HAL open science

The Flow Observational Grid: an Observation-Based Solution to Assess Flow States

Camille Tordet, Séverine Erhel, Eric Wodey, Eric Jamet, Nicolas Nardi, Corentin Gonthier

To cite this version:

Camille Tordet, Séverine Erhel, Eric Wodey, Eric Jamet, Nicolas Nardi, et al.. The Flow Observational Grid: an Observation-Based Solution to Assess Flow States. Journal of Happiness Studies, 2021, 22 (7), pp.3069-3089. 10.1007/s10902-021-00356-7 . hal-03155873

HAL Id: hal-03155873

https://hal.science/hal-03155873v1

Submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Flow Observational Grid: an observation-based solution to assess flow states

Camille Tordet [(1)], Séverine Erhel [(1)], Eric Wodey [(2)], Eric Jamet [(1)], Nicolas Nardi [(2)] & Corentin Gonthier [(1)]

(1) Univ Rennes, LP3C (Laboratoire de Psychologie : Cognition, Comportement, Communication) – EA 1285, 35000 Rennes, France

(2) Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, 35000 Rennes, France

Correspondence concerning this article should be addressed to Camille Tordet, Laboratoire LP3C, Campus Villejean, Place du Recteur Henri Le Moal, CS 24307, 35043 Rennes Cedex, France. E-mail: camille.tordet@univ-rennes2.fr

This version of the article has been accepted for publication after peer review and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s10902-021-00356-7

Abstract

The flow experience is a state in which people are completely concentrated and immersed in an activity. This positive psychology concept is relevant to both performance and subjective well-being in a range of activities, but it is difficult to measure: the usual methods of questionnaires and physiological measurements are inappropriate for many applied settings and may interfere with the flow state itself. The aim of the present study was to demonstrate the feasibility of a third approach, observation-based flow measurement, by developing and validating the Flow Observational Grid (FOG). The FOG is composed of three dimensions scored by an independent observer: concentration, joy and frustration. The psychometric properties of the grid were assessed in a sample of 50 participants each playing nine short video game sessions. Recordings of the 450 video game sessions were coded by two observers with the FOG, and the convergence between FOG scores and a flow questionnaire was assessed. The FOG demonstrated very good inter-rater reliability, as well as convergent validity with the short version of the FSS. Our results show that it is possible to assess flow using observation-based methods as an alternative to more accurate measures, which may be useful in many contexts where physiological and self-reported measures are not practical.

Keywords

Flow Theory; Observational Grid; Psychometric validation

The concept of flow (defined as a state in which people are completely concentrated and immersed in an activity; Csikszentmihalyi, 1990) is especially relevant because being in a flow state is related to positive psychological functioning. For example, the flow state is associated with satisfaction with life and subjective well-being (Asakawa, 2004; Fritz & Avsec, 2007); flow improves well-being in the workplace (Bakker, 2005), and a flow state bolsters positive emotions in (video)games (Rankin et al., 2018). It is also associated with higher performance in a range of activities such as sports (Jackson et al., 1998; Swann et al., 2012), music (Iusca, 2015; Stavrou et al., 2007) or education (Erhel & Jamet, 2019).

Despite its relevance to many aspects of positive psychological functioning, measuring flow is not straightforward. To date, two major methods are available to assess flow states: self-reports and physiological measures; both have significant limitations in applied contexts. The objective of the current study was to test the possibility of using a third method to assess flow: direct observation of the participant and scoring on a Flow Observational Grid (FOG) (Addessi et al., 2006, 2012, 2015; Borderie & Michinov, 2016; Custodero, 1998, 2005).

Flow Theory relevant to Measurement and Observation

Flow is often described with nine components (Csikszentmihalyi, 1990, 1997; Csikszentmihalyi et al., 2005), on which flow measurement could be based. Three of these components are proximal conditions of flow, necessary for a flow state to occur, and the other six are consequences of the flow state (see Kawabata & Mallett, 2011; Nakamura & Csikszentmihalyi, 2014). The proximal conditions are (i) balance between challenge and skills, which allows the person to feel that the situation is both challenging and manageable. When there is an imbalance then it can create a state of anxiety (skills below demands), a state of boredom (skills higher than demands) or even apathy; (ii) the activity must provide specific goals to keep the person's attention focused; and (iii) clear feedback must be provided to allow the person to adjust their progression. The following six dimensions are aspects of the flow

state itself: (iv) loss of self-consciousness: the person is focused only on the activity, forgetting their usual concerns; (v) merging of action and awareness; the person and the activity are not dissociated; (vi) altered sense of time: time goes more quickly or more slowly; (vii) intense concentration: all attentional resources are invested in the task, while other stimuli are ignored; (viii) sense of control: the person has the perception of being able to control the situation; and (ix) the activity becomes autotelic: performing the activity is a goal in itself. Among these nine aspects of flow, some are highly subjective and would be impossible to observe directly (e.g. altered sense of time, sense of control), but some key components of flow can lend themselves to observation. In particular, two components constitute the core of the original definition of flow (Csikszentmihalyi, 1975) and both are observable: absorption in the activity and enjoyment of the activity.

Absorption in the activity in particular is usually viewed as the main feature of the flow state (Nakamura & Csikszentmihalyi, 2014; Pace, 2004; Swann et al., 2012). The idea of absorption is intrinsically related with attentional resources, which are central to flow theory (Weber et al., 2009): based on the "fixed capacity" model of attentional processing (Kahneman, 1973), the attentional resources of individuals are limited. Thus, as soon as an individual is focused on a task and experiences a flow state, their resources are fully invested and they no longer have attentional resources to perform another task. This in turn can impact certain self-reflective processes, such as time perception and self-consciousness. This state of attention is peculiar, in that people in a flow state have no subjective feeling of mental effort (Csikszentmihalyi, 1990). The activity seems less demanding in terms of attention and efforts, despite the fact that attention is being fully invested in the task; the positive valence of the activity can explain this paradox (Moller et al., 2010). Critically, attention to the activity can be observed by behavioral indices such as location of the gaze and facial expressions (e.g. Borderie & Michinov, 2016).

The second essential aspect of flow is the autotelic nature of the experience, which is often equated with enjoyment with the activity (e.g. Csikszentmihalyi, 1975; Jackson et al., 2008). The state of flow is usually associated with positive emotions (Michailidis et al., 2018), and there is a link between the level of flow and the level of positive affects, such as satisfaction and enjoyment, confirming that they are components of the flow experience (Asakawa, 2004; Rogatko, 2009). It is also possible to observe certain external manifestations of enjoyment or pleasure: facial expressions, such as smiles, and exclamations of joy.

Although flow is often associated with enjoyment, different kinds of emotion are valid components of the flow experience (Borderie & Michinov, 2016). This idea stems from the component of "balance between challenges and skills", one of the proximal conditions of flow: given that a flow experience requires a balanced challenge (neither boredom nor anxiety), it seems necessary to experience positive emotions related to success, but it also seems necessary to experience frustration related to unsuccessful outcomes and to keep trying (Kiili, 2006). Frustration can increase the person’s commitment to the activity and help them stay focused. In other words, negative feedback during an activity is not necessarily detrimental to the flow experience: if the subject detects that they have the skills to meet the challenge, the usefulness of the feedback is more important to elicit pleasure than the valence of the feedback (Csikszentmihalyi et al., 2005). In this view, a feeling of frustration can help remain concentrated and provide excitation (Lazzaro, 2009). This suggests frustration as a third possible observable aspect of the flow experience.

Frustration has rarely been studied in the context of flow, but it can be viewed as a credible aspect of the flow experience (Kaye et al., 2018), and it has been considered an observable behavior contributing to flow in at least one observational study (Borderie & Michinov, 2016). In the context of video games (the activity used in the current study), failure and frustration are necessary for a feeling of overall satisfaction (Lyons, 2015). In addition to

helping to remain focused, the emotion of frustration can ultimately create positive emotions when it is overcome during the game (Birk et al., 2015), and feelings of frustration are necessary to generate a feeling of overall satisfaction at the end of the game (Lyons, 2015). For all these reasons, observable behaviors of frustration can contribute to the proximal condition of balance between challenges and skills[1].

