BIROn - Birkbeck Institutional Research Online

Journal Pre-proofs

Mokken scale analysis of the Internet Gaming Disorder Scale–Short-Form and the Gaming Disorder Test

Laura Maldonado-Murciano, Halley M. Pontes, Maite Barrios, Juana Gómez-Benito, Georgina Guilera

Please cite this article as: L. Maldonado-Murciano, H.M. Pontes, M. Barrios, J. Gómez-Benito, G. Guilera, Mokken scale analysis of the Internet Gaming Disorder Scale–Short-Form and the Gaming Disorder Test, *Addictive Behaviors Reports* (2024), doi: https://doi.org/10.1016/j.abrep.2024.100567

Title: Mokken Scale Analysis of the Internet Gaming Disorder Scale–Short-Form and the Gaming Disorder Test

Running title: **MSA of the IGDS9-SF and the GDT**

Authors: Laura Maldonado-Murciano[1,2,3] , Halley M. Pontes[4] , Maite Barrios[1,2] , Juana Gómez-Benito[1,2] , and Georgina Guilera[1,2]

Affiliations:

[1] Faculty of Psychology, University of Barcelona

[2] Institute of Neurosciences, University of Barcelona

[3] Department of Psychology, International University of Catalonia

[4] Department of Organizational Psychology, Birkbeck, University of London

Authors' notes

Laura Maldonado-Murciano https://orcid.org/0000-0001-5427-729X

Halley M. Pontes https://orcid.org/0000-0001-8020-7623

Maite Barrios https://orcid.org/0000-0003-1979-0818

Juana Gómez-Benito https://orcid.org/0000-0002-4280-3106

Georgina Guilera https://orcid.org/0000-0002-4941-2511

Abstract:

In recent years, research on Gaming Disorder (GD) has grown substantially with researchers developing different psychometric tools for assessing disordered gaming. Two of the most prominent tools are the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF) and the Gaming Disorder Test (GDT), which assess disordered gaming under the American Psychiatric Association (APA) and the World Health Organisation (WHO) frameworks

respectively. The main aim of this study was to assess and compare the scalability, reliability, and validity of both scales to determine if the tools effectively assess GD in a normative sample, through the Mokken Scale Analysis (MSA). A sample of 605 participants (42.31% female, mean$_{age}$ = 23.98 years, SD = 9.21 years) was recruited for the present study. Results showed that both the IGDS9-SF and GDT were unidimensional, with all items presenting latent monotonicity fitting in the Monotone Homogeneity Model (MHM). Item characteristic curves did not intersect and presented with adequate fit in the Double Monotonicity Model (DMM). These findings further support the psychometric adequacy of the IGDS9-SF and GDT, attesting to their suitability to assess disordered gaming.

Corresponding Author:

Laura Maldonado-Murciano, Faculty of Psychology, University of Barcelona, Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain. Email: lmaldonado@ub.edu

Informed Consent: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants included in the study.

All authors have read and agreed to the published version of the manuscript.

Declaration of Interest: Laura Maldonado-Murciano, Halley M. Pontes, Maite Barrios, Juana Gómez-Benito, and Georgina Guilera declare that they have no conflict of interest.

Abstract

In recent years, research on Gaming Disorder (GD) has grown substantially, with researchers developing different psychometric tools for assessing disordered gaming. Two of the most prominent tools are the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF) and the Gaming Disorder Test (GDT), which assess disordered gaming under the American Psychiatric Association (APA) and the World Health Organisation (WHO) frameworks, respectively. The main aim of this study was to assess and compare the scalability, reliability, and validity of both scales to determine if the tools effectively assess GD in a normative sample, through the Mokken Scale Analysis (MSA). A sample of 605 participants (42.31% female, mean$_{age}$ = 23.98 years, SD = 9.21 years) was recruited for the present study. Results showed that both the IGDS9-SF and GDT were unidimensional, with all items presenting latent monotonicity, fitting within the Monotone Homogeneity Model (MHM). Item characteristic curves did not intersect and presented an adequate fit in the Double Monotonicity Model (DMM). These findings further support the psychometric adequacy of the IGDS9-SF and GDT, attesting to their suitability for assessing disordered gaming.

Introduction

Scientific interest in disordered gaming has grown substantially over the last decade (Pontes & Griffiths, 2020), leading to an increase in the global scientific production in behavioural addiction research, including disordered gaming (Sixto-Costoya et al., 2021). The inclusion of 'Internet Gaming Disorder' (IGD) as a tentative disorder in Section 3 of the 5th revision of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; APA 2013) by the American Psychiatric Association (APA) and the formal recognition of 'Gaming Disorder' (GD) as an official mental health disorder by the World Health Organization (WHO) in the 11th version of the *International Classification of Diseases* (ICD-11; WHO, 2022) have likely contributed to the broader scientific interest in disordered gaming.

Disordered gaming is currently conceptualised under both the APA and WHO frameworks (Montag et al., 2021; Pontes et al., 2022; Zhou et al., 2020). Within the APA framework, IGD is defined as 'persistent and recurrent use of the internet to engage in games, often with other players, leading to clinically significant impairment or distress' (APA, 2013, p. 795). Moreover, under the WHO framework, GD is understood as 'an excessive pattern of gaming behaviour (online and/or offline) reflected by lack of control, greater priority given to gaming over other life interests and activities, and continuation of the behaviour despite adverse outcomes' (WHO, 2022, Section 6C51).

