

Abstract

In view of the growing interest regarding binge-watching (i.e., watching multiple episodes of television (TV) series in a single sitting) research, two measures were developed and validated to assess binge-watching involvement (“Binge-Watching Engagement and Symptoms Questionnaire”, BWESQ) and related motivations (“Watching TV Series Motives Questionnaire”, WTSMQ). To promote international and cross-cultural binge-watching research, the present article reports on the validation of these questionnaires in nine languages (English, French, Spanish, Italian, German, Hungarian, Persian, Arabic, Chinese). Both questionnaires were disseminated, together with additional self-report measures of happiness, psychopathological symptoms, impulsivity and problematic internet use among TV series viewers from a college/university student population ($N = 12,616$) in 17 countries. Confirmatory factor, measurement invariance and correlational analyses were conducted to establish structural and construct validity. The two questionnaires had good psychometric properties and fit in each language. Equivalence across languages and gender was supported, while construct validity was evidenced by similar patterns of associations with complementary measures of happiness, psychopathological symptoms, impulsivity and problematic internet use. The results support the psychometric validity and utility of the WTSMQ and BWESQ for conducting cross-cultural research on binge-watching.

Keywords: binge-watching, TV series, questionnaires, cross-cultural, confirmatory factor analysis, measurement invariance

Towards a cross-cultural assessment of binge-watching: psychometric evaluation of the
“Watching TV Series Motives” and “Binge-Watching Engagement and Symptoms”
questionnaires across nine languages

Viewers of television (TV) series are currently enjoying unprecedented levels of choice and convenience. No longer dependent on linear TV programming, they can now access as many TV series episodes as they want, regardless of time and place, due to the expansion of on-demand viewing services (e.g., Netflix, Hulu, Amazon Prime) widely available on internet-connected devices. In this context, online TV series watching is increasingly becoming a major part of many individuals' daily lives (Deloitte's digital media trends survey 2018, 2019). However, this **major shift in TV series viewing patterns** has also led to the emergence of binge-watching, which in the absence of a consensual definition, may be referred to as watching multiple episodes of TV series **in a single sitting** (Exelmans & Van den Bulck, 2017; **Flayelle et al., 2020**). Binge-watching has evolved into a common practice, especially among young viewers (Exelmans & Van den Bulck, 2017; Panda & Pandey, 2017; Spangler, 2016; YouGov Omnibus, 2017): recent market reports revealed binge-watching habits among 91% of 14- to 20-year-old and 86% of 21- to 34-year-old individuals (Deloitte's digital media trends survey, 2018).

While binge-watching may provide an enhanced viewing experience due to a deeper sense of immersion (Erickson, Dal Cin, & Byl, 2019; Matrix, 2014; Merrill & Rubenking, 2019; Petersen, 2016; Shim & Kim, 2018; Steiner & Xu, 2018), social inclusion or group affiliation (Flayelle, Maurage, & Billieux, 2017; Panda & Pandey, 2017; Pittman & Sheehan, 2015; Ramayan, Munsayac Estella, & Abu Bakar, 2018; Steiner & Xu, 2018), and personal enrichment (Adachi, Ryan, Frye, McClurg, & Rigby, 2017; Mikos, 2016; Perks, 2015; Tukachinsky & Eyal, 2018), there have been academic and clinical concerns about the

potential development of heavier viewing patterns that may generate negative consequences for some individuals. These concerns have prompted a recent proliferation of studies of binge-watching, identifying potentially deleterious effects on academic and professional performance (De Feijter, Khan, & Van Gisbergen, 2016; Petersen, 2016; Rubenking, Bracken, Sandoval, & Rister, 2018), sleep hygiene (Brookes & Ellithorpe, 2017; Exelmans & Van den Bulck, 2017; Kruger, Karmakar, Elhai, & Kramer, 2015a), physical activity and healthy eating (Kubota, Cushman, Zakai, Rosamond, & Folsom, 2018; Morris, Bradbury, Cross, Gunter, & Murphy, 2018; Vaterlaus, Spruance, Frantz, & Kruger, 2019), and quality of social life (De Feijter et al., 2016; Hernández Pérez & Martínez Díaz, 2016; Vaterlaus et al., 2019). Given these data, along with other findings reporting associations between binge-watching and mental health concerns like anxiety and depression (Ahmed, 2017; Kruger, Karmakar, Elhai, & Kramer, 2015b; Sung, Kang, & Wee, 2015; Tukachinsky & Eyal, 2018), and the potential predictive role of poor self-control in its onset and maintenance (Hasan, Kumar Jha, & Liu, 2018; Merrill & Rubenking, 2019; Tukachinsky & Eyal, 2018), binge-watching is increasingly viewed as an addiction-like phenomenon (e.g., Granow, Reinecke, & Ziegele, 2018; Orosz, Bóthe, & Tóth-Király, 2016; Riddle, Peebles, Davis, Xu, & Schroeder, 2017; Shim, Lim, Jung, & Shin, 2018; Starosta, Izydorczyk, & Lizińczyk, 2019; Steiner & Xu, 2018; Sung et al., 2015; Tukachinsky & Eyal, 2018) that should be further investigated and characterized (Brookes & Ellithorpe, 2017; Flayelle, Maurage, Vögele, Karila, & Billieux, 2019a; Merikivi, Bragge, Scornavacca, & Verhagen, 2019; Shim et al., 2018; Spruance, Karmakar, Kruger, & Vaterlaus, 2017; Starosta et al., 2019; Sung, Kang, & Wee, 2018; Walton-Pattison, Dombrowski, & Pesseau, 2018).

A key concern currently limiting the expansion of this field is the lack of standardized measurement instruments across research teams for quantifying binge-watching behaviors and motivations (Erickson et al., 2019; Exelmans & Van den Bulck, 2017; Granow et al., 2018;

Riddle et al., 2017). Initial efforts were arguably inconclusive, as illustrated by a review of several preliminary measurement tools, which consisted of: 1) exploratory measurement items without proper psychometric validation (e.g., Granow et al., 2018; Panda & Pandey, 2017; Pittman & Sheehan, 2015; Shim & Kim, 2018; Shim et al., 2018); 2) pre-adaptations of existing TV scales (“TV Addiction Scale”; Horvath, 2004; “Viewing Motivation Scale”; Rubin, 1983), which were limited by their lack of direct reference to binge-watching of TV series (Riddle et al., 2017; Starosta et al., 2019; Sung et al., 2018); and 3) quantitative tools evaluating problematic binge-watching from a “confirmatory” approach (e.g., through adopting assessments of the core features of substance-use disorders (SUDs); “Problematic Series Watching Scale”, PSWS; Orosz et al., 2016; “Questionnaire of Excessive Binge-Watching Behaviors”; Starosta et al., 2019). The use of the latter in the framework of recreational activities has been subject to considerable criticism for the potential risk of over-pathologization (Billieux, Schimmenti, Khazaaal, Maurage, & Heeren, 2015; James & Tunney, 2016; Kardefelt-Winther et al., 2017; Starcevic, Billieux, & Schimmenti, 2018), particularly because applying SUD criteria to such behaviors may not appropriately discriminate between addiction and high engagement or passion (Billieux, Flayelle, Rumpf, & Stein, 2019; Kardefelt-Winther, 2015; Charlton & Danforth, 2007). Finally, to our knowledge, the only measure of TV-series-watching engagement that has arguably overcome these limitations, the “Series Watching Engagement Scale” (SWES; Tóth-Király et al., 2017), has other weaknesses. In particular, this instrument does not address the specific construct of binge-watching, but rather focuses on auxiliary and supporting factors (e.g., motivational aspects of “social interaction” and “self-development”), and facets with a relatively controversial status in the media psychology literature (e.g., “identification”, which is not empirically supported and considered by some too simplistic to report on connectedness with media; Hoffner & Buchanan, 2005; Konijn, 1999; Konijn & Hoorn, 2005; Oatley, 1994; Zillmann, 1994;

Zillmann, Hezel, & Medoff, 1980).

By contrast, recent work (Flayelle, Canale et al., 2019) on the development and validation of the “Watching TV Series Motives Questionnaire” (WTSMQ) and the “Binge-Watching Engagement and Symptoms Questionnaire” (BWESQ) constitutes a step towards meeting the need for valid and sound assessments of binge-watching-related phenomena. These questionnaires assess two constructs, whose analysis of their relationships is hypothesized to be central in early-stage investigation of binge-watching behaviors (Flayelle, Maurage et al., 2019a; Pittman & Sheehan, 2015; Rubenking & Bracken, 2018; Shim & Kim, 2018; Sung et al., 2018). First, the WTSMQ was developed to assess TV series watching motivations, which are likely key for the understanding of the development and maintenance of binge-watching behaviors (e.g., Uses-and-Gratifications and Selective Exposure theories; Katz, Blumler, & Gurevitch, 1973; Knobloch-Westerwick, 2015; Rubin, 2009; Zillmann & Bryant, 1985). Second, the BWESQ was developed to assess the type of binge-watching engagement experienced (from non-problematic to unregulated and deleterious binge-watching). In particular, this questionnaire allows dissociating high (but not unhealthy) binge-watching involvement from problematic involvement. Building upon prior qualitative focus-group research of binge-watching (Flayelle et al., 2017), both scales were disseminated in a large sample of French-speaking viewers of TV series. Exploratory and confirmatory factor analyses were undertaken in two independent subsamples to assess the structural validity of the scales. Findings indicated sound factorial designs with good psychometric properties and fits for both questionnaires (Flayelle, Canale et al., 2019). The WTSMQ involves the following four-factor model: (1) *social* (i.e., interest in bonding with others through watching TV series); (2) *emotional enhancement* (i.e., desire to watch TV series to experience intense affective states); (3) *enrichment* (i.e., interest in developing one’s intellectual experiences and knowledge through watching TV series); and (4) *coping/escapism* (i.e., desire to watch TV

series to avoid thinking about real-life problems or to cope with negative affect). The BWESQ consists of the following seven-factor model: (1) *engagement* (i.e., extent of involvement in watching TV series); (2) *positive emotions* (i.e., emotional benefits derived from watching TV series); (3) *pleasure preservation* (i.e., use of strategies aimed at maintaining or enhancing pleasure relating to watching TV series); (4) *desire/savouring* (i.e., amount of desire for and appreciation of watching TV series); (5) *binge-watching* (i.e., severity of continued viewing); (6) *dependency* (i.e., difficulty abstaining from watching TV series); and (7) *loss of control* (i.e., negative consequences associated with binge-watching). Their construct validity was reflected in shared positive relationships, as well as associations with supplementary measures of affect and problematic internet use, attesting to the discriminatory ability of the BWESQ in distinguishing high (but healthy) involvement from problematic involvement in binge-watching. Building on the strength of this psychometric validation as well as a firm anchoring in prior phenomenological knowledge of binge-watching, the WTSMQ and BWESQ therefore appear valid and reliable assessment instruments, that are particularly relevant for developing knowledge about binge-watching. On the one hand the WTSMQ may facilitate additional research into key determinants of and motives for binge-watching. On the other hand, by avoiding *a priori* consideration of binge-watching as an addictive disorder while acknowledging elevated involvement in itself, the BWESQ allows problem binge-watching research to move forward without passionate watching of TV series being inappropriately pathologized.

Nevertheless, given the widespread availability of on-demand viewing and online streaming technology (e.g., Netflix, the leading service in this area, currently reaches over 190 countries with 148 million subscribers worldwide; Netflix Media Center, 2019), the investigation of binge-watching should also consider cross-cultural factors, using measurement invariant assessment instruments to integrate and compare findings. The aim of

the current study was, therefore, to test the psychometric properties of the WTSMQ and BWESQ across nine languages (i.e., Spanish, French, English, Hungarian, Italian, German, Arabic, Persian, and Chinese) in a large international sample of viewers of TV series, and to examine their measurement equivalence according to language and gender. Additionally, drawing on the known correlates of binge-watching (i.e., diverse mental health issues, poor self-control) and the proposal that binge-watching may be problematic, relationships with relevant independent measures (e.g., self-reported happiness, psychopathological symptoms, impulsivity and problematic internet use) were investigated to assess construct validity in the nine translated versions.

Method

Participants and procedure

An online survey was disseminated mainly among a college/university student population ($N = 12,616$) across seventeen countries and nine languages: Spanish ($n = 3,312$), French ($n = 3,088$), English ($n = 2,580$), Hungarian ($n = 777$), Italian ($n = 673$), German ($n = 652$), Arabic ($n = 540$), Persian ($n = 512$), and Chinese ($n = 482$). The respondents' countries of residence for each sub-sample are shown in Table 1, and their sociodemographic characteristics are reported in Table 2. Following an identical structure across languages, the online survey successively included: (1) a short demographic questionnaire and questions about TV-series-watching behaviors (i.e., viewing frequency, average time spent watching during a typical working day/day off, number of episodes usually watched in one viewing session); (2) the "Watching TV Series Motives Questionnaire" and the "Binge-Watching Engagement and Symptoms Questionnaire" (WTSMQ and BWESQ; Flayelle, Canale et al.,

2019); (3) the “Subjective Happiness Scale” (SHS; Lyubomirsky & Lepper, 1999); (4) the “Brief Symptom Inventory” (BSI-18; Derogatis, 2001); (5) the “Short Impulsive Behavior Scale” (s-UPPS-P; Billieux et al., 2012); and (6) the “Compulsive Internet Use Scale” (CIUS; Meerkerk, Van Den Eijnden, Vermulst, & Garretsen, 2009). The original validated French versions of the WTSMQ and BWESQ were first translated into English, in accordance with the conventional translation and back-translation procedure (Beaton, Bombardier, Guillemin, & Ferraz, 2000), and all **discrepancies¹** that emerged from the comparison between the back-translated and initial French versions were deliberated (between the first and last authors of this study and the French-English translator) until optimal agreement was found. The English versions of both scales were then shared with each national coordinator who replicated the same standardized process with the help of bilingual translators on site to adapt them into the remaining languages. The majority of the additional validated questionnaires included in the survey were already available in all languages and, if not, another round of translation² was conducted by the local investigator.

All language-specific surveys were hosted on the same online platform (Qualtrics) and each national coordinator was responsible for distributing them in their respective academic environments (e.g., through advertisements during lectures, emails to students, announcements among university research participant pools and university social networks)³. Data were collected between May 2018 and January 2019. Inclusion criteria were identical to those applied in the initial validation study (Flayelle, Canale et al., 2019): being at least 18 years of age, being fluent in the targeted language and having watched TV series episodes on a regular basis or more intensively (several episodes in one session) on DVD, computers, digital platforms or streaming devices, over the last six months. Participants provided informed consent before completing the survey with an average response time of 20 minutes. **Although the online survey participation was entirely voluntary,** some study sites (Australia,

South Africa, and the United States) provided participants with incentives (course credits or prize drawing) to boost participation rates. Anonymity and confidentiality were ensured throughout the survey completion as no data allowing the identification of participants were collected (e.g., internet protocol [IP] address), with the sole exception of email addresses when incentives were put in place. In such cases, the email contact list was only used for the draw purpose or the attribution of academic credits. This study obtained approval from the Ethics Review Panel⁴ of the University of Luxembourg in addition to receiving clearance from the local Institutional Review Boards of some partner universities (those in Australia, Egypt, Hungary, South Africa, the United Kingdom, and the United States).

[INSERT TABLE 1 HERE]

[INSERT TABLE 2 HERE]

Measures

Watching TV Series Motives Questionnaire (WTSMQ)

The WTSMQ (Flayelle, Canale et al., 2019) is a 22-item scale assessing TV series watching motivations with four core dimensions: *social* (e.g., “I watch TV series to relate to others more easily, because TV series give me something to discuss.”), *emotional enhancement* (e.g., “I watch TV series to be captivated and experience extraordinary adventures by proxy.”), *enrichment* (e.g., “I watch TV series to develop my personality and broaden my views.”), and *coping/escapism* (e.g., “I watch TV series to escape reality and seek shelter in fictional worlds.”). Items are scored on a 4-point Likert scale ranging from 1 (*not at all*) to 4 (*to a great extent*), with an average score calculated for each subscale. The internal consistencies for all language-specific samples are presented in the following results section.

Binge-Watching Engagement and Symptoms Questionnaire (BWESQ)

The BWESQ (Flayelle, Canale et al., 2019) is a 40-item scale assessing binge-watching engagement and features of problematic binge-watching. The questionnaire consists of seven scales: *engagement* (e.g., “Watching TV series is one of my favorite hobbies.”), *positive emotions* (e.g., “Watching TV series is a cause for joy and enthusiasm in my life.”), *pleasure preservation* (e.g., “I worry about getting spoiled.”), *desire/savoring* (e.g., “I look forward to the moment I will be able to see a new episode of my favorite TV series.”), *binge-watching* (e.g., “When an episode comes to an end, and because I want to know what happens next, I often feel an irresistible tension that makes me push through the next episode.”), *dependency* (e.g., “I get tense, irritated or agitated when I can’t watch my favorite TV series.”), and *loss of control* (e.g., “I sometimes try not to spend as much time watching TV series, but I fail every time.”). Items are scored on a 4-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*), with an average score calculated for each subscale. The internal consistencies for all language-specific samples are presented in the following results section.

Subjective Happiness Scale (SHS)

The SHS (original English version; Lyubomirsky & Lepper, 1999) is a 4-item measure of global self-report happiness with respondents rating the extent to which they feel happy and unhappy (e.g., “In general, I consider myself a very happy person.”). Participants evaluated each item on a 7-point rating scale, a mean total score (ranging from 1 to 7) being then computed. The internal consistency of the SHS ranged from .65 (Chinese version) to .88 (German version).

Brief Symptom Inventory-18 (BSI-18)

The BSI-18 (original English version; Derogatis, 2001) assesses general psychological distress with 18 descriptions of physical and emotional complaints distributed over three facets: *depression* (e.g., “Feeling no interest in things.”), *anxiety* (e.g., “Feeling tense.”), and *somatization* (e.g., “Trouble getting breath.”). Respondents have to specify on a scale from 0 (*not at all*) to 4 (*very much*) to what extent they are troubled by such experiences. A total score is computed for each of the three subscales. The internal consistencies for all language-specific samples were high, ranging from .76 (Persian version; *somatization*) to .89 (Spanish version; *depression*).

Short Impulsive Behavior Scale (s-UPPS-P)

The s-UPPS-P (original French version; Billieux et al., 2012) is a 20-item scale evaluating five facets of impulsivity: *negative urgency* (e.g., “When I am upset I often act without thinking.”), *positive urgency* (e.g., “When I am really excited, I tend not to think on the consequences of my actions.”), *lack of premeditation* (e.g., “I usually think carefully before doing anything.” – the item is reverse scored), *lack of perseverance* (e.g., “I generally like to see things through to the end.”), and *sensation-seeking* (e.g., “I sometimes like doing things that are a bit frightening.”). Items are scored on a 4-point Likert scale ranging from 1 (*strongly agree*) to 4 (*strongly disagree*). A total score is calculated for each of the five subscales. The internal consistencies of the s-UPPS-P subscales ranged from .60 (German version; *positive urgency*) to .92 (Italian version; *lack of perseverance*).

Compulsive Internet Use Scale (CIUS)

The CIUS (original English version; Meerkerk et al., 2009) is a 14-item scale assessing problematic internet use on five scales: *loss of control* (e.g., “Do you find it difficult

to stop using the internet when you are online?”), *preoccupation* (e.g., “Do you think about the internet, even when not online?”), *withdrawal symptoms* (e.g., “Do you feel restless, frustrated, or irritated when you cannot use the internet?”), *coping or mood modification* (e.g., “Do you go on the internet when you are feeling down?”), and *conflict* [e.g., “Do you neglect your daily obligations (work, school, or family life) because you prefer to go on the internet?”]. Items are scored on a 5-point scale ranging from 0 (*never*) to 4 (*very often*), and are summed to yield a total single score. Internal consistencies were high across all language-specific samples, ranging between .86 (Arabic version) and .93 (Spanish version).

Statistical analyses

For data analyses, only full sets of responses⁵ were explored, explaining sample size variations within the same language-based sample. In a first step, descriptive statistics concerning sociodemographic characteristics and TV series viewing patterns were computed to compile a profile of the whole and individual samples using SPSS statistical package (version 24.0). Confirmatory factor analyses (CFAs) were then conducted for each language-specific sample, as well as for the overall sample to examine the adequacy of fit of the 4-factor and 7-factor models derived from the initial WTSMQ and BWESQ validation (Flayelle, Canale et al., 2019). The software used to perform these analyses was EQS 6.4 (Bentler, 2006). Non-normal distributions of items from the WTSMQ and BWESQ scales (see Supplemental Table 1 available from: <https://osf.io/pxzw8/>) were addressed by applying robust estimation methods (robust Maximum Likelihood, ML; Finney & DiStefano, 2013). In line with best practice in Structural Equation Modeling (Kline, 2015; Hooper, Coughlan, & Mullen, 2008) to respect original factorial integrity of both scales and to ensure the comparability between countries, we did not apply any modification to the models based on

modification indices, even when minor changes (e.g., correlations between error terms) significantly increased the models' fit. Goodness of fit for the CFA models was assessed through the following indices: the root mean square error of approximation (RMSEA), the comparative and incremental fit indices (CFI and IFI, respectively), and the standardized root mean square residual (SRMR). An excellent model fit was identified when the CFI and the IFI were $\geq .95$, the RMSEA $\leq .05$, and the SRMR $\leq .05$ (Bagozzi & Yi, 2011; Schermelleh-Engel & Müller, 2003). Using less restrictive criteria, values $\geq .90$ for the CFI and the IFI, $\leq .08$ for the RMSEA, and $\leq .10$ for the SRMR were considered acceptable (Hooper et al., 2008). For the sake of transparency, Satorra-Bentler chi-square (X^2), general model significance (p), and relative chi-square (X^2/df) were reported; however, given that X^2 is highly sensitive to sample size (Jöreskog & Sörbom, 1993; Markland, 2007), which in our study far exceeds the standards required for conducting this type of analysis (Hair, Black, & Babin, 2010), these indices were not employed to assess the adequacy of the CFA models.

To assess whether the factor structures of the WTSMQ and BWESQ were valid for their use across different languages and in both genders⁶, multi-group CFAs according to language and gender were conducted. Specifically, we tested four levels of measurement invariance: 1) configural (test whether items load on the same factor across groups), 2) metric (test whether item factorial loadings are equal across groups), 3) scalar (test whether item intercepts are equal across groups) and 4) error variance invariance (test whether items measurement error are equal across groups). The adequacy of the increasingly constrained models was assessed through the difference between pairs of nested models (Δ) in the RMSEA, CFI and SRMR. A change $\geq .01$ in the CFI, $\geq .015$ in the RMSEA, and $\geq .03$ in the SRMR indicates a significant decrease in the model fit when testing for measurement invariance (Chen, 2007). This procedure was also used to assess the adequacy of merging into a single dataset the data obtained in different countries for the same language (these results

can be found in Supplemental Tables 2 and 3 at: <https://osf.io/pxzw8/>), a procedure that was performed before conducting the individual CFAs in each language-based dataset.

Reliability of the WTSMQ and BWESQ total scores and factors was assessed through the ordinal Cronbach's alpha (α) and the McDonald's omega (ω). Both indices were calculated using the R package “userfriendlyscience” (Peters, 2014). According to the criteria proposed by Hunsley and Mash (2008), reliability indices between .70 and .79 were considered appropriate, between .80 and .89 good, and \geq .90 excellent. Finally, the construct validity of the WTSMQ and the BWESQ was appraised by investigating their relationships with age and SHS, BSI-18, s-UPPS-P and CIUS scores across all samples by means of Spearman's correlational analyses⁷, while Pearson point-biserial correlations were used to explore links with gender⁸. To account for multiple comparisons, the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) was also performed to hold the false discovery rate at 5%.

Results

Descriptive statistics

TV-series-watching characteristics and average scores for all questionnaire study variables are reported in Table 3.

