Journal of Gambling Issues
Volume 44, Spring 2020
http://igi.camh.net/doi/pdf/10.4309/jgi.2020.44.6
DOI: http://dx.doi.org/10.4309/jgi.2020.44.6

Evaluating the Reliability and Validity of the Short Gambling Harm Screen: Are Binary Scales Worse Than Likert Scales at Capturing Gambling Harm?

James McLauchlan,[1] Matthew Browne,[1] Alex M. T. Russell,[2] & Matthew Rockloff[1]

[1] Experimental Gambling Research Laboratory, School of Health, Medical and Applied Sciences, CQUniversity, Bundaberg, Queensland, Australia
[2] Experimental Gambling Research Laboratory, School of Health, Medical and Applied Sciences, CQUniversity, Sydney, New South Wales, Australia

Abstract

Gambling-related harm has become a key metric for measuring the adverse consequences of gambling on a population level. Yet, despite this renewed understanding in contemporary research, little exploration has been conducted to evaluate which instrument is best suited to capture the harmful consequences of gambling. This study was designed with the aim of determining whether Likert scales were better suited to capture gambling harm than binary scales. We hypothesized that the Short Gambling Harm Screen (SGHS), initially scored using a binary scale, would perform similarly to the alternate form that was Likertized for the purpose of this study. A corresponding comparison in the reverse direction was executed for the Problem Gambling Severity Index. The SGHS's performance was assessed via a repeated-measures design in combination with three other measures of validity administered at the conclusion of the survey. In the end, we found that changing the scoring format (i.e., from binary to Likert) had negligible impact on the SGHS's psychometric performance. We conclude that the original scoring method of the SGHS is not only appropriate but also no less suitable than Likert scales in measuring gambling harm.

Keywords: gambling harm, Short Gambling Harm Screen (SGHS), forced-choice binary, dichotomous scale, binary scale, Likert scale comparison, Problem Gambling Severity Index (PGSI)

Résumé

Les dommages liés au jeu sont devenus une mesure clé pour évaluer les conséquences néfastes du jeu à l'échelle de la population. Pourtant, malgré cette compréhension

renouvelée dans la recherche contemporaine, on effectue très peu d'exploration pour évaluer quel instrument est le mieux adapté pour comprendre les conséquences néfastes du jeu. Cette étude a été conçue dans le but de déterminer si les échelles de Likert étaient mieux adaptées que les échelles binaires pour saisir les dommages liés au jeu. Nous avons émis l'hypothèse que le dépistage rapide du jeu problématique (Short Gambling Harm Screen ou SGHS), initialement évalué à l'aide d'une échelle binaire, ne fonctionnera pas différemment de la forme de Likert alternative qui a été créée aux fins de cette étude. Une comparaison correspondante dans la direction inverse a été effectuée pour l'indice de gravité du jeu excessif (PGSI). Les performances du SGHS ont été évaluées par un plan de mesures répétées, combinés à trois autres mesures de validité administrées à la fin du sondage. En fin de compte, nous avons constaté que le changement du format de pointage (c.-à-d. du binaire au Likert) avait un impact négligeable sur le rendement psychométrique du SGHS. Nous concluons que la méthode de pointage originale du SGHS est non seulement appropriée, mais également non moins appropriée que les échelles de Likert pour évaluer les dommages liés au jeu.

Introduction

Contemporary research has focused on gambling-related harm as a key metric of the negative impacts of gambling at the population level (Blaszczynski, 2009; Browne, Greer, Rawat, & Rockloff, 2017; Rodgers, Caldwell, & Butterworth, 2009; Sproston, Erens, & Orford, 2000). The emphasis on harm, rather than gambling disorders, recognizes that traditional measures such as the Problem Gambling Severity Index (PGSI; Ferris & Wynne, 2001) are not well suited to measure the impact of harm on a population level. The need for a new measure of harm was met by a new 10-item screen dedicated to measuring harm—the Short Gambling Harm Screen (SGHS; Browne, Goodwin, & Rockloff, 2017). However, Delfabbro and King (2017) have raised concerns regarding the use of binary scoring of each of the harm symptomology indicators. This dispute raises the question of whether a count of the presence of symptoms, as used by the SGHS, is inferior to measures that elicit degree of frequency or intensity with respect to gambling harm. The present study aimed to evaluate this question via a repeated measures design, in which the performance of the two response formats are compared on several psychometric criteria.

Harm-centred measurement approaches

A population health approach to gambling problems implies that harm, understood as a decrement to health and wellbeing, is the key outcome to be addressed. A corollary to this is that harm can occur on a continuum from mild to severe; and a practical observation is that prevalence is much lower at the severe end of the

spectrum (Browne, Greer, et al., 2017). For instance, Raisamo, Mäkelä, Salonen, and Lintonen's (2014) found that considerable harms were reported even at the lower end of gambling frequency and expenditure levels. A population study conducted in the UK has revealed similar trends, reporting individuals experiencing harms were most prevalent in the lower gambling consumption groups (Canale, Vieno, & Griffiths, 2016). In Australia, Browne and Rockloff (2018) conducted a study assessing the prevalence of harmful consequences across four problem-gambling risk categories, including no-risk, low-risk, moderate-risk, and problem gamblers. The data, again, showed that most gambling-related harms are much more common in combined categories of low-risk gamblers than the high-risk problem gamblers. Together, the evidence suggests there is merit in gauging population-level impact across the spectrum of harm, rather than relying solely on prevalence of problem gamblers as a proxy for harm.