Existing Flow Measures and their Limitations

Self-report is the most common way to measure flow. A particular form of self-report is the experience sampling method. Csikszentmihalyi and colleagues (1977) used this approach by equipping participants with a pager that ringed at random times during the day. As soon as they received an alert, participants had to immediately report their activity and respond to a short flow questionnaire. This method is used in studies performed in ecological contexts, when the participants evolve in their natural environment. However, this approach can interrupt an ongoing flow state, and it does not allow to study the flow during a specific activity. Moreover, there is only a low probability that the participant will be in a flow state when the signal is sent.

Another possibility is self-report questionnaires, completed during or after an activity. Many scales exist; some of the most frequently used are the Flow State Scale (Jackson & Marsh, 1996) and the Flow Short Scale (Rheinberg et al., 2002). For example, the Flow State Scale (FSS) is a self-reported questionnaire to be completed by the participant directly after an activity has been completed. The nine dimensions of flow originally described by Csikszentmihalyi (1990) are evaluated by 4 items each. A short version (9 items, 1 item for each 9 dimensions) was also developed (Jackson et al., 2008). This scale has been used in various domains, ranging from physical activity to video games (e.g. Iusca, 2015; Kivikangas,

[1] It is also worth noting that both positive and negative emotions draw the player's attention and refocus it on the game (Lazzaro, 2004, 2009); both types of emotions demonstrate the player's commitment (Borderie & Michinov, 2016). The interaction between positive and negative emotions can be perceived as a way to keep the player concentrated and engaged (Lazzaro, 2009), and frustration behaviors could thus also possibly reflect attention to the activity.

2006; J. J. Martin & Cutler, 2002). A large number of other scales have been developed to assess flow in specific activities; examples include the use of social networks (Kaur et al., 2016), video games (Fang et al., 2013; Fu et al., 2009), education (Heutte et al., 2014) or musical creation (Wrigley & Emmerson, 2013).

The method of self-report is effective, allows for the measurement of even highly subjective aspects of the flow state, and has been used in countless past studies to understand psychological determinants of the flow state and its relations with other psychological variables. However, questionnaires involve two serious limitations, which are also shared with the experience sampling method. First, some people may not have access to their subjective experiences or may not be able to report them; this is especially an issue with children. In fact, children tend to have weak metacognitive skills leading to difficulty in assessing their subjective states (Fisher, 1998), and limited communication skills that can prevent them from reporting on these subjective states (Shamir et al., 2009). Second, questionnaires require in-depth introspection, which is not compatible with the flow state (intense concentration, loss of self-consciousness: (Brewer et al., 1991; Nakamura & Csikszentmihalyi, 2014). In other words, asking individuals whether they are in a flow state can be expected to disrupt the flow state. In certain contexts it is not possible to interrupt people during or even after the activity to complete a self-report, precluding the use of self-reports entirely. In other cases, questionnaires can be used retrospectively, but this requires individuals to reconstruct a memory of their past experience, which can be unreliable (e.g. Ericsson & Simon, 1980).

Assessing the flow state during the activity itself (online) would be an appropriate solution in certain contexts to avoid some of the limits of the questionnaire method. To this end, a few studies have assessed flow with physiological measurements, such as electroencephalography (Nacke et al., 2011), salivary cortisol level (Tozman et al., 2017), or cardiovascular function and electromyography (de Manzano et al., 2010). For example, heart

rate variability is reduced when there is a match between skills and challenge level (Keller et al., 2011). Another study showed a link between flow state (measured by self-reported questionnaire) and decrease in heart rate variability with pianists (de Manzano et al., 2010). An inverted U relation was found between salivary cortisol level (a stress hormone) and flow state: flow experience is related to moderate increase of cortisol (Keller et al., 2011; Peifer et al., 2014, 2015).

Physiological indicators enable an objective evaluation of flow; with this method, flow states can also be evaluated online, without directly interfering with the person activity. This method too has important limitations, however. First, physiological data are often difficult to analyze and to interpret. Relatively few studies have used these methods, and how they perform in different contexts is still poorly known. A systematic review of psychophysiological indicators of the flow state, especially based on the peripheral nervous system, reported contradictory results, in particular concerning heart rate and cortisol level (Knierim et al., 2018); the results are often unstable and contradictory within the same study (Mansfield et al., 2012). The measurement equipment is also costly and can be very constraining for the participant (e.g. sensors in electroencephalography, electromyography), which makes it unsuitable for certain contexts of study (e.g. at work, during mobile activities).

Observing Flow States

The third possible method of flow measurement, and the focus of the present study, is the method of direct observation: observing the subject engaged in a task, either directly or in video recordings, and scoring the observed behavior as a reflect of the flow state. Very few studies have used this method, most of them with children and with musical activities (Addessi et al., 2006, 2012, 2015; Custodero, 1998, 2005). For example, Custodero (1998) used an observational grid composed of a semantic differential scale of 9 affective indicators and 10 behavioral indicators to examine flow. After observing the videotapes of children playing with

a musical instrument, the observer completed the observational grid, which contained items such as “Was the child aware of peers? Was the child in flow?”. In another study, Custodero (2005) proposed three observable indices of flow: challenge seeking (e.g. activity initiated by the child, quality of movement), challenge monitoring (e.g. anticipation), and social context (e.g. awareness of adults and peers). Addessi and colleagues (2006) relied on video analyses conducted by observers, who used an observational grid that attempted to include the nine aspects of flow theory. Observers had to score each dimension based on occurrences of various criteria.

A slightly different approach was used in a case study conducted to define observable indicators of a flow episode (Borderie & Michinov, 2016). The authors developed a coding scheme composed of three categories: concentration, enjoyment, and discomfort, all assessed based on two types of clues: postural / facial clues (e.g. fixed gaze, smiles, mumbles) and verbal clues (e.g. verbalization expressing joy or anger). The authors considered that "a phase of intense concentration followed by a certain degree of joy constitutes an episode of flow" (Borderie & Michinov, 2016, p. 23). In this case study, four participants played a video game and were filmed during this phase. An observer then watched this game session and coded each indicator by occurrence, relatively to the timeline of the session: that is, they reported each observable occurrence of enjoyment, discomfort, and period of concentration. Each participant then watched their video game replay and was asked to report on a timeline the mental state they had experienced during the video game session. Finally, these two timelines were compared, showing that the results collected with the observational grid were closely related to the participants' subjective experience of flow. Through interviews, participants noticed they felt frustration during flow episode, supporting the idea that frustration can reflect concentration and challenge during a flow episode.

These few studies show that it is possible to use other-report observational measures of flow, which avoids interfering directly with the ongoing activity. However, generalizing the observational grids used in these studies is not straightforward. Grids developed for the field of music education are highly specific (Addessi et al., 2006, 2012, 2015; Custodero, 1998, 2005), and their psychometric qualities have never been assessed. The various coding criteria used to evaluate the flow state in these studies were sometimes very imprecise (e.g. “Was the child in flow?” in Custodero, 1998), very difficult to observe due to their highly subjective nature (e.g. “Change in the perception of time" in Addessi et al., 2006), or only allowed for qualitative descriptions of the flow state (Custodero, 2005). As for the research of Borderie and Michinov (2016), although these results are helpful in laying the groundwork for developing an observational grid, this was a case study, using a qualitative approach and including very few subjects. It is also important to note that the coding scheme used by the authors provided information about the number of flow episodes during a game session, but not about intensity of the flow state; according to Csikszentmihalyi (1992), considering the flow as an all-or-nothing state is not the right approach, as optimal experience can be viewed as a continuum between micro-flow and deep-flow. These limitations highlight the need for the development of a more general observational grid.

Rationale for the present study

Given the features of the flow state and given the importance of assessing the flow state online and with a non-invasive and non-disruptive method, it seems that developing a method to allow observation-based assessment of flow would be relevant. Investigating an observational measure would also be a way to develop cross-method assessment, as called for by prior studies (e.g. Jackson & Marsh, 1996). A few studies have shown that such an observation-based assessment is possible, and it is now necessary to develop a more general observational grid that can complement physiological measures and self-report questionnaires.

The aim of this study was to develop such a Flow Observational Grid (FOG), which could be used to assess flow states in a wide range of activities and participant profiles.

Development of the FOG was inspired by the work of Borderie and Michinov (2016) and by recent flow research, and was designed to include three observable dimensions: (1) Intense concentration on the task, which constitutes the principal characteristic of a flow state (Nakamura & Csikszentmihalyi, 2014; Pace, 2004), and is the source of three other characteristics of the flow state: time distortion, loss of self-consciousness and merging of action and awareness (Kawabata & Mallett, 2011; Nakamura & Csikszentmihalyi, 2014). (2) Enjoyment associated with the activity, as the flow state results in autotelic pleasure (Sweetser & Wyeth, 2005), and is frequently associated with positive emotions and with expressions of enjoyment about the activity (Kiili, 2006; Michailidis et al., 2018). (3) Frustration, which can be seen as a contributor of the challenge-skills balance proximal condition of flow: as a force to become more involved in the task and to keep trying (Kiili, 2006). These three dimensions have been considered together to create an index of flow by some authors, even in the context of self-report scales assessing other components of flow (see de Manzano et al., 2010).

To assess the psychometric properties of the FOG, it was necessary to record behaviors during a situation that could elicit various levels of flow. We designed an experimental study where participants were invited to play 9 video game sessions, each two-minutes long. Video games have been recognized as a conductive activity to induce flow experience (Sherry, 2004), and were used in past research on observable features of the flow (Borderie & Michinov, 2016). We varied the difficulty level of the game across gaming sessions for each participant, so as to keep sessions interesting and possibly elicit more diversity in flow states by manipulating the balance between challenge and skills (for other uses of this approach, see Keller et al., 2011; Rheinberg & Vollemeyer, 2003). Two independent observers rated participant behaviors during the gaming sessions, and the reliability of the FOG was assessed through the inter-rater

agreement between the two observers. Lastly, validity of the FOG was tested by assessing its convergent (or concurrent) validity with a self-report flow state questionnaire (the Short Flow State Scale or S-FSS; Jackson et al., 2008), completed by the participant after each session.

Method

Flow Observational Grid (FOG)

This observational instrument includes 3 dimensions: concentration, enjoyment and frustration. Observers evaluate the participant on these three dimensions by using point scales: the behavior of the subject throughout the activity receives a score on each dimension (e.g. 0 on the concentration dimension for a complete lack of concentration during the task). Intensity of the flow state during the activity can then be summarized with a total score based on the three dimension scores.

To ensure reliability of coding by the observers, precise criteria were created to guide coding on each dimension (we were concerned that a rating scale without descriptive anchors, such as a visual analogue scale, would elicit highly subjective ratings as to what constitutes 100% frustration or enjoyment, and thus low inter-rater reliability). Rating scales were initially developed with 4-point scales for all three dimensions, which appeared to be a good compromise to ensure that descriptive anchors did not overlap and were sufficiently discriminative. However, pilot testing with a first subject (collected with the same procedure as the rest of the study, but not included in the final dataset) suggested that the rating scale for Attention lacked granularity: small fluctuations of attention were observed in the task but our initial criteria led to coding them all as maximal attention. This led to the use of a 5-point scale for this dimension to improve discrimination. The final coding criteria are summarized in Table 1.

Table 1

Coding criteria for the Flow Observational Grid

Dimension	Possible scores
Concentration	0- the person's attention is mostly focused on the environment, not the activity 1- the person's attention alternates between the environment and the activity 2- the person's attention is mostly on the activity but can be momentarily captured by outside distractions 3- the person's attention is entirely focused on the activity 4- the person's attention is entirely focused on the activity and the person is absorbed in the activity, he "talks to the activity" (e.g. 'Go go, Faster !')
Enjoyment	0- the person doesn't express any enjoyment 1- the person expresses few (one or two) signs of joy, and these signs are of limited intensity (e.g. sketch a smile) 2- the person expresses several (three or more) signs of joy during or immediately after the activity (e.g. smile, signs of satisfaction or victory) 3- the person exclaims during the activity or immediately after (e.g. 'Yes, Yes !; Cool; Good Game').
Frustration	0- the person doesn't express any sign of frustration 1- the person expresses few (one or two) signs of frustration, and these signs are of limited intensity (e.g. quick pouts, frowns) 2- the person expresses several (three or more) signs of frustration during or immediately after the activity (e.g. grimaces, hands on the face) 3- the person exclaims during the activity or immediately after (e.g. 'No, No ! ', swearing)

Note. Total score computed as [(concentration / 4 + enjoyment / 3 + frustration / 3) / 3] x 100.

To complete the FOG, the observer has to watch the entire activity sequence. During the viewing, the observer controls whether concentration of participant is directed on the activity, and counts occurrences of enjoyment and frustration. Note that the observer is required to observe the emotions expressed both during the activity and at the end of the activity (see Table 1): indeed, task-related emotions may be expressed not during the flow experience, but at the end (Walker, 2010). Lastly, the observer chooses a score for the three dimensions based on the criteria provided in Table 1. After reporting the three dimension scores, a total score is

computed, giving equal weight to the three dimensions[2]: total score = [(Concentration / 4 + Enjoyment / 3 + Frustration / 3) / 3] x 100. Total scores thus range between 0 and 100.

Validation Procedure

Participants.

A sample of 50 students participated in the study. Students were recruited through the university's social network and came from different majors. Inclusion criteria were being a native French speaker, and playing video games for less than 25 hours a week (see Gentile, 2009). The sample was composed of 31 males and 19 females, with an average age of 23.06 years (SD = 3.12). 31% of participants played video games every day, 29% regularly (2-3 times per week), 24% sometimes (1 to 4 times per month) and 16% rarely (2 to 5 times per year). 90% of participants had already played the video game system used in this study (a Nintendo Wii) and 24% had already played the video game used in this study. Participants received a gift voucher worth 10€ in exchange for participation after the study. All participants provided informed consent, including consent to be filmed for scientific purposes, and were informed that they were free to withdraw from the experiment at any time. The experiment was approved by the University ethics board.

Materials.

Video game. The video game used was Super Smash Bros. Brawl, published by Nintendo in 2008 for the Nintendo Wii console. This game has received positive reviews by the video game community. This is a fighting game taking place in an arena. The objective is to knock an opponent off the stage by using several attacks. At the end of the game, the winner is the character who has knocked its opponent out of the arena the most. Participants played

[2] As noted by a reviewer, it could be theoretically meaningful to assign different weights to the three dimensions, but this is not typically done in questionnaire studies and such differential weights would have been arbitrary. In the current study, experimenting with different weights either reduced or did not change correlations with the S-FSS.

with the single-player mode. To make the testing sessions more interesting, the game was played under three degrees of difficulty (in-game labels of 1/9, 4/9 and 9/9, corresponding roughly to easy, difficult and hard; note that these difficulty levels were chosen arbitrarily and did not take the player's skills into account, making them unsuitable to test the effect of difficulty on flow). Various characters and arenas were available; for this study, participants played with only one character (Link), against three different opponents (Meta-Knight, Yoshi, Fox) and in three different arenas (Castle siege, Mario circuit, Pokemon stadium). Overall, 59.4% of the game sessions were won, including 89.5% of the easy game sessions, 80% of the normal sessions and only 10.1% of the difficult game sessions.

Flow Questionnaire. As a convergent validity measure, the participants' subjective state of flow was assessed with a self-report questionnaire, the Short Flow State Scale (S-FSS; Jackson et al., 2008, based on the longer version of the FSS-2: Jackson & Eklund, 2002). The S-FSS was translated into French for the purposes of this study (translation/back translation procedure; Van de Vijver & Hambleton, 1996). The S-FSS comprises nine items corresponding to the nine major components of flow: balance between challenge and skills, clear goal, unambiguous feedback, merging of action and awareness, concentration on the task, sense of control, loss of self-consciousness, transformation of time, and autotelic experience (Csikszentmihalyi, 1990). We also included an item to evaluate core flow: 'I was totally involved in this activity' (A. J. Martin & Jackson, 2008). Items were rated on a 5-point scale (from 1 = *Strongly disagree*, to 5 = *Strongly agree*). For the total score based on summing the 10 items, internal consistency was α = .75, very similar to the result of α = .77 reported by Jackson and colleagues (2008)[3].