[INSERT TABLE 3 HERE]

Watching TV Series Motives Questionnaire (WTSMQ)

Structural analysis and measurement invariance across language and gender

The adequacy of the four-factor model from the preliminary WTSMQ validation was tested through CFA. This model proposes that the 22 items comprising this scale may be grouped into four correlated first-order factors (for a comprehensive description of the factorial structure and items distribution, see Flayelle, Canale et al., 2019). Given the confirmatory nature of this study, other competing models were not tested (e.g., unifactorial models, second-order factors). Results from individual CFAs for each language and across all samples are reported in Table 4. As expected, given the datasets' sample sizes, the Satorra-Bentler X^2 value of significance did not exceed the .05 value to consider the models' fit as satisfactory. In addition, the CFI and IFI were consistently under the .90 threshold in all the assessed models, except for the Arabic sample and the whole dataset, in which both indices were near an acceptable value (.89). As for the X^2 , CFI and IFI are sensitive to sample size (Rigdon, 1996), as well as to the item response scale (in particular, ordered categorical answer scales; Finney & DiStefano, 2013). As a result, Rigdon (1996) advised that the CFI is better suited to assess the adequacy of exploratory research designs (i.e., studies comprising small sample sizes) whereas alternative indices such as the RMSEA are better suited to confirmatory contexts (i.e., studies comprising large samples). Furthermore, Kenny and McCoach (2003) argue that the CFI tends to deteriorate in models comprising a large number of variables and indicators, especially for correctly specified models (note that the models described in this paper for the WTSMQ and BWESQ comprise 203 and 719 *df* respectively). In contrast, the RMSEA consistently demonstrates an opposite pattern: i.e., a systematic decrease in models comprising an increasing number of variables (Kenny & McCoach, 2003). Given these limitations, we analysed the goodness of fit of our CFA models by relying on the recommendation made by Kenny and McCoach (2003), who suggest that complex models involving lower Tucker-Lewis index (TLI) and CFI values give no real cause for concern

insofar as the RMSEA seems better. In our CFA models, the RMSEA and the SRMR were below the thresholds of .08 and .10 in all the language-based datasets as well as in the whole sample. The best adjustment according to these indices was obtained for the whole sample (RMSEA = .060; SRMR = .051) whereas the worst was obtained for the Persian dataset (RMSEA and SRMR of .079).

[INSERT TABLE 4 HERE]

To test measurement invariance of the WTSMQ according to language and gender, we conducted a series of multi-group CFAs. As displayed in Table 5, language and gender configural invariance of the WTSQM was supported (RMSEA = .065; SRMR = .067 [according to language]; RMSEA = .060; SRMR = .051 [according to gender]), so we subsequently estimated models with increasing levels of constraints to test higher levels of invariance. Regarding metric invariance, changes in the RMSEA and SRMR did not show a significant worsening in the model fit neither for language (Δ RMSEA = .001; Δ SRMR = .010) nor for gender invariance (Δ RMSEA = .001; Δ SRMR = .005). Similarly, the models' fit did not significantly decrease when subsequent levels of gender invariance were tested (Δ in RMSEA and SRMR were always below .015 and .03, respectively), thus supporting a complete equivalence of the WTSMQ in males and females. However, the significant Δ in SRMR when scalar and error invariance according to language was tested (.117 and .116) suggested the presence of differences at these levels of measurement according to the language of administration.

For language (not for gender) invariance, values for the Δ in CFI exceeded the threshold of .015 (Δ CFI of .017, .012, and .022 for metric, scalar and error invariance).

However, following the same approach as individual CFAs, this CFI-based index was not considered to assess the adequacy of the invariance models.

[INSERT TABLE 5 HERE]

Internal consistency

Reliability indices for the WTSMQ total score and factors are displayed in Table 6. Few differences between ordinal Cronbach's alpha (α) and McDonald's omega (ω) were observed. Convergence between both indices was considered as a good indicator of scale reliability under different conditions (Zinbarg, Revelle, Yovel, & Li, 2005). For the whole sample as well as for the majority of the different language-based samples, both indices clearly exceed the criterion of .70 established by Hunsley and Mash (2008) to consider the reliability of a scale appropriate. The only exception was found in the Chinese dataset, where reliability for factor 4 was below .70 (α and ω of .60). Reliability for the other language-based datasets and for the whole sample ranged between .71-.92 and .82-.90 respectively, with most values indicating good to excellent scale reliability. Thus, the WTSMQ can be considered a reliable measure in each language-based sample.

[INSERT TABLE 6 HERE]

Binge-Watching Engagement and Symptoms Questionnaire (BWESQ)

Structural analysis and measurement invariance across language and gender

The adequacy of the seven-factor model from the preliminary BWESQ validation was tested through CFA (following a similar data-analytic approach to the one used for the

WTSMQ). This model proposes that the 40 items comprising this scale may be grouped into seven correlated first-order factors. As displayed in Table 4, goodness of fit indices for the BWESQ individual CFAs were acceptable for all the language-based dataset (RMSEA ranging between .056-.062 and SRMR ranging between .057-.074) and in the whole sample (RMSEA = .059; SRMR = .063). Consistent with our expectations that the low CFI and IFI values were linked to the degree of complexity of our CFA models (in terms of number of indicators and latent variables) and not to a truly poor fitting factorial structure, we observed a significant decrease of these indices in the results for this scale (note that the BWESQ has 516 *df* more than previously); conversely, results for the RMSEA are slightly better (the tendency documented by Kenny and McCoach in increasingly complex models; Kenny & McCoach, 2003).

Results from measurement invariance of the BWESQ across languages and gender are displayed in Table 5. Results are notably similar to those reported for the WTSMQ. Configural invariance according to language (RMSEA = .058; SRMR = .067) and gender (RMSEA = .059; SRMR = .063) was confirmed during the first step of the multi-group CFAs. The small changes in the fit indices at the next steps also supported metric invariance according to language (Δ RMSEA < .000; Δ SRMR = .012) and gender (Δ RMSEA = .001; Δ SRMR = .006). Furthermore, the increase in the level of measurement constraints at the subsequent steps did not result in a significant deterioration of the models' fit (Δ RMSEA = .001; Δ SRMR < .000 [scalar invariance]; Δ RMSEA = .001; Δ SRMR = .006 [error invariance]) across gender groups, providing strong evidence that the BWESQ operates similarly in males and females. However, scalar invariance according to language was only partially supported (Δ RMSEA = .007 and Δ SRMR = .031; i.e., extremely near to .03 threshold) and error variance invariance rejected (Δ SRMR = .037). Even when Δ in CFI was not considered to assess the adequacy of multi-group models, all the values except for the

language error variance invariance ($\Delta CFI = .011$) were below .01, thus supporting different levels of measurement equivalence between the language versions of the BWESQ and in both genders.

Internal consistency

Reliability indices for the BWESQ total score and factors are displayed in Table 6. Once again, few differences between ordinal Cronbach's alpha (α) and McDonald's omega (ω) were observed, and the majority of reliability values were good to excellent (even better than for the WTSMQ). Apart from the Cronbach's alpha from factor 7 in the Chinese dataset ($\alpha = .68$; $\omega = .71$) and from factor 5 in the German dataset ($\alpha = .67$; $\omega = .71$), reliability was always above .70. In particular, reliability for the rest of the language-based datasets and for the whole sample ranged between .72-.97 and .75-.96 respectively, once again with a clear preponderance of values indicating excellent scale reliability. As a result, the BWESQ can be considered a reliable measure for each language-based sample, even more reliable than the WTSMQ (which might be due to the higher number of items comprising each scale as well as the whole scale).

Scale inter-correlations and convergent validity

The correlation ranges obtained among all samples between the WTSMQ and BWESQ with one another, and between each of them with additional measures (i.e., age, gender, and scores on the SHS, BSI-18, s-UPPS-P and CIUS) are reported in Tables 7, 8 and 9. The comprehensive review of language-specific correlations together with the nine language-versions of the WTSMQ and BWESQ can be found at: <https://osf.io/pxzw8/>.

On the whole, positive relationships emerged in all samples between the various subscales of the WTSMQ and BWESQ. In this regard, the *emotional enhancement* and *coping-escapism* motivations systematically encompassed the largest associations with all BWESQ-related dimensions, with non-problematic binge-watching factors (i.e., *engagement*, *positive emotions*, *pleasure preservation*, *desire/savoring*) being more strongly related to *emotional enhancement*, whereas problematic-binge-watching-related facets (i.e., *dependency*, *loss of control*) were more strongly connected to *coping-escapism*.