Binary scales or Likert scales?

It is perhaps intuitively appealing to suppose that Likert scales are generally more reliable and accurate than a binary response format because of their potential for capturing more information. However, the extant research suggests this is not generally the case. Grassi et al.'s (2007) study provides a useful illustration. The authors replaced the Likert scales in the 36-item short-form health survey (SF-36) with forced-choice binary scales, and found that the answering format had "no substantial effect" on the test-retest reliability or internal consistency. In another study, Geldhof et al. (2015) compared the responses collected using both binary and Likert format of the Selection Optimisation and Compensation (SOC) questionnaire and concluded that the answering formats were practically interchangeable. Further, in a study published by Litong-Palima, Albers and Glückstad's (2018) the binary format outperformed its Likert counterparts on measures of reliability. Considering research in the marketing context, binary scales have consistently demonstrated similar reliability to Likert scales (Dolnicar & Grün, 2013a; Dolnicar & Grün, 2013b Dolnicar, Grün, & Leisch, 2011; Dolnicar & Leisch, 2012). A common thread running through these studies is the findings that binary scale do not perform significantly differently from their Likert counterparts.

The lack of evidence for the superiority of Likert over binary response formats is counterintuitive considering the greater potential for informational content in an interval scale. Likert scales provide participants with the opportunity to choose from a range of responses to denote a degree of agreement, frequency, or severity. These ordered responses, typically between four to seven points (Adelson & McCoach, 2010), provide the potential for participants to indicate a more precise response to the probe. Nevertheless, there is an absence of guidelines on the way in which Likert scales are to be designed. For instance, there are several options for answer stems (e.g., likely, agree, most of the time, etc.). There is also no definitive way by which the resulting scores should be aggregated. For example, certain researchers advocate for the use of neutral mid-points (Raaijmakers, Van Hoof, 't Hart, Verbogt, &

Vollebergh, 2000; Velez & Ashworth, 2007), while others warn against them (Guy & Norvell, 1977; Wakita, Ueshima, & Noguchi, 2012). The optimal number of rating categories vary from two (McCallum, Keith, & Wiebe, 1988) to eleven (Cummins & Gullone, 2000; Leung, 2011). Certain researchers argue that reliability increases with the number of scale points (Lozano, García-Cueto, & Muñiz, 2008; Weng, 2004), while others have found evidence suggesting that reliability is largely independent of the number of scale points (Bendig, 1954; Komorita, 1963; Matell & Jacoby, 1971).

Theoretical considerations may also explain why Likert response formats do not, in practice, tend to perform better than their binary counterparts for many applications. Given that Likert items typically yield scores (e.g., 0, 1, 2, 3) that are then summed across items to create a scale score, this format requires the strong (item-response theoretic) assumption that each step in the ordered response represents an identical difference of degree on the hypothesized latent construct (Michell, 2012). Binary scales involve only the weaker assumption that the various items are similarly related to, or load onto, the underlying construct. It is also worth considering the higher degree of cognitive effort employed by respondents in answering with a Likert scale, and the degree to which differences in ordered responses might therefore reflect either noise, or a systematic bias in terms of minimising or maximising responses. Binary responses, such as reporting whether an event happened or alternatively whether a symptom is present, are arguably inherently more concrete and less ambiguous, and may therefore be less vulnerable to these forms of error.

Despite the heavy reliance on surveys as the main method for data collection on gambling harm, the question of response format has not yet been explored within gambling research. Even though the SGHS and the FocaL Adult Gambling Screen (FLAGS; Schellinck, Schrans, Schellinck, & Bliemel, 2015) are both scored using a binary response format, neither has been subject to a similar analysis in response to the concerns raised by Delfabbro and King (2017). The aim of the present study is, therefore, to examine the influence of different response formats have on the psychometric properties of the SGHS. More specifically, the research objective is to compare the reliability of the SGHS, initially scored using a binary scale, against a Likert version of SGHS to determine which scale format is more suited for capturing gambling harm. The present study hypotheses that psychometric performance of the binary SGHS will not differ substantially (i.e., the difference will be below the $p < .05$ threshold), in both reliability and validity, from the alternate Likert form.