Since its definition, disordered gaming has been found to have a prevalence of 3.3% in the global population, according to a recent meta-analysis (Kim et al., 2022). It significantly impacts quality of life (Phan et al., 2019) and has notable comorbidity with other disorders such as depression (Ostinelli et al., 2021). In terms of health-related outcomes, the meta-analysis by Männikkö et al. (2020) showed that disordered gaming is linked to depression, anxiety, obsessive-compulsive disorder, and somatization. Several important risk factors have been identified by Gao et al. (2022) in their meta-analysis, including, but not limited to, stress, higher time spent gaming, family disfunction, interpersonal problems, emotional distress, low self-esteem, and existing mental disorders (e.g., hyperactivity/inattention, anxiety, and depression).

Due to public health relevance, researchers have recently developed standardised psychometric tools to assess disordered gaming under the APA and WHO diagnostic frameworks (see King et al., 2020). Two of the most widely used tools are the Internet Gaming Disorder Scale–Short-Form (IGDS9-SF; Pontes & Griffiths, 2015) and the Gaming Disorder Test (GDT; Pontes et al., 2021), which assess disordered gaming under the APA and the WHO framework, respectively. Despite the relatively large body of evidence supporting the sound psychometric properties of these two tools (Evren, Dalbudak, et al., 2020; Gomez et al., 2019; Maldonado-Murciano et al., 2020, 2021; Pontes et al., 2021; Pontes & Griffiths, 2014, 2015; Schivinski et al., 2018), no previous research has yet investigated whether there is a stable hierarchy in terms of item difficulty within the two scales across different participants, nor whether the sum score of each item can be used to order respondents on the latent trait being measured (i.e., disordered gaming).

The Mokken Scale Analysis (MSA) is suitable for investigating the relationship between items and latent traits and has evolved from the Guttman method of examining hierarchies of items within a scale (Watson et al., 2012). The MSA is a type of non-parametric item response theory (NIRT) model, a more flexible model which present fewer restrictions on the data compared to parametric IRT models (such as Rasch models). This flexibility increases the number of items whose behaviour can be accurately explained in the

analysis (Meijer et al., 1990). In NIRT models, if the model fits the data, the sum score of each participant can be used to order respondents based on the severity of the latent trait measured (see Koopman et al., 2022; and van der Ark & Bergsma, 2010).

The literature supports the use of MSA, as this method has been shown to be useful in assessing the item hierarchy in several scales (Abdelhamid et al., 2020; Friedrich et al., 2015; Muncer & Speak, 2016). MSA explores whether items of a scale are hierarchically ordered by degree of difficulty and detects any violations (e.g., test-takers agreeing with a 'difficult' item while not agreeing with an 'easier' one). Moreover, the scalability coefficients allow us to assess item quality within a given set of items or, when creating a new tool, to use them as an item selection procedure (Sijtsma & Van der Ark, 2017). Practically speaking, MSA allows us to evaluate the degree to which individuals can be meaningfully ordered on the construct such that individuals with higher scores exhibit more symptoms. Likewise, MSA enables us to assess the degree to which symptoms have a consistent interpretation between participants, such that the relative order of symptoms is the same for those with relatively low or high levels of addiction (Sijtsma et al., 2011).

As for the construct of disordered gaming, to the best of our knowledge, only one study has been conducted using MSA (Finserås et al., 2019). Nevertheless, the present study extends the findings reported by Finserås et al. (2019) by addressing two main points. Firstly, this study will focus on the two most prominent scales for disordered gaming (i.e., IGDS9-SF and GDT) developed under the current APA and WHO diagnostic frameworks. Secondly, this study will provide findings not tied to a specific developmental stage as the IGDS9-SF and GDT evaluate disordered gaming more broadly across the lifespan and not only on adolescents.

The primary objective of this study is to evaluate and compare the psychometric properties of the IGDS9-SF and GDT in measuring GD based on different diagnostic frameworks. Specifically, this study aims to assess and compare the scalability, reliability, and validity of these instruments to determine how and whether these two instruments effectively capture GD in a normative sample of adults. By applying MSA (see Mokken, 1971; and Mokken & Lewis, 1982) to both instruments, the study seeks to identify the extent to which the items in each instrument, as well as the instruments themselves, are effective for use in psychological research and practice. The results of this comparative psychometric study will help researchers and health practitioners in the selection of the most adequate tool for measuring GD, ensuring reliable and valid assessments that can inform effective interventions and research outcomes.

Methods and materials

Participants and procedures

Similar to other validation studies of the IGDS9-SF and GDT (e.g., Beranuy et al., 2020; Maldonado-Murciano et al., 2020) the target population for this study was Spanish-speaking individuals aged 16 or older who played video games at least once a week. Participants were recruited through an online survey, which was advertised via multiple social media platforms (i.e., Facebook, Instagram, Reddit, and Twitter) and the online course management system of a second-year course in the Psychology degree program at the University of Barcelona. All study procedures were conducted in accordance with the Declaration of Helsinki and approved by the Bioethics Committee of the University of Barcelona. All participants provided their consent online. In accordance with Spanish

legislation (i.e., Article 7.1 of the Spanish Law 3/2018 of December 5, on the Protection of Personal Data and Guarantee of Digital Rights, 2018), parental consent was not required for minors, as for this type of study, minors can provide their own consent starting at the age of 14.

A sample of 1,023 participants was initially recruited for the study. However, a total of 418 participants were removed due to the following reasons: not having played any video games in the last 12 months ($n = 297$, 29.03%), playing more hours a week than hours in the week ($n = 79$, 7.72%), reporting not having played video games during the entire week (0h per week, $n = 11$, 1.07%), and not completing the whole survey ($n = 31$, 3.03%). As a result, a final sample of 605 participants was deemed eligible and used in the statistical analyses (available on request), with 23.97% of the participants recruited through the psychology course and 76.03% through the advertisements. In terms of sample characteristics, 256 participants were female (42.31%), and the mean age was 23.98 years ($SD = 9.21$ years; range: 16-73 years). Regarding educational level, 57.85% of the participants had completed secondary studies, followed by 28.92% who had achieved higher education. Additionally, 11.74% had primary education, and the remaining 1.49% did not have any formal education (see Table 1 for detailed sociodemographic and gaming behaviour characteristics).