As for external correlates, although exhibiting a small effect size (Cohen, 1988), what particularly stands out across all languages is a stronger positive association between gender and the *coping/escapism* motivation. *Coping/escapism* also consistently presented the strongest small to moderate negative relationships with happiness (i.e., SHS total score), and a similar relationship was observed with *dependency* in the BWESQ. Similarly, all the BSI-18 domains (i.e., *depression*, *anxiety*, *somatization*) displayed more pronounced small to medium relationships with *coping/escapism* and *dependency*, followed by *binge-watching* and *loss of control*. In all samples, although small in magnitude, the association between impulsivity and motivations for viewing TV series was higher for *coping/escapism* with *negative urgency*, *positive urgency*, *lack of premeditation* and *lack of perseverance*, whereas *sensation-seeking* was more related to the *enrichment* motive. Among the BWESQ-related domains, the s-UPPS-P subscales' scores were repeatedly associated to a greater extent (small to medium effects) with problematic binge-watching factors (i.e., *binge-watching*, *dependency*, *loss of control*), with *negative urgency* and *sensation-seeking* being more specifically connected to *dependency*, *positive urgency* to *binge-watching*, and both *lack of premeditation* and *lack of perseverance* to *loss of control*. Finally, and concurrent with the afore-mentioned relationships, the CIUS total score was in all instances more strongly related to problematic

binge-watching factors (i.e., *binge-watching*, *dependency*, *loss of control*), as well as to the *coping/escapism* motivation, involving mainly moderate to large positive associations.

[INSERT TABLE 7 HERE]

[INSERT TABLE 8 HERE]

[INSERT TABLE 9 HERE]

Discussion

The present study investigated the psychometric properties of the “Watching TV Series Motives Questionnaire” (WTSMQ) and the “Binge-Watching Engagement and Symptoms Questionnaire” (BWESQ), two recently developed quantitative instruments measuring TV series watching motivations and binge-watching engagement and symptoms, among nine language-specific samples (i.e., Spanish, French, English, Hungarian, Italian, German, Arabic, Persian, and Chinese) in 17 countries.

This work is particularly relevant in the context of the rapidly growing body of research on binge-watching worldwide, where the provision of valid and reliable instruments that perform well across different languages has become a central requirement to ensure accurate and meaningful comparisons of findings across studies. From this perspective, the goodness of fit of each measurement model was tested in all languages by means of individual CFAs, followed by the examination of the language and gender factor equivalence of both instruments using multi-group CFAs. Finally, the construct validity of the nine language-versions of the WTSMQ and BWESQ was considered through the correlational

patterns identified with additional measures of happiness, psychopathological symptoms, impulsivity and problematic internet use.

Consistent with the initial validation study (Flayelle, Canale et al., 2019), the factorial structures of both scales were found to replicate appropriate adjustments across all languages in the light of the fit indices (e.g., RMSEA, SRMR) considered better suited in view of our confirmatory framework and the complexity of the assessed models (Kenny & McCoach, 2003; Rigdon, 1996). As such, the theoretical factor models underlying these two instruments hold across languages/cultures represented in this study. Additionally, overall measurement invariance according to language and gender was supported for both, thus implying that, whichever the language spoken, male and female TV series viewers interpreted the WTSMQ and BWESQ items in a conceptually similar manner. Beyond indicating their validity for use across the nine languages at hand, in both genders, this statistical property ensures that potential comparisons of results based on these quantitative tools express genuine differences in the constructs being measured. Finally, as further evidence of their high reliability, both scales were consistently characterized by good to excellent internal consistency, sharing very close coefficients' values from (language) version to version. Backed by the present evidence of their good psychometric properties, both the WTSMQ and BWESQ thus prove to be reliable invariant measures in the nine different languages investigated.

The construct validity of all translated versions of the WTSMQ and BWESQ was supported by the nature of their relationships with each other, as well as with extra measures, showing similar patterns of associations across the different language-versions of the scales. Importantly, the BWESQ domains considered as non-problematic (i.e., *engagement, positive emotions, pleasure preservation, desire/savoring*) consistently displayed stronger connections to the *emotional enhancement* motivation as assessed by the WTSMQ. This seems consistent with evidence that the main reason many individuals binge-watch is simply because this is

entertaining (Panda & Pandey, 2017; Pittman & Sheehan, 2015; Ramayan et al., 2018; Shao & Beneza, 2018; Sung et al., 2018); this motive, in turn, most typically promotes pursuit of leisure activities.

In contrast, the *coping-escapism* factor of the WTSMQ showed stronger links to the BWESQ domains, which are considered to reflect problematic binge-watching (i.e., *dependency, loss of control*), just as in their initial validation. This not only resonates with recent findings highlighting the incentive role played by escapism motivation in binge-watching behaviors (Panda & Pandey, 2017; Rubenking et al., 2018; Starosta et al., 2019), but also may relate to relationships to problematic involvement in recreational behaviors that are often implemented to face adverse emotional states (e.g., problematic internet use or gaming; Ballabio et al., 2017; Bowditch, Chapman & Naweed, 2018; Kardefelt-Winther, 2014; Tang et al., 2014; Whang, Lee, & Chang, 2003; Yee, 2007). In this respect, it is worth noting the stronger association identified across samples between *coping/escapism* and being female, which is somewhat reminiscent of the higher rates of depression in women (Albert, 2015; Cyranowski, Frank, Young, & Shear, 2000; Nolen-Hoeksema, 1990). Furthermore, other potentially addictive behaviors (e.g., gambling) are more strongly related to negative reinforcement motivations in females as compared to males (Zakariaeiz & Potenza, 2018). The current findings therefore suggest problematic binge-watching may involve maladaptive coping or emotion-regulation strategies, as in other potentially addictive behaviors (Flayelle, Maurage et al., 2019a, 2019b; Rubenking & Bracken, 2018; Tukachinsky & Eyal, 2018).

Finally, the reciprocal stronger positive relationships that systematically were observed between *coping/escapism* and problematic binge-watching factors (i.e., *binge-watching, dependency, loss of control*) on the one hand, and self-reported unhappiness, psychopathological symptoms (i.e., *depression, anxiety, somatization*), impulsivity domains and problematic internet use on the other, are further suggestive of the construct validity of

the nine language-versions of the WTSMQ and BWESQ, and highlight important clinical relationships across cultures. These findings are in accordance with previous studies reporting associations between binge-watching and depression (Ahmed, 2017; Sung et al., 2015; Tukachinsky & Eyal, 2018), anxiety (Kruger et al., 2015b; Sung et al., 2015; Tefertiller & Maxwell, 2018), and heightened impulsivity (Flayelle, Maurage et al., 2019b; Riddle et al., 2017). Therefore, beyond supporting the construct validity of both scales, such patterns of correlations, that are seen across all samples, suggest the potential ability of the BWESQ to distinguish problematic from elevated but non-harmful binge-watching in each of its translations.

This unique feature of the BWESQ instrument thus represents an important added value to the assessment of binge-watching behaviors, given the relevance of discriminating between high and problematic engagement for establishing “disordered” use of technology (Billieux et al., 2019; Brockmeyer et al., 2009; Charlton & Danforth, 2007, 2010; Deleuze, Long, Liu, Maurage, & Billieux, 2018; Gentile, Coyne, & Bricolo, 2013). Such a notion applied to the context of TV series watching resonates with recent work drawing on the *Dualistic Model of Passion* (Vallerand, 2015; Vallerand et al., 2003), which has emphasized that harmonious passion (i.e., significant involvement performed in harmony with other aspects of one’s life) is especially related to adaptive correlates of TV series watching, while obsessive passion (i.e., excessive involvement that generates conflict with other activities) is more specifically linked to maladaptive ones (Orosz, Vallerand, Bőthe, Tóth-Király, & Paskuj, 2016; Tóth-Király et al., 2019). Taken together, the current results emphasize the reliability and validity of the WTSMQ and BWESQ measurement instruments over the nine languages, and provide evidence of their utility for future cross-cultural research on problematic binge-watching that is able to avoid pathologizing such a popular leisure activity.

Several limitations should be underlined. First, from a methodological standpoint, the means employed to collect data varied between sites (notably with some relying on the use of incentives), thereby generating certain gaps in the local sample sizes obtained. Still, no major differences exist as for the models' goodness of fit between the samples where incentives were offered or not. Second, as the data are cross-sectional and self-reported, biases related to social desirability, lack of introspection or memory recall might be present, potentially reducing their temporal and ecological validity. Third, some Cronbach's alpha values for the WTSMQ/BWESQ subscales were slightly below the recommended threshold of .70 (Hunsley & Mash, 2008) in their Chinese and German language-versions, while McDonald's omega values were considered appropriate (with the exception of factor 4 of the Chinese version of the WTSMQ). This may reflect methodological issues (e.g., language adaptation of the scales). Fourth, one weakness of the WTSMQ and BWESQ psychometric structures across all languages is that CFI and IFI values were also systematically below the optimal recommended thresholds (Hooper et al., 2008). These particular indices were, however, not the most suitable to evaluate the appropriateness of the currently assessed models. Finally, our sampling of mainly university students may limit the generalizability of the results. Future studies aimed at continuing the assessment effort of the cross-cultural psychometric validity of both quantitative instruments should therefore be undertaken in other populations.

Conclusion

Overall, the cumulative positive results of this study confirm the cross-cultural robustness of the WTSMQ and BWESQ assessment instruments examined across nine languages in a multinational sample of 12,616 TV-series viewers from Africa, Asia, Europe, the Middle East, North America, Oceania, and South America. The study not only

demonstrated the psychometric validity of the instruments across widely distributed geographic locations, but also provided evidence of similar patterns of relationships between motivational and behavioral aspects of binge-watching and negative health measures, suggesting that common features may be linked to problematic binge-watching across cultures. At a time when binge-watching is a popular activity warranting research across jurisdictions, valid measures enabling comparability of data are key to promote an understanding of binge-watching across cultures. The WTSMQ and BWESQ will allow the further examination of binge-watching and the underlying motivations, helping to ensure the integrity and coherence of such research.

¹ 11% of both WTSMQ and BWESQ items were concerned.

² These additional translations concerned the SHS (Hungarian, Persian), BSI-18 (Arabic, Chinese, Persian) and CIUS (Chinese, Hungarian, Persian).

³ Note that the study was also advertised in the popular press in France.

⁴ Project identification code: ERP 18-008.

⁵ A total number of 14,672 respondents started to fill in the questionnaires, with 73% of them completing the entire survey.

⁶ Given the very low prevalence of participants having reported “transgender” and “other” about their gender identity, only male and female data were considered in such analyses.

⁷ Spearman’s correlations were used to address non-normal distribution of data.

⁸ In line with the above-mentioned reason, only two categories of data (i.e., male and female) were included in the correlational analyses.

References

- Adachi, P.J.C., Ryan, R.M., Frye, J., McClurg, D., & Rigby, C.S. (2017). "I can't wait for the next episode!": Investigating the motivational pull of television dramas through the lens of Self-Determination Theory. *Motivation Science*, 4, 78–94. doi:10.1037/mot0000063
- Ahmed, A. (2017). New era of TV-watching behavior: Binge-watching and its psychological effects. *Media Watch*, 8, 192–207. doi:10.15655/mw/2017/v8i2/49006
- Albert, P.R. (2015). Why is depression more prevalent in women? *Journal of Psychiatry & Neuroscience*, 40, 219–221. doi:10.1503/jpn.150205
- Bagozzi, R. P., & Yi, Y. (2011). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40, 8–34. doi:10.1007/s11747-011-0278-x
- Ballabio, M., Griffiths, M. D., Urbán, R., Quattiroli, A., Demetrovics, Z., & Király, O. (2017). Do gaming motives mediate between psychiatric symptoms and problematic gaming? An empirical survey study. *Addiction Research & Theory*, 25, 397–408. doi:10.1080/16066359.2017.1305360
- Beaton, D.E., Bombardier, C., Guillemin, F., & Ferraz, M.B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25, 3186–3191. doi:10.1097/00007632-200012150-00014
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57, 289–300. doi:10.2307/2346101

- Bentler, P. M. (2006). EQS structural equations program manual. Encino, CA: Multivariate Software.
- Billieux, J., Flayelle, M., Rumpf, H. J., & Stein, D. (2019). High involvement versus pathological involvement in video games: A crucial distinction for ensuring the validity and utility of gaming disorder. *Current Addiction Reports*, 6, 323–330. doi:10.1007/s40429-019-00259-x
- Billieux, J., Rochat, L., Ceschi, G., Carré, A., Offerlin-Meyer, I., Defeldre, A.C., Khazaal, Y., Besche-Richard, C., & Van der Linden, M. (2012). Validation of a short French version of the UPPS-P Impulsive Behavior Scale. *Comprehensive Psychiatry*, 53, 609–615. doi:10.1016/j.comppsy.2011.09.001
- Billieux, J., Schimmenti, A., Khazaal, Y., Maurage, P., & Heeren, A. (2015). Are We Overpathologizing Everyday Life? A Tenable Blueprint for Behavioral Addiction Research. *Journal of Behavioral Addictions*, 4, 119–123. doi:10.1556/2006.4.2015.009
- Bowditch, L., Chapman, J., & Naweed, A. (2018). Do coping strategies moderate the relationship between escapism and negative gaming outcomes in World of Warcraft (MMORPG) players? *Computers in Human Behavior*, 86, 69–76. doi:10.1016/j.chb.2018.04.030
- Brockmeyer, J.F., Fox, C.M., Curtiss, K.A., McBroom, E., Burkart, K.M., & Pidruzni, J.N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game playing. *Journal of Experimental Social Psychology*, 45, 624–634. doi:10.1016/j.jesp.2009.02.016

- Brookes, S., & Ellithorpe, M. (2017). *Good for your mood, bad for your health: Narrative involvement, health behaviors, and binge watching*. Paper presented at the 67th ICA Annual Conference, San Diego, CA.
- Charlton, J.P., & Danforth, I.D.W. (2007). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior*, 23, 1531–1548. doi:10.1016/j.chb.2005.07.002
- Charlton, J.P., & Danforth, I.D.W. (2010). Validating the distinction between computer addiction and engagement: Online game playing and personality. *Behaviour & Information Technology*, 29, 601–613. doi:10.1080/01449290903401978
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504. doi:10.1080/10705510701301834
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum, Lawrence, Associates.
- Cyranowski, J.M., Frank, E., Young, E., & Shear, M.K. (2000). Adolescent onset of the gender difference in lifetime rates of major depression: A theoretical model. *Archives of General Psychiatry*, 57, 21–27. doi:10.1001/archpsyc.57.1.21
- De Feijter, D., Khan, J.V., & Van Gisbergen, M.S. (2016). Confessions of a 'guilty' couch potato: Understanding and using context to optimize binge-watching behavior. *TVX '16 Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, Chicago, Il: ACM, 2016. doi:10.1145/2932206.2932216
- Deleuze, J., Long, J., Liu, T. Q., Maurage, P., & Billieux, J. (2018). Passion or addiction?

Correlates of healthy versus problematic use of videogames in a sample of French-speaking regular players. *Addictive Behaviors*, 82, 114–121. doi:10.1016/j.addbeh.2018.02.031

Deloitte’s digital media trends survey, 12th edition (2018). “A new world of choice for digital consumers”. Retrieved from https://www2.deloitte.com/content/dam/insights/us/articles/4479_Digital-media-trends/4479_Digital_media%20trends_Exec%20Sum_vFINAL.pdf

Deloitte’s digital media trends survey, 13th edition (2019). “Piecing it together”. Retrieved from https://www2.deloitte.com/content/dam/insights/us/articles/4782_digital-media-trends-13th-edition/DI_Digital-media-trends-13th-edition.pdf

Derogatis, L. R. (2001). Brief Symptom Inventory (BSI)-18: Administration, scoring and procedures manual. Minneapolis, MN: NCS Pearson

Erickson, S. E., Dal Cin, S., & Byl, H. (2019). An experimental examination of binge watching and narrative engagement. *Social Sciences*, 8, 19. doi:10.3390/socsci8010019

Exelmans, L., & Van den Bulck, J. (2017). Binge viewing, sleep, and the role of pre-sleep arousal. *Journal of Clinical Sleep Medicine*, 13, 1001–1008. doi:10.5664/jcsm.6704

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation models. In *Structural Equation Modeling: A Second Course* (2nd Editio, p. 703). Charlotte, NC: Information Age Publishing.