Methods

Participants

Adult gamblers (n = 618) who gamble at least two to four times a month were recruited for this study through TurkPrime, a North American online research panel

recruitment service. Participants who had missing answers (n = 42), showed pattern responding (n = 17), or scored greater than 2 standard deviations apart in their responses between the repeated measures were excluded (n = 4). Additional multivariate outliers (n = 23) were identified using Mahalanobis distance with a $p < .05$ threshold, and subsequently removed from the sample. A total of 532 (female = 204) participants aged from 18 to 87 (M = 42.07, SD = 13.13) were included for analysis. See Table 1 for the participant demographic characteristic summary.

Design

Participants completed two tests with alternative forms of SGHS and PGSI over a one-week test-retest interval. They were randomly allocated to either complete the same form (i.e., Likert-Likert or Binary-Binary) or the alternative forms (i.e., Likert-Binary or Binary-Likert) at the one-week follow-up. Though the participants might complete different forms of the SGHS and PGSI across time one and time two, the forms did not differ in the same testing (i.e., if a participant received the Binary SGHS at time-one, they will also receive the binary PGSI at time-one). See Table 2 for a summary of the different permutations. Approximately 44% (n = 234) completed repeat assessment of the same form, while the remaining 56% (n = 298) completed the alternate form at follow-up. Participants also completed several other validation measures (described below) at the end of the one-week follow-up survey.

Table 1
Ethnicity, Marital status, Income, and Employment Status Summary

Demographic variable	*n*	%
Ethnicity		
White	372	69.9%
Hispanic, Latino or Spanish	42	7.9%
African American	96	18.0%
Asian	17	3.2%
American Indian or Alaskan Native	2	.4%
Middle Eastern or North African	1	.2%
Other	2	.4%
Marital Status		
Married	238	44.7%
De facto	3	.6%
Separated	16	3.0%
Divorced	45	8.5%
Widowed	11	2.1%
Never married	219	41.2%
Income		
Less than $24,999	82	15.4%
$25,000 to $49,999	165	31.0%
$50,00 to $74,999	125	23.5%
$75,000 to $99,999	88	16.5%
$100,000 to $124,999	44	8.3%
$125,000 or more	28	5.3%

Table 2
Measures used in each permutation and the number of participants randomly allocated into each set

Permutation	Type	T1	T2	*N*	%
Permutation A (Likert-Likert)	Test-retest	SGHS L & PGSI L	SGHS L, PGSI L, PWI, BIS & K6	119	22.4
Permutation B (Binary-Binary)	Test-retest	SGHS B & PGSI B	SGHS B, PGSI B, PWI, BIS & K6	115	21.6
Permutation C (Likert-Binary)	Alternate form	SGHS L & PGSI L	SGHS B, PGSI B, PWI, BIS & K6	150	28.2
Permutation A (Binary-Likert)	Alternate form	SGHS B & PGSI B	SGHS L, PGSI L, PWI, BIS & K6	148	27.8

Note. L = Likert form; B = Binary form.

Table 3
Test-retest reliability, means and standard deviation of the SGHS and PGSI in each form

	Time 1		Time 2		Test-retest reliability	
Test	*M*	*SD*	*M*	*SD*	*r*	*p*
SGHS L	9.94	9.10	10.55	9.47	.89	< .001
PGSI L	9.09	8.21	9.62	8.75	.89	< .001
SGHS B	4.11	3.47	4.30	3.46	.85	< .001
PGSI B	3.78	3.23	3.87	3.21	.86	< .001

Note. L = Likert form; B = Binary form.

Procedure

Participants were recruited to participate in two online surveys. They were compensated in the form of either reward points, cash or gift cards of their choice. This study was approved by the Central Queensland University Ethics Committee (approval number 0000021464), and informed consent was obtained at the outset of the first survey.

Analysis

Each measure's internal consistency was calculated using Cronbach's alpha calculated on either the tetrachoric (for binary) or polychoric (for Likert) item correlation matrix. The SGHS's test-retest reliability, alternate-form reliability, convergent validity and discriminant validity was computed using Spearman correlations. Comparisons between two forms of SGHS was done using Fisher's Z test (Myers & Sirois, 2006), which provides a test of significance between non-parametric correlation coefficients by converting them into standardized (z) scores (Zar, 2005).

Table 4
Correlation between SGHS, PGSI, BIS-Brief, PWI and K6

	PGSI L T1	SGHS L T1	PGSI B T1	SGHS B T1	PGSI L T2	SGHS L T2	PGSI B T2	SGHS B T2
PGSI L T1	-	-	-	-	-	-	-	-
SGHS L T1	.94**	-	-	-	-	-	-	-
PGSI B T1	-	-	-	-	-	-	-	-
SGHS B T1	-	-	.83**	-	-	-	-	-
PGSI L T2	.91**	.88**	.73**	.72**	-	.-	-	-
SGHS L T2	.87**	88**	.70**	.75**	.91**	-	-	-
PGSI B T2	.69**	.71**	.86**	.77**	-	-	-	-
SGHS B T2	.68**	.74**	.83**	.85**	-	-	.84**	-
BIS	.29**	.23**	.20**	.22**	.25**	.28**	.15*	.17*
PWI	.04	-.04	-.09	-.12*	-.07	-.11	-.04	-.10
K6	.64**	67**	54**	.57**	.66**	.68**	.55**	.60**

Note. * = correlation is significant at the .05 level ($p < .05$, two-tailed); ** = correlation is significant at the .001 level ($p < .001$, two-tailed).

