

An Item Response Theory Analysis of the Problem Gambling Severity Index

Carla Sharp, Lynne Steinberg, Ilya Yaroslavsky, Andre Hofmeyr, Andrew Dellis, Don Ross and Harold Kincaid
Assessment 2012 19: 167 originally published online 19 August 2011
DOI: 10.1177/1073191111418296

The online version of this article can be found at:
<http://asm.sagepub.com/content/19/2/167>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Assessment* can be found at:

Email Alerts: <http://asm.sagepub.com/cgi/alerts>

Subscriptions: <http://asm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://asm.sagepub.com/content/19/2/167.refs.html>

>> [Version of Record](#) - May 14, 2012

[OnlineFirst Version of Record](#) - Aug 19, 2011

[What is This?](#)

An Item Response Theory Analysis of the Problem Gambling Severity Index

Assessment
19(2) 167–175
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191111418296
http://asm.sagepub.com



Carla Sharp¹, Lynne Steinberg¹, Ilya Yaroslavsky¹, Andre Hofmeyr², Andrew Dellis², Don Ross^{2,3}, and Harold Kincaid⁴

Abstract

Increases in the availability of gambling heighten the need for a short screening measure of problem gambling. The Problem Gambling Severity Index (PGSI) is a brief measure that allows for the assessment of characteristics of gambling behavior and severity and its consequences. The authors evaluate the psychometric properties of the PGSI using item response theory methods in a representative sample of the urban adult population in South Africa ($N = 3,000$). The PGSI items were evaluated for differential item functioning (DIF) due to language translation. DIF was not detected. The PGSI was found to be unidimensional, and use of the nominal categories model provided additional information at higher values of the underlying construct relative to a simpler binary model. This study contributes to the growing literature supporting the PGSI as the screen of choice for assessing gambling problems in the general population.

Keywords

problem gambling severity index, item response theory analysis, population screening

Problem gambling refers to gambling behavior that causes negative consequences for the gambler, others in the social network of the gambler, or for the community (Ferris & Wynne, 2001). Against the background of growing concerns about the increasing availability of gambling, several self-report population-based screens of problem gambling have been developed (Holtgraves, 2009). One such screen, the Problem Gambling Severity Index (PGSI), which is the scored component of the Canadian Problem Gambling Index (Ferris & Wynne, 2001), was designed to provide an alternative to the more frequently used South Oaks Gambling Screen (SOGS; Lesieur & Blume, 1987). The SOGS has received much criticism for taking a binary categorical and “medical” view of problem gambling and for insufficient focus on the social and environmental aspects of problem gambling. Because the SOGS was developed specifically for use in clinical settings, it does not include items for less severe behavioral indicators. The SOGS has therefore been criticized as underidentifying individuals with subthreshold problem gambling (Holtgraves, 2009; Strong, Breen, Lesieur, & Lejuez, 2003). The SOGS also fails to perform well in determining prevalence rates in the general population (Culleton, 1989; Holtgraves, 2009) and typically fails to demonstrate an underlying single factor that explains at least 50% of the variance characteristic of most population screens (Arthur et al., 2008).

In contrast to the SOGS, the PGSI was developed specifically to measure problem gambling in the general population. Instead of categorizing individuals as nonproblem

gamblers or pathological gamblers (a dichotomous 0/1 classification), the PGSI is able to identify different subgroups of problem gamblers with different levels of risk status (no, low, moderate, and high). This feature is especially important in epidemiological research where the aim is often the identification of those at risk for developing a disorder (Sharp, Goodyer, