participant pool in the pre-test phase, initial scale validation phase, and advanced validation phase are 624, 1671, and 1950 respectively.

In this research, we paid much attention to item redundancy, which is one of the theoretical bases of scale development (DeVellis, 2016). Although in the final scale, redundancy is undesirable, during the scale development phases, two items are worth keeping even if they differ by one word (DeVellis, 2016).

During the scale development, if an item did not perform well in the quantitative phase, we would create more items to reword the problematic item. Additionally, during scale development, items are often added, dropped, or reworded (Netemeyer et al., 2003). Therefore, re-estimating the measurement model using a new sample of data is important (Mackenzie et al., 2011). In this research, all the samples in the quantitative phases, including pre-test, initial scale validation, and advanced scale validation, are mutually independent, which helps to mitigate the common method variance (Hinkin, 1995).

The literature generally agrees that the response rate is the key metric to identify nonresponse (Callegaro et al., 2015). For the research using the non-probability-based panel, Callegaro and Disogra (2008) suggest reporting the completion rate, break-off rate, screening completion rate, and study-specific eligibility rate. Table 3 summarises the number of participants and the response rates for all the questionnaires that we launched in this research.

In this research, we applied several procedural methods to mitigate common method variance using. First, respondents' anonymity needs to be protected, which reduces evaluation apprehension (Podsakoff et al., 2003). In the instructions, we informed the participants that their responses were entirely anonymous, and there were no right or wrong answers. Second, another approach to control common method bias is to counterbalance question order (Hulland et al., 2018; Podsakoff et al., 2003). In this research, we randomised the items at two levels. On the one hand, we randomised the items inside the pages of constructs. On the other hand, we randomised the orders of the pages of constructs in the questionnaire. Third, it is also possible to reduce common method bias through careful construction of the items themselves (Podsakoff et al., 2003). Following the suggestions of Tourangeau et al. (2000), we (1) kept questions simple, specific, and concise; (2) avoided double-barrelled questions; (3) decomposed questions relating to more than one possibility into simpler, more focused questions; (4) avoided complicated syntax.

In our research, we also used a screening strategy to minimise the effects of survey satisficing. Survey satisficing refers to "*an alternative behaviour choice for respondents who lack survey engagement, but who for whatever reasons (such as a contract, incentives, embarrassment, habit or curiosity), still participate in a survey* (Callegaro et al., 2015, p. 102)" Detecting satisficing behaviour helps to reduce the common method bias. Previous empirical evidence (J. L. Huang et al., 2012) has shown that several psychometric properties improve, including item interrelatedness, facet dimensionality, and measurement structure, after removing the observations with satisficing behaviour. As a result, before conducting any data analyse, we needed to apply strategies to detect the participants who conducted satisficing behaviour during the survey. The detailed screening strategy is described in Appendix 1. The mentioned strategy was applied to all the quantitative phases (Phases 3–5).

With respect to the software environment, we used NVivo 12 for the qualitative data analysis. To conduct the quantitative data analysis, we used the R programming language in the RStudio environment, and we used *lavaan*[5] (Rosseel, 2014) to conduct the latent variable modelling.

4. Phase 1: conceptualisation of the constructs and item generation

Literature review and qualitative evidence are two ways to enhance the accuracy and comprehensiveness of construct definition (Netemeyer et al., 2003). In this study, we first reviewed the literature for an *a priori* specification of the dimensionality of the constructs of flow and further refined the conceptualisation by interviewing video game players. After conceptualising the dimensions of flow in the video game context, we generated the items for the corresponding constructs. There are two main sources to generate items: literature and the population of interest (Hinkin, 1995; Netemeyer et al., 2003). We first reviewed the previous studies of dispositional flow to explore how researchers operationalised the constructs, which helped us to develop the theoretical sensibility. Then, we applied a specific qualitative approach, grounded theory (Charmaz, 2014; Corbin and Strauss, 2015; Glaser and Strauss, 1967), to conduct the theoretical sampling and qualitative data analysis. Several researchers have suggested grounded theory as a useful tool in the initial phase of scale development (Bearss et al., 2016; Hinkin, 1995; Rowan and Wulff, 2007). Moreover, interviewing with members of the population can provide insights into item wording and response formats (Netemeyer et al., 2003).

[2] Prolific (https://www.prolific.co/) was launched in 2014, by a group of graduate students from Oxford and Sheffield Universities, as a software incubator company.

[3] URL: https://researcher-help.prolific.co/hc/en-gb/articles/360010165173-Can-I-screen-participants-within-my-survey-

[4] A two-step sampling process seperates the questionnaire into two parts: While the first questionnaire only charges the function to screen the target population, the second questionnaire contains the questions for the primary research. In the pre-screening questionnaire, apart from the demographic questions, we asked the participants the following question: Have you played videogames in the last six months?In the pre-screening questionnaire for both the qualitative and quantitative phase, we selected the participants with the following demographic characteristics to take part in the study: U.S. citizen, age ranging from 18 to 60, and they should have played video games in the last six months.

[5] URL: https://lavaan.ugent.be/

Table 2
Research process.

Research steps	Objectives	Details	Methodology
1. Conceptualisation of the constructs and item generation			
• Literature review	Finding the established conceptualisation of the flow measures	Reviewing literature related to flow experience.	Literature review
• In-depth interviews	Conceptualising the dimensions of flow in the video game context	13 in-depth online interviews with 11 informants	Grounded theory
• Item generation	Generating the initial item pool	Using both inductive and deductive reasoning to generate items	Grounded theory and literature review
Number of items: 64			
2. Expert judging			
• Expert judging	Assessing face validity	5 specialists from marketing and consumer research	Expert judging
Number of items: 64			
3. Pre-test			
• Survey 1		Total sample = 103; Valid sample = 96	
• Pilot data analysis	Pilot assessment of reliability	Assessing the reliability in the whole sample and in subgroups	Examination of Cronbach's alpha
	Pilot examination of measurement structure	Examining the measurement structure of the nine-dimension flow model	Exploratory factor analysis
	Pilot assessment of validity	Assessing the nomological validity	Correlation test
Number of items: 51			
4. Initial development and validation			
Number of items: 62			
• Survey 2		Total sample = 313; Valid sample = 289	
• Initial data analysis	Initial examination of measurement structure	Examining the measurement structure of the nine-dimension flow model	Exploratory factor analysis (All observations and subgroup analysis)
	Initial assessment of validity	Assessing the convergent validity, discriminant validity, and nomological validity	Confirmatory factor analysis
Number of items: 47			
5. Advanced development and validation			
Number of items: 71			
• Survey 3		Total sample = 637; Valid sample = 593	
• Advanced data analysis	Advanced examination of measurement structure	Examining the measurement structure of the nine-dimension flow model	Exploratory factor analysis and confirmatory factor analysis
	Advanced assessment of validity (I)	Assessing the convergent validity of the reduce item pool.	Confirmatory factor analysis
	Assessment of measurement invariance	Assessing the configural invariance, weak invariance, strong invariance, and strict invariance	Omnibus Test
	Advanced assessment of validity (II)	Assessing the convergent validity and discriminant validity of the final item pool.	Confirmatory factor analysis
	Examination of hierarchical structure	Examining the different specification of flow structure	Confirmatory factor analysis and structural equation modelling
	Advanced assessment of validity (III)	Assessing the nomological validity of different specification of flow structure	Structural equation modelling (All observations and subgroup analysis)
Number of items: 28			

Table 3
Summary of response rates.

Questionnaire	Introduction breakoff rate (IBR)	Questionnaire breakoff rate (QBR)	Total breakoff rate (TBR)	Completion rate (CR)	Screening completion rate (S_COMP)	Study specific eligibility rate (S_ELIG)
Pre-screening questionnaire ($n = 624$)	3.69%	0.00%	3.69%	96.30%	96.30%	99.50%
Pre-screening questionnaire ($n = 1047$)	3.53%	0.89%	4.39%	95.60%	95.60%	99.00%
Pre-screening questionnaire ($n = 279$)	8.96%	0.39%	9.32%	90.70%	90.70%	99.20%
Integrated pre-screening questionnaire ($n = 1671$)	3.59%	0.56%	4.13%	95.90%	95.90%	99.20%
Integrated pre-screening questionnaire ($n = 1950$)	4.36%	0.54%	4.87%	95.10%	95.10%	99.20%
Pre-test questionnaire ($n = 103$)	0.00%	1.94%	1.94%	98.00%	98.10%	99.00%
Initial scale validation questionnaire ($n = 313$)	1.92%	2.28%	4.15%	95.80%	96.50%	99.30%
Advanced scale validation questionnaire ($n = 637$)	0.94%	1.43%	2.35%	97.60%	97.80%	99.40%

We conducted 13 semi-structured Interviews with 11 video game players, who were recruited on Prolific. Appendix 2 demonstrates the demographic and gaming profiles of the informants.

Reviewing the relevant literature of flow experience helped us to cultivate theoretical sensitivity (Charmaz, 2014; Corbin and Strauss, 2015) when conducting fieldwork and analysing qualitative data using grounded theory (Charmaz, 2014). After we conducted initial and focus coding, nine categories emerged through inductive reasoning, which was in line with the nine-dimensions conceptualisation of flow experience (Jackson, 1996). We summarise the definitions of each construct in Table 4.

Researchers advocate that scale developers should consult several sources when generating the item pool (Netemeyer et al., 2003). Previous studies and members of the target population are the two main sources to generate items (Netemeyer et al., 2003). Therefore, in this research, we generated the initial item pool based on the items of previous scales (Abbasi et al., 2017, 2019; Brockmyer et al., 2009; Fu et al., 2009; Jackson et al., 2011; Jennett et al., 2008; O'Brien et al., 2018; Wiebe et al., 2014) and the qualitative results from phase 1. Although there are no hard-and-fast rules for the size of an initial item pool (Netemeyer et al., 2003), we followed the advice of DeVellis (2016) to generate a pool with at least twice the size of the final scale. Eventually, we generated an initial item pool with 64 items, which is shown in Appendix 3. The items in the initial pool as well as the questionnaires in the rest of the scale development phases were measured using a 7-point ordinal scale with the following anchor labels: 1-Never, 2-Almost never, 3-Rarely, 4-Sometimes, 5-Frequently, 6-Almost always, 7-Always.

Regarding the conceptualisation of flow experience at the level of second-order factors, factor analyses warranted the components represented a unidimensional structure (Beard and Hoy, 2010; Engeser and Rheinberg, 2008; Jackson et al., 2011, 2008; Jackson and Eklund, 2002; Jackson and Marsh, 1996) However, this conclusion may be premature (Engeser and Schiepe-Tiska, 2012). The components of flow experience can be highly correlated but at the same time dissociated (Engeser and Schiepe-Tiska, 2012). Researchers proposed that three of the nine dimensions, Clear goals, Unambiguous feedback, and Challenge-skill balance, were the antecedents of flow experience (Jackson, 2012; Keller and Landhäußer, 2012; Nakamura and Csikszentmihalyi, 2012). According to this conceptualisation of flow experience, the three antecedents may affect flow experience separately, which we named it as independent antecedent model. Moreover, the three antecedents can also be conceptualised as a second-order factor that leads to flow experience. Keller and Landhäußer (2012) considered that it is not meaningful to consider the three antecedents as distinct factors, because it is simply not possible to perceive a fit of skills and task demands when engaging in an activity without clear task instructions or without diagnostic information regarding one's progress or success in the activity. Following the proposition of the previous researchers (Landhäußer and Keller, 2012), we conceptualised that the three antecedents of flow formed a reflective second-order factor, which we named it as Perceived fit of Goal-Feedback-Balance (PFGFB), and we called this conceptualisation approach as hierarchical antecedent model. Moreover, we also took the methodological considerations into the account when conceptualising the PFGFB. We conceptualised the PFGFB as reflective instead of formative because the reflective measurement approach aims at maximising the overlap between the indicators (Hair Jr, Hult, Ringle, and Sarstedt, 2016). Moreover, formative indicators have no individual measurement error terms and assume error-free in a conventional sense (Hair Jr et al., 2016), which is unlikely to happen in survey research. We concluded the three conceptualisations of flow experience at the edges of Table 4.

5. Phase 2: expert judging

As part of the scale development process, expert judging helps to control the content and face validity of an emerging scale (DeVellis, 2016; Netemeyer et al., 2003). We developed a questionnaire for the purpose of expert judging. In this questionnaire, all the items, response formats, number of scale points, and instructions were listed for judging, as suggested by previous researchers (Netemeyer et al., 2003). On the qualitative side, we created an open-ended question for each construct of the scale as well as for the whole scale, so that experts could give their

Table 4
Definitions of the flow constructs and different approaches of conceptualisation.

		Constructs	Definitions		
Hierarchical antecedent model	Perceived fit of Goal-Feedback-Balance (PFGFB)	Clear goals (CG)	Players know clearly which operations they are supposed to do in the next phases. The clarity of purpose keeps players fully connected to the in-game tasks and responsive to appropriate cues.	Antecedent 1	Independent antecedents model
		Unambiguous feed back (UF)	Players receive immediate and unambiguous feedback about how well they are processing towards the in-game goals.	Antecedent 2	
		Challenge-skill balance (CSB)	In the video game contexts, challenges are the in-game tasks to be completed, and skills are the subjective belief or confidence that players have to overcome the challenges. When players experience the balance between challenge and skill, they may enter the flow experience.	Antecedent 3	
	Flow experience	Concentration (CON)	When players are in the flow state, they totally focus on the specific gaming tasks, and they concentrate on the task at hand in the game world.	Flow experience	
		Action-awareness merging (AAM)	When players are in the flow state, their action and awareness would be merged. Through total absorption in the game world, video game players are associated with the game world and feel that they are a part of the game world.		
		Sense of control (SC)	When players are in the flow state, they have a sense of natural control in the game world.		
		Loss of self-consciousness (LSS)	When video game players have flow state, their ego disappears in the real world. The impact of stimulus from the real world on video game players' self-consciousness is reduced.		
		Transformation of time (TT)	When players are in the flow state, players' perception of time is affected. Players cannot perceive the time or lose the awareness of time.		
		Autotelic experience (AE)	Autotelic experience describes the intrinsically rewarding experience that flow brings to the individual. Autotelic experience is the source of enjoyment that flow experience brings to the individual.		
		Unidimensional model			

commentary at both item level and scale level. On the quantitative side, we applied the method of Zaichkowsky (1985), which is the most common way to conduct quantitative expert judging (Hardesty and Bearden, 2004). We invited seven professors from two Spanish universities to form the expert panel, who were specialised in consumer behaviour. Among the experts, five completed the questionnaire, and two engaged in the questionnaire but not finished it.

We used three rules for the evaluation of expert judging results: sum score rule, complete rule and not representative rule (Hardesty and Bearden, 2004). Table 5 demonstrates the descriptive statistics of the quantitative results of expert judging. According to these results, the face validity of the overall initial item pool is acceptable. The complete quantitative results of expert judging can be found in Appendix 3.

Combining the quantitative and qualitative feedbacks from the expert, we did not trim any items from the initial pool.

6. Phase 3: pre-test

103 participants recruited from Prolific participated in the pre-test. After applying the screening strategy that we mentioned in Appendix 1, 96 samples remained in our dataset. Tables 6 and 7 demonstrate the demographic and gaming profile of the participants, respectively. We applied exploratory factor analysis (EFA) to examine the measurement structure and assess the validity of the scale. EFA can be used for two primary purposes in scale development (Netemeyer et al., 2003). On the one hand, it reduces the number of items in a scale so that the remaining items maximise the explained variance and reliability in the scale. On the other hand, it helps to identify the underlying dimensions in a scale.

Internal consistency reliability should be assessed in the pre-test when developing a new scale (Netemeyer et al., 2003). The estimates of alpha coefficient across all the observations and each subgroup are summarised in Appendix 4. According to the results, we observed that all the coefficients were acceptable except for the measure of Challenge-skill balance (0.66) in the female group.

Then, we applied EFA using a recommended procedure in the literature (Hair et al., 2013, p. 104). We used principal axis factoring (PAF) as the factor extraction method and the PROMAX method when there was a need to obtain a rotated solution.. PROMAX, as an oblique rotation method, was used because this rotation method looks for the degree to which multiple scales/dimensions correlate, which reveals more meaningful theoretical factors in scale development (Netemeyer et al., 2003). We followed *a priori* criterion to set the number of factors as nine, as this number of flow dimensions were theoretically (Jackson, 1996, 2012; Nakamura and Csikszentmihalyi, 2012) and empirically (Jackson et al., 2011, 2008; Jackson and Eklund, 2002; Jackson and Marsh, 1996) supported in previous research. As suggested in the literature (Hair et al., 2013), we conducted unrotated solution before PROMAX solutions to trim the items in the pool. Appendix 4 demonstrates the full results of EFA. After dropping the unqualified items (communality value < 0.5; factor loading < 0.5), 51 items remained in the pool. Then, we re-ran the PROMAX solution using the remaining items. We found that no communality value was lower than 0.5, all factor loadings exceeded 0.5 without a cross-loading problem, and all eigenvalues were greater than 1.

7. Phase 4: initial development and validation

At the end of the pre-test, only two items remained in the factor Challenge-skill balance, indicating violation of the three-indicator rule (Hair et al., 2013). In addition, the spearman coefficient between Challenge-skill balance and Loss of self-consciousness was low (0.057), suggesting a potential problem against the previous measurement model (Jackson et al., 2011). Therefore, in this study, we modified the item pool and assessed the modified scale using a larger sample. In the modified item pool, there were 62 items. Additionally, we needed a larger sample to initially validate the generalisability of the scale across gender and age groups.

We recruited 313 participants recruited from Prolific participated in the initial scale validation. After applying the screening strategy that we mentioned in Appendix 1, 289 samples were left in our dataset. Tables 6 and 7 show the demographic and gaming profiles of the participants, respectively. Later, we conducted a series of assessments suggested in the literature, including an initial examination of measurement structure and initial assessment of validity (Netemeyer et al., 2003). We repeated the procedure of conducting EFA as we did in the pre-test using the new dataset.

After conducting the EFA, 51 items remained in the pool. Later, we repeated the PROMAX solution using the remaining items, and no communality or loading problem was found, and the eigenvalues were greater than 1. Appendix 5 shows the full results of the factor matrix. Moreover, when attempting to define the underlying structure among items, validation of any factor analysis result is essential (Hair et al., 2013). In this study, we used the split sample analysis. We classified the observations into four groups according to their gender and age (under 30 and over 30). The results of multiple group EFA can be found in Appendix 5. Later, we dropped the items that had either communality or loading problems (communality value < 0.5; factor loading < 0.5) in any two of the four groups, and 47 items were left in the pool.

We also conducted an initial assessment of convergent and discriminant validity. To assess these validities, we specified a confirmatory factor analysis (CFA) model using the remaining 47 items. To estimate the CFA model with ordered categorical data with more than six points, we used maximum likelihood with Satorra-Bentler scaling, as suggested by Finney and DiStefano (2013). We used the pre-defined thresholds of model fit indices to check the fit of the models: the model fit is acceptable when χ^2 statistic is significant, CFI > 0.90, TLI > 0.9, RNI > 0.90, SRMR < 0.1, RMSEA < 0.08) (Hair et al., 2013). The chi-squared statistic ($\chi^2 = 2064.028$, df $= 1188$, $p < 0.001$) and the model fit indices (CFI $= 0.921$, TLI $= 0.915$, RNI $= 0.921$, SRMR $= 0.057$, RMSEA $= 0.051$) demonstrate that the CFA model is acceptable. Moreover, all the standardised loading estimates and the values of average variance extracted (AVE) were higher than 0.5, which suggests adequate convergence (Hair et al., 2013). As for assessing the discriminant validity, a robust approach to test the discriminant validity is to compare

Table 5
Descriptive statistics of quantitative results of expert judging.

	Mean	SD	Minimum	1st quartile	Median	3rd quartile	Maximum	Not representative	Somewhat representative	Clearly representative
Expert 1	2.58	0.56	1	2	3	3	3	2	23	39
Expert 2	2.62	0.6	1	2	3	3	3	4	16	44
Expert 3	2.53	0.69	1	2	3	3	3	7	16	41
Expert 4	2.56	0.56	1	2	3	3	3	2	24	38
Expert 5	2.19	0.83	1	1	2	3	3	17	18	29
Sumscore rule	12.48	1.73	8	12	12	14	15			
Complete rule	2.98	1.33	1	2	3	4	5			
Not representative rule	0.5	0.69	0	0	0	1	3			

Table 6
Demographic profile of the participants.