Measures

In addition to the SGHS and PGSI, several measures were included to assess external validity for each version of the scale. The Kessler Psychological Distress Scale (K6; Kessler et al., 2002) and the Personal Wellbeing Index (PWI; Cummins, 1997) served as outcome-oriented validity; that is, harms and problems should be associated with increased distress and lower wellbeing. Additionally, impulsivity is a known risk factor for both gambling problems and harm (Browne et al., 2019; Russell, Hing, Li, & Vitartas, 2019), and therefore trait-impulsivity was measured using the Barratt Impulsiveness Scale-Brief (BIS-Brief; Steinberg, Sharp, Stanford, & Tharp, 2013).

Gambling Harm. The SGHS (Browne, Goodwin et al., 2017) is designed to assess a respondent's degree of gambling-related harm. The original SGHS is scored on a forced-choice binary scale and contains 10 items derived from a comprehensive 72-item checklist (Browne, Goodwin et al., 2017, Langham et al., 2015). The SGHS includes harmful consequences that are more prevalent (e.g., "decreased savings" or "sold personal items") and uses a binary response scale. It has been shown to be a good proxy of the full checklist (r = .94) and enjoys a negative linear relationship to wellbeing (Browne, Goodwin et al., 2017). This scale was compared with a four-point Likert version of the same measure with 0 being "never," 1 denoting "sometimes," 2 representing "most of the time," and 3 indicating "almost always." Alpha reliability in the present study for the original binary SGHS was .95 at both times one and two. The Likert-scored SGHS alpha reliability was also identical across surveys: .97 at time one and two. Refer to Table 6 for the combined response distribution of both forms of SGHS.

Problem Gambling. Participants completed the PGSI (Ferris and Wynne, 2001), a standard tool for assessing degree of gambling problems in surveys. The nine-item PGSI contains questions such as "has gambling caused you any health problems, including stress or anxiety?" It is scored on a four-point Likert scale, with 0 representing "never" and 3 indicating "almost always." An alternate binary form of the PGSI was also calculated for this study. Alpha reliability in this study for the standard (Likert) PGSI was .96 at time one and .98 at time two; the alternative binary version had a reliability of .89 at both time one and two. See Table 5 for the response distribution of the PGSI.

Psychological Distress. The Kessler Screening Scale for Psychological Distress (K6; Kessler et al., 2002) was chosen to measure the presence of distress among the participants. The K6 consists of six items scored on a five-point Likert scale from 0 ("none of the time") to 4 ("all the time"). Coefficient alpha in the current study was .94. As noted above, the SGHS should predict greater psychological distress (Brown, Oldenhoff, Allen, & Dowling, 2016).

Personal Wellbeing. The PWI, adapted from the Comprehensive Quality of Life Scale (Cummins, 1997) was used to measure general life satisfaction. It is an eight-item questionnaire designed to measure multiple domains associated with quality of

Table 5
Combined response distribution (T1 and T2) for Likert and binary PGSI

	Likert				Binary	
Item	0	1	2	3	0	1
Have you bet more than you could really afford to lose?	185 (34.52%)	201 (37.5%)	81 (15.11%)	69 (12.69%)	272 (51.52%)	256 (48.48%)
Have you needed to gamble with larger amount of money to get the same feeling of excitement?	202 (37.62%)	150 (27.93%)	116 (21.6%)	69 (12.85%)	278 (52.06%)	250 (47.94%)
When you gambled, did you go back another day to try to win back the money you lost	142 (26.49%)	176 (32.84%)	126 (23.51%)	92 (17.16%)	251 (47.54%)	277 (52.46%)
Have you borrowed money or sold anything to get money to gamble?	283 (51.36%)	99 (17.96%)	86 (15.61%)	68 (12.34%)	282 (53.41%)	246 (46.59%)
Have you felt that you might have a problem with gambling?	230 (42.91%)	137 (25.56%)	86 (16.05%)	83 (15.49%)	360 (65.34)	191 (34.66)
Have people criticized your betting or told you that you had a gambling problem, regardless of whether or not you thought it was true?	254 (47.39%)	119 (22.20%)	87 (16.23%)	76 (14.18%)	311 (58.9%)	217 (41.1%)
Have you felt guilty about the way you gamble or what happens when you gamble?	205 (38.25%)	140 (26.12%)	109 (20.34%)	82 (15.3%)	276 (52.27%)	252 (47.73%)
Has gambling caused you any health problems, including stress or anxiety?	281 (51.85%)	113 (20.85%)	76 (14.02%)	66 (12.18%)	342 (64.77%)	186 (35.23%)
Has your gambling caused any financial problems for you or your household?	264 (49.25%)	97 (18.1%)	90 (16.79%)	85 (15.86%)	347 (64.5%)	191 (35.5%)

Note. For the Likert response, 0 = never, 1 = sometimes, 2 = most of the time, 3 = almost always. In the binary format, 0 = no, 1 = yes.