[Insert Table 1]

Measures

Sociodemographic and gaming behaviour data: The survey collected sociodemographic and gaming behaviour data aligned with previous similar psychometric studies (e.g., Maldonado-Murciano et al., 2020, 2021), including participants' gender, age, weekly time spent gaming, and maximum hours ever played in a row. It is important to note that the sociodemographic information collected differed slightly between the social media and the course questionnaires (e.g., birthplace was only asked of participants recruited through social media).

Internet Gaming Disorder Scale–Short-Form (IGDS9; Pontes & Griffiths, 2015): The IGDS9-SF is a nine-item standardised test designed to assess disordered gaming under the nine criteria of the APA framework: (i) *preoccupation;* (ii) *withdrawal*; (iii) *tolerance*; (iv) *loss of control*; (v) *loss of interests*; (vi) *continuation*; (vii) *deception*; (viii) *escape*; and (ix) *negative consequences*. The items are rated using a five-point response scale ranging from 1 (never) to 5 (very often), with higher scores indicating greater levels of disordered gaming symptoms. The Spanish version of the IGDS9-SF used in this study has shown excellent psychometric properties (Beranuy et al., 2020; Maldonado-Murciano et al., 2020; Sánchez-Iglesias et al., 2020). In the present study, the IGDS9-SF scores showed high internal consistency ($\alpha = .84$, $\omega = .86$).

Gaming Disorder Test (GDT; Pontes et al., 2021): The GDT is a brief, four-item psychological test developed to assess disordered gaming under the WHO framework. All items are rated on a five-point scale ranging from 1 (never) to 5 (very often), yielding total scores ranging from 4 to 20 points, with higher scores indicating a greater degree of disordered gaming symptoms. The first three items of the GDT reflect (1) *impaired control over gaming,* (2) *increased priority given to gaming*, and (3) *continuation despite negative consequences*, while the fourth item assesses potential functional impairments by evaluating gamers' (4) *experience of significant problems in life* due to GD. For this study, the Spanish version of the GDT was used (Maldonado-Murciano et al., 2021) as it has been shown to

exhibit adequate psychometric properties. In the present study, the GDT scores showed high internal consistency ($\alpha = .81$, $\omega = .83$).

Statistical analysis

The NIRT models typically assessed with MSA are the Monotone Homogeneity Model (MHM), which implies ordering individuals on a scale based on their total scores, and the Double Monotonicity Model (DMM), which implies ordering both persons (as in the MHM) and items in terms of difficulty, and tests whether this item ordering is invariant across the latent variable continuum (Mokken & Lewis, 1982). As originally proposed by Mokken and Lewis (1982), the MHM is sufficient for test construction. However, if the test is to be administered, the DMM is also necessary to ensure that the order of item difficulties remains consistent for all participants. This stability can be studied by comparing the relative item ordering between participants with different overall levels on the construct. The present study assessed whether the IGDS9-SF and GDT fit both models. As these scales include polytomous items, the MSA for polytomous scales proposed by Molenaar (1991, 1997) was adopted.

Four assumptions are used to define the most popular NIRT models: unidimensionality, local independence, latent monotonicity, and no intersection of Item Response Functions (IRFs), with the latter being only used in the case of the DMM (van der Ark, 2012). Unidimensionality signifies that all the items in the scale measure a single latent variable (Sijtsma & Van der Ark, 2017). Local independence implies that the responses of the items are unrelated between them when the level of the latent variable is controlled for (van der Ark, 2012). Latent monotonicity refers to the increasing probability of a higher item score as the level of the latent trait increases (Watson et al., 2012). Finally, non-intersection assumes that IRFs do not intersect (Watson et al., 2012), indicating that items are ordered by difficulty.

To explore whether the MHM and DMM models fit the IGDS9-SF and GDT items, various analyses were conducted. Two confirmatory factor analyses (CFA) were fitted to assess the unidimensionality of the IGDS9-SF and GDT (Smits et al., 2012). Given that the items on both scales are on a Likert-point scale, models were estimated using the weighted least squares mean and variance adjusted (WLSMV) estimator. The fit of the models was evaluated using the following indices: Tucker-Lewis index (TLI), comparative fit index (CFI), and standardised root mean square residual (SRMR). The suggested guidelines of Hu and Bentler (1999) were followed when interpreting the goodness of fit, with values of CFI ≥ .95, TLI ≥ .95, and SRMR ≤ .08 considered satisfactory fit.

The Item Step Response Functions (ISRF) and Item Response Functions (IRF) were obtained to observe the monocity of the items individually.

Three scalability coefficients (van der Ark, 2012), based on Loevinger's coefficient H of homogeneity, were obtained: the item scalability coefficient (H_j), the item pairs scalability coefficient (H_{jk}), and the scale total scalability coefficient (H). When a scale fits the MHM, the H_j can be interpreted as an index of item 'discrimination', that is, how well the item can distinguish between respondents with low and high scores on the latent trait (McGrory et al., 2014), with higher H_j values indicating greater discrimination (Sijtsma et al., 2011). Similarly, coefficient H is a weighted mean of the item coefficients and can be interpreted as an average discrimination power, indicating the precision of ordering persons by their total score on the latent trait (Sijtsma et al., 2011).