Group		Pre-test (n = 96)		Initial scale validation (n = 289)		Advanced scale validation (n = 593)	
		Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Gender	Male	51	53.12%	142	49.13%	296	49.92%
	Female	45	46.88%	147	50.87%	297	50.08%
Age	≤30 years old	47	48.96%	138	47.75%	289	48.74%
	>30 years old	49	51.04%	151	52.25%	304	51.26%
Student	Yes (Full time)	22	22.92%	61	21.11%	113	19.06%
	Yes (Part time)	6	6.25%	11	3.81%	30	5.06%
	No	68	70.83%	217	75.09%	450	75.89%
Education	Middle school and below	1	1.04%	2	0.69%	3	0.51%
	Vocational school	0	0.00%	4	1.38%	15	2.53%
	High school	26	27.08%	54	18.69%	136	22.93%
	Community college	13	13.54%	52	17.99%	101	17.03%
	Undergraduate degree	46	47.92%	146	50.52%	252	42.50%
	Master's degree	6	6.25%	28	9.69%	70	11.80%
	PhD	4	4.17%	3	1.04%	16	2.70%
Employment	Yes (Full time)	33	34.38%	119	41.18%	255	43.00%
	Yes (Part time)	17	17.71%	51	17.65%	116	19.56%
	No	46	47.92%	119	41.18%	222	37.44%
Income	I don't have an income	21	21.88%	55	19.03%	115	19.39%
	Less than 1000 USD	21	21.88%	53	18.34%	98	16.53%
	1001 USD −2000 USD	10	10.42%	42	14.53%	97	16.36%
	2001 USD −3000 USD	13	13.54%	30	10.38%	86	14.50%
	3001 USD −5000 USD	6	6.25%	53	18.34%	72	12.14%
	5001 USD −7000 USD	5	5.21%	22	7.61%	36	6.07%
	More than 7001 USD	17	17.71%	25	8.65%	74	12.48%
	Don't Know/No Answer	3	3.12%	9	3.11%	15	2.53%

Table 7
Gaming profile of the participants.

Group		Pre-test (n = 96)		Initial scale validation (n = 289)		Advanced scale validation (n = 593)	
		Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Gaming time per week	Less than 4 hour a week	21	21.88%	66	22.84%	107	18.04%
	5–7 h a week	17	17.71%	65	22.49%	119	20.07%
	8–12 h a week	17	17.71%	66	22.84%	127	21.42%
	13–20 h a week	12	12.50%	48	16.61%	111	18.72%
	More than 20 h a week	29	30.21%	44	15.22%	129	21.75%
Gaming time per session	Less than 1 h at a time	19	19.79%	78	26.99%	125	21.08%
	1–2 h at a time	33	34.38%	102	35.29%	214	36.09%
	2–3 h at a time	24	25.00%	68	23.53%	149	25.13%
	3–5 h at a time	12	12.50%	31	10.73%	78	13.15%
	More than 5 h at a time	8	8.33%	10	3.46%	27	4.55%
Platform	Windows	80	83.33%	225	77.85%	469	79.09%
	MAC	7	7.29%	17	5.88%	45	7.59%
	PS4/PS4 Pro	40	41.67%	112	38.75%	256	43.17%
	Xbox One/Xbox One X	21	21.88%	78	26.99%	143	24.11%
	WII U	7	7.29%	17	5.88%	37	6.24%
	Nintendo Switch	39	40.62%	141	48.79%	289	48.74%
	PS Vita	0	0.00%	5	1.73%	20	3.37%
	3DS	15	15.62%	46	15.92%	85	14.33%
	Android	43	44.79%	155	53.63%	288	48.57%
	IOS	27	28.12%	98	33.91%	207	34.91%

the values of AVE for any two factors with the square of the correlation estimate between these two factors (Fornell and Larcker, 1981). In our case, all the values of AVE were greater than the squared correlation estimates of the corresponding factors, which suggests a good discriminant validity. All the mentioned results are summarised in Appendix 5.

In conclusion, the evidence shows that the 47-item scale had both convergent and discriminant validity.

8. Phase 5: advanced development and validation

After conducting the first four studies, we had already had a prototype scale. Later, we re-worded the problematic items that we detected in the previous study. Then, we sent the modified item pool to a native English speaker for proofreading. In this modified and proofread item pool, there were 71 items.

In this study, we finalise the scale and further establish its psychometric properties. To do so, we conducted a series of advanced examinations to assess the measurement structure at the item level, the measurement invariance, the convergent validity, the discriminant validity, the hierarchical measurement structure, and the nomological validity. 637 participants recruited from Prolific participated in the initial scale validation. After applying the screening strategy that we mentioned in Appendix 1, 593 samples were left in our dataset. Table 6 and Table 7 give the demographic and gaming profiles of the participants, respectively.

We first applied the EFA to examine the measurement structure of the flow experience. After trimming the unqualified items in the unrotated and PROMAX solutions (communality value < 0.5; factor loading < 0.5), 62 items remained in the item pool.

Then, we conducted the CFA using the 62 items to further assess the validity of the scale. We specified a nine-factor first-order model using maximum likelihood with S-B scaling. The chi-squared statistic (χ^2 =

4693.595, $p < 0.001$) and model fit index (CFI = 0.902, TLI = 0.897, RNI = 0.902, SRMR = 0.055, RMSEA = 0.052) demonstrate that the CFA model is acceptable. However, as the factor loading should be higher than 0.5, ideally higher than 0.7 (Hair et al., 2013), we removed six items from the pool. Although the ultimate goal of confirmatory factor analysis (CFA) is to obtain an answer as to whether a given measurement model is valid, the process of CFA provides additional diagnostic information that may suggest modifications for either addressing unsolved problems or improving the model's test of measurement theory (Hair et al., 2013). There are a series of *post hoc* tools available to improve the model performance, such as standardised residuals, modification indices, and specification searches (Hair et al., 2013). In the scale development process, these *post hoc* tools are useful for assessing items that have correlated measurement errors or load strongly on a factor other than their intended factor (Netemeyer et al., 2003). Standardised residuals greater than |4.0| suggest a potentially unacceptable degree of error that may call for the deletion of an offending item (Hair et al., 2013). After examining the standardised residuals, we found that 23 items were affected. Additionally, according to the results of the modification index power test (Saris et al., 2009), 25 items were affected. When deciding to drop the unqualified items, we took all the mentioned results into consideration. For instance, the standardised residual between SC7P and SC3P is −4.057, which is greater than |4.0|. Meanwhile, the MI power test shows that the relationship between SC7P and SC5 is misspecified. Additionally, the factor loading of SC7P (0.739) is much lower in comparison to other items in the same factor: SC3P (0.745), SC4 (0.772), SC5 (0.775). Taking all these pieces of information together, we decided to drop the SC7P. We generalised this logic to all the pool to trim items, and 43 of them were left. Later, we used the 43 items to specified again the nine-factor first-order model. According to the model fit indices (CFI = 0.969, TLI = 0.966, RNI = 0.969, SRMR = 0.037, RMSEA = 0.035), the performance of 43-items model is much better than that of 63 items model.

In this research, we paid great attention to assess the measurement invariance. Measurement invariance is a logical prerequisite when studying differences across groups (Jiang et al., 2017). Measurement invariance refers to the consistency of a measurement instrument across groups (Nimon and Reio, 2011), which concerns whether scores from the operationalisation of a construct have the same meaning under different conditions (Meade and Lautenschlager, 2004). If evidence supporting a measure's invariance is lacking, conclusions based on that scale are at best ambiguous and at worst erroneous (Steenkamp and Baumgartner, 1998). Thus, unless measurement invariance is established, conducting cross-group comparisons of a mean difference or other structural parameters is meaningless (Schmitt and Kuljanin, 2008). Moreover, if there is empirical evidence for measurement invariance, the generalisability of the scale is enhanced (Marsh, 1994; Netemeyer et al., 2003; Steenkamp and Baumgartner, 1998).

We assessed the measurement invariance using the Omnibus test approach (Fischer and Karl, 2019; Kline, 2015; Steenkamp and Baumgartner, 1998). In this study, we concerned about the measurement invariance in two basic demographic variables of video game players: gender and age. In the video game context, gender and age are two critical variables that serve to separate the market (ESA, 2019; Newzoo, 2019b). The results are shown in Table 8. According to the results, the 43-items model has weak invariance across both the gender groups (p = 0.617) and the age groups (p = 0.674). However, results show that this model does not prove strong invariance in gender groups (0.674 < 0.001) and in the age groups (0.674 < 0.001). Therefore, we repeated the approach of Saris et al. (2009) in the models where the factor loadings and intercepts are constrained to be equal across groups to further refine the scale. The results showed that the relationships of 17 items were misspecified in the gender groups, and the same number of items were also misspecified in the age groups. Meanwhile, we refined the scale using the modification index. Items with a higher modification index were trimmed until there were three items in each factor, which complies with the three-item rule that avoids identification problems (Hair et al., 2013). Besides, controlling the length of the scale enhances the brevity of the questionnaire and limits fatigue for the participants (Netemeyer et al., 2003). Finally, 28 items were left on the scale, which are shown in Table 9.

We then specified a CFA model using the mentioned 28 items. According to the model fit index (CFI = 0.988, TLI = 0.985, RNI = 0.988, SRMR = 0.027, RMSEA = 0.026), the 28-item model reached the best performance amongst the models that we had specified. Additionally, all the factor loadings were greater than 0.7, which reach an ideal level (Hair et al., 2013).

All the mentioned results of factor loadings, standardised residuals (results greater than |4.0|), modification index, and expected parameter change are summarised in Appendix 6.

Later, we re-run the tests of measurement invariance using the remaining 28 items. The results are shown in Table 8. We found that although the 28-items model still failed to reach strong invariance across the gender groups for a significance level of 0.05, the measurement invariance was improved significantly considering that the p-value was approaching the acceptable level (p = 0.04). Additionally, we found that the 28 items model reached strict invariance across the age group.

Regarding the convergent and discriminant validity of the 28-items model, we repeated the assessment procedure as we did in the previous study. According to the results shown in Table 10, all the AVEs are greater than 0.5, with all the estimates of composite reliability and Cronbach's alpha higher than 0.7. These results suggest that the 28-item model has solid convergent validity. Moreover, all the values of AVE are greater than the squared correlation estimates of the corresponding

Table 8
Results of measurement invariance tests.

Model	Group	Constraints	DF	AIC	BIC	Chi-square statistic	Chi-square difference	DF difference	p value
43 items model	Gender	Configural	1648	60,778	62,226	2570.9			
		Weak	1682	60,749	62,047	2609.9	30.969	34	0.617
		Strong	1716	60,731	61,880	2659.1	67.128	34	<0.001 ***
		Strict	1759	60,731	61,692	2745.8	38.508	43	0.666
	Age	Configural	1648	60,856	62,303	2483.4			
		Weak	1682	60,831	62,129	2526.2	29.803	34	0.674
		Strong	1716	60,818	61,967	2581.1	78.619	34	<0.001 ***
		Strict	1759	60,810	61,771	2659.7	36.836	43	0.735
28 items model	Gender	Configural	628	41,469	42,522	874.94			
		Weak	647	41,448	42,417	891.27	13.673	19	0.802
		Strong	666	41,436	42,322	917.71	31.072	19	0.040 *
		Strict	694	41,438	42,201	975.16	32.026	28	0.273
	Age	Configural	628	41,577	42,630	846.27			
		Weak	647	41,560	42,529	867.3	15.35	19	0.700
		Strong	666	41,546	42,432	890.91	27.702	19	0.089
		Strict	694	41,527	42,290	928.09	19.755	28	0.873

Table 9
Items in the video game dispositional flow scale.