Flayelle, M., Canale, N., Vögele, C., Karila, L., Maurage, P., & Billieux, J. (2019). Assessing binge-watching behaviors: Development and validation of the “Watching TV Series

- Motives” and “Binge-Watching Engagement and Symptoms” Questionnaires. *Computers in Human Behavior*, 90, 26–36. doi:10.1016/j.chb.2018.08.022
- Flayelle, M., Maurage, P., & Billieux, J. (2017). Toward a qualitative understanding of binge-watching behaviors: A focus group approach. *Journal of Behavioral Addictions*, 6, 457–471. doi:10.1556/2006.6.2017.060
- Flayelle, M., Maurage, P., Karila, L., Vögele, C., & Billieux, J. (2019b). Overcoming the unitary exploration of binge-watching: A cluster analytical approach. *Journal of Behavioral Addictions*, 8, 586–602. doi:10.1556/2006.8.2019.53
- Flayelle, M., Maurage, P., Ridell Di Lorenzo, K., Vögele, C., Gainsbury, S.M., & Billieux, J. (2020). Binge-watching: What do we know so far? A first systematic review of the evidence. *Current Addiction Reports*, 7, 44–60. doi:10.1007/s40429-020-00299-8
- Flayelle, M., Maurage, P., Vögele, C., Karila, L., Billieux, J. (2019a). Time for a plot twist: Beyond confirmatory approaches to binge-watching research. *Psychology of Popular Media Culture*, 8, 308–318. doi:10.1037/ppm0000187
- Gentile, D. A., Coyne, S. M., & Bricolo, F. (2013). Pathological technology addictions: What is scientifically known and what remains to be learned. In K. E. Dill (Ed.). *The Oxford handbook of media psychology* (pp. 382–402). Oxford: Oxford University Press.
- Granow, V., Reinecke, L., & Ziegele, M. (2018). Binge-watching & psychological well-being: Media use between lack of control and perceived autonomy. *Communication Research Reports*, 35, 392–401. doi:10.1080/08824096.2018.1525347
- Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate Data Analysis: A Global Perspective*. Boston: Pearson.

- Hasan, R., Kumar Jha, A., & Liu, Y. (2018). Excessive use of online video streaming services: Impact of recommender system use, psychological factors, and motives. *Computers in Human Behavior*, 80, 220–228. doi:10.1016/j.chb.2017.11.020
- Hernández Pérez, J.F., & Martínez Díaz, M.A. (2016). Nuevos modelos de consumo audiovisual: los efectos del binge-watching sobre los jóvenes universitarios [New forms of audiovisual consumption: Binge watching effects on university students] En *adComunica. Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*, no. 13. Castellón: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, 201–221. doi:10.6035/21740992.2017.13.11
- Hoffner, C., & Buchanan, N. (2005). Young adult's wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, 7, 325–351. doi:10.1207/S1532785XMEP0704_2
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modeling: Guidelines for Determining Model Fit. *Electronic Journal on Business Research Methods*, 6, 53–60.
- Horvath, C. W. (2004). Measuring television addiction. *Journal of Broadcasting and Electronic Media*, 48, 378–398. doi:10.1207/s15506878jobem4803_3
- Hunsley, J., & Mash, E. J. (2008). A guide to assessments that work. Oxford University Press.
- IBM Corp. Released (2015). IBM SPSS statistics for windows, version 24.0. Armonk, NY: IBM Corp.
- James, R.J.E., & Tunney, R.J. (2016). The need for a behavioural analysis of behavioural addictions. *Clinical Psychology Review*, 52, 69–76. doi:10.1016/j.cpr.2016.11.010

Jöreskog, K., & Sörbom, D. (1993). LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Chicago, IL: Scientific Software International Inc.

Kardefelt-Winther, D., Heeren, A., Schimmenti, A., van Rooij, A., Maurage, P., Carras, M., ... Billieux, J. (2017). How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction*, *112*, 1709–1715. doi:10.1111/add.13763

Kardefelt-Winther D. (2015). Problems with atheoretical and confirmatory research approaches in the study of behavioral addictions. *Journal of Behavioral Addictions*, *4*, 126–129. doi:10.1556/2006.4.2015.019

Kardefelt-Winther, D. (2014). A conceptual and methodological critique of internet addiction research: Towards a model of compensatory internet use. *Computers in Human Behavior*, *31*, 351–354. <https://doi.org/10.1016/j.chb.2013.10.059>.

Katz, E., Blumler, J.G., & Gurevitch, M. (1973). Uses and gratifications research. *Public Opinion Quarterly*, *37*, 509–523. doi:10.1086/268109

Kenny, D. A., & McCoach, D. B. (2003). Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*, 333–351. doi:10.1207/S15328007SEM1003_1

Kline, R. B. (2015). Principles and Practice of Structural Equation Modeling (4th edition). New York: Guilford University Press.

Knobloch-Westerwick, S. (2015). *Choice and preference in media use*. New York: Routledge.

Konijn, E. A. (1999). Spotlight on spectators: Emotions in the theater. *Discourse Processes*,

28, 169–194. doi:10.1080/01638539909545079

Konijn, E. A., & Hoorn, J. F. (2005). Some like it bad. *Media Psychology*, 7, 107–144. doi:10.1207/S1532785XMEP0702_1

Kruger, J. S., Karmakar, M., Elhai, J., & Kramer, A. (2015a). Screening for sleep problems: Binge watching in the internet era and its relationship to sleep habits. 143rd APHA Annual Meeting and Exposition, Chicago, Illinois.

Kruger, J. S., Karmakar, M., Elhai, J., & Kramer, A. (2015b). Looking into screen time: Mental health and binge watching. 143rd APHA Annual Meeting and Exposition, Chicago, Illinois.

Kubota, Y., Cushman, M., Zakai, N., Rosamond, W.D., & Folsom, A.R. (2018). TV viewing and incident venous thromboembolism: The atherosclerotic risk in communities study. *Journal of Thrombosis and Thrombolysis*, 45, 353–359. doi:10.1007/s11239-018-1620-7

Lyubomirsky, S., & Lepper, H. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, 46, 137–155. doi:10.1023/A:1006824100041

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling. *Personality and Individual Differences*, 42, 851–858. doi:10.1016/j.paid.2006.09.023

Matrix, S. (2014). The Netflix effect: Teens, binge watching, and on-demand digital media trends. *Jeunesse: Young People, Texts, Cultures*, 6, 119–138. doi:10.1353/jeu.2014.0002

- Meerkerk, G.J., Van Den Eijnden, R.J., Vermulst, A.A., & Garretsen, H.F. (2009). The compulsive Internet use scale (CIUS): Some psychometric properties. *CyberPsychology & Behavior*, 12, 1–6. doi: 10.1089/cpb.2008.0181
- Merikivi, J., Bragge, J., Scornavacca, E., & Verhagen, T. (2019). Binge-watching serialized video content: A transdisciplinary review. *Television & New Media*, online first publication. doi: 10.1177/1527476419848578
- Merrill, K., & Rubenking, B. (2019). Go long or go often: Influences on binge-watching frequency and duration among college students. *Social Sciences*, 8, 10. doi: 10.3390/socsci8010010
- Mikos, L. (2016). Digital media platforms and the use of TV content: Binge watching and video-on-demand in Germany. *Media and Communication*, 4, 154–161. doi:10.17645/mac.v4i3.542
- Morris, J.S., Bradbury, K.E., Cross, A.J., Gunter, M.J., & Murphy, N. (2018). Physical activity, sedentary behaviour and colorectal cancer risk in the UK Biobank. *British Journal of Cancer*, 118, 920–929. doi: 10.1038/bjc.2017.496
- Netflix Media Center (2019), “About Netflix.” Retrieved from <https://media.netflix.com/en/about-netflix>
- Nolen-Hoeksema, S. (1990). Sex differences in depression. Stanford University Press; Stanford, CA.
- Oatley, K. (1994). A taxonomy of the emotions of literary response and a theory of identification in fictional narrative. *Poetics*, 23, 53–74. doi:0.1016/0304-422X(94)P4296-S

- Orosz, G., Bőthe, B., & Tóth-Király, I. (2016). The development of the Problematic Series Watching Scale (PSWS). *Journal of Behavioral Addictions*, 5, 144–150. doi:10.1556/2006.5.2016.011
- Orosz, G., Vallerand, R.J., Bőthe, B., Tóth-Király, I., & Paskuj, B. (2016). On the correlates of passion for screen-based behaviors: The case of impulsivity and the problematic and nonproblematic Facebook use and TV series watching. *Personality and Individual Differences*, 101, 167–176. doi: 10.1016/j.paid.2016.05.368
- Panda, S., & Pandey, S.C. (2017). Binge-watching and college students: Motivations and outcomes. *Young Consumers*, 18, 425–438. doi:10.1108/YC-07-2017-00707
- Perks, L.G. (2015). *Media marathoning: Immersions in morality*. Lanham, MD: Lexington Books.
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16, 56–69.
- Petersen, T.G. (2016). To binge or not to binge: A qualitative analysis of college students' binge watching habits. *Florida Communication Journal*, 44, 77–88.
- Pittman, M., & Sheehan, K. (2015). Sprinting a media marathon: Uses and gratifications of binge-watching television through Netflix. *First Monday*, 20. doi:10.5210/fm.v20i10.6138
- Ramayan, S., Munsayac Estella, A. L., & Abu Bakar, I. A. (2018). The effects of binge watching on interpersonal communication among Department of Communication and Liberal Arts (DCLA) students. *Ideology*, 3, 127–143.

