Reconsidering item response categories in gaming disorder symptoms measurement

DANIEL L. KING[1]* , ABEL NOGUEIRA-LÓPEZ[2,3] ,
CHRISTINA R. GALANIS[1] , TOSHITAKA HAMAMURA[1] ,
CHRISTIAN BÄCKLUND[4] , ALESSANDRO GIARDINA[3] ,
JOËL BILLIEUX[3,5] and PAUL H. DELFABBRO[6]

[1] College of Education, Psychology & Social Work, Flinders University, Australia

[2] University of León, León, Spain

[3] Institute of Psychology, University of Lausanne, Lausanne, Switzerland

[4] Luleå University of Technology, Sweden

[5] Centre for Excessive Gambling, Addiction Medicine, Lausanne University Hospitals (CHUV), Lausanne, Switzerland

[6] School of Psychology, The University of Adelaide, Australia

ABSTRACT

Gaming disorder (GD) screening often involves self-report survey measures to detect the presence of symptoms. Studies have shown that gamers' responses vary greatly across survey items. Some symptoms, such as preoccupation and tolerance, are frequently reported by highly engaged but non-problematic gamers, and therefore these symptoms are thought to lack specificity and are suggested to be less important in classification decisions. We argue that the influence of response categories (e.g., dichotomous responses, such as 'yes' or 'no'; or frequency categories, such as 'rarely' and 'often') on item responses has been relatively underexplored despite potentially contributing significantly to the psychometric performance of items and scales. In short, the type of item response may be just as important to symptom reporting as the content of survey questions. We propose some practical alternatives to currently used item categories across GD tools. Research should examine the performance of different response categories, including whether certain response categories aid respondents' comprehension and insight, and better capture pathological behaviours and harms.

KEYWORDS

gaming disorder, screening, assessment, response category, validation, psychometric

Despite being a recent addition to health classification systems relative to other addictive disorders, there is an abundance of self-report scales for gaming disorder (GD). By comparison, gambling disorder, the condition that has the most symptoms in common with GD, tends to be assessed by only a few screening tools (i.e., the *Problem Gambling Severity Index, NODS-CLiP,* and the *South Oaks Gambling Screen*) from a selection of about a dozen major instruments used internationally (Caler, Garcia, & Nower, 2016). A systematic review by King et al. (2020a) identified 32 tools for GD and noted that at least two new tools had been developed each year since 'Internet gaming disorder' was listed in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) in 2013. With the inclusion of gaming disorder in the Eleventh Revision of the International Classification of Diseases (ICD-11), the field now has two major taxonomic references to GD, each with its own clinical description and slightly different emphasis on symptoms. It seems likely that more GD scales will be developed until such time that the field has an accepted 'gold standard' instrument or otherwise rallies support for a shortlist of tools with meaningfully distinct properties and advantages for different applications (King & Delfabbro, 2019).

In the meantime, the field continues to debate the most appropriate approach to assessing GD and its symptoms (Castro-Calvo et al., 2021), with implications for epidemiological research (e.g., prevalence rates based on symptom scores) and clinical research (e.g., cut-off eligibility for treatment and measuring recovery). In this paper, we draw attention to a psychometric issue that has received much less attention and empirical scrutiny: the influence of item response categories on scale scores and classifications. In our view, the importance of the content of survey items is often discussed, but the influence of items' corresponding response categories (e.g., dichotomous responses, such as 'yes' or 'no'; or frequency categories, such as 'rarely' and 'often') has been relatively underexplored despite potentially contributing significantly to the psychometric performance of items and scales. We suggest that the type of item response may be just as important to symptom reporting as the content of survey questions.

Although, as noted above, there are more than 30 screening tools for GD, there are perhaps more similarities than differences across these measures (NB: Readers should be aware that many tools have highly similar names and acronyms). Many measures in the GD literature have attempted to capture the nine criteria listed in the DSM-5 (also, the new DSM-5-TR) classification of Internet gaming disorder, which is not currently listed as an official diagnosis in this nomenclature. Some measures include one item for each symptom (e.g., the Clinical Video game Addiction Test [C-VAT] 2.0: Van Rooij, Schoenmakers, & Van de Mheen, 2017; the Internet Gaming Disorder Short Form-9 [IGDSF-9]: Pontes & Griffiths, 2016; the Internet Gaming Disorder Scale-9 [IGD Scale-9]: Lemmens, Valkenburg, & Gentile, 2015), whereas other scales have employed two or more items for a symptom (e.g., the Internet Gaming Disorder-20 [IGD-20]; Pontes, Kiraly, Demetrovics, & Griffiths, 2014; the Internet Gaming Disorder Scale-27 [IGD Scale-27]: Lemmens et al., 2015; the Problematic Online Gaming Questionnaire [POGQ]; Demetrovics et al., 2012). Most commonly, the criterion for functional impairment has been measured by more than one item, with a separate item for each specific domain or area of life affected by problematic gaming (e.g., school, work, health, family and other social relationships). The Behavioral Addiction Measure-Video Gaming (BAM-VG), for example, has nine items (out of 19) that refer to gaming-related problems (Sanders & Williams, 2016). Of importance to this discussion is the observation that, although these scales have slight variations in their wording for each symptom item (e.g., tolerance items tend to refer to a need for "increasing time" or "more exciting" games; King, Herd, & Delfabbro, 2017; Razum, Baumgartner, & Glavak-Tkalić, 2023), these scales have generally been quite consistent and similar in their use of response categories.

Based on psychometric reviews (King, Haagsma, Delfabbro, Gradisar, & Griffiths, 2013; King, Billieux, Carragher, & Delfabbro, 2020; King, Chamberlain et al., 2020), there are three conventional approaches in GD scale response categories. The first is the dichotomous "Yes/No" option, with each "Yes" response scored equivalently as 1 point. Some authors have employed an additional response category (e.g., "Sometimes") as an intermediate option between "Yes" and "No", which adds some complexity to scale scoring and interpretation (e.g., Király et al., 2017). As a sidenote, another complication is that some items in GD measures refer to "sometimes" in their framing (e.g., *Do you sometimes skip household chores in order to spend more time playing video games?*; Gentile, 2009), and similar frequency-related qualifiers (e.g., 'regularly' and 'often') are commonly used in GD items. Gentile's (2009) study of 1,178 adolescent gamers, for example, included a "Sometimes" response along with "Yes" and "No" in their 11-point problematic gaming scale. Gentile compared how scoring "Sometimes" as either being worth one point (1), half a point (0.5), or as zero (0) affected overall scale scores and classification. Gentile reported prevalence rates of 19.8%, 8.5% and 7.9%, for each of these scoring options, respectively, thereby showing that treating "sometimes" as an affirmative response could make certain items much more sensitive and greatly increase the prevalence rate of the condition (i.e., by a factor of 2). However, putting aside the complication of "Sometimes" and other intermediate options, the 'black-and-white' "Yes/No" format may still not provide a clear indication of any given symptom. A "Yes" response on most GD items does not necessarily indicate the severity of the symptom (i.e., how much dysfunction or harm it generates), nor does it indicate whether the symptom was transient or recurrent.

The second convention in GD scale response categories is the frequency scale, which is typically a 5-point Likert scale (i.e., 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always). This was the most common response category type across the tools identified in recent reviews (King, Billieux et al., 2020; King, Chamberlain et al., 2020). For scoring purposes, however, researchers will often convert "Often" and "Always" to affirmative responses (i.e., the equivalent of "Yes") (e.g., Montag et al., 2019). Although this approach is useful because it enables greater sensitivity, which may help in identifying "at risk" respondents, the frequency approach has its limitations. One issue is that the frequency of a symptom does not necessarily indicate the level of harm, even when that symptom is reported as having occurred "Often" in the past 12 months. For example, an "Often" response to an item such as "*Do you feel preoccupied with gaming?*" could mean that the respondent experienced minor daily distractions due to gaming but not in an intrusive and interfering manner. It may also be a common occurrence for those who work in game development, retail gaming, or the esports industry. Similarly, in response to an item about deceiving a close family member, or losing an important work opportunity, a "Rarely" response may indicate that this has occurred only once or twice in the past year, but may have nevertheless had serious (and continuing) negative consequences. These examples highlight, particularly for behaviors such as gaming, that pure frequency response categories may be poor at distinguishing between experiences associated with pathology and harm and those which are minor or relatively inconsequential.

In short, the clinical relevance of the symptom may not be consistently related to its frequency.

Another complication is that, on a Likert scale, a "Sometimes" response may be represented as twice as "large" as a "Rarely", when in reality the symptom may have occurred many more times than that. This would occur when a "Sometimes" response is weighted as 2 points and a "Rarely" response is rated 1 point, but the symptom or behavior that occurs "sometimes" may occur much more often – or has much greater impact on functioning – than this weighting would suggest. In this way, the "frequency" scale may give a distorted impression of respondents' experiences, with more extreme symptom profiles brought closer to non-problematic gamers and add to challenges in distinguishing problem and non-problem gamers. A related issue is that that options like "rarely" and "sometimes" might be more affected by subjective interpretation of time. For example, some participants may have difficulty in choosing between options such as "Rarely" and "Sometimes" when they feel that their personal experience falls somewhere within the significant gulf between these two options and feel compelled to select the higher (or lower) option in lieu of a more suitable intermediate option. An alternative may be to offer respondents more specific frequency formulations to choose from (e.g., 'never', 'less than monthly', '1–3 times a month', '1–3 times a week', 'at least 4 times a week'; or, 'never', 'monthly', 'weekly', 'daily or almost daily'), as used in the Alcohol Use Disorders Identification Test (AUDIT), for example. However, unlike the AUDIT, only limited research has examined the psychometric performance of different response categories for distinguishing the different risk or problem levels of gaming disorder symptoms.

The third and final convention in response categories is the agreement scale, which refers to the type (i.e., agree vs disagree) and extent of agreement with self-referential statements about symptoms (e.g., "*I have tried to cut back on my playing, but with no success*": Vadlin, Åslund, & Nilsson, 2015). Like the frequency scale, the agreement response categories are usually presented on a 5-point Likert scale (i.e., 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree or Disagree, 4 = Agree, 5 = Strongly Agree). In the same way, the agreement scale approach enables greater sensitivity, but it is also limited in its capacity to determine the severity of the symptom. For some items, too, it is not clear whether "Strongly Agree" and "Agree" provide different information; for example, "*Sometimes I neglect my school work or skip class in order to play*" may indicate the same thing regardless of conviction. Additionally, agreement ratings are vulnerable to central tendency bias, where respondents tend to avoid selecting the endpoint or extreme option of the scale (Stevens, 1971).

So far, we have reviewed the three most common types of response categories in GD screening and identified some common limitations. Another general difficulty for scales lies in defining the boundary between normative and problematic gaming, given that even a 'healthy passion' (see Vallerand et al., 2003) can sometimes take priority in an individual's life. Video gaming is a recreational activity that is enjoyed by many millions of people, and players form strong motivations to play games. Some aspects of a healthy player-game relationship may be unintentionally highlighted in items that attempt to isolate only the negative and harmful aspects of gaming (King, Chamberlain et al., 2020). One general way of reducing "over-pathologization" (i.e., erroneously treating normal gaming as problematic) may be to prioritise items related to functional impairment in scoring decisions (Billieux et al., 2017). Another method of improving all scales would be to validate scores in clinical samples (e.g., Higuchi et al., 2021), rather than use the conventional approach of recruiting from online convenience samples.

In our view, the GD field could also examine alternative response categories in tools, drawing from approaches used for other mental disorders. For example, the Yale-Brown Obsessive-Compulsive Scale (Goodman et al., 1989) employs response categories tailored to each item and with an emphasis on distress and dysfunction. For example, the item on obsessive thoughts asks "*How much do your obsessive thoughts interfere with your work, school, social, or other important role functioning?*" The two most severe response categories are "substantial impairment in social or occupational performance" and "Incapacitating". For the purpose of screening, these responses are arguably providing more useful information than would be obtained by knowing, for example, the frequency of obsessive thoughts. This approach could be implemented and tested in GD scales, toward improving the specificity of items that are known to be oversensitive, such as preoccupation (Charlton & Danforth, 2007; Infanti, Valls-Serrano, Perales, Vögele, & Billieux, 2023). For example, the IGDT-10 (Király et al., 2017) item "*When you were not playing, how often have you fantasized about gaming, thought of previous gaming sessions, and/or anticipated the next game? Never/Sometimes/Often*" could be modified to "*When not playing, how much does your fantasizing about gaming, thoughts of previous gaming sessions, and/or anticipation to game interfere with your work, school, social, or other important role functioning?*" with accompanying response categories that refer to impairment.

In the addictions field, the Penn Alcohol Craving Scale (PACS: Flannery, Volpicelli, & Pettinati, 1999) may be another useful point of reference for improving GD scales. The PACS item referring to craving experiences provides response categories that refer to urge and the difficulty associated with controlling the urge. This approach could be accommodated within GD tools that refer to withdrawal and other negative mood states associated with gaming (NB: studies using item response theory have shown that items referring to negative mood states have particularly low specificity; Brand, Rumpf, King, Potenza, & Wegmann, 2020). For example, the C-VAT2.0 (Van Rooij et al., 2017) item "*Did you have a strong urge (need) to play video games? Yes/No*" could be modified to "*At its most severe point, how strong was your urge (need) to play games?*" with accompanying response categories that include "strong urge, but easily controlled", "strong urge and difficult to control", and "strong urge and would have played games if it was

available". For these examples, empirical studies could compare the test performance in paired samples, including individuals diagnosed with GD.