Number	Second order factor	First order factor	Codes during development	Item description
				When playing video games,…
Item 1	Perceived fit of Goal-Feedback-Balance (PFGFB)	Clear goals (CG)	CG2NMP	…I know how to proceed during the gaming session.
Item 2			CG5P	…I clearly understand the goals.
Item 3			CG6M	…I know which operations to do in the game world.
Item 4		Unambiguous feed back (UF)	UF2	…I perceive immediate feedback from the game mechanics.
Item 5			UF3	…I receive immediate feedback on my gaming progress.
Item 6			UF5N	…I perceive immediate feedback on my actions in the game world.
Item 7		Challenge-skill balance (CSB)	CSB1MM	…I feel that my gaming skills are proportional to the in-game challenges.
Item 8			CSB2NP	…I feel that my gaming skills are at a similar level to the in-game challenges.
Item 9			CSB5NP	…I feel that my gaming skills are balanced with the in-game challenges.
Item 10			CSB7NP	…I feel that my gaming skills are up to the in-game challenges.
Item 11	Flow experience (FE)	Concentration (CON)	CON2	…I focus on the game.
Item 12			CON3	…I remain concentrated.
Item 13			CON4N	…I concentrate on the task at hand in the game world.
Item 14		Action-awareness merging (AAM)	AAM5N	…I am associated with the game world.
Item 15			AAM7	…I feel that I am the character in the game.
Item 16			AAM10NP	…I feel that I am part of the game world.
Item 17		Sense of control (SC)	SC3P	…I can perceive the natural control of the game.
Item 18			SC4	…I feel a sense of control in the game.
Item 19			SC5	…I fully control my operations in the game world.
Item 20		Loss of self-consciousness (LSS)	LSS2R	…I forget about things in the real world.
Item 21			LSS3RP	…I tune out everything else around me.
Item 22			LSS5R	…I forget about what is occurring in the real world.
Item 23		Transformation of time (TT)	TT4N	…I cannot perceive the flow of time.
Item 24			TT6P	…I lose my awareness of time.
Item 25			TT7M	…I forget about time.
Item 26		Autotelic experience (AE)	AE1P	…I enjoy each gaming session to the full.
Item 27			AE7M	…the gaming session makes me feel great.
Item 28			AE8	…I feel rewarded.

factors. This evidence demonstrates that the 28-item model has good discriminant validity.

After confirming the first-order factor structure, we proceeded to explore the hierarchical structure of the scale, because theoretically, the nine dimensions of flow do not exist separately. Therefore, we specified three models using different conceptualisations of flow experience, including the unidimensional model, the independent antecedent model, and the hierarchical antecedent model. We applied CFA to estimate the unidimensional model, as there was no path estimate according to the conceptualisation. Meanwhile, we used structural equation modelling (SEM) to estimate the independent antecedent model and the hierarchical antecedent model, as there are paths between the antecedents of flow and flow per se. The statistical significance of the path coefficients provides key information of nomological validity (Mackenzie et al., 2011). If these paths are significant, it means that other factors (in this case, the antecedents of flow) and related to the focal factor (in this case, the flow experience) as specified in the nomological network, which therefore enhances the confidence of the nomological validity (Mackenzie et al., 2011).

The estimated results of the unidimensional model are shown in Table 11 (See also Fig. 1 for the visualisation). The results show that all the factor loadings of the first-order factors are significant ($p < 0.001$). Moreover, all the loadings are greater than 0.5, except for LSS (0.367) and TT (0.291). The model fit indices (CFI = 0.954, TLI = 0.949, RNI = 0.954, SRMR = 0.086, RMSEA = 0.048) demonstrate that the CFA model is acceptable. The Akaike's information criterion (AIC) and Bayesian information criterion (BIC) are 41,759.168 and 42,044.205, respectively.

The estimated results of the independent antecedent model are shown in Table 11 (See also Fig. 2 for the visualisation). According to the results, all the factor loadings of the first-order factors are significant ($p < 0.001$). Besides, all the loadings are greater than 0.5, except for LSS (0.411) and TT (0.331). The path estimates show that all the coefficients are positive and significant: CG ($\beta = 0.300$, $p < 0.001$), UF ($\beta = 0.253$, $p < 0.001$), and CSB ($\beta = 0.376$, $p < 0.001$). The r-squared of flow experience is 0.553. The model fit indices (CFI = 0.956, TLI = 0.95, RNI = 0.956, SRMR = 0.083, RMSEA = 0.048) demonstrate that the SEM model is acceptable. The AIC and BIC are 41,746.903 and 42,045.096,

Table 10
Correlation matrix with AVE, composite reliability, and Cronbach's alpha.

	CG	UF	CSB	CON	AAM	SC	LSS	TT	AE	Composite reliability	Cronbach's alpha
CG	**0.65**	0.16	0.28	0.25	0.07	0.37	0	0	0.16	0.94	0.94
UF	0.4	**0.71**	0.16	0.16	0.05	0.2	0.05	0.02	0.13	0.88	0.88
CSB	0.53	0.41	**0.68**	0.19	0.09	0.38	0.03	0.03	0.19	0.94	0.94
CON	0.5	0.39	0.44	**0.6**	0.14	0.28	0.12	0.05	0.28	0.9	0.9
AAM	0.26	0.23	0.31	0.38	**0.58**	0.14	0.22	0.15	0.26	0.93	0.92
SC	0.61	0.45	0.62	0.53	0.38	**0.57**	0.03	0.01	0.26	0.84	0.84
LSS	0.05	0.23	0.16	0.35	0.47	0.19	**0.72**	0.36	0.06	0.95	0.95
TT	0	0.14	0.16	0.22	0.38	0.11	0.6	**0.67**	0.06	0.95	0.95
AE	0.4	0.36	0.43	0.53	0.51	0.51	0.24	0.25	**0.58**	0.92	0.92

Table 11
Summary of the results of CFA and SEM in advanced scale validation.

Model	Second order factor	First order factor	Factor loading estimates All observations	Male group	Female group	Under 30 group	Over 30 group	Path estimates All observations	Male group	Female group	Under 30 group	Over 30 group
Unidimensional model	Flow experience (FE)	Clear goals (CD)	0.663 ***	0.765 ***	0.569 ***	0.681 ***	0.655 ***	NA				
		Unambiguous feedback (UF)	0.555 ***	0.596 ***	0.534 ***	0.574 ***	0.542 ***					
		Challenge-skill balance (CSB)	0.690 ***	0.686 ***	0.670 ***	0.735 ***	0.641 ***					
		Concentration (CON)	0.707 ***	0.724 ***	0.658 ***	0.661 ***	0.747 ***					
		Action-awareness merging (AAM)	0.544 ***	0.374 ***	0.730 ***	0.461 ***	0.600 ***					
		Sense of control (SC)	0.790 ***	0.950 **	0.653 ***	0.846 ***	0.734 ***					
		Loss of self-consciousness (LSS)	0.367 ***	0.230 **	0.510 ***	0.318 ***	0.388 ***					
		Transformation of time (TT)	0.291 ***	0.233 **	0.395 ***	0.232 **	0.331 ***					
		Autotelic experience (AE)	0.686 ***	0.627 ***	0.755 ***	0.612 ***	0.734 ***					
Independent antecedents model	NA	Clear goals (CD)	NA					0.300 ***	0.431 ***	0.192 *	0.280 **	0.350 ***
		Unambiguous feedback (UF)						0.253 ***	0.252 ***	0.285 ***	0.228 **	0.274 ***
		Challenge-skill balance (CSB)						0.376 ***	0.318 ***	0.407 ***	0.437 ***	0.307 **
	Flow experience (FE)	Concentration (CON)	0.720 ***	0.735 ***	0.651 ***	0.675 ***	0.760 ***	NA				
		Action-awareness merging (AAM)	0.582 ***	0.390 ***	0.766 ***	0.494 ***	0.633 ***					
		Sense of control (SC)	0.763 ***	0.946 *	0.621 ***	0.835 ***	0.705 ***					
		Loss of self-consciousness (LSS)	0.411 ***	0.249 **	0.559 ***	0.355 ***	0.428 ***					
		Transformation of time (TT)	0.331 ***	0.249 **	0.441 ***	0.263 **	0.368 ***					
		Autotelic experience (AE)	0.711 ***	0.640 ***	0.757 ***	0.643 ***	0.744 ***					
	R squared							0.553	0.677	0.49	0.6	0.534
Hierarchical antecedent model	Perceived fit of Goal-Feedback-Balance (PFGFB)	Clear goals (CD)	0.702 ***	0.791 ***	0.633 ***	0.711 ***	0.700 ***	0.873 ***	0.942 **	0.830 ***	0.894 ***	0.873 ***
		Unambiguous feedback (UF)	0.576 ***	0.606 ***	0.559 ***	0.589 ***	0.565 ***					
		Challenge-skill balance (CSB)	0.736 ***	0.705 ***	0.752 ***	0.772 ***	0.691 ***					
	Flow experience (FE)	Concentration (CON)	0.721 ***	0.734 **	0.656 ***	0.674 ***	0.758 ***	NA				
		Action-awareness merging (AAM)	0.581 ***	0.390 **	0.763 ***	0.493 ***	0.633 ***					
		Sense of control (SC)	0.766 ***	0.948 •	0.628 ***	0.838 ***	0.710 ***					
		Loss of self-consciousness (LSS)	0.407 ***	0.246 *	0.550 ***	0.353 **	0.422 ***					
		Transformation of time (TT)	0.327 ***	0.247 *	0.432 ***	0.259 **	0.365 ***					
		Autotelic experience (AE)	0.711 ***	0.639 **	0.758 ***	0.641 ***	0.746 ***					
	R squared							0.763	0.887	0.688	0.799	0.762

respectively.

The estimated results of the hierarchical antecedent model are shown in Table 11 (See also Fig. 3 for the visualisation). The results demonstrate that all the factor loadings of the first-order factors are significant ($p < 0.001$). In addition, all the factor loadings of the first-order factors that form PFGFB are greater than 0.5. However, there are two factor loadings of the first-order factors that form flow experience are lower than 0.5, which are LSS (0.407) and TT (0.327). The path estimate shows that the coefficient between the PFGFB and flow experience is positive and significant ($\beta = 0.873$, $p < 0.001$). The r-squared of flow experience is 0.763. The model fit indices (CFI = 0.956, TLI = 0.951, RNI = 0.956, SRMR = 0.083, RMSEA = 0.047) demonstrate that the SEM model is acceptable. The AIC and BIC are 41,744.727 and 42,034.150, respectively.