For a set of items to form a Mokken scale, two necessary conditions must be met: all values of H_{jk} must be greater than 0, and H_{jk} must be greater than an a priori chosen criterion c (Watson et al., 2012). In this study, the recommendation from previous studies specifying $c = .30$ was adopted (Mokken, 1971; Mokken & Lewis, 1982). Moreover, all values of H need to exceed .30 to show evidence of an adequate fit of the item in the scale. If H is between .30 and .40, it is considered a weak scale, medium scale if it is between .40 and .50, and strong scale if it is above .50 (Mokken & Lewis, 1982).

Furthermore, the automated item selection procedure (AISP) was applied to identify items that satisfy the Mokken scale conditions. The AISP is a sequential method that assesses whether all items fit within the scale by evaluating their scalability indices, aiming to achieve the maximum total H (Smits et al., 2012). In the present study, four different lower bounds (c) were tested as recommended in the literature (Mokken, 1971): .05, .10, .30, and .50. For each analysis, the inter-item covariances need to be positive and exceed these lower bounds to confirm the unidimensionality of the items ($H_j \geq c \geq 0$). However, while MSA and the AISP procedure can provide evidence of unidimensionality, they are not definitive tests, and additional analyses (e.g., CFA) are often required to comprehensively test unidimensionality.

The Manifest Item Invariant Ordering (MIIO; Ligtvoet et al., 2010) was also analysed to test whether items can be ordered in terms of difficulty, forming a hierarchical structure, and whether this ordering is consistent across the latent trait. The method involves evaluating the IRFs for each pair of items to ensure that they do not intersect (van der Ark, 2012). In cases where item pairs do not hold MIIO (i.e., violations of the MIIO), they need to be removed from the set of items, and the scalability coefficient H^T must be assessed with the remaining items, as it expresses the degree to which the order of the items remains invariant (Sijtsma & Van der Ark, 2017). The H^T coefficient was interpreted using the established guidelines proposed by Mokken (1971): weak ordering ($.3 < H^T \leq .4$), moderate ordering ($.4 < H^T \leq .5$), and strong ordering ($H^T > .5$).

Finally, the reliability of the IGDS9-SF and GDT was assessed using Cronbach's alpha, the lambda-2 statistic (λ^2) (Guttman, 1945), the Molenar-Sijtsma statistic (MS) (Sijtsma & Molenaar, 2002), which has a more negligible bias for MSA, and the Latent Class Reliability Coefficient (LCRC), an unbiased statistic of test-score reliability (van der Ark et al., 2011).

Statistical analyses were performed using R version 4.1.2 (R Core Team, 2022) and the following packages: *dplyr* V 1.0.7 (Wickham et al., 2018), *psych* V 2.1.9 (Revelle, 2015), and *Mokken* V 3.0.6 (van der Ark et al., 2021).

Results

The fit indices of the two CFAs confirmed the unidimensionality of both the IGDS9-SF (CFI = 1.0, TLI = 1.0, SRMR = .02) and the GDT (CFI = 1.0, TLI = 1.0, SRMR = .05), suggesting in turn that the assumption of local independence is met (Sijtsma et al., 2011). The resulting path diagrams are shown in Figure S1 of the Supplementary Material.

Figures S2 and S3 of the Supplementary Material present the ISRFs and IRFs for the IGDS9-SF and the GDT, respectively, while Table 2 shows the results of the MSA for the items of the scales.

The scalability of the complete scales, measured with Loevinger's coefficient of homogeneity, was medium for the IGDS9-SF ($H = .45$) and strong for the GDT ($H = .62$). All item coefficients on both scales exceeded .30, suggesting that the items reflect a single latent trait. Specifically, for the IGDS9-SF items, H_j values ranged from .51 (item 2) to .37 (item 7), including items with low (i.e., items 7 and 8), medium (items 1, 3, 4, 5, 6, and 9), and high (item 2) discrimination values. As for the GDT items, they showed values above .62, suggesting strong accuracy, with item 4 being slightly more discriminative than the rest ($H_j = .63$). Furthermore, all the item pair scalability (H_{jk}) indices for both scales were significantly different from 0 (see Tables S1 and S2 of the Supplementary Material), suggesting that Guttman errors were not detrimental to the overall interpretation of the scores.

[Insert Table 2]

The AISP showed that all the items conform to a single scale (AISP = 1) under the four prespecified lower bounds (i.e., .05, .10, .30, .55). No violations of the MIIO were found (see Table 2), suggesting that the order of the items was consistent for all respondents, regardless of their total scale score. Furthermore, for both scales, in terms of the degree to which the order of the items remains invariant, the results showed a weak ordering of the items ($H_{IGDS9-SF}^{T} = 0.31$; $H_{GDT}^{T} = 0.41$).

It can be concluded that there is a hierarchical order of the items in terms of difficulty. As shown in Table 2, for the IGDS9-SF, the most 'difficult' item was item 9 (*negative consequences*; mean score = 1.21), while the 'easiest' was item 8 (*escape*; mean score = 2.34). Moreover, for the GDT, the most 'difficult' was item 4 (*experience of significant problems in life*; mean score = 1.29), while the 'easiest' was item 2 (*increased priority given to gaming*; mean score = 1.98).

Finally, reliability was high for both scales, as the IGDS9-SF showed values of $\alpha = .84$ $\lambda^2 = .85$, MS = .85, and LCRC = .82, while the GDT showed values of $\alpha = .81$, $\lambda^2 = .82$, MS = .83, and LCRC = .82, suggesting a high level of internal consistency.