There are many possible modifications to GD item response categories that could bring a greater focus on the qualitative nature of the symptom and its negative impact. Some alternative response categories highlighted in this paper do not appear to be any more difficult or time-consuming for respondents than existing approaches. We suggest researchers consider evaluating different response categories to identify whether they aid respondents' comprehension and insight, and better capture pathological behaviours and harms. Further, more studies should employ item response theory to evaluate gaming disorder tools, particularly newly designed ones (of which there are many), to examine how different response categories (e.g., dichotomous vs polytomous) may affect the inherent difficulty and performance of items. This research would contribute to debates on whether some symptoms represent 'core' versus 'peripheral' (or essential and additional) features of GD, and whether some symptoms (e.g., preoccupation) have greater clinical utility when screened more effectively.

REFERENCES

Billieux, J., King, D. L., Higuchi, S., Achab, S., Bowden-Jones, H., Hao, W., … Poznyak, V. (2017). Functional impairment matters in the screening and diagnosis of gaming disorder: Commentary on: Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal (Aarseth et al.). *Journal of Behavioral Addictions*, 6(3), 285–289. https://doi.org/10.1556/2006.6.2017.036.

Brand, M., Rumpf, H. J., King, D. L., Potenza, M. N., & Wegmann, E. (2020). Clarifying terminologies in research on gaming disorder and other addictive behaviors: Distinctions between core symptoms and underlying psychological processes. *Current Opinion in Psychology*, 36, 49–54. https://doi.org/10.1016/j.copsyc.2020.04.006.

Caler, K., Garcia, J. R. V., & Nower, L. (2016). Assessing problem gambling: A review of classic and specialized measures. *Current Addiction Reports*, 3, 437–444. https://doi.org/10.1007/s40429-016-0118-7.

Castro-Calvo, J., King, D. L., Stein, D. J., Brand, M., Carmi, L., Chamberlain, S. R., … Billieux, J. (2021). Expert appraisal of criteria for assessing gaming disorder: An international Delphi study. *Addiction*, 116, 2463–2475. https://doi.org/10.1111/add.15411.

Charlton, J. P., & Danforth, I. D. (2007). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior*, 23, 1531–1548. https://doi.org/10.1016/j.chb.2005.07.002.

Demetrovics, Z., Urbán, R., Nagygyörgy, K., Farkas, J., Griffiths, M. D., Pápay, O., … Oláh, A. (2012). The development of the problematic online gaming questionnaire (POGQ). *Plos One*, 7(5), e36417. https://doi.org/10.1371/journal.pone.0036417.

Flannery, B. A., Volpicelli, J. R., & Pettinati, H. (1999). Psychometric properties of the Penn alcohol craving scale. *Alcoholism: Clinical and Experimental Research*, 23(8), 1289–1295. https://doi.org/10.1097/00000374-199908000-00001.

Gentile, D. (2009). Pathological video-game use among youth ages 8 to 18: A national study. *Psychological Science*, 20(5), 594–602. https://doi.org/10.1111/j.1467-9280.2009.02340.x.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., … Charney, D. S. (1989). The Yale-Brown obsessive-compulsive scale: I. Development, use, and reliability. *Archives of General Psychiatry*, 46, 1006–1011. https://doi.org/10.1001/archpsyc.1989.01810110048007.

Higuchi, S., Osaki, Y., Kinjo, A., Mihara, S., Maezono, M., Kitayuguchi, T., … Saunders, J. B. (2021). Development and validation of a nine-item short screening test for ICD-11 gaming disorder (GAMES test) and estimation of the prevalence in the general young population. *Journal of Behavioral Addictions*, 10, 263–280. https://doi.org/10.1556/2006.2021.00041.

Infanti, A., Valls-Serrano, C., Perales, J.-C., Vögele, C., & Billieux, J. (2023). Gaming passion contributes to the definition and identification of problematic gaming. *Addictive Behaviors*, 147, 107805. https://doi.org/10.1016/j.addbeh.2023.107805.