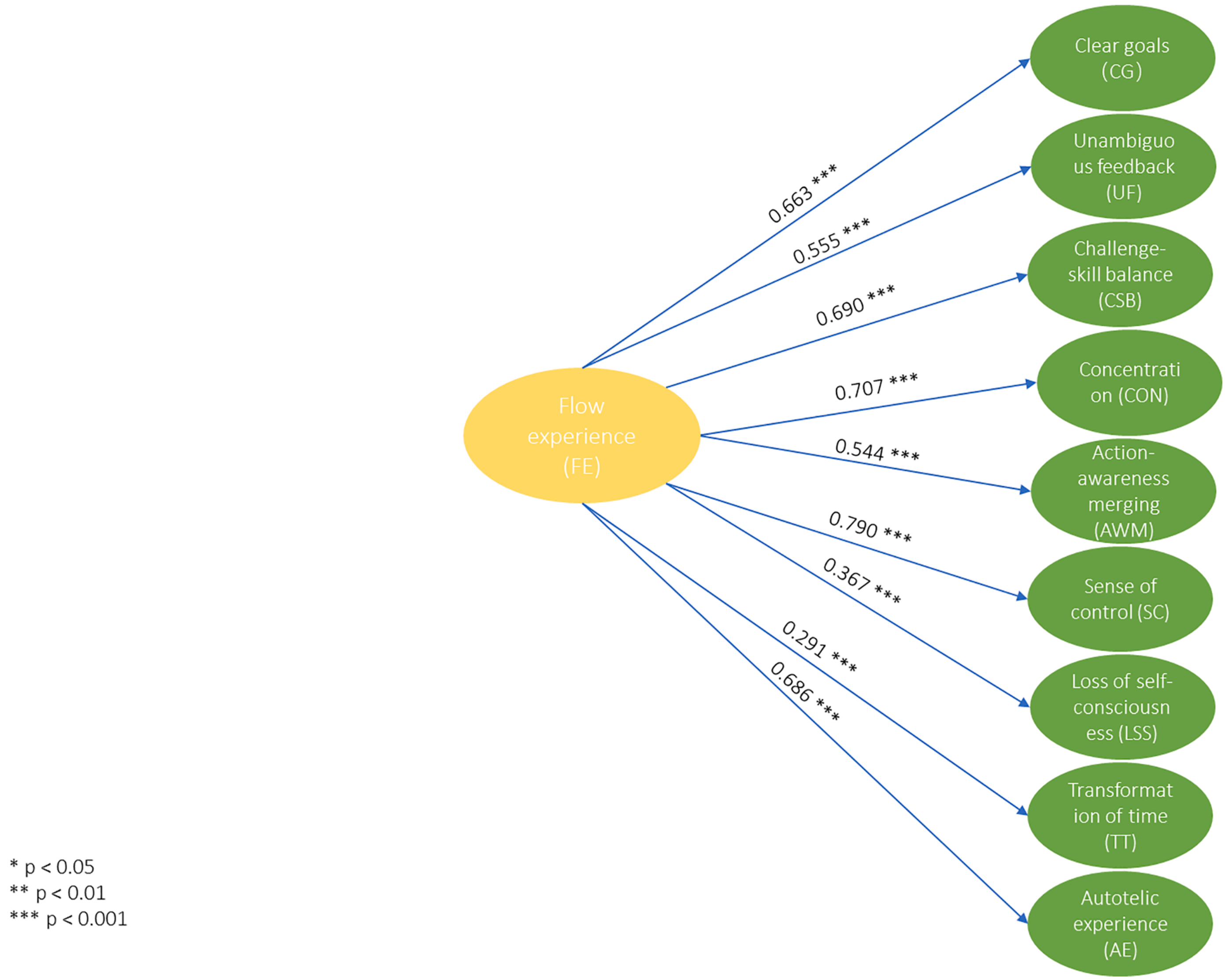

Fig. 1. Estimated results of the unidimensional model.

Finally, after confirming the measurement invariance and measurement structure, we conducted the subgroup analysis across the four groups: male group ($n = 296$), female group ($n = 297$), under 30 group ($n = 289$), over 30 group ($n = 304$). The main purpose of conducting subgroup analysis is to assess whether values of model parameters of substantive interest vary appreciably across different samples (Kline, 2015). We estimated the three mentioned models using observations from each of these groups. The results are shown in Table 11. According to the results, all the estimates of factor loadings are significant at a p-value of 0.05, except for the estimates between sense of control and flow experience in the hierarchical antecedent model in the male group, which is significant at a p-value of 0.1 (0.948, $p = 0.088$). Meanwhile, all the path estimates are significant across all the subgroups at a p-value of 0.05. Finally, all the model fit indices that we have mentioned in this section are summarised in Table 12.

9. Discussion

In this research, we conceptualised the flow experience in the video game context and developed a new scale named *Video Game Dispositional Flow Scale* (VGDFS) through five phases. VGDFS is the first scale to measure the psychometric properties of dispositional flow experience in the video game context, and its target population are adult players between 18 and 60. The birth of the VGDFS is a response to the conceptual flaw and mixing empirical findings in the video game flow literature. On the one hand, although the EGameFlow scale (Fu et al., 2009), the Game engagement questionnaire (Brockmyer et al., 2009), the User engagement scale (Wiebe et al., 2014), and the Consumer video engagement scale (Abbasi et al., 2017, 2019), were developed and verified in the video game context, the content domains overlapped, and the problem of construct proliferation was noted. On the other hand, while the DFS-2 (Jackson et al., 2011; Jackson and Eklund, 2002) seems suitable to assess the flow experience amongst adolescent players (Wang et al., 2009), the results failed to be replicated among adult players (Procci et al., 2012). We also responded to the initiative of Hays et al. (2018) to remove the copyright restrictions of self-report measures, which promotes the advancement of flow research in the video game context. In the following paragraphs, we introduce several results that are highlighted in this research.

First, we conceptualised nine flow dimensions in the video game context. Our conceptualisation of flow dimensions was based on the original flow theory (Csikszentmihalyi, 1975, 1990), the flow dimensions in the sports context (Jackson, 1996), and the qualitative data from the interviews. We conceptualised the flow dimensions at the beginning of the scale development, not only because this step clarifies the content domain (DeVellis, 2016; Netemeyer et al., 2003), but also

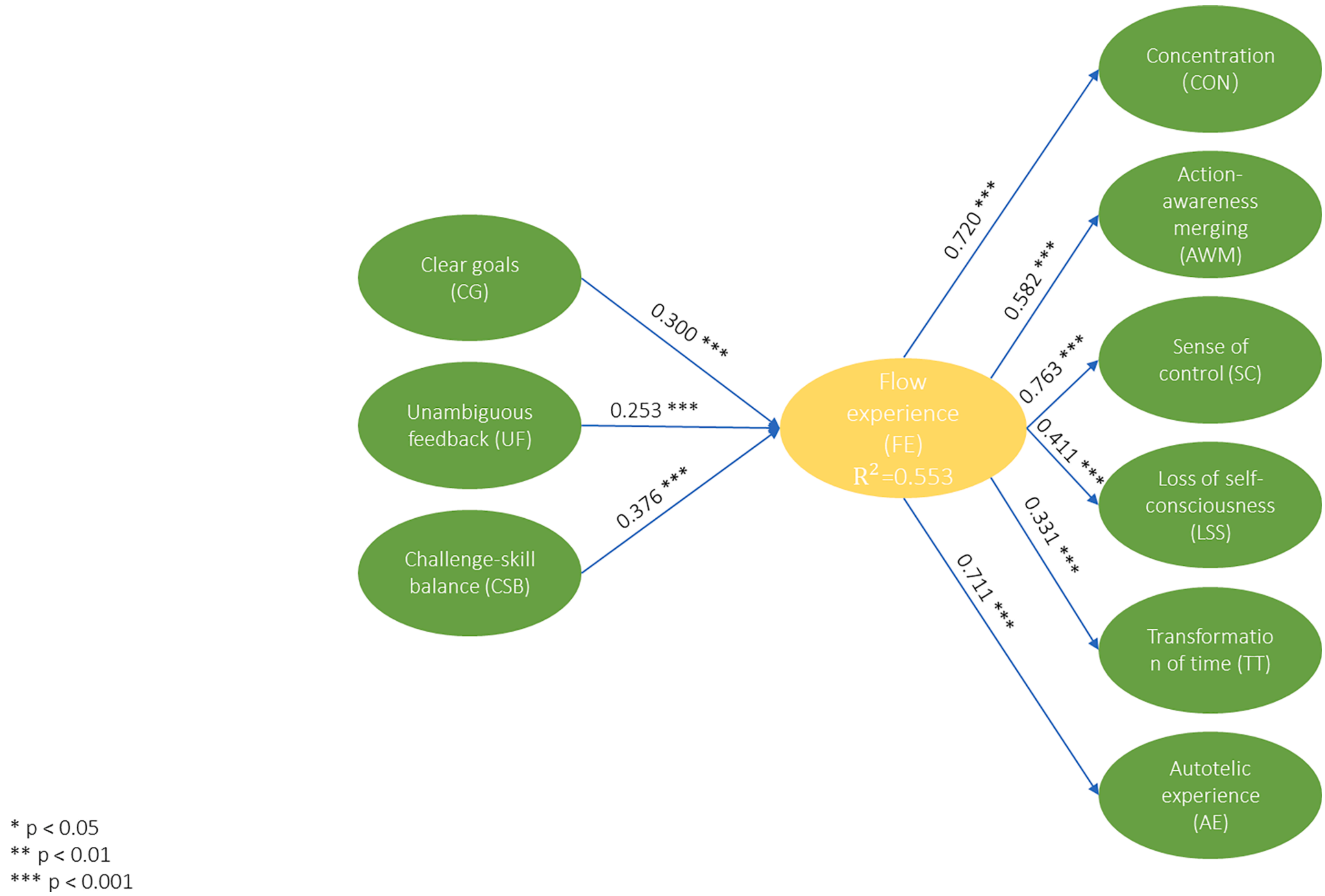

Fig. 2. Estimated results of the independent antecedent model.

because it is a response to the prevalent construct proliferation problem in the video game literature. The current literature on video games assimilates multiple terms that emulate flow experiences (Michailidis et al., 2018), such as immersion (Brown and Cairns, 2004; Ermi and Mäyrä, 2005; Jennett et al., 2008; Procci and Bowers, 2011) and engagement (Abbasi et al., 2017, 2019; Brockmyer et al., 2009; Wiebe et al., 2014). We recognise that the content domains of flow experience, immersion, and engagement have a certain degree of intersection. However, we also agree that any attempts to take some of the components of flow experience as the definitional aspect of flow will consequently disregard essential parts (Engeser and Schiepe-Tiska, 2012). Additionally, compared with the previous conceptualisation of flow experience in the video game context, the GameFlow (Sweetser and Wyeth, 2005), our conceptualisation of flow experience is closer to the original flow theory (Csikszentmihalyi, 1975, 1990). For instance, instead of dividing challenging and skill into two distinct dimensions, we consider that the balance of these two components forms a critical dimension to enter the flow experience. Moreover, we did not include social interaction into the content domain of flow experience, as the own researchers who conceptualised GameFlow (Sweetser and Wyeth, 2005, p. 10) stated that "*social interaction is not an element of flow, and often can even interrupt immersion in games*". Therefore, we consider that social interaction is more likely to be an extraneous factor to affect flow experience in the nomological network instead of being a component of flow per se. In conclusion, our conceptualisation of flow experience is highly consistent with the original flow experience (Csikszentmihalyi, 1975, 1990), which delineates the content domain of flow for future video game researchers to avoid the construct proliferation problem. However, we also encourage future researchers to provide more behavioural and neurophysiological evidence in experimental settings to refine the conceptualisation of flow experience in the video game context.