Discussion

The present study sought to investigate the psychometric properties of the IGDS9-SF and the GDT using MSA, testing different models (i.e., MHM and DMM). The main findings suggested that both scales were unidimensional, presenting local independence, latent monotonicity, and non-intersecting IRFs. Additionally, both scales obtained high values for the reliability coefficients, further supporting the robustness of reliability of their scores, as shown in previous research (e.g., Evren, Pontes, et al., 2020; Maldonado-Murciano et al., 2020, 2021; Schivinski et al., 2018; T'ng & Pau, 2020; Yam et al., 2019).

Accordingly, both the IGDS9-SF and the GDT fit the MHM model, suggesting that the sum score of these scales is a robust indicator of disordered gaming levels (Sijtsma & Molenaar, 2002) further allowing: (1) the ordering of the respondents across this latent trait, (2) the use of item scores to assess the difficulty, and (3) observing the hierarchy of the items (Watson et al., 2012). Both scales also fit in the DMM, which signifies that the item difficulty hierarchy remained invariant across all the participants.

In summary, item 2 of the IGDS9-SF (*withdrawal*) and item 4 of the GDT (*experience of significant problems in life*) were the most discriminative items on their respective scales. This finding suggests that these items are particularly effective in differentiating individuals

based on the latent trait of disordered gaming. Specifically, higher scores on these items are more strongly associated with higher overall disordered gaming scores, indicating a greater symptom load of disordered gaming.

In terms of difficulty, the items of the IGDS9-SF were ordered from item 9 (*negative consequences*) as the most difficult item to item 8 (*escape*) as the easiest. This finding aligns well with previous IRT studies showing that item 9 requires relatively higher levels of GD to be endorsed (Gomez et al., 2019; Maldonado-Murciano et al., 2020), and that escapism is highly correlated with GD (Montag et al., 2019, 2022). Within the APA framework, the potential implications of these findings imply that the experience of negative consequences associated with problematic gaming behaviour may be a critical indicator of GD, therefore, clinicians should pay particular attention when individuals exhibit this symptom, as they may be experiencing a more severe form of GD. With this information, clinicians may also consider escapism as an early warning sign to identify individuals who may be at risk of developing more severe symptoms of GD.

From an intervention standpoint, focusing on escapism as an important symptom may be misleading, as video games are often used to escape real-life problems and therefore may not necessarily indicate disordered gaming (Laconi et al., 2017). Furthermore, Stavropoulos et al. (2019) found that individuals with ADHD use video games to escape reality, while Myrseth et al. (2017) concluded that gaming serves as a coping mechanism for depression, loneliness, or an unpleasant reality.

In the GDT, item 4 (*experience of significant problems in life*) presented slightly greater difficulty than the rest of the items. The least 'difficult' item was item 2 (*increased priority given to gaming*). This means that within the WHO framework, the experience of significant life problems represents a more severe symptom of GD. As such, clinicians should pay particular attention to this symptom, as individuals who report experiencing significant life problems due to gaming may be facing a more severe form of the disorder. This symptom may constitute a critical indicator of the severity of GD, warranting immediate and focused treatment. Furthermore, regarding assigning increased priority to gaming activities, clinicians can consider this symptom as an early warning sign of GD. Consequently, large-scale interventions in different settings, such as schools, may identify and target individuals who prioritise gaming over other important life activities, as this symptom may aid in early detection and intervention, potentially preventing the escalation of GD. At the psychometric level, the findings suggest that the items of the GDT could better distinguish between persons with high and low levels of disordered gaming than those of the IGDS9-SF, as the GDT items appear to be slightly more discriminative. This finding is paramount to health practitioners and clinicians when considering which diagnostic framework and tool to use in the assessment of disordered gaming.

Taken together, the present findings add to the growing body of evidence supporting the psychometric properties of the IGDS9-SF and GDT (Beranuy et al., 2020; Maldonado-Murciano et al., 2020, 2021; Sánchez-Iglesias et al., 2020), further attesting to the fact that both tools are adequate to assess disordered gaming in research and clinical settings.

These findings have important implications for the use of the IGDS9-SF and GDT in preventive interventions and educational settings. For instance, in high-school environments, these scales can be utilized by counsellors and educators to identify students who may be at risk of disordered gaming. Early identification can facilitate timely intervention, helping to prevent the escalation of gaming-related issues. For preventive interventions, the scales can

be integrated into routine health assessments to monitor individuals' gaming behaviours. This approach allows for the detection of early warning signs and the implementation of targeted interventions aimed at promoting healthy gaming habits and mitigating potential negative outcomes associated with excessive gaming.

Nevertheless, it is important to consider potential limitations in the present research. The main limitation pertains to the sample selection strategy, as participants were self-selected, making the reported results not generalisable to the Spanish population. Another limitation is the limited representativeness of the sample, as it consists predominantly of young adults and includes a minimum age of 16, excluding younger gamers from the study. Finally, we did not collect information about the type of games the participants play (i.e., online or offline), which also affects the generalisability of results.

Future studies may help advance the existing literature by further investigating the diagnostic accuracy of the IGDS9-SF and GDT in clinical samples in terms of sensitivity and specificity, and by using a valid and reliable gold standard that allows for the estimation of cut-off points, particularly for the GDT, as this area of research has been explored within the APA framework using the IGDS9-SF (Qin et al., 2020). Another potential line of future research is the validation of both tools across different age ranges, as there is no evidence regarding how they perform in adolescent and elderly populations.