Procedure.

[3] A confirmatory factor analysis, performed on data averaged per participant to account for non-independence, showed that the factor structure of the S-FSS had similar fit in our study (CFI = .87, RMSEA = .13, SRMR = .108) and in the original study of Jackson et al. (2008: CFI = .87, RMSEA = .124, SRMR = .08). All items loaded on the same general factor.

The experiment was conducted in a lab dedicated to the observation of participants and their interactions with technological devices. The lab includes a reconstructed apartment living room, which makes it possible to conduct observations in a context that is as close as possible to ecological conditions. Participants were filmed throughout each game session, with a camera centered on the head and the upper part of the bust. Subjects were informed that they were being filmed (as mentioned on the consent form), but the recordings were made with a non-intrusive camera integrated in an eye-tracking system embedded in the living room television.

Upon their arrival, participants were invited to read and sign the consent form and were informed that they were free to withdraw at any time, then completed the preliminary questionnaire (age, genre, degrees of skills in video-game, experience with different games and game consoles). Participants began the study with two training sessions on the game (each 2-minutes long) to familiarize themselves with the controls. Each participant then completed nine successive game sessions, of two minutes each. A separate video recording was made for each session. Participants completed the S-FSS (Jackson et al., 2008) after each session. Other data was collected (physiological and eye-tracking measurements) in a broader experimental framework, but they are irrelevant to the present study and will not be examined here.

In order to maintain participant interest and possibly induce variable levels of flow, levels of difficulty, opponents and arenas of play were varied throughout the sessions. For all participants, sessions 1, 4 and 7 were easy conditions, sessions 2-5-8 were normal conditions and sessions 3-6-9 were hard conditions; sessions 1-2-3 were performed with the first opponent, sessions 4-5-6 with the second and sessions 7-8-9 with the third; sessions 1-4-7 were performed in the first arena, sessions 2-5-8 in the second and sessions 3-6-9 in the third.

Data processing.

In total, 450 game sessions were completed by the 50 participants. Out of these 450 sessions, we collected 436 usable recordings: 14 videos were excluded from the analysis

because they were either unusable (stopped video recording) or the player made an improper manipulation (e.g. quitting the game). For each session, the videos were analyzed from the beginning of the game until the moment where the final result was displayed, as per the FOG criteria. All videos were scored by the first author (Observer 1; $N = 436$). A subset of the videos was also analyzed independently by a second observer (Observer 2; 30% of the total, $n = 133$) to allow for computation of inter-rater consistency. Neither observer had participated in data collection. After each video, the observers reported the scores for each dimension. Ratings of the second observer were only used for the computation of inter-rater reliability.

Statistical analysis.

To investigate the psychometric properties of the FOG, we examined its discriminating power (Ferguson, 1949; Kline, 2015; Thurlow, 1950), which requires total scores to be distributed over a reasonable range, without a floor or ceiling effect; this was achieved by examining the distribution, range, skewness and kurtosis of scores. Inter-rater reliability was tested as the intra-class correlation coefficient (ICC) between ratings of the two observers using the *psych* package (Revelle, 2018) for R. We report ICC(2, 1), i.e. two-way random effects for a single rating (see Koo & Li, 2016; McGraw & Wong, 1996) for both absolute agreement and consistency between raters. Validity was assessed by examining the correlation between scores on the FOG and scores on the S-FSS flow questionnaire.

As multiple game sessions were collected within the same participants, a potential source of bias for some analyses, especially the analysis of validity, was non-independence of the data. To confirm that this did not bias the results, we also examined properties of the FOG by averaging the nine sessions together, yielding a single data point per participant. The results of these analyses are reported in the Appendix. For the correlation between FOG and S-FSS, we also performed a multilevel analysis, using the *lme4* package (Bates et al., 2015) for R, with participants treated as a random factor, S-FSS scores and level of difficulty as independent

variables, and total FOG scores as the dependent variable. Both measures were standardized to allow for estimation of the standardized slope. The results of this analysis are included in the main text.

Statistical analyses are reported without controlling for variance related to difficulty level of the gaming sessions, both because it was varied specifically in an effort to elicit variability, and because it appeared to have little impact on total scores on the FOG (easy: $M = 48.29$, $SD = 14.68$; normal: $M = 49.96$, $SD = 15.09$; hard: $M = 50.79$, $SD = 14.76$; $F(2, 433) = 1.06$, $p = .349$, $\eta^2_p = .00$). Controlling for difficulty did not substantially change the results. Another possible concern is the effect of time, as performing nine consecutive gaming sessions could have been repetitive enough to disrupt flow. A multilevel analysis testing the effect of time (with game session treated as a numeric predictor) found a marginal increase of scores on the FOG with time, $b = .07$, $p = .062$, and a significant but descriptively weak increase of scores on the S-FSS with time, $b = .23$, $p < .001$. Thus, flow tended to slightly increase rather than decrease over successive sessions, possibly due to better handling of the game controls. Adding time as a covariate in analyses did not substantially change the results either.

Results

Descriptive Statistics

Descriptive statistics for the FOG and the S-FSS are summarized in Table 2. Responses on the S-FSS revealed that, on average, participants experienced a high level of flow during our study ($M = 3.85$ out of 5, $SD = 0.56$; more specifically, 64.67% of testing sessions had scores above 3 on the three items of the S-FSS measuring proximal conditions of flow; see Kawabata & Mallett 2011; Kawabata and Evan, 2016). A total of 91% of all sessions had S-FSS scores above the mid-range point of 3.0. Total scores on the FOG and its subdimensions were normally distributed and covered the range of possible values, except for the

Concentration dimension, where all participants had high scores (further confirming the need for a 5-point rather than 4-point scale). This result reflects intense concentration of players during this video gaming activity. Despite that, total scores on the FOG demonstrated a normal distribution with satisfying range (see Figure 1), indicating adequate discriminating power.

Reliability Analysis

The analysis of inter-rater agreement based on ICCs showed very good reliability for the FOG. Total scores were very similar for the two observers, not only in terms of correlation but also in terms of absolute value (consistency= .85, absolute agreement = .84), indicating that our coding criteria elicited precise ratings from the observer. Scores on the three FOG dimensions, when considered separately, also showed satisfying inter-rater agreement: this was true for Concentration (consistency = .77, absolute agreement = .76); for Enjoyment (consistency = .73, absolute agreement = .73); and for Frustration (consistency = .77, absolute agreement = .77).

Validity Analysis

Correlations between the various measures are summarized in Table 3. The analysis of inter-correlations between dimensions of the FOG showed a significant positive relationship between concentration and frustration, and a negative relation between frustration and enjoyment. There was no correlation between concentration and enjoyment, however; $r = .00$, $p = .962$.

Convergent validity was tested by examining the correlation between the FOG and the S-FSS self-report questionnaire. As indicated in Table 3, correlations between dimensions of the FOG and the S-FSS were of similar size for concentration, enjoyment and frustration, suggesting that all three dimensions reflected aspects of the flow state. Critically, total scores on the FOG and S-FSS were significantly correlated, $r(434) = .45$, $p < .001$, indicating good

convergent validity of the observational grid. The multilevel analysis showed results very similar to the correlational analysis, with a significant relation between S-FSS and total scores on the FOG, $t = 7.35$, $p < .001$, $b = .40$, thus confirming that the observational grid demonstrated convergent validity with the self-reported flow measure.

Table 2

Descriptive statistics for the S-FSS and for the FOG.

Measures		*N*	*M*	*SD*	Skewness	Kurtosis	Range
S-FSS		436	3.85	0.56	-0.23	-0.45	2.20 – 5.00
FOG Observer 1	Concentration	436	3.24	0.46	0.78	-0.29	2 – 4
	Enjoyment	436	0.66	0.88	1.08	0.05	0 – 3
	Frustration	436	1.38	1.00	0.07	-1.07	0 – 3
	Total FOG	436	49.68	14.85	0.50	0.38	16.67 – 100
FOG Observer 2	Concentration	133	3.35	0.48	0.62	-1.64	3 – 4
	Enjoyment	133	0.77	0.90	0.84	-0.36	0 – 3
	Frustration	133	1.12	0.95	0.13	-1.26	0 – 3
	Total FOG	133	49.00	15.94	0.84	1.10	25.00 – 100

Note. The range of possible scores was 1 – 5 for the S-FSS, 0 – 4 for Concentration, 0 – 3 for Enjoyment and Frustration, and 0 – 100 for total FOG scores.

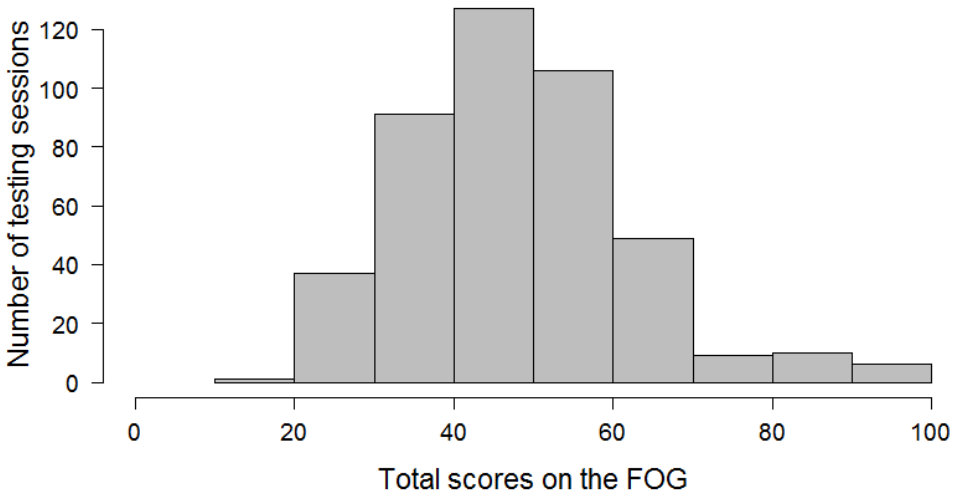

Figure 1. Distribution of total scores on the FOG for Observer 1.

Table 3

Matrix of bivariate correlations for the FOG and S-FFS.

	FOG Concentration	FOG Enjoyment	FOG Frustration	FOG Total
FOG Concentration	-			
FOG Enjoyment	.00	-		
FOG Frustration	.15*	-.12*	-	
Total	.37*	.57*	.70*	-
S-FSS	.30*	.30*	.24*	.45*

Note. Correlations computed based on the complete data set provided by Observer 1; N = 436. All correlations marked with * were significant at the p <=.001 level.

Discussion

The main goal of the present study was to develop and test an observational grid to assess flow states. The FOG demonstrated satisfying discriminating power, and very good inter-rater reliability; furthermore, its significant correlation with a self-report measure, the S-FSS, showed that it demonstrated convergent validity with participants' subjective flow states. Overall, our findings demonstrate that flow-related cues can lend themselves to observation, in line with recent work (Borderie & Michinov, 2016), and the satisfying psychometric qualities of the FOG indicate that this grid may be used to assess flow states.

Psychometric validation revealed two unexpected results in the present study. The first is that the concentration variable had only limited range, demonstrating a ceiling effect: based on the results of the FOG, all participants appeared to be very focused on the task. Although unexpected, we do not view this as an issue with the grid; in fact, this result is consistent with the protocol designed for this study. Indeed, the methods were specifically chosen to elicit high concentration: participants were required to play a fun and engaging video game in a laboratory context, were paid to do so, and the game sessions remained short at two minutes per session. It is likely that using the grid in an ecological context would yield lower concentration due to distractions in the environment, interruptions or attentional lapses from the participants. Therefore, this aspect of the results should disappear in non-laboratory studies.

The second unexpected result was that the concentration dimension did not correlate with the enjoyment dimension. These two dimensions have typically been observed to correlate in prior studies using questionnaires (e.g. Jackson & Eklund, 2002; Kawabata & Mallett, 2011). The reasons for this null correlation may be both theoretical and methodological. Theoretically, intense focus may not be necessary for enjoyment – and conversely, that being intensely focused does not necessarily result in enjoyment. Methodologically, the ceiling effect on concentration may have produced a restriction of range leading to underestimation of correlations with this measure, and the ratings for each dimension may be relatively imprecise assessments of the underlying construct (with each dimension score was based on a single rating, these scores are closer analogues to single items in a questionnaire than to the dimensional scores obtained from a sum of items in questionnaires such as the FSS). In practice, this non-correlation is not especially problematic, as all three dimensions of the FOG correlated significantly with scores on the S-FSS, suggesting that all three dimensions measured aspects of the flow state. However, this serves to remind that the three dimensions constitute complementary aspects of both the flow experience and the flow measurement, and only total scores on the FOG should be used to index flow states, not dimension scores considered separately.

The correlation between the FOG and the S-FSS, at $r = .45$, appeared to be satisfying from a psychometric standpoint. It is rare to observe correlations higher than .60 between different instruments when testing convergent validity (e.g. Cronbach, 1960), and a .45 correlation was higher than could have been expected for substantially different instruments using widely different methods – one based on observer coding without any participant input and the other on self-reports from participants. It is also worth noting that the short scale used here for convergent validity has more limited psychometric qualities than the full-length scale (see Jackson et al., 2008), which mechanically limits correlations with other measures. These

results can be usefully compared to studies assessing the convergence between physiological and self-report measures of flow: across a range of psychophysiological studies (Barros et al., 2018; Bian et al., 2016; Chin & Kales, 2019; Kivikangas, 2006; Mansfield et al., 2012; Peifer et al., 2012; Peifer et al., 2014; Stoll & Pithan, 2016; Tozman et al., 2015, 2017), the median observed correlation between physiological and self-report measures was in the .10 to .20 range; only a handful of results reached .45. In other words, observation seems to perform at least as well as physiological measures of flow. This being said, our .45 correlation between the FOG and S-FSS is still far below unity and suggests that the two types of measures could be differentially sensitive to different aspects of the flow state. This further reinforces the need to combine observation with other measures when possible, and the need for further research into cross-method assessment of flow (Jackson & Marsh, 1996).

Possible Applications

The observational method proposed here relies on three components of the flow experience: concentration (indexing absorption in the activity), enjoyment (indexing autotelic experience) and frustration related to performing an activity (as an aspect of the balance between challenges and skills). As a consequence, this grid may be useful in many contexts where participants have to deliver a performance, where they experience emotional feedbacks of joy and frustration, and where these emotions can be observed. These aspects depend on the specific activity being considered. For example, the grid could be used in sports, or in musical practice with musicians-in-training; however, it would be more difficult to use in expert musical performance, where musicians do not typically demonstrate observable manifestations of frustration. In this study, we used video games to test the FOG because this activity is recognized as being conducive to flow states. Other studies will be beneficial to generalize this grid to other activities. It is possible that frustration in particular could be unrelated to flow in different contexts, such as meditation, where the balance between challenges and skills might

take a different from and not benefit from negative feedback. Scoring criteria used for the FOG should apply to most similar situations, although some adjustments may be needed in highly specific task contexts.

The observational approach used here made it possible to overcome some of the difficulties encountered with self-administered questionnaires and physiological indicators methods. Contrary to self-report questionnaires, the observation-based method does not require the participant to retrospectively reconstruct the mental states that he felt during the activity, which is preferable since the characteristics of the flow state (intense concentration, loss of self-consciousness) make it difficult to perform accurate retrospection (Brewer et al., 1991; Nakamura & Csikszentmihalyi, 2014). The observation-based method is not affected by the introspection bias encountered with questionnaires, as the subjectivity of the participant does not come into play. In addition, this method is simple to implement and does not require specific equipment to measure flow states, unlike physiological methods that can require a long preparation, specific or invasive devices.