King, D. L., Billieux, J., Carragher, N., & Delfabbro, P. H. (2020a). Face validity evaluation of screening tools for gaming disorder: Scope, language, and overpathologizing issues. *Journal of Behavioral Addictions*, 9, 1–13. https://doi.org/10.1556/2006.2020.00001.

King, D. L., Chamberlain, S. R., Carragher, N., Billieux, J., Stein, D., Mueller, K., … Delfabbro, P. H. (2020b). Screening and assessment tools for gaming disorder: A comprehensive systematic review. *Clinical Psychology Review*, 77, 101831. https://doi.org/10.1016/j.cpr.2020.101831.

King, D. L., & Delfabbro, P. H. (2019). *Internet gaming disorder: Theory, assessment, treatment, and prevention*. Cambridge, MA: Elsevier Academic Press. ISBN: 9780128129241.

King, D. L., Haagsma, M. C., Delfabbro, P. H., Gradisar, M., & Griffiths, M. D. (2013). Toward a consensus definition of pathological video-gaming: A systematic review of psychometric assessment tools. *Clinical Psychology Review*, 33, 331–342. https://doi.org/10.1016/j.cpr.2013.01.002.

King, D. L., Herd, M. C. E., & Delfabbro, P. H. (2017). Tolerance in internet gaming disorder: A need for increasing gaming time or something else? *Journal of Behavioral Addictions*, 6, 525–533. https://doi.org/10.1556/2006.6.2017.072.

Király, O., Sleczka, P., Pontes, H. M., Urbán, R., Griffiths, M. D., & Demetrovics, Z. (2017). Validation of the ten-item internet gaming disorder test (IGDT-10) and evaluation of the nine DSM-5 internet gaming disorder criteria. *Addictive Behaviors*, *64*, 253–260. https://doi.org/10.1016/j.addbeh.2015.11.005.

Lemmens, J. S., Valkenburg, P. M., & Gentile, D. A. (2015). The internet gaming disorder scale. *Psychological Assessment*, *27*(2), 567–582. https://doi.org/10.1037/pas0000062.

Montag, C., Schivinski, B., Sariyska, R., Kannen, C., Demetrovics, Z., & Pontes, H. M. (2019). Psychopathological symptoms and gaming motives in disordered gaming—A psychometric comparison between the WHO and APA diagnostic frameworks. *Journal of Clinical Medicine*, *8*, 1691. https://doi.org/10.3390/jcm8101691.

Pontes, H. M., & Griffiths, M. D. (2016). Portuguese validation of the internet gaming disorder scale–short-form. *Cyberpsychology, Behavior, and Social Networking*, *19*(4), 288–293. https://doi.org/10.1089/cyber.2015.0605.

Pontes, H. M., Kiraly, O., Demetrovics, Z., & Griffiths, M. D. (2014). The conceptualisation and measurement of DSM-5 internet gaming disorder: The development of the IGD-20 test. *Plos One*, *9*(10), e110137. https://doi.org/10.1371/journal.pone.0110137.

Razum, J., Baumgartner, B., & Glavak-Tkalić, R. (2023). Psychometric validity and the appropriateness of tolerance as a criterion for internet gaming disorder: A systematic review. *Clinical Psychology Review*, *101*, 102256. https://doi.org/10.1016/j.cpr.2023.102256.

Sanders, J. L., & Williams, R. J. (2016). Reliability and validity of the behavioral addiction measure for video gaming. *Cyberpsychology, Behavior, and Social Networking*, *19*(1), 43–48.

Stevens, S. S. (1971). Issues in psychophysical measurement. *Psychological Review*, *78*, 426–450. https://doi.org/10.1037/h0031324.

Vadlin, S., Åslund, C., & Nilsson, K. W. (2015). Development and content validity of a screening instrument for gaming addiction in adolescents: The Gaming Addiction Identification Test (GAIT). *Scandinavian Journal of Psychology*, *56*(4), 458–466. https://doi.org/10.1111/sjop.12196.

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., … Marsolais, J. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, *85*(4), 756–767. https://doi.org/10.1037/0022-3514.85.4.756.

Van Rooij, A. J., Schoenmakers, T. M., & Van de Mheen, D. (2017). Clinical validation of the C-VAT 2.0 assessment tool for gaming disorder: A sensitivity analysis of the proposed DSM-5 criteria and the clinical characteristics of young patients with 'video game addiction'. *Addictive Behaviors*, *64*, 269–274. https://doi.org/10.1016/j.addbeh.2015.10.018.