References

Abdelhamid, G. S. M., Gómez-Benito, J., Abdeltawwab, A. T. M., Abu Bakr, M. H. S., & Kazem, A. M. (2020). A Demonstration of Mokken Scale Analysis Methods Applied to Cognitive Test Validation Using the Egyptian WAIS-IV. *Journal of Psychoeducational Assessment*, *38*(4), 493–506. https://doi.org/10.1177/0734282919862144

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Beranuy, M., Machimbarrena, J. M., Vega-Osés, M. A., Carbonell, X., Griffiths, M. D., Pontes, H. M., & González-Cabrera, J. (2020). Spanish validation of the internet gaming disorder scale–short form (IGDS9-SF): Prevalence and relationship with online gambling and quality of life. *International Journal of Environmental Research and Public Health*, *17*(5), 1562. https://doi.org/10.3390/ijerph17051562

Evren, C., Dalbudak, E., Topcu, M., Kutlu, N., Evren, B., Pontes, H. M., Dalbudak, E., Evren, B., Topcu, M., & Kutlu, N. (2020). Psychometric validation of the Turkish nine-item Internet Gaming Disorder Scale–Short Form (IGDS9-SF). *Psychiatry Research*, *265*(2), 144. https://doi.org/10.5455/pcp.20200429072430

Evren, C., Pontes, H., Dalbudak, E., Evren, B., Topcu, M., & Kutlu, N. (2020). Psychometric validation of the Turkish Gaming Disorder Test: A measure that evaluates disordered gaming according to the World Health Organization framework. *Psychiatry and Clinical Psychopharmacology*, *30*(2), 144. https://doi.org/10.5455/pcp.20200429072430

Finserås, T. R., Pallesen, S., Mentzoni, R. A., Krossbakken, E., King, D. L., & Molde, H. (2019). Evaluating an Internet Gaming Disorder scale using Mokken scaling analysis. *Frontiers in Psychology*, *10*(APR), 911. https://doi.org/10.3389/fpsyg.2019.00911

Friedrich, O., Sipötz, J., Benzer, W., Kunschitz, E., & Höfer, S. (2015). The dimensional structure of the MacNew Health Related Quality of Life questionnaire: A Mokken Scale Analysis. *Journal of Psychosomatic Research*, *79*(1), 43–48. https://doi.org/10.1016/j.jpsychores.2015.04.007

Gao, Y. X., Wang, J. Y., & Dong, G. H. (2022). The prevalence and possible risk factors of internet gaming disorder among adolescents and young adults: Systematic reviews and meta-analyses. *Journal of Psychiatric Research*, *154*, 35–43. https://doi.org/10.1016/J.JPSYCHIRES.2022.06.049

Gomez, R., Stavropoulos, V., Beard, C., & Pontes, H. M. (2019). Item Response Theory Analysis of the Recoded Internet Gaming Disorder Scale-Short-Form (IGDS9-SF). *International Journal of Mental Health and Addiction*, *17*(4), 859–879. https://doi.org/10.1007/s11469-018-9890-z

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282. https://doi.org/10.1007/BF02288892

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Irwing, P., Booth, T., & Hughes, D. J. (Eds.). (2018). *The Wiley handbook of psychometric testing : a multidisciplinary reference on survey, scale, and test development*. Wiley Blackwell.

Kim, H. S., Son, G., Roh, E.-B., Ahn, W.-Y., Kim, J., Shin, S.-H., Chey, J., & Choi, K.-H. (2022). Prevalence of gaming disorder: A meta-analysis. *Addictive Behaviors*, *126*, 107183. https://doi.org/10.1016/j.addbeh.2021.107183

King, D. L., Chamberlain, S. R., Carragher, N., Billieux, J., Stein, D., Mueller, K., Potenza, M. N., Rumpf, H. J., Saunders, J., Starcevic, V., Demetrovics, Z., Brand, M., Lee, H. K., Spada, M., Lindenberg, K., Wu, A. M. S., Lemenager, T., Pallesen, S., Achab, S., … Delfabbro, P. H. (2020). Screening and assessment tools for gaming disorder: A comprehensive systematic review. *Clinical Psychology Review*, *77*. https://doi.org/10.1016/j.cpr.2020.101831

Koopman, L., Zijlstra, B. J. H., & van der Ark, L. A. (2022). A two-step, test-guided Mokken scale analysis, for nonclustered and clustered data. *Quality of Life Research*, *31*(1), 25–36. https://doi.org/10.1007/s11136-021-02840-2

Laconi, S., Pirès, S., & Chabrol, H. (2017). Internet gaming disorder, motives, game genres and psychopathology. *Computers in Human Behavior*, *75*, 652–659. https://doi.org/10.1016/j.chb.2017.06.012

Law 3/2018 of December 5, on the Protection of Personal Data and Guarantee of Digital Rights, Boletín Oficial del Estado 119544 (2018). https://www.boe.es/diario_boe/txt.php?id=BOE-A-2018-16673

Ligtvoet, R., van der Ark, L. A., te Marvelde, J. M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, *70*(4), 578–595. https://doi.org/10.1177/0013164409355697

Maldonado-Murciano, L., Pontes, H. M., Barrios, M., Gómez-Benito, J., & Guilera, G. (2021). Psychometric Validation of the Spanish Gaming Disorder Test (GDT): Item Response Theory and Measurement Invariance Analysis. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-021-00704-x

Maldonado-Murciano, L., Pontes, H. M., Griffiths, M. D., Barrios, M., Gómez-Benito, J., & Guilera, G. (2020). The Spanish version of the Internet Gaming Disorder Scale-Short Form (IGDS9-SF): Further examination using item response theory. *International Journal of Environmental Research and Public Health*, *17*(19), 7111. https://doi.org/10.3390/ijerph17197111

Männikkö, N., Ruotsalainen, H., Miettunen, J., Pontes, H. M., & Kääriäinen, M. (2020). Problematic gaming behaviour and health-related outcomes: A systematic review and meta-analysis. *Journal of Health Psychology*, *25*(1), 67–81. https://doi.org/10.1177/1359105317740414/SUPPL_FILE/SUPPLEMENTARY_DOCU MENT_F.PDF