For these reasons, the FOG or similar observation grids may prove a useful tool to assess flow states in different contexts, especially when the population or the situation makes it impossible to use the conventional methods of self-reported measures and physiological measurements. For example, young children may be unable to communicate their mental states through questionnaires or self-reported measures, either because of their developmental level in reading ability, or because of an inability to accurately put their feelings into words. An example is the study of Inal and Cagiltay (2007), who examined children's flow experiences through interviews because of the age of participants, and in which they could not use all items of a scale because children could not understand all the questions.

Activities that cannot be interrupted to complete a questionnaire constitute another application case for the FOG. For example, it would be inappropriate to have a questionnaire

completed during a pressing situation or an anxiety-provoking situation. Likewise, interrupting people during a lesson at school or during a work task may interfere with the activity. This grid could also find an application when measuring the qualities of video games, enabling to explore the flow state of participants without interfering with the activity during game sessions. In such cases, measuring the flow state online, without interrupting the person, can be an alternative to self-report questionnaires and physiological indicators. A side benefit is that the FOG does not require additional time for the procedure, as would be the case with questionnaires. This is especially helpful for studies with a repeated measures design (as was the case here, with nine consecutive sessions), where having participants fill out questionnaires is time-consuming and can lead to participant disengagement.

Critically, we view this observation-based method as complementary with the other approaches to flow measurement. Observation is objective, but it is by definition unable to access the details of participant experience: given that flow is a subjective state, the self-report method can be expected to provide a more accurate window into flow experience in most cases. Besides, the FOG only assesses three dimensions and thus a restricted part of the flow experience, as many out of the nine components of flow theory initially described by Csikszentmihalyi (1990) cannot be accessed through observation. The limited correlation with the S-FSS also questions whether observation assesses exactly the same aspects of flow as self-report. To improve the quality of flow assessment, the FOG should thus be coupled with other measurement methods (ideally, self-report) whenever possible. When other methods are not available, however, our results show that the FOG is a valuable alternative even when used alone, allowing to widen the field of flow research.

Limitations

The validation procedure we used for the FOG had several limitations. First, this study was conducted with a student population. Although the core features of the grid should remain

comparable in other samples (answering observation criteria mostly depends on the observer, contrary to self-report questionnaires), more research is needed to test this grid with other populations such as children for whom observation may prove to be an interesting alternative method. However, while it would be beneficial to test the grid with other populations and in other contexts, we do not believe it would be critical to perform a new psychometric validation when translating the grid to another language – as would be necessary with a questionnaire. This is because the participants never directly interact with the grid, and scoring criteria are relatively objective, allowing for use by trained researchers in a way that could be comparable across languages.

A separate issue is that the duration of the activity session in the current study was relatively short, with two minutes of game time per session. This may have influenced flow experience, as players could need longer game sessions to enter a deeper flow state, or conversely could exit flow if the duration of the task is too long; it could therefore be useful to assess the functioning of the FOG with a longer activity. Another possible benefit to using longer time periods is that it would make it possible to use a longer self-report scale to assess flow. The short version of the FSS was used for the present study because game sessions were short: a longer scale would have taken longer to complete than the game sessions themselves, which would certainly have interfered with the validity of the measure and/or the flow experience itself; however, the S-FSS has more limited psychometric qualities than the longer version (Jackson et al., 2008). Moreover, the S-FSS uses a single item for each of the nine components of flow, which makes it impossible to examine them separately; conversely, a longer scale would make it possible to verify how observable behavior relates to the various components of flow.

Furthermore, there is another reason why task duration is an important issue in flow research: flow may not be constant throughout the whole duration of a task. Flow research uses

variable activity durations, ranging from minutes, for example a game session of 3 minutes (de Sampaio Barros et al., 2018), to tens of minutes, for example chess players playing for 25 minutes (Tozman et al., 2017). The FOG was designed and is appropriate to evaluate flow states during short periods of activity; evaluating the flow state for a longer period with this method, which yields a single total score, would certainly make the results imprecise. Indeed, the FOG yields a single global score and informs about the general flow state of the participant during the activity, but does not allow for precise estimation of peak flow periods. The problem is exactly the same with self-report questionnaires, but the observation-based method has an added benefit in this case: it allows the activity to be split in brief sessions, without interrupting the participant. If intending to use the FOG for a longer task, it would thus be preferable to decompose the activity in sub-sessions (possibly based on sub-goals or precise interval lengths), and to assess flow state separately for each of these sub-sessions. This method can be used to study the flow state with a much greater temporal resolution than self-report measures, and even to study the effect of time directly. The observational method could thus provide a window into the precise timecourse of the flow experience.

Acknowledgements

This research was supported technologically and logistically by the LOUSTIC userlab on human factors and observation of the uses of information and communication technology, for which we would like to thank Echraf Mekacher and Ludivine Gueho. Special thanks to Oumaïma Hathroubi, Énora David and Amélie Ferré who helped us to collect the data for this study during their Master's degree. We would also like to thank Ophélie Zampatti who helped us code the videos.

References

Addessi, A. R., Ferrari, L., Carlotti, S., & Pachet, F. (2006). Young children musical experience with a flow machine. *Proceedings of the 9th International Conference of Music Perception and Cognition*, 1658–1665.

Addessi, A. R., Ferrari, L., & Carugati, F. (2012). Observing and measuring the Flow emotional state in children interacting with the MIROR platform. *Proc. 12th Int. Conf. Music Percept. Cogn. 8th Trienn. Conf. Eur. Soc. Cogn. Sci. Music, E. Cambouropoulos, C. Tsougras, P. Mavromatis, and K. Pastiadis, Eds., Thessaloniki, Greece*, 20–30.

Addessi, A. R., Ferrari, L., & Carugati, F. (2015). The Flow Grid: A Technique for Observing and Measuring Emotional State in Children Interacting with a Flow Machine. *Journal of New Music Research*, *44*(2), 129–144. https://doi.org/10.1080/09298215.2014.991738

Asakawa, K. (2004). Flow Experience and Autotelic Personality in Japanese College Students: How do they Experience Challenges in Daily Life? *Journal of Happiness Studies*, *5*(2), 123–154. https://doi.org/10.1023/B:JOHS.0000035915.97836.89

Bakker, A. B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior*, *66*(1), 26–44. https://doi.org/10.1016/j.jvb.2003.11.001

Barros, M. F. de S., Araújo-Moreira, F. M., Trevelin, L. C., & Radel, R. (2018). Flow experience and the mobilization of attentional resources. *Cognitive, Affective & Behavioral Neuroscience*, *18*(4), 810–823. doi:10.3758/s13415-018-0606-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., Sun, X., & Meng, X. (2016). A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. *Personal and Ubiquitous Computing*, *20*, 821–832. doi:10.1007/s00779-016-0953-5

Birk, M. V., Iacovides, I., Johnson, D., & Mandryk, R. L. (2015). The False Dichotomy between Positive and Negative Affect in Game Play. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, 799–804. https://doi.org/10.1145/2793107.2810258

Borderie, J., & Michinov, N. (2016). Identifying Flow in Video Games: Towards a New Observation-Based Method. *International Journal of Gaming and Computer-Mediated Simulations*, *8*(3), 19–38. https://doi.org/10.4018/IJGCMS.2016070102

Brewer, B. W., Van Raalte, J. L., Linder, D. E., & Van Raalte, N. S. (1991). Peak Performance and the Perils of Retrospective Introspection. *Journal of Sport and Exercise Psychology*, *13*(3), 227–238. https://doi.org/10.1123/jsep.13.3.227

Chin, M. S., & Kales, S. N. (2019). Is there an optimal autonomic state for enhanced flow and

executive task performance? *Frontiers in Psychology*, *10*. doi:10.3389/fpsyg.2019.01716

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). Harper.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. Jossey-Bass Publishers.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal performance*. New York: Harper and Row.