Table 6
Combined Response distribution (T1 and T2) for binary and Likert format of SGHS

	Likert				Binary	
Item	0	1	2	3	0	1
Reduction of my available spending money	129 (24.07%)	223 (41.6%)	105 (19.59%)	79 (14.74%)	215 (40.72%)	313 (59.28%)
Reduction of my savings	211 (39.34%)	154 (28.73%)	98 (18.49%)	73 (13.77%)	290 (54.92%)	238 (45.08%)
Less spending on recreational expenses	155 (28.92%)	195 (36.38%)	112 (18.28%)	74 (13.81%)	237 (44.89%)	291 (55.11%)
Had regrets that made me feel sorry about my gambling	197 (36.75%)	155 (28.92%)	104 (19.4%)	80 (14.93%)	259 (49.05%)	269 (50.95%)
Feel ashamed of my gambling	246 (44.48%)	132 (23.91%)	88 (15.91%)	87 (15.73%)	330 (62.5%)	198 (37.5%)
Sold personal items	301 (56.37%)	97 (18.16%)	69 (12.92%)	67 (12.55%)	378 (71.59%)	150 (28.41%)
Increased credit card debt	282 (52.61%)	101 (18.84%)	85 (15.86%)	68 (12.69%)	356 (67.42%)	172 (32.58%)
Spent less time with people I care about	247 (46.08%)	117 (21.83%)	105 (19.59%)	67 (12.5%)	329 (62.31%)	199 (37.68%)
Felt distressed about my gambling	252 (47.01%)	118 (22.01%)	89 (16.60%)	77 (14.37%)	336 (63.64%)	192 (36.74%)
Felt like a failure (545)	261 (47.89%)	129 (23.67%)	85 (15.6%)	70 (12.84%)	330 (62.5%)	198 (37.5%)

Note. For the Likert response, 0 = never, 1 = sometimes, 2 = most of the time, 3 = almost always. In the binary format, 0 = no, 1 = yes.

life, including living standards, health, achievements, safety, sense of belonging and future prospects. These items are scored on an 11-point scale, with 0 denoting "no satisfaction at all" and 10 indicating "completely satisfied." Alpha reliability in this study was .93. The SGHS should predict lower wellbeing (Blackman, Browne, Rockloff, Hing, & Russell, 2019).

Impulsiveness. Impulsiveness was assessed via the Barratt Impulsiveness Scale-Brief (BIS-Brief; Steinberg et al., 2013), an abbreviated version of its 30-item predecessor (Barratt, 1959). The BIS-brief consists of eight items and uses a four-point Likert scale to capture the degree to which the participants agreed with statements such as "I don't pay attention" or "I act on the spur of the moment." Cronbach's alpha in the present study was .64. Behavioural impulsivity is a risk factor for gambling problems and harm (Russell et al., 2019).

Results

Test-retest reliability of the SGHS

The test-retest results are summarized and presented in Table 3. Both forms of SGHS showed strong test-retest reliability, being strongly correlated between time one and time two (all $p < .001$): .86 for the binary form, and .88 for the Likert form. These correlations were not significantly different, $Z = -1.27$, $p = .10$.

Alternate-form reliability of the SGHS

Alternate form reliability was assessed by comparing the correlations between different forms of the SGHS across time one and time two. The similar form test-retest correlations mentioned above (.86 / .88) formed the benchmark with which to evaluate the alternative forms. When comparing alternate forms across Time One and Time Two, the correlations were .75 and .74 for the Binary-Likert and Likert-Binary administration, respectively. Comparing test-retest reliability across forms (approx .745) and within forms (approximately .87) allows us to estimate the variance uniquely attributable to varying the form along. Squaring these correlations to approximate the proportion of shared variance, this finding corresponds to approximately 76% of shared variance, or 24% of the variance attributable to random effects over time. In the case of alternate forms, in which error is attributable to *both* random time effects and differing response formats, approximately 56% of variance was shared. Thus, 20% of variance in responses can be attributed to the differing forms.

Convergent validity of the SGHS

The results of the correlations between each form of SGHS, K6, PWI and BIS are summarized in Table 5.

Time one. There were no significant differences between the correlations of the binary and Likert response formats with external measures. The correlation with the K6 was .57 for the binary form and .67 for the Likert form. Although the Likert form performed better, the difference between these correlations were not significantly different, $Z = 1.87$, $p = .06$. With regards to the BIS, the correlation was .22 for the binary scale and .23 for the Likert scale, $Z = -.12$, $p = .90$. The correlation between SGHS and the PWI was -.12 for the binary scale and -.04 for the Likert scale. This difference in correlation between the two scales was also not statistically significant, $Z = .92$, $p = .36$.