McGrory, S., Shenkin, S. D., Austin, E. J., & Starr, J. M. (2014). Lawton IADL scale in dementia: Can item response theory make it more informative? *Age and Ageing*, *43*(4), 491–495. https://doi.org/10.1093/ageing/aft173

Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and Empirical Comparison of the Mokken and the Rasch Approach to IRT. *Applied Psychological Measurement*, *14*(3), 283–298. https://doi.org/10.1177/014662169001400306

Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis*. De Gruyter Mouton. https://doi.org/10.1515/9783110813203

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, *6*(4), 417–430. https://doi.org/10.1177/014662168200600404

Molenaar, I. W. (1991). A Weighted Loevinger H-Coefficient Extending Mokken Scaling to Multicategory Items. *Kwantitatieve Methoden*, *12*(37), 97–117.

Molenaar, I. W. (1997). Nonparametric Models for Polytomous Responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 369–380). Springer-Verlag.

Montag, C., Kannen, C., Schivinski, B., & Pontes, H. M. (2021). Empirical evidence for robust personality-gaming disorder associations from a large-scale international investigation applying the APA and WHO frameworks. *PLOS ONE*, *16*(12), e0261380. https://doi.org/10.1371/journal.pone.0261380

Muncer, S. J., & Speak, B. (2016). Mokken scale analysis and confirmatory factor analysis of the Health of the Nation Outcome Scales. *Personality and Individual Differences*, *94*, 272–276. https://doi.org/10.1016/j.paid.2016.01.051

Myrseth, H., Kjellevold Olsen, O., Kristian Borud, E., & Åge Strand, L. (2017). Predictors of Gaming Behavior among Military Peacekeepers – Exploring the Role of Boredom and Loneliness in Relation to Gaming Problems. *Journal of Military Studies*, *9*(1), 1–10. https://doi.org/10.1515/jms-2017-0001

Ostinelli, E. G., Zangani, C., Giordano, B., Maestri, D., Gambini, O., D'Agostino, A., Furukawa, T. A., & Purgato, M. (2021). Depressive symptoms and depression in individuals with internet gaming disorder: A systematic review and meta-analysis. *Journal of Affective Disorders*, *284*, 136–142. https://doi.org/10.1016/j.jad.2021.02.014

Pontes, H. M., & Griffiths, M. D. (2014). Assessment of internet gaming disorder in clinical research: Past and present perspectives. *Clinical Research and Regulatory Affairs*, *31*(2–4), 35–48. https://doi.org/10.3109/10601333.2014.962748

Pontes, H. M., & Griffiths, M. D. (2015). Measuring DSM-5 internet gaming disorder: Development and validation of a short psychometric scale. *Computers in Human Behavior*, *45*, 137–143. https://doi.org/10.1016/j.chb.2014.12.006

Pontes, H. M., & Griffiths, M. D. (2020). A new era for gaming disorder research: Time to shift from consensus to consistency. *Addictive Behaviors*, *103*. https://doi.org/10.1016/j.addbeh.2019.106059

Pontes, H. M., Schivinski, B., Kannen, C., & Montag, C. (2022). The interplay between time spent gaming and disordered gaming: A large-scale world-wide study. *Social Science & Medicine*, *296*, 114721. https://doi.org/10.1016/j.socscimed.2022.114721

Pontes, H. M., Schivinski, B., Sindermann, C., Li, M., Becker, B., Zhou, M., & Montag, C. (2021). Measurement and conceptualization of gaming disorder according to the World Health Organization framework: The development of the Gaming Disorder Test. *International Journal of Mental Health and Addiction*, 1–21. https://doi.org/10.1007/s11469-019-00088-z

Qin, L., Cheng, L., Hu, M., Liu, Q., Tong, J., Hao, W., Luo, T., & Liao, Y. (2020). Clarification of the Cut-off Score for Nine-Item Internet Gaming Disorder Scale–Short Form (IGDS9-SF) in a Chinese Context. *Frontiers in Psychiatry*, *11*, 470. https://doi.org/10.3389/fpsyt.2020.00470

R Core Team. (2022). *R: A language and environment for statistical computing* (4.1.2). https://www.r-project.org/

Revelle, W. (2015). psych: Procedures for Personality and Psychological Research. v 1.5.6. In *Software*. https://personality-project.org/r/psych

Sánchez-Iglesias, I., Bernaldo-De-Quirós, M., Labrador, F. J., Estupiñá Puig, F. J., Labrador, M., & Fernández-Arias, I. (2020). Spanish validation and scoring of the Internet Gaming Disorder Scale - Short-Form (IGDS9-SF). *Spanish Journal of Psychology*, *23*, 1–11. https://doi.org/10.1017/SJP.2020.26

Schivinski, B., Brzozowska-Woś, M., Buchanan, E. M., Griffiths, M. D., & Pontes, H. M. (2018). Psychometric assessment of the internet gaming disorder diagnostic criteria: An item response theory study. *Addictive Behaviors Reports*, *8*, 176–184. https://doi.org/10.1016/j.abrep.2018.06.004

Sijtsma, K., Meijer, R. R., & Andries van der Ark, L. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences*, *50*(1), 31–37. https://doi.org/10.1016/j.paid.2010.08.016

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. SAGEA.