Csikszentmihalyi, M. (1992). A reponse to the Kimiecik & Stein and Jackson papers. *Journal of Applied Sport Psychology*, *4*(2), 181–183. https://doi.org/10.1080/10413209208406460

Csikszentmihalyi. (1997). *Finding flow: The psychology of engagement with everyday life* (1st ed). Basic Books.

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2005). Flow. In *Handbook of competence and motivation* (pp. 598–608). Guilford Press.

Csikszentmihalyi, M., Larson, R., & Prescott, S. (1977). The ecology of adolescent activity and experience. *Journal of Youth and Adolescence*, *6*(3), 281–294. https://doi.org/10.1007/BF02138940

Custodero, L. A. (1998). Observing flow in young children's music learning. *General Music Today*, *12*(1), 21–27.

Custodero, L. A. (2005). Observable indicators of flow experience: A developmental perspective on musical engagement in young children from infancy to school age. *Music Education Research*, *7*(2), 185–209. https://doi.org/10.1080/14613800500169431

de Manzano, Ö., Theorell, T., Harmat, L., & Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion*, *10*(3), 301–311. https://doi.org/10.1037/a0018432

de Sampaio Barros, M. F., Araújo-Moreira, F. M., Trevelin, L. C., & Radel, R. (2018). Flow experience and the mobilization of attentional resources. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(4), 810–823. https://doi.org/10.3758/s13415-018-0606-4

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. doi:10.1037/0033-295X.87.3.215

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*, 106–114. https://doi.org/10.1016/j.chb.2018.09.020

Fang, X., Zhang, J., & Chan, S. S. (2013). Development of an Instrument for Studying Flow in Computer Game Play. *International Journal of Human-Computer Interaction*, *29*(7), 456–470. https://doi.org/10.1080/10447318.2012.715991

Ferguson, G. A. (1949). On the theory of test discrimination. *Psychometrika*, *14*(1), 61–68. https://doi.org/10.1007/BF02290141

Fisher, R. (1998). Thinking About Thinking: Developing Metacognition in Children. *Early Child Development and Care*, *141*(1), 1–15. https://doi.org/10.1080/0300443981410101

Fritz, B. S., & Avsec, A. (2007). The experience of flow and subjective well-being of music students. *Horizons of Psychology*, *16*, 5–17.

Fu, F.-L., Su, R.-C., & Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101–112. https://doi.org/10.1016/j.compedu.2008.07.004

Gentile, D. (2009). Pathological Video-Game Use Among Youth Ages 8 to 18: A National Study. *Psychological Science*, *20*(5), 594–602. https://doi.org/10.1111/j.1467-9280.2009.02340.x

Heutte, J., Fenouillet, F., Boniwell, I., Martin-Krumm, C., & Csikszentmihalyi, M. (2014, October). Optimal learning experience in digital environments: Theoretical concepts, measure and modelisation. *Symposium "Digitial Learning in 21st Century Universities"*. https://hal.archives-ouvertes.fr/hal-01470855

Inal, Y., & Cagiltay, K. (2007). Flow experiences of children in an interactive social game environment. *British Journal of Educational Technology*, *38*(3), 455–464. https://doi.org/10.1111/j.1467-8535.2007.00709.x

Iusca, D. (2015). The Relationship between Flow and Music Performance Level of Undergraduates in Exam Situations: The Effect of Musical Instrument. *Procedia - Social and Behavioral Sciences*, *177*, 396–400. https://doi.org/10.1016/j.sbspro.2015.02.376

Jackson, S. A., & Eklund, R. C. (2002). Assessing Flow in Physical Activity : The Flow State Scale–2 and Dispositional Flow Scale–2. *Journal of Sport and Exercise Psychology*, *24*(2), 133-150. https://doi.org/10.1123/jsep.24.2.133

Jackson, S. A., Ford, S. K., Kimiecik, J. C., & Marsh, H. W. (1998). Psychological Correlates of Flow in Sport. *Journal of Sport and Exercise Psychology*, *20*(4), 358–378. https://doi.org/10.1123/jsep.20.4.358

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

Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts. *Journal of Sport and Exercise Psychology*, *30*(5), 561–587. https://doi.org/10.1123/jsep.30.5.561

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kaur, P., Dhir, A., Chen, S., & Rajala, R. (2016). Flow in context: Development and validation of the flow experience instrument for social networking. *Computers in Human Behavior*, *59*, 358–367. https://doi.org/10.1016/j.chb.2016.02.039

Kawabata, M., & Evans, R. (2016). How to Classify Who Experienced Flow from Who Did Not Based on the Flow State Scale-2 Scores: A Pilot Study of Latent Class Factor Analysis. *The Sport Psychologist*, *30*(3), 267–275. https://doi.org/10.1123/tsp.2014-0053

Kawabata, M., & Mallett, C. J. (2011). Flow experience in physical activity: Examination of

the internal structure of flow from a process-related perspective. *Motivation and Emotion*, *35*(4), 393–402. https://doi.org/10.1007/s11031-011-9221-1

Kaye, L. K., Monk, R. L., Wall, H. J., Hamlin, I., & Qureshi, A. W. (2018). The effect of flow and context on in-vivo positive mood in digital gaming. *International Journal of Human-Computer Studies*, *110*, 45–52. https://doi.org/10.1016/j.ijhcs.2017.10.005

Keller, J., Bless, H., Blomann, F., & Kleinböhl, D. (2011). Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. *Journal of Experimental Social Psychology*, *47*(4), 849–852. https://doi.org/10.1016/j.jesp.2011.02.004

Kiili, K. (2006). Evaluations of an Experiential Gaming Model. *Human Technology*, *2*(2), 187–201. doi:10.17011/ht/urn.2006518

Kivikangas, J. M. (2006). *Psychophysiology of flow experience: An explorative study* [Unpublished master's thesis]. University of Helsinki. https://helda.helsinki.fi/bitstream/handle/10138/19953/psychoph.pdf

Kline, P. (2015). *A Handbook of Test Construction (Psychology Revivals): Introduction to Psychometric Design* (1st ed.). Routledge. https://doi.org/10.4324/9781315695990

Knierim, M. T., Rissler, R., Dorner, V., Maedche, A., & Weinhardt, C. (2018). The Psychophysiology of Flow: A Systematic Review of Peripheral Nervous System Features. *Information Systems and Neuroscience*, 109–120. https://doi.org/10.1007/978-3-319-67431-5_13

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lazzaro, N. (2004). *Why We Play Games: Four Keys to More Emotion Without Story*. Presented at the 2004 Game Developers Conference. http://www.xeodesign.com/xeodesign_whyweplaygames.pdf

Lazzaro, N. (2009). Why we play: Affect and the Fun of Games. Designing Emotions for Games, Entertainment Interfaces, and Interactive Products. In *Human-Computer Interaction: Designing for Diverse Users and Domains*.