Time two. At time two, the correlation between SGHS and K6 was .60 for the binary scale and .68 for the Likert scale. This difference in the size of the correlation between the two scales was not statistically significant, $Z = 1.56$, $p = .12$. The correlation with the BIS was .17 for the binary scale and .28 for the Likert scale. There was no significant difference between the two correlations, $Z = 1.33$, $p = .18$. As for the PWI, the correlation was .10 for the binary scale and .11 for the Likert scale. This difference in the size of the correlation between the two scales was again not statistically significant, $Z = .12$, $p = .90$.

Concurrent Validity of the SGHS

At time one, the correlation between the SGHS and PGSI was .84 for the binary (binary SGHS-binary PGSI) and .94 for the Likert form (Likert SGHS-Likert PGSI). This difference in the size of the correlation between the two scales was statistically significant, $Z = -6.29$, $p < .001$. For time two, the correlation between the SGHS and PGSI was .84 for the binary and .91 for the Likert form. This difference in the size of the correlation between the two scales was also statistically significant, $Z - 3.51$, $p < .001$.

Discussion

This study investigated the psychometric properties of the Short Gambling Harm Screen and compared whether scoring methods had an influence on said properties. We hypothesized that the performance of the binary SGHS would not differ significantly compared to its alternate, Likertized form in terms of reliability and validity. As predicted, the type of scales employed did not have substantial effects on the SGHS's internal consistency, test-retest, and alternate-form reliability. We found that both the binary and Likert format produced similar results with respect to convergent validity and discriminant validity. We note a general pattern of the Likert scale performing slightly better than the binary scale, but also note that differences between the performance of the scales was not statistically significant for any analysis, and that effect sizes were small.

In terms of concurrent validity, however, our data did suggest that the Likert format correlated significantly higher with the PGSI than the binary version, although the difference in correlations was not large. This may be because of the fact that the

PGSI is designed to detect problem gamblers, who lie at the extreme end of the spectrum of harm, while the SGHS was intended to capture gambling harm across a broader scale. As such, it is possible that the Likert version correlated better with the PGSI because it allows for the detection of more extreme states of harm compared to those who experience it only occasionally.

Binary scales come with several other practical advantages beyond reliability and validity estimates. As discussed above, force-choice questionnaires, such as the SGHS, are quicker to administer and less ambiguous than Likert responses. Its concise nature helps to mitigate certain of the common artefacts observed in health surveys, particularly with respondent fatigue (O'Reilly-Shah, 2017). Moreover, unlike Likert scales, which involve strong psychometric assumptions (Michell, 2012), binary scales involves the more limited assumption that each item should provide independent information about the underlying construct it is measuring. Finally, the interpretation of Likert scales in this context is more straightforward: the score reflects the number of distinct symptoms the individual is presenting. Our study is primarily limited by a relatively small sample size. Our conclusion suggests that there are no differences in psychometric properties between the two forms, but our significance tests are sensitive to sample size. That is, small samples tend to produce null results. A larger sample size could detect more subtle differences in item performance between the two forms. Nevertheless, our current results suggest that differences between the forms, if they exist, are likely to be small. Another limitation of our results is that we only compared two response scales: force-choice binary and four-point Likert scales. Future studies could perhaps explore other formats to ensure these findings can be generalized in another context as well.

Overall, our findings suggest that response format did not yield a major impact on measures of reliability and validity. These results appear to resolve Delfabbro and King's (2017) concern that the binary scoring might have a negative impact on the validity of the SGHS. Finally, the present study also offers some practical advice for the use of forced-choice binary response scales in psychological testing in general; that is, at least in the context of measuring gambling harm, there is no reason to assume that a binary response format is any less suited than other answering scales at capturing participant responses.

Conclusion

The question surrounding whether binary scales are suited to measure gambling harm was raised as a key concern for the use of SGHS in population surveys. In this study, we hypothesized that the scoring format should not have a substantial impact on the SGHS's reliability or validity. Our data demonstrated that while there was one slight difference in concurrent validity, the binary version of the SGHS did not generally perform significantly differently to the Likert scales on several measures of scale performance. Consequently, we tentatively conclude that the binary format used to score SGHS is just as effective as Likert-type scales.

References

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement, 70*, 796–807. https://doi.org/10.1177/0013164410366694

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills*, *9*, 191–198. https://doi.org/10.2466/pms.1959.9.3.191

Bendig, A. W. (1954). Reliability and the number of rating-scale categories. *Journal of Applied Psychology, 38*, 38–40. https://doi.org/10.1037/h0055647

Blackman, A., Browne, M., Rockloff, M., Hing, N., & Russell, A. M. T. (2019). Contrasting effects of gambling consumption and gambling problems on subjective wellbeing. *Journal of Gambling Studies*, *35*, 773–792. https://doi.org/10.1007/s10899-019-09862-z

Blaszczynski, A. (2009). Problem gambling: We should measure harm rather than "cases." *Addiction, 104*, 1072–1074. https://doi.org/10.1111/j.1360-0443.2009.02505.x