Sijtsma, K., & Van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology*, *70*(1), 137–158. https://doi.org/10.1111/bmsp.12078

Sixto-Costoya, A., Castelló-Cogollos, L., Aleixandre-Benavent, R., & Valderrama-Zurián, J. C. (2021). Global scientific production regarding behavioral addictions: An analysis of the literature from 1995 to 2019. *Addictive Behaviors Reports*, *14*, 2352–8532. https://doi.org/10.1016/j.abrep.2021.100371

Smits, I. A. M., Timmerman, M. E., & Meijer, R. R. (2012). Exploratory Mokken Scale Analysis as a Dimensionality Assessment Tool: Why Scalability Does Not Imply Unidimensionality. *Applied Psychological Measurement*, *36*(6), 516–539. https://doi.org/10.1177/0146621612451050/FORMAT/EPUB

Stavropoulos, V., Adams, B. L. M., Beard, C. L., Dumble, E., Trawley, S., Gomez, R., & Pontes, H. M. (2019). Associations between attention deficit hyperactivity and internet gaming disorder symptoms: Is there consistency across types of symptoms, gender and

countries? *Addictive Behaviors Reports*, *9*, 100158. https://doi.org/10.1016/J.ABREP.2018.100158

T'ng, S. T., & Pau, K. (2020). Validation of a translated internet gaming disorder scale (short form) and measurement invariance across sex groups in Malaysian samples. *Current Psychology*. https://doi.org/10.1007/s12144-020-00668-6

van der Ark, L. A. (2012). New Developments in Mokken Scale Analysis in R. *Journal of Statistical Software*, *48*(5). https://doi.org/10.18637/jss.v048.i05

van der Ark, L. A., & Bergsma, W. P. (2010). A Note on Stochastic Ordering of the Latent Trait Using the Sum of Polytomous Item Scores. *Psychometrika*, *75*(2), 272–279. https://doi.org/10.1007/S11336-010-9147-7

van der Ark, L. A., Koopman, L., Straat, J. H., & van den Bergh, D. (2021). *Conducts mokken scale analysis* (pp. 3–65).

van der Ark, L. A., van der Palm, D. W., & Sijtsma, K. (2011). A latent class approach to estimating test-score reliability. *Applied Psychological Measurement*, *35*(5), 380–392. https://doi.org/10.1177/0146621610392911

Watson, R., van der Ark, L. A., Lin, L. C., Fieo, R., Deary, I. J., & Meijer, R. R. (2012). Item response theory: How Mokken scaling can be used in clinical practice. *Journal of Clinical Nursing*, *21*(19–20), 2736–2746. https://doi.org/10.1111/j.1365-2702.2011.03893.x

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. https://cran.r-project.org/package=dplyr

World Health Organization. (2022). *International classification of diseases for mortality and morbidity statistics (11th Revision)*. https://icd.who.int/browse11/l-m/en

Yam, C. W., Pakpour, A. H., Griffiths, M. D., Yau, W. Y., Lo, C. L. M., Ng, J. M. T., Lin, C. Y., & Leung, H. (2019). Psychometric testing of three Chinese online-related addictive behavior instruments among Hong Kong university student. *Psychiatric Quarterly*, *90*(1), 117–128. https://doi.org/10.1007/s11126-018-9610-7

Zhou, X., Wu, R., Liu, C., Kou, J., Chen, Y., Pontes, H. M., Yao, D., Kendrick, K. M., Becker, B., & Montag, C. (2020). Higher levels of (Internet) Gaming Disorder symptoms according to the WHO and APA frameworks associate with lower striatal volume. *Journal of Behavioral Addictions*, *9*(3), 598–605. https://doi.org/10.1556/2006.2020.00066

Tables

Table 1

Descriptive sociodemographic and gaming behaviour data from the sample ($n = 650$)

Variables	Mean	SD	Min.	Max.
Age	23.98	9.21	16	73
Hours playing video games on working days	1.92	2.07	0	16
Hours playing video games on free days	3.39	2.85	0	24
Max. hours ever played	**39.86**	**28.65**	**1**	**89**

Note: SD = standard deviation; Min. = minimum value; Max. = maximum value,

Table 2

Descriptive statistics and Mokken Scale Analysis results of the IGDS9-SF and GDT ($n = 650$)

Scale/Items	Mean	SD	H_j	SE	AISP				v. MIIO
					.05	.10	.30	.55	
IGDS9-SF									
Item 1: *preoccupation*	1.96	1.09	.48*	.03	1	1	1	1	0
Item 2: *withdrawal*	1.52	0.79	.51**	.03	1	1	1	1	0
Item 3: *tolerance*	1.54	0.87	.48*	.03	1	1	1	1	0
Item 4: *loss of control*	1.57	0.85	.47*	.03	1	1	1	1	0
Item 5: *loss of interests*	1.64	0.97	.43*	.03	1	1	1	1	0
Item 6: *continuation*	1.40	0.84	.44*	.04	1	1	1	1	0
Item 7: *deception*	1.45	0.88	.37*	.04	1	1	1	1	0
Item 8: *scape*	2.34	1.23	.38*	.04	1	1	1	1	0
Item 9: *opportunity loss*	1.21	0.59	.43*	.04	1	1	1	1	0
GDT									
Item 1: *impaired control over gaming*	1.76	0.88	.62**	.03	1	1	1	1	0
Item 2: *increased priority given to gaming*	1.98	1.03	.62**	.03	1	1	1	1	0

Item 3: *continuation despite negative consequences*	1.89	1.05	.62**	.03	1	1	1	1	0
Item 4: *experience of significant problems in life*	1.26	0.62	.63**	.03	1	1	1	1	0

Note: SD = Standard Deviation, H_j = Loevinger's Coefficient, SE = Standard Error, v. MIIO = Violations of the Manifest Invariant Item Ordering, AISP = Automated Item Selection Procedure, * Medium Discrimination, **Strong Discrimination.