- Riddle, K., Peebles, A., Davis, C., Xu, F., & Schroeder, E. (2017). The addictive potential of television binge-watching: Comparing intentional and unintentional binges. *Psychology of Popular Media Culture*, 7, 589–604. doi:10.1037/ppm0000167
- Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling*, 3, 369–379. doi:10.1080/10705519609540052
- Rubenking, B., & Bracken, C.C. (2018). Binge-watching: A suspenseful, emotional, habit. *Communication Research Reports*, 35, 381–391. doi:10.1080/08824096.2018.1525346
- Rubenking, B., Bracken, C.C., Sandoval, J., & Rister, A. (2018). Defining new viewing behaviours: What makes and motivates TV binge-watching? *International Journal of Digital Television*, 9, 69–85. doi:10.1386/jdtv.9.1.69_1
- Rubin, A.M. (2009). Uses-and-gratifications perspective on media effects. In Bryant, J., & Oliver, M.B. (Eds.), *Media effects: Advances in theory and research* (pp.165–184). New York: Routledge.
- Rubin, A. M. (1983). The uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting*, 27, 37–47. doi:10.1080/08838158309386471
- Schermelleh-Engel, K., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8, 28–74.
- Shao, C., & Beneza, P.M. (2018, August). *Binge-watching: Social and psychological factors behind audience's binging behaviour*. Paper presented at the AEJMC Annual Conference, Washington.

- Shim, H., & Kim, K. J. (2018). An exploration of the motivations for binge-watching and the role of individual differences. *Computers in Human Behaviors*, 82, 94–100. doi:10.1016/j.chb.2017.12.032
- Shim, H., Lim, S., Jung, E.E., & Shin, E. (2018). I hate binge-watching but I can't help doing it: the moderating effect of immediate gratification and need for cognition on binge-watching attitude-behavior relation. *Telematics and Informatics*, 35, 1971–1979. doi:10.1016/j.tele.2018.07.001
- Spangler, T. (2016). "Binge nation: 70% of Americans engage in marathon TV viewing." Retrieved from <https://variety.com/2016/digital/news/binge-watching-us-study-deloitte-1201737245/>
- Spruance, L.A., Karmakar, M., Kruger, J.S., & Vaterlaus, J.M. (2017). "Are you still watching?": Correlations between binge TV watching, diet and physical activity. *Journal of Obesity & Weight Management*, online first publication.
- Starcevic, V., Billieux, J., & Schimmenti, A. (2018). Selfitis, selfie addiction, Twitteritis: Irresistible appeal of medical terminology for problematic behaviours in the digital age. *Australian and New Zealand Journal of Psychiatry*, 52, 408–409. doi:10.1177/0004867418763532
- Starosta, J., Izydorczyk, B., & Lizińczyk, S. (2019). Characteristics of people's binge-watching behavior in the "entering into early adulthood" period of life. *Health Psychology Report*, 7, 149–164. doi:10.5114/hpr.2019.83025
- Steiner, E., & Xu, K. (2018). Binge-watching motivates change: Uses and gratifications of streaming video viewers challenge traditional TV research. *Convergence*, online first

publication. doi:10.1177/1354856517750365

Sung, Y.H., Kang, E.Y., & Wee, L. (2015). *A bad habit for your health? An exploration of psychological factors for binge-watching behavior*. Paper presented at the 65th ICA Annual Conference, Puerto Rico.

Sung, Y.H., Kang, E.Y., & Wee, L. (2018). Why do we indulge? Exploring motivations for binge watching. *Journal of Broadcasting & Electronic Media*, 62, 408–426. doi:10.1080/08838151.2018.1451851

Tang, J., Yu, Y., Du, Y., Ma, Y., Zhang, D., & Wang, J. (2014). Prevalence of internet addiction and its association with stressful life events and psychological symptoms among adolescent internet users. *Addictive Behaviors*, 39, 744–747. doi:10.1016/j.addbeh.2013.12.010

Tefertiller, A.C., & Maxwell, L.C. (2018). Depression, emotional states, and the experience of binge-watching narrative television. *Atlantic Journal of Communication*, 26, 278–290. doi:10.1080/15456870.2018.1517765

Tóth-Király, I., Böthe, B., Tóth-Fáber, E., Gyöző, H., & Orosz, G. (2017). Connected to TV series: Quantifying series watching engagement. *Journal of Behavioral Addictions*, 6, 472–489. doi:10.1556/2006.6.2017.083

Tóth-Király, I., Böthe, B., Neszta Márki, A., Rigó, A., & Orosz, G. (2019). Two sides of the same coin: The differentiating role of need satisfaction and frustration in passion for screen-based activities. *European Journal of Social Psychology*, 49, 1190–1205. doi:10.1002/ejsp.2588

Tukachinsky, R., & Eyal, K. (2018). The psychology of marathon television viewing:

- Antecedents and viewer involvement. *Mass Communication and Society*, 21, 275–295.
doi:10.1080/15205436.2017.1422765
- Vallerand, R. J. (2015). *The psychology of passion: A dualistic model*. New York, NY: Oxford University Press.
- Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., ...Marsolais, J. (2003). Les passions de l'âme: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, 85, 756–767. doi:10.1037/0022-3514.85.4.756
- Vaterlaus, J.M., Spruance, L.A., Frantz, K., & Kruger, J.S. (2019). College student television binge watching: Conceptualization, gratifications, and perceived consequences. *The Social Science Journal*, 56, 470–479. doi:10.1016/j.soscij.2018.10.004
- Whang, L. S. M., Lee, S., & Chang, G. (2003). Internet over-users' psychological profiles: A behavior sampling analysis on internet addiction. *CyberPsychology and Behavior*, 6, 143–150. <https://doi.org/10.1089/109493103321640338>
- Walton-Pattison, E., Dombrowski, S.U., & Presseau, J. (2018). “Just one more episode”: Frequency and theoretical correlates of television binge watching. *Journal of Health Psychology*, 23, 17–24. doi:10.1177/1359105316643379
- Yee, N. (2007). Motivations of play in online games. *Journal of CyberPsychology and Behavior*, 9, 772–775. <https://doi.org/10.1089/cpb.2006.9.772>
- YouGov Omnibus (2017). “58% of Americans binge-watch TV show”. Retrieved from <https://today.yougov.com/news/2017/09/13/58-americans-binge-watch-tv-shows/>

Zillmann, D. (1994). Mechanisms of emotional involvement with drama. *Poetics*, 23, 33–51.

doi:10.1016/0304-422X(94)00020-7

Zakariaeiz, Y., & Potenza, M.N. (2018). Gender-related differences in addiction: A review of human studies. *Current Opinion in Behavioral Sciences*, 23, 171–175.

doi:10.1016/j.cobeha.2018.08.004

Zillmann, D., & Bryant, J. (1985). Affect, mood, and emotion as determinants of selective exposure. In D. Zillmann & J. Bryant (Eds.), *Selective exposure to communication* (pp. 157–190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zillmann, D., Hezel, R. T., & Medoff, N. J. (1980). The effect of affective states on selective exposure to televised entertainment fare. *Journal of Applied Psychology*, 10, 323–339.

doi:10.1111/j.1559-1816.1980.tb00713.x

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70, 123–133. doi:10.1007/s11336-003-0974-7