Brown, M., Oldenhof, E., Allen, J. S., & Dowling, N. A. (2016). An empirical study of personality disorders among treatment-seeking problem gamblers. *Journal of Gambling Studies*, *32*, 1079–1100. https://doi.org/10.1007/s10899-016-9600-3

Browne, M., Goodwin, B. C., & Rockloff, M. J. (2017). Validation of the Short Gambling Harm Screen (SGHS): A tool for assessment of harms from gambling. *Journal of Gambling Studies, 34*, 499–512. https://doi.org/10.1007/s10899-017-9698-y

Browne, M., Greer, N., Rawat, V., & Rockloff, M. (2017). A population-level metric for gambling-related harm. *International Gambling Studies, 17*, 163–175. https://doi.org/10.1080/14459795.2017.1304973

Browne, M., Hing, N., Rockloff, M., Russell, A. M. T., Greer, N., Nicoll, F., & Smith, G. (2019). A multivariate evaluation of 25 proximal and distal risk-factors for gambling-related harm. *Journal of Clinical Medicine*, *8*, 509–524. https://doi.org/10.3390/jcm8040509

Browne, M., & Rockloff, M. J. (2018). Prevalence of gambling-related harm provides evidence for the prevention paradox. *Journal of Behavioral Addictions 7*, 410–422. https://doi.org/10.1556/2006.7.2018.41

Canale, N., Vieno, A., & Griffiths, M. D. (2016). The extent and distribution of gambling-related harms and the prevention paradox in a British population survey. *Journal of Behavioral Addictions, 5*, 204–212. https://doi.org/10.1556/2006.5.2016.023

Cummins, R. A. (1997). *Comprehensive quality of life scale: Adult: Manual* (5th ed., pp. 1–51). Melbourne, Australia: School of Psychology, Deakin University.

Cummins, R. A., & Gullone, E. (2000, March). *Why we should not use 5-point Likert scales: The case for subjective quality of life measurement*. Paper presented at the Proceedings, second international conference on quality of life in cities. https://www.worldcat.org/title/icqolc-2000-proceedings-of-the-second-international-conference-on-quality-of-life-in-cities-21-century-qol-8-10-march-2000-westin-stamford-hotel-singapore/oclc/52617774

Delfabbro, P., & King, D. (2017). Prevention paradox logic and problem gambling: Does low-risk gambling impose a greater burden of harm than high-risk gambling? *Journal of Behavioral Addictions, 6*, 163–167. https://doi.org/10.1556/2006.6.2017.022

Dolnicar, S., & Grün, B. (2013a). "Translating" between survey answer formats. *Journal of Business Research, 66*, 1298–1306. https://doi.org/10.1016/j.jbusres.2012.02.029

Dolnicar, S., & Grün, B. (2013b). Validly measuring destination image in survey studies. *Journal of Travel Research, 52*, 3–14. https://doi.org/10.1177/0047287512457267

Dolnicar, S., Grün, B., & Leisch, F. (2011). Quick, simple and reliable: Forced binary survey questions. *International Journal of Market Research, 53*, 231–252. https://doi.org/10.2501/ijmr-53-2-231-252

Dolnicar, S., Rossiter, J. R., & Grün, B. (2012). "Pick Any" measures contaminate brand image studies. *International Journal of Market Research, 54*, 821–834. https://doi.org/10.2501/ijmr-54-6-821-834

Ferris, J. A., & Wynne, H. J. (2001). *The Canadian Problem Gambling Index: Final report*. Ottawa, ON: Canadian Centre on Substance Abuse.

Geldhof, G. J., Gestsdottir, S., Stefansson, K., Johnson, S. K., Bowers, E. P., & Lerner, R. M. (2015). Selection, optimization, and compensation: The structure, reliability, and validity of forced-choice versus Likert-type measures in a sample of late adolescents. *International Journal of Behavioral Development, 39*, 171–185. https://doi.org/10.1177/0165025414560447

Grassi, M., Nucera, A., Zanolin, E., Omenaas, E., Anto, J. M., & Leynaert, B. (2007). Performance comparison of Likert and binary formats of SF-36 version 1.6 across ECRHS II adults populations. *Value in Health, 10*, 478–488. https://doi.org/10.1111/j.1524-4733.2007.00203.x

Guy, R. F., & Norvell, M. (1977). The Neutral point on a Likert scale. *The Journal of Psychology, 95*, 199–204. https://doi.org/10.1080/00223980.1977.9915880

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*, 959–976. https://doi.org/10.1017/S0033291702006074

Komorita, S. S. (1963). Attitude content, intensity, and the neutral point on a Likert scale. *The Journal of Social Psychology, 61*, 327–334. https://doi.org/10.1080/00224545.1963.9919489

Langham, E., Thorne, H., Browne, M., Donaldson, P., Rose, J., & Rockloff, M. (2015). Understanding gambling related harm: A proposed definition, conceptual framework, and taxonomy of harms. *BMC Public Health, 16*, 80. https://doi.org/10.1186/s12889-016-2747-0