Second, we developed a 28-item dispositional flow sale, VGDFS, which is applicable in the video game context. Empirical results suggest that the VGDFS is a reliable and valid scale to measurement dispositional flow experience among Adult American video games players. Moreover, results also suggest that our scale has weak invariance across gender groups and strict invariance across age groups. The existence of measurement invariance enhances the generalisability of a scale (Netemeyer et al., 2003). However, we would still urge future researchers to conduct replication studies to verify the applicability of VGDFS in a more border population outside the United States, such as players from other English-speaking countries (e.g. Canada and the United Kingdom) and developing countries where English is one of the official languages (e.g. India and the Philippines).

Third, we assessed three operationalisations of flow experience, which are the unidimensional model, independent antecedent model, and hierarchical antecedent model. The AIC and BIC suggest that the hierarchical antecedent model performs slightly better than the other two models. Additionally, when specifying clear goals, unambiguous feedback, and challenge-skill balance are grouped into a second-order factor, they explain much more variance than when specified individually. Therefore, we recommend future researchers specify the flow experience using the hierarchical antecedent model, as it is better supported both theoretically (Keller and Landhäußer, 2012) and empirically. However, we also noted that the factor loadings of loss of self-consciousness and transformation of time are relatively low even in the hierarchical antecedent model. Considering the good performance

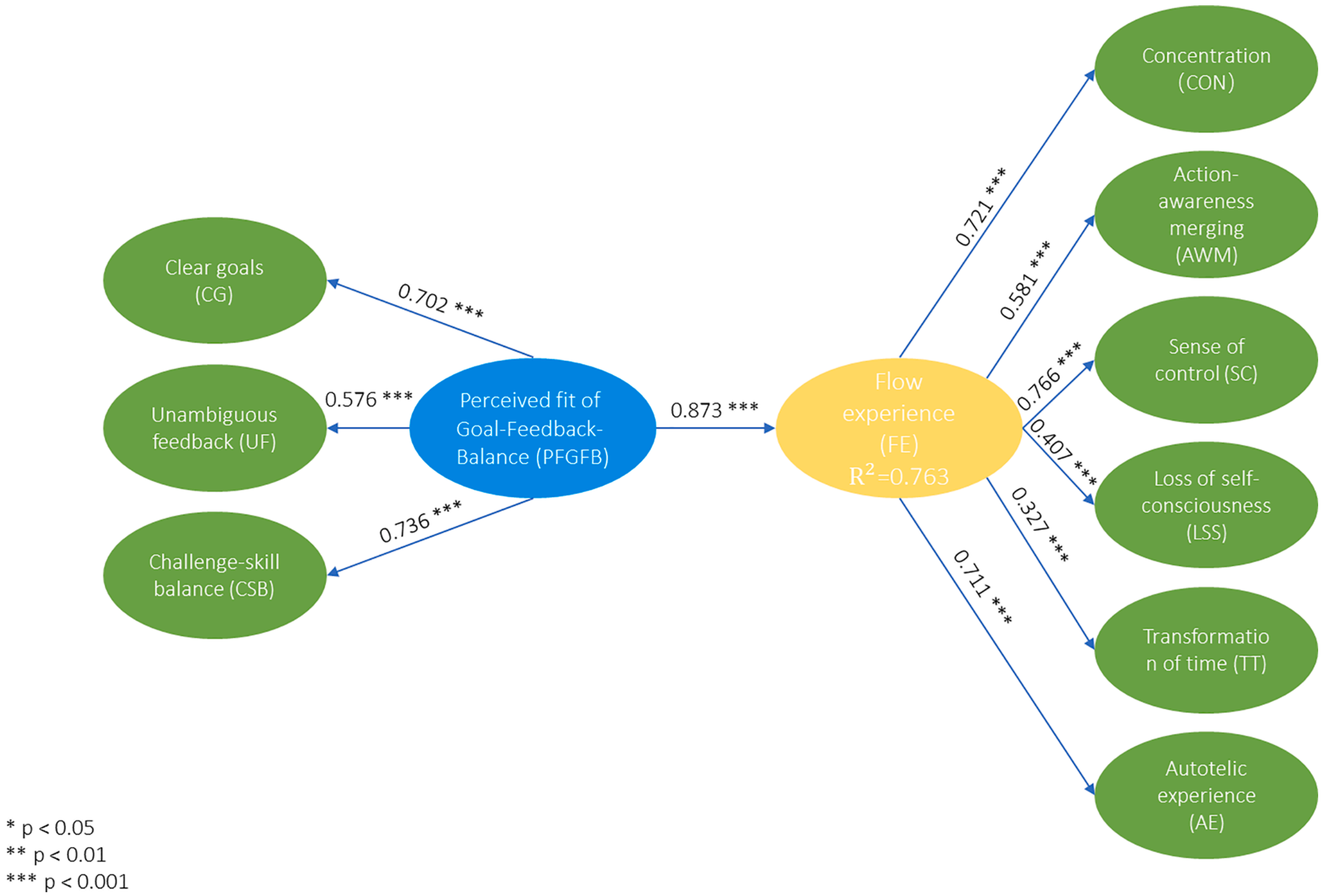

Fig. 3. Estimated results of the hierarchical antecedent model.

Table 12
Model fit index.

Model specification	Group	χ2	df	p value	CFI	TLI	RNI	SRMR	RMSEA
First-order model (62 items)	All observations (n = 593)	4693.595	1793	<0.001	0.902	0.897	0.902	0.055	0.052
First-order model (43 items)		1409.09	824	<0.001	0.969	0.966	0.969	0.037	0.035
First-order model (28 items)		438.475	314	<0.001	0.988	0.985	0.988	0.027	0.026
Unidimensional model (28 items)	All observations (n = 593)	811.026	341	<0.001	0.954	0.949	0.954	0.086	0.048
	Male group (n = 296)	682.996	341	<0.001	0.933	0.926	0.933	0.099	0.058
	Female group (n = 297)	590.562	341	<0.001	0.953	0.948	0.953	0.084	0.05
	Under 30 group (n = 289)	607.388	341	<0.001	0.946	0.94	0.946	0.088	0.052
	Over 30 group (n = 304)	640.849	341	<0.001	0.944	0.938	0.944	0.094	0.054
Independent antecedents model (28 items)	All observations (n = 593)	792.76	338	<0.001	0.956	0.95	0.956	0.083	0.048
	Male group (n = 296)	678.714	338	<0.001	0.933	0.925	0.933	0.097	0.058
	Female group (n = 297)	574.094	338	<0.001	0.955	0.95	0.955	0.081	0.048
	Under 30 group (n = 289)	600.089	338	<0.001	0.947	0.941	0.947	0.085	0.052
	Over 30 group (n = 304)	631.28	338	<0.001	0.945	0.939	0.945	0.092	0.053
Hierarchical antecedent model (28 items)	All observations (n = 593)	794.584	340	<0.001	0.956	0.951	0.956	0.083	0.047
	Male group (n = 296)	680.092	340	<0.001	0.933	0.926	0.933	0.097	0.058
	Female group (n = 297)	577.503	340	<0.001	0.955	0.95	0.955	0.082	0.048
	Under 30 group (n = 289)	600.73	340	<0.001	0.947	0.941	0.947	0.086	0.052
	Over 30 group (n = 304)	632.964	340	<0.001	0.945	0.939	0.945	0.092	0.053

of the first-order factors of the VGDFS, we speculate that there are other possible operationalisations of flow experience. For example, it is possible that loss of self-consciousness and transformation of time are actually the consequences of flow experience rather than the flow per se. Therefore, we encourage future researchers to explore the operationalisation of flow experience more deeply using both conceptual and experimental approaches.

10. Implications

This research has both theoretical and practical implications.

On the theoretical side, first, we first adapted the conceptualisation of the nine flow dimensions (Jackson, 1996) to the video game context. We then defined each dimension of the flow experience to delineate the content domain. Clearly defining the constructs, including dimensions and domains, is an essential step when developing scale (Churchill, 1979; Netemeyer et al., 2003). Our conceptualisation of the nine flow dimensions guides future researchers to clarify the dimensions and domains according to the flow theory in the video game context, which helps to moderate the prevalent construct proliferation problem in the video game literature where dimensions of flow experience are used. Second, we empirically tested three operationalisations of flow experience. From both the theoretical and empirical sides, the hierarchical antecedent model is better supported. We, therefore, recommend future researchers operationalise the flow experience using the hierarchical antecedent model and explore its theoretical generalisability in other contexts.

On the practical side, the VGDFS is the first scale that faithfully conceptualises the dimensions of the original flow theory (Csikszentmihalyi, 1975, 1990) in the video game context. The appearance of the VGDFS makes up for the deficiency of the application of DFS-2 (Jackson et al., 2011; Jackson and Eklund, 2002) among adult players in the video game context (Procci et al., 2012). Therefore, practitioners in the video game industry, such as game developers and project managers, are encouraged to use the VGDFS to measure video game players' dispositional flow experience. For instance, video game developers could use the VGDFS to measure players' dispositional flow during the alpha and beta tests in the pre-launching period, and evaluate its correlates with other essential factors, such as the intention to play the game, attitude toward playing the game, satisfaction, loyalty, and purchase intention. The results of the mentioned psychological correlates will give insights regarding the game balance to the video game development process, which improves players' gaming experience. Moreover, in many cases, researchers not only incorporate the flow experience as a whole in their empirical models (e.g. Animesh et al., 2017) but also include exclusively certain dimensions of flow (e.g. Patanasiri and Krairit, 2019). Therefore, the VGDFS is expected not only to be used as a whole to measure the flow experience but also can be disassembled and used separately to measure the dimensional facets of flow.

Finally, unlike copyrighted dispositional flow scales, such as DFS-2 (Jackson et al., 2011; Jackson and Eklund, 2002), S DFS (Jackson et al., 2011, 2008; Martin and Jackson, 2008), and C DFS (Jackson et al., 2011, 2008; Martin and Jackson, 2008), we responded the initiative of Hays et al. (2018) to remove the financial obstructs in using psychometric measures. We believe an open assess dispositional flow scale helps to build a more equitable, accessible, and innovative world for both academic researchers and industrial practitioners.

11. Limitations

Despite the implications that our research has, several limitations are noted.

First, we used Prolific, a non-probability-based panel, to recruit the participants. Non-probability-based panels, or volunteer opt-in panels, involve a self-selection process by the people who want to join the panel (Callegaro and Disogra, 2008; Callegaro et al., 2015). Non-probability-based panels do not include the non-internet population (Callegaro et al., 2015). Thus, although Prolific has several advantages (Palan and Schitter, 2018) that we have introduced and nowadays there are fewer video game players in developed countries lack Internet access, we strongly recommend future researchers to use probability-based panels to recruit their participants if they intend to conduct a replication study in developing countries with limited internet access, which serves to reduce the coverage error.

Second, in this research, we tested the measurement invariance of the VGDFS across gender and age groups, which are two key demographic profiles among video game players (ESA, 2019; Newzoo, 2019b). However, there are more variables that can be tested, such as the measurement invariance across nations (Steenkamp and Baumgartner, 1998) and over time (Kline, 2015, p. 396). Therefore, we encourage future researchers to verify the measurement invariance of VGDFS across different countries and at different time points, which are important for the future cross-cultural and longitudinal studies of video game flow experience.

Third, the VGDFS was developed in the general video game context. However, there exist different video games according to different classification methods, such as action/strategy games, online/offline games, console/portable games, paid/freemium games, etc. Therefore, future researchers are encouraged to conduct replication studies to test the applicability in the mentioned sub-contexts of video games. Replication studies are likely to generate new insights (Evanschitzky and Scott Armstrong, 2013) and they are critical to improve the external validity (Easley et al., 2000).

Forth, the VGDFS is a composite scale with 28 items. Like other composite scales, VGDFS also suffers from the length problem. Lengthy scale increases not only the nonresponse rate (Vicente and Reis, 2010) but also the common method bias (MacKenzie and Podsakoff, 2012). Therefore, we encourage future researchers to conduct scale shortening studies (Coste et al., 1997) to control the number of items in the VGDFS.

Fifth, in this research, we assessed the nomological validity of the VGDFS by estimating the path between the antecedents of flow and flow experience per se. However, researchers should note that the results were derived from the correlational study. To further confirm the causal relationship between antecedents of flow and flow experience, researchers should employ experimental designs. Moreover, to further verify the nomological validity, it is important to explore the nature of lawful relationships between the focal construct and other constructs, apart from testing whether the indicators of the focal construct relate to measures of other constructs in the manner expected (Mackenzie et al., 2011). Therefore, we encourage future researchers to conduct factor analytic studies, in which measures of other constructs are involved, such as intention to play video games, attitude towards playing video games, satisfaction, loyalty, and purchase intention of in-game goods. This approach helps to further explore the measurement structure as well as the nomological validity of the VGDFS.

12. Conclusions

In this research, we conceptualised flow experience and developed a dispositional flow scale in the video game context. We named our 28-item scale *Video Game Dispositional Flow Scale* (VGDFS). We found that the VGDFS fitted well in the commonly operationalised structure, including unidimensional model, independent antecedent model, and hierarchical antecedent model, although the last of these gives the best degree of model fit. The VGDFS empowers both the academic and industrial research of dispositional flow experience in the video game context. video game researchers and practitioners are therefore encouraged to use the VGDFS to explore the correlates between flow experience and other variables. We also mentioned the limitations of this research and provided the directions for future researchers.

CRediT authorship contribution statement

Xiaowei Cai: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Javier Cebollada:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Mónica Cortiñas:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We acknowledge financial support for this research from the Spanish Ministry of Education and Science [Project number: ECO2015-65393-R and PID2019-108554RB-I00] and the Public University of Navarre.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijhcs.2021.102746.

References

Abbasi, A.Z., Ting, D.H., Hlavacs, H., 2017. Engagement in games: developing an instrument to measure consumer videogame engagement and its validation. Int. J. Comput. Games Technol. 2017, 1–10.

Abbasi, A.Z., Ting, D.H., Hlavacs, H., Costa, L.V., Veloso, A.I., 2019. An empirical validation of consumer video game engagement: a playful-consumption experience approach. Entertain. Comput. 29, 43–55.

Animesh, Pinsonneault, Yang, Oh, 2017. An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. MIS Q. 35, 789–810.

Beard, K.S., Hoy, W.K., 2010. The nature, meaning, and measure of teacher flow in elementary schools : a test of rival hypotheses. Educ. Adm. Q. 46, 426–458.

Bearss, K., Taylor, C.A., Aman, M.G., Whittemore, R., Lecavalier, L., Miller, J., Scahill, L., 2016. Using qualitative methods to guide scale development for anxiety in youth with autism spectrum disorder. *Autism* 20, 663–672.

Bergkvist, L., Eisend, M., 2021. The dynamic nature of marketing constructs. J. Acad. Market. Sci. 49, 521–541.

Bosman, S. (2020). The way consumers interact with games is changing; nearly half of gamers are female; over half are 30+. Retrieved from https://newzoo.com/insights/articles/consumer-engagement-with-games-is-changing-gamer-segmentation-personas-gender-age-demographics/.

Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N., 2009. The development of the Game Engagement Questionnaire: a measure of engagement in video game-playing. J. Exp. Soc. Psychol. 45, 624–634.

Brown, E., Cairns, P., 2004. A grounded investigation of game immersion. In: Conference on Human Factors in Computing Systems, pp. 1297–1300.

Cai, X., Cebollada, J., & Cortiñas, M. (2020). Video game players' purchase behaviour of different types of in-game goods. Retrieved from https://ssrn.com/abstract=3535180.

Callegaro, M., Disogra, C., 2008. Computing response metrics for online panels. Public Opin. Q. 72, 1008–1032.

Callegaro, M., Manfreda, K.L., Vehovar, V., 2015. Web Survey Methodology. SAGE, Thousands Oaks.

Casler, K., Bickel, L., Hackett, E., 2013. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Comput. Hum. Behav. 29, 2156–2160.

Chang, C.C., 2013. Examining users′ intention to continue using social network games: a flow experience perspective. Telematics Inform. 30, 311–321.

Charmaz, K., 2014. Constructing Grounded Theory, 2nd ed. Sage, Thousands Oaks.

Chen, S.-.H., Wu, W.-.Y., Dennison, J., 2018. Validation of EGameFlow: a self-report scale for measuring user experience in video game play. Comput. Entertain. 16, 1–15.

Churchill, G.A., 1979. A paradigm for developing better measures of marketing constructs. J. Market. Res. 16, 64–73.

Clack, L.A., Watson, D., 1995. Constructing validity: basic issues in objective scale development. Psychol. Assess. 7, 309–319.

Corbin, J., Strauss, A., 2015. Basics of Qualitative Research, 4th ed. SAGE Publications, Thousands Oaks.

Coste, J., Guillemin, F., Pouchot, J., Fermanian, J., 1997. Methodological approaches to shortening composite measurement scales. J. Clin. Epidemiol. 50, 247–252.

Csikszentmihalyi, M., 1975. Beyond Boredom and Anxiety: Experiencing Flow in Work and Play. Jossey-Bass, San Francisco.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper & Row, New York.

Csikszentmihalyi, M., Csikszentmihalyi, I.S., 1988. Optimal Experience: Psychological Studies of Flow in Consciousness. Cambridge university press.

Csikszentmihalyi, M., Csikszentmihalyi, I.S., 1992. Psychological Studies of Flow in Consciousness. Cambridge university press.

DeVellis, R.F., 2016. Scale Development: Theory and Applications, 4th ed. SAGE Publications, Thousands Oaks.

Easley, R.W., Madden, C.S., Dunn, M.G., 2000. Conducting marketing science: the role of replication in the research process. J. Bus. Res. 48, 83–92.

Engeser, S., Rheinberg, F., 2008. Flow, performance and moderators of challenge-skill balance. Motiv. Emot. 32, 158–172.

Engeser, S., Schiepe-Tiska, A, 2012. Historical lines and an overview of current research on flow. In: Engeser, S. (Ed.), Advances in Flow Research. Springer, New York.

Ermi, L., Mäyrä, F., 2005. Fundamental components of the gameplay experience: analysing immersion. In: Proceedings of DiGRA 2005 Conference: Changing Views - Worlds in Play.

ESA, 2019. Entertainment software association (2019) Essential facts about the computer and video game industry. Entertain. Softw. Assoc. Retrieved from https://www.theesa.com/wp-content/uploads/2019/05/ESA_Essential_facts_2019_final.pdf%0Ahttps://www.theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/.

Evanschitzky, H., Scott Armstrong, J., 2013. Research with In-built replications: comment and further suggestions for replication research. J. Bus. Res. 66, 1406–1408.

Fang, X., Chan, S., Brzezinski, J., Nair, C., 2010. Development of an instrument to measure enjoyment of computer game play. Int. J. Hum. Comput. Interact. 26, 868–886.

Fang, X., Zhang, J., Chan, S.S., 2013. Development of an instrument for studying flow in computer game play. Int. J. Hum. Comput. Interact. 29, 456–470.

Finneran, C.M., Zhang, P., 2003. A person-artefact-task (PAT) model of flow antecedents in computer-mediated environments. Int. J. Hum. Comput. Stud. 59, 475–496.

Finney, S.J., DiStefano, C., 2013. Nonnormal and categorical data in structural equation modeling. In: Hancock, G.R., Mueller, R.O. (Eds.), Structural Equation modeling: A second Course, 2nd ed. Information Age.

Fischer, R., Karl, J.A., 2019. A primer to (cross-cultural) multi-group invariance testing possibilities in R. Front. Psychol. 10, 1–18.

Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. J. Market. Res. 18, 39–50.

Fu, F.L., Su, R.C., Yu, S.C., 2009. EGameFlow: a scale to measure learners' enjoyment of e-learning games. Comput. Educ. 52, 101–112.

Glaser, B.G., Strauss, A.L., 1967. Discovery of Grounded theory: Strategies for Qualitative Research. Routledge.

Ha, I., Yoon, Y., Choi, M., 2007. Determinants of adoption of mobile games under mobile broadband wireless access environment. Inf. Manage. 44, 276–286.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2013. Multivariate Data Analysis, 7th ed. Pearson, Harlow.

Jr, Hair, F, J., Hult, G.T.M., Ringle, C., Sarstedt, M, 2016. A Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM), 2nd ed. SAGE Publications, Thousands Oaks.

Hamari, J., 2015. Why do people buy virtual goods? Attitude toward virtual good purchases versus game enjoyment. Int. J. Inf. Manage. 35, 299–308.

Hamari, J., Keronen, L., 2017. Why do people buy virtual goods: a meta-analysis. Comput. Hum. Behav. 71, 59–69.

Hamari, J., Koivisto, J., 2014. Measuring flow in gamification: dispositional Flow Scale-2. Comput. Hum. Behav. 40, 133–143.

Hardesty, D.M., Bearden, W.O., 2004. The use of expert judges in scale development. Implications for improving face validity of measures of unobservable constructs. J. Bus. Res. 57, 98–107.

Hays, R.D., Weech-Maldonado, R., Teresi, J.A., Wallace, S.P., Stewart, A.L., 2018. Commentary: copyright restrictions versus open access to survey instruments. Med. Care 56, 107–110.

Hinkin, T.R., 1995. A review of scale development practices in the study of organizations. J. Manage. 21, 967–988.

Hoffman, D.L., Novak, T.P., 1996. Marketing in hypermedia computer-mediated environments: conceptual foundations. J. Mark. 60, 50–68.

Hsiao, C.H., Tang, K.Y., 2016. On the post-acceptance of mobile movie-themed games. Electron. Commer. Res. Appl. 18, 48–57.

Hsu, C.L., Lu, H.P., 2004. Why do people play on-line games? An extended TAM with social influences and flow experience. Inf. Manage. 41, 853–868.

Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., DeShon, R.P., 2012. Detecting and deterring insufficient effort responding to surveys. J. Bus. Psychol. 27, 99–114.

Huang, T., Bao, Z., Li, Y., 2017. Why do players purchase in mobile social network games? An examination of customer engagement and of uses and gratifications theory. Program 51, 259–277.

Hulland, J., Baumgartner, H., Smith, K.M., 2018. Marketing survey research best practices: evidence and recommendations from a review of JAMS articles. J. Acad. Market. Sci. 46, 92–108.

Jackson, S.A., 1996. Toward a conceptual understanding of the flow experience in elite athletes. Res. Q. Exerc. Sport 67, 76–90.

Jackson, S.A., 2012. Flow. In: Ryan, R.M. (Ed.), The Oxford Handbook of Human Motivation. Oxford, pp. 127–140.

Jackson, S.A., Eklund, B., Martin, A., 2011. The flow manual the manual for the flow scales manual. Mind Garden. Mind Garden.

Jackson, S.A., Eklund, R.C., 2002. Assessing flow in physical activity: the flow state scale-2 and dispositional flow scale-2. J. Sport Exerc. Psychol. 24, 133–150.

Jackson, S.A., Ford, S.K., Kimiecik, J.C., Marsh, H.W., 1998. Psychological correlates of flow in sport. J. Sport Exerc. Psychol. 20, 358–378.

Jackson, S.A., Marsh, H.W., 1996. Development and validation of a scale to measure optimal experience: the flow state scale. J. Sport Exerc. Psychol. 18, 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. J. Sport Exerc. Psychol. 30, 561–587.

Jacoby, J., 1978. Consumer research: a state of the art review. J. Mark. 42, 87–96.

Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A., 2008. Measuring and defining the experience of immersion in games. Int. J. Hum. Comput. Stud. 66, 641–661.

Jiang, G., Mai, Y., Yuan, K.H., 2017. Advances in measurement invariance and mean comparison of latent variables: equivalence testing and a projection-based approach. Front. Psychol. 8, 1–13.

Keller, J., Landhäußer, A., 2012. The flow model revisited. In: Engeser, S. (Ed.), Advances in Flow Research. Springer.

Kim, D., Ko, Y.J., 2019. The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction. Comput Human Behav 93, 346–356.

Klarkowski, M., Johnson, D., Wyeth, P., Smith, S., Phillips, C., 2015. Operationalising and measuring flow in video games. In: OzCHI 2015: Being Human - Conference Proceedings, pp. 114–118.
Kline, R.B., 2015. Principles and Practice of Structural Equation Modelling, 4th ed. Guilford Press, New York.
Landhäußer, A., Keller, J., 2012. Flow and Its Affective, cognitive, and performance-related consequences. In: Engeser, S. (Ed.), Advances in Flow Research. Springer, New York, pp. 65–85.
MacKenzie, S.B., Podsakoff, P.M., 2012. Common method bias in marketing: causes, mechanisms, and procedural remedies. J. Retail. 88, 542–555.
Mackenzie, S.B., Podsakoff, P.M., Podsakoff, N.P., Mackenzie, S.B., 2011. Construct measurement and validation procedures in mis and behavioral research: integrating new and existing techniques. MIS Q. 35, 293–334.
Marsh, H.W., 1994. Confirmatory Factor Analysis Models of Factorial Invariance: a Multifaceted Approach. Struct. Equ. Model. 1, 5–34.
Martin, A.J., Jackson, S.A., 2008. Brief approaches to assessing task absorption and enhanced subjective experience: examining "short" and "core" flow in diverse performance domains. Motiv. Emot. 32, 141–157.
Meade, A.W., Lautenschlager, G.J., 2004. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organ. Res. Methods 7, 361–388.
Michailidis, L., Balaguer-Ballester, E., He, X., 2018. Flow and immersion in video games: the aftermath of a conceptual challenge. Front. Psychol. 9, 1–8.
Moneta, G.B., 2012. On the measurement and conceptualization of flow. In: Engeser, S. (Ed.), Advances in Flow Research. Springer, New York, pp. 23–50.
Nah, F.F., Hall, F., 2014. Flow in gaming: literature synthesis and framework development. Int. J. Inf. Syst. Manage. 1, 83–124.
Nakamura, J., Csikszentmihalyi, M., 2012. Flow theory and research. In: Lopez, S.J., Snyder, C.R. (Eds.), The Oxford Handbook of Positive Psychology, 2nd ed. Oxford, pp. 195–206.
Netemeyer, R.G., Bearden, W.O., Sharma, S., 2003. Scaling Procedures: Issues and Applications. SAGE, Thousands Oaks.
Newzoo. (2019a). *Global games market report*. Retrieved from https://newzoo.com/insights/trend-reports/newzoo-global-games-market-report-2019-light-version/.
Newzoo. (2019b). Male and *female gamers: how their similarities and differences shape the games m*arket. Retrieved from https://newzoo.com/insights/articles/male-and-female-gamers-how-their-similarities-and-differences-shape-the-games-market/.
Nimon, K., Reio, T.G., 2011. Measurement invariance: a foundational principle for quantitative theory building. Hum. Resour. Dev. Rev. 10, 198–214.
O'Brien, H.L., Cairns, P., Hall, M., 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int. J. Hum. Comput. Stud. 112, 28–39.
Palan, S., Schitter, C., 2018. Prolific.ac—a subject pool for online experiments. J. Behav. Exp. Finance 17, 22–27.
Patanasiri, A., Krairit, D., 2019. A comparative study of consumers' purchase intention on different internet platforms. Mob. Netw. Appl. 24, 145–159.
Peer, E., Brandimarte, L., Samat, S., Acquisti, A., 2017. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J. Exp. Soc. Psychol. 70, 153–163.
Podsakoff, P.M., Mackenzie, S.B., Lee, J.-.Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88, 879–903.
Podsakoff, P.M., MacKenzie, S.B., Podsakoff, N.P., 2016. Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. Organ. Res. Methods 19, 159–203.
Procci, K., Bowers, C., 2011. An examination of flow and immersion in games. Proc. Hum. Factors Ergon. Soc. 2183–2187.
Procci, K., Singer, A.R., Levy, K.R., Bowers, C., 2012. Measuring the flow experience of gamers: an evaluation of the DFS-2. Comput. Hum. Behav. 28, 2306–2312.
Rosseel, Y., 2014. lavaan : an R package for structural equation modeling. Presentation: Lond. Sch. Econ. 48, 1–21.
Rowan, N., Wulff, D., 2007. Using qualitative methods to inform scale development. Qual. Rep. 12, 450–466.
Saris, W.E., Satorra, A., van der Veld, W.M., 2009. Testing structural equation models or detection of misspecifications? Struct. Equ. Model. 16, 561–582.
Schmitt, N., Kuljanin, G., 2008. Measurement invariance: review of practice and implications. Hum. Resour. Manage. Rev. 18, 210–222.
Sepehr, S., Head, M., 2018. Understanding the role of competition in video gameplay satisfaction. Inf. Manage. 55, 407–421.
Sherry, J.L., 2004. Flow and media enjoyment. Commun. Theory 14, 328–347.
Shin, D.H., Shin, Y.J., 2011. Why do people play social network games? Comput. Hum. Behav. 27, 852–861.
Soutter, A.R.B., Hitchens, M., 2016. The relationship between character identification and flow state within video games. Comput. Hum. Behav. 55, 1030–1038.
Steenkamp, J.E.M., Baumgartner, H., 1998. Assessing measurement invariance in cross-national consumer research. J. Consum. Res. 25, 78–107.
Sweetser, P., Johnson, D., Wyeth, P., Anwar, A., Meng, Y., Ozdowska, A., 2017. GameFlow in different game genres and platforms. Comput. Entertain. 15, 1–24.
Sweetser, P., Wyeth, P., 2005. GameFlow: a model for evaluating player enjoyment in games. Comput. Entertain. 3, 1–24.
Tourangeau, R., Rips, L.J., Rasinski, K., 2000. The Psychology of Survey Response. Cambridge university press, Cambridge.
Vicente, P., Reis, E., 2010. Using questionnaire design to fight nonresponse bias in web surveys. Soc. Sci. Comput. Rev. 28, 251–267.
Wang, C.K.J., Liu, W.C., Khoo, A., 2009. The psychometric properties of dispositional flow scale-2 in internet gaming. Curr. Psychol. 28, 194–201.
Wiebe, E.N., Lamb, A., Hardy, M., Sharek, D., 2014. Measuring engagement in video game-based environments: investigation of the user engagement scale. Comput. Hum. Behav. 32, 123–132.
Zaichkowsky, J.L., 1985. Measuring the involvement construct. J. Consum. Res. 12, 341–352.
Zhou, T., 2013. Understanding the effect of flow on user adoption of mobile games. Pers. Ubiquitous Comput. 17, 741–748.