Lyons, E. J. (2015). Cultivating Engagement and Enjoyment in Exergames Using Feedback, Challenge, and Rewards. *Games for Health Journal*, *4*(1), 12–18. https://doi.org/10.1089/g4h.2014.0072

Mansfield, B. E., Oddson, B. E., Turcotte, R. T., & Couture, R. (2012). A possible physiological correlate for mental flow. *The Journal of Positive Psychology*, *7*(4), 327–333. doi:10.1080/17439760.2012.691982

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. *Motivation and Emotion*, *32*(3), 141–157. https://doi.org/10.1007/s11031-008-9094-0

Martin, J. J., & Cutler, K. (2002). An Exploratory Study of Flow and Motivation in Theater Actors. *Journal of Applied Sport Psychology*, *14*(4), 344–352. https://doi.org/10.1080/10413200290103608

McGraw, K. O., & Wong, S. P. (1996). "Forming inferences about some intraclass correlations coefficients": Correction. *Psychological Methods*, *1*(4), 390–390. https://doi.org/10.1037/1082-989X.1.4.390

Michailidis, L., Balaguer-Ballester, E., & He, X. (2018). Flow and Immersion in Video Games: The Aftermath of a Conceptual Challenge. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01682

Moller, A. C., Meier, B. P., & Wall, R. D. (2010). Developing an experimental induction of flow: Effortless action in the lab. In *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 191–204). MIT Press. https://doi.org/10.7551/mitpress/9780262013840.003.0010

Nacke, L. E., Stellmach, S., & Lindley, C. A. (2011). Electroencephalographic Assessment of Player Experience: A Pilot Study in Affective Ludology. *Simulation & Gaming*, *42*(5), 632–655. https://doi.org/10.1177/1046878110378140

Nakamura, J., & Csikszentmihalyi, M. (2014). The Concept of Flow. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 239–263). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_16

Pace, S. (2004). A grounded theory of the flow experiences of Web users. *International Journal of Human-Computer Studies*, *60*(3), 327–363. https://doi.org/10.1016/j.ijhcs.2003.08.005

Peifer, C. (2012). Psychophysiological correlates of flow-experience. In S. Engeser (Ed.), *Advances in flow research* (pp. 139–164). Springer Science + Business Media. doi:10.1007/978-1-4614-2359-1_8

Peifer, C., Schächinger, H., Engeser, S., & Antoni, C. H. (2015). Cortisol effects on flow-experience. *Psychopharmacology*, *232*(6), 1165–1173. https://doi.org/10.1007/s00213-014-3753-5

Peifer, C., Schulz, A., Schächinger, H., Baumann, N., & Antoni, C. H. (2014). The relation of flow-experience and physiological arousal under stress—Can u shape it? *Journal of Experimental Social Psychology*, *53*, 62–69. https://doi.org/10.1016/j.jesp.2014.01.009

Rankin, K., Walsh, L. C., & Sweeny, K. (2018). A better distraction: Exploring the benefits of flow during uncertain waiting periods. *Emotion*. https://doi.org/10.1037/emo0000479

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research* (Version 1.8.12). Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych

Rheinberg, F., & Vollmeyer, R. (2003). Flow-Erleben in einem Computerspiel unter experimentell variierten Bedingungen [Flow experience in a computer game under

experimentally varied conditions]. *Zeitschrift Für Psychologie*, *211*, 161–170.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2002). Die Erfassung des Flow-Erlebens. In *Diagnostik von Motivation und Selbstkonzept* (pp. 261–279). Hogrefe Verlag. https://publishup.uni-potsdam.de/frontdoor/index/index/docId/551

Rogatko, T. P. (2009). The Influence of Flow on Positive Affect in College Students. *Journal of Happiness Studies*, *10*(2), 133–148. https://doi.org/10.1007/s10902-007-9069-y

Shamir, A., Mevarech, Z. R., & Gida, C. (2009). The assessment of meta-cognition in different contexts: Individualized vs. peer assisted learning. *Metacognition and Learning*, *4*(1), 47–61. https://doi.org/10.1007/s11409-008-9032-2

Sherry, J. L. (2004). Flow and Media Enjoyment. *Communication Theory*, *14*(4), 328–347. https://doi.org/10.1111/j.1468-2885.2004.tb00318.x

Stavrou, N. A., Jackson, S. A., Zervas, Y., & Karteroliotis, K. (2007). Flow Experience and Athletes' Performance with Reference to the Orthogonal Model of Flow. *The Sport Psychologist*, *21*(4), 438–457. https://doi.org/10.1123/tsp.21.4.438

Stoll, O., & Pithan, J. M. (2016). Running and flow: Does controlled running lead to flow-states? Testing the Transient Hypofontality Theory. In L. Harmat, F. Ø. Andersen, F. Ullén, J. Wright, & G. Sadlo (Eds.), *Flow experience: Empirical research and applications* (pp. 65–75). Springer International Publishing. doi:10.1007/978-3-319-28634-1_5

Swann, C., Keegan, R. J., Piggott, D., & Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psychology of Sport and Exercise*, *13*(6), 807–819. https://doi.org/10.1016/j.psychsport.2012.05.006

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment*, *3*(3), 3. https://doi.org/10.1145/1077246.1077253

Thurlow, W. R. (1950). Direct Measures of Discriminations Among Individuals Performed by Psychological Tests. *The Journal of Psychology*, *29*(2), 281–314. https://doi.org/10.1080/00223980.1950.9916033

Tozman, T., Magdas, E. S., MacDougall, H. G., & Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. *Computers in Human Behavior*, *52*, 408–418. doi:10.1016/j.chb.2015.06.023

Tozman, T., Zhang, Y. Y., & Vollmeyer, R. (2017). Inverted U-Shaped Function Between Flow and Cortisol Release During Chess Play. *Journal of Happiness Studies*, *18*(1), 247–268. https://doi.org/10.1007/s10902-016-9726-0

Van de Vijver, F., & Hambleton, R. K. (1996). Translating Tests. *European Psychologist*, *1*(2), 89–99. https://doi.org/10.1027/1016-9040.1.2.89

Walker, C. J. (2010). Experiencing flow: Is doing it together better than doing it alone? *The Journal of Positive Psychology*, *5*(1), 3–11. https://doi.org/10.1080/17439760903271116

Weber, R., Tamborini, R., Westcott-Baker, A., & Kantor, B. (2009). Theorizing Flow and Media Enjoyment as Cognitive Synchronization of Attentional and Reward Networks. *Communication Theory*, *19*(4), 397–422. https://doi.org/10.1111/j.1468-2885.2009.01352.x

Wrigley, W. J., & Emmerson, S. B. (2013). The experience of the flow state in live music performance. *Psychology of Music*, *41*(3), 292–305. https://doi.org/10.1177/0305735611425903

Appendix

This appendix reports the results obtained when averaging the nine gaming sessions together to account for non-independence of the data. Descriptive statistics are represented in Table A1; they were very similar to the results obtained for separate gaming sessions. Inter-rater agreement was excellent for total scores (consistency = .94, absolute agreement = .92), and was also very good for all dimensions separately (concentration: consistency = .97, absolute agreement =.94; enjoyment: consistency = .87, absolute agreement = .87; frustration: consistency = .84, absolute agreement = .83). Bivariate correlations between dimensions of the FOG and the S-FSS are reported in Table A2. They were all comparable to or higher than correlations obtained for separate gaming sessions, with the exception of the correlation between frustration and enjoyment which was slightly weaker. Overall, these results confirm adequate psychometric properties for the FOG.

Table A1

Descriptive statistics for the S-FSS and for the FOG, at the subject level.

Measures		*N*	*M*	*SD*	Skewness	Kurtosis	Range
S-FSS		436	3.85	0.48	-0.21	0.07	2.53 – 4.87
FOG Observer 1	Concentration	50	3.25	0.34	1.09	0.18	2.67 – 4
	Enjoyment	50	0.67	0.52	0.71	-0.09	0 – 2
	Frustration	50	1.38	0.65	0.52	-0.39	0.44 – 3
	Total FOG	50	49.86	10.50	0.50	0.28	29.63 – 75.31
FOG Observer 2	Concentration	15	3.35	0.41	0.60	-1.62	3 – 4
	Enjoyment	15	0.78	0.55	0.69	0.04	0 – 1.88
	Frustration	15	1.12	0.57	0.10	-1.07	0 – 2.11
	Total FOG	15	49.13	10.83	0.82	2.20	29.94 – 76.54

Note. The range of possible scores was 1 – 5 for the S-FSS, 0 – 4 for Concentration, 0 – 3 for Enjoyment and Frustration, and 0 – 100 for total FOG scores.

Table A2

Matrix of bivariate correlations for the FOG and S-FFS, at the subject level.

	FOG Concentration	FOG Enjoyment	FOG Frustration	FOG Total
FOG Concentration	-			
FOG Enjoyment	.12	-		
FOG Frustration	.24	.04	-	
Total	.50***	.60***	.78***	-
S-FSS	.42*	.30*	.44*	.58***

Note. Correlations computed based on the complete data set provided by Observer 1; $N = 50$. All correlations marked with * are significant at the $p <= .05$ level; correlations marked with *** are significant at the $p < .001$ level.