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-Point Likert scales. *Journal of Social Service Research, 37*, 412–421. https://doi.org/10.1080/01488376.2011.580697

Litong-Palima, M., Albers, K. J., & Glückstad, F. K. (2018, June). *Stability and similarity of clusters under reduced response data* [Paper presentation]. 32nd Annual Conference of the Japanese Society for Artificial Intelligence, Kagoshima, Japan.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*, 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*, 657–674. https://doi.org/10.1177/001316447103100307

McCallum, D. M., Keith, B. R., & Wiebe, D. J. (1988). Comparison of response formats for Multidimensional Health Locus of Control Scales: Six levels versus two levels. *Journal of Personality Assessment, 52*, 732–736. https://doi.org/10.1207/s15327752jpa5204_12

Michell, J. (2012). "The constantly recurring argument": Inferring quantity from order. *Theory & Psychology, 22*, 255–271. https://doi.org/10.1177/0959354311434656

Myers, L., & Sirois, M. (2006). Spearman correlation coefficients, differences between. In S. Kotz, C. B. Read, N. Balakrishnan, & B. Vidakovic (Eds.), *Encyclopedia of statistical sciences* (2nd ed. Pp. 7901-7903). Hoboken, NJ: Wiley-Interscience. https://doi.org/10.1002/0471667196.ess5050.pub2

Raaijmakers, Q. A. W., Van Hoof, J. T. C., 't Hart, H., Verbogt, T. F. M. A., & Vollebergh, W. A. M. (2000). Adolescents' midpoint responses on Likert-type scale

items: Neutral or missing values? *International Journal of Public Opinion Research, 12*, 208–216. https://doi.org/10.1093/ijpor/12.2.209

Raisamo, S. U., Mäkelä, P., Salonen, A. H., & Lintonen, T. P. (2014). The extent and distribution of gambling harm in Finland as assessed by the Problem Gambling Severity Index. *The European Journal of Public Health, 25*, 716–722. https://doi.org/10.1093/eurpub/cku210

O'Reilly-Shah, V. (2017). Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. *Journal of Life and Environmental Sciences, 5*,1–17. https://doi.org/10.7717/peerj.3785

Rodgers, B., Caldwell, T., & Butterworth, P. (2009). Measuring gambling participation. *Addiction, 104*, 1065–1069. https://doi.org/10.1111/j.1360-0443.2008.02412.x

Russell, A. M. T., Hing, N., Li, E., & Vitartas, P. (2019). Gambling risk groups are not all the same: Risk factors amongst sports bettors. *Journal of Gambling Studies*, *35*, 225–246. https://doi.org/10.1007/s10899-018-9765-z

Schellinck, T., Schrans, T., Schellinck, H., & Bliemel, M. (2015). Construct development for the FocaL Adult Gambling Screen (FLAGS): A risk measurement for gambling harm and problem gambling associated with electronic gambling machines. *Journal of Gambling Issues*, *30*, 140–173. https://doi.org/10.4309/jgi.2015.30.7

Sproston, K., Erens, B., & Orford, J. (2000). *Gambling behaviour in Britain: Results from the British gambling prevalence survey*. London, UK: National Centre for Social Research.

Steinberg, L., Sharp, C., Stanford, M. S., & Tharp, A. T. (2013). New tricks for an old measure: The development of the Barratt Impulsiveness Scale–Brief (BIS-Brief). *Psychological Assessment*, *25*, 216–226. https://doi.org/10.1037/a0030550

Velez, P., & Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods*, *1*, 69–74. https://doi.org/10.18148/srm/2007.v1i2.76

Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement, 72*, 533–546. https://doi.org/10.1177/0013164411431162

Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64*, 956–972. https://doi.org/10.1177/0013164404268674

Zar, J. H. (2005). Spearman rank correlation: Overview. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics* (2nd ed., Vol. 7, 5095–5101). Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9781118445112.stat05964

****

Submitted May 27, 2020; accepted June 8, 2020. This article was peer reviewed. All URLs were available at the time of submission.

For correspondence: James Robert Bell McLauchlan, BPsych, Experimental Gambling Research Laboratory, 38 Regent Avenue, Springvale, 3171, Victoria, Australia. E-mail: james_mclauchlan@outlook.com

Competing interests: None declared (all authors).

Ethics approval: This study was approved by the Central Queensland University Ethics Committee (Approval number - 0000021464) on May 10, 2019.

Acknowledgements: This research was supported in part by grants from the 2018 Summer Research Scholarship program at CQUniversity. I would like to express my deepest gratitude to my supervisor, Professor Matthew Browne, for his patience, unwavering support, and constructive feedback. His willingness to offer his time so generously has been very much appreciated. My grateful thanks also extend to Dr Alex Russell for assisting with data analysis and to Professor Matthew Rockloff, who has helped with the design of the study and editing the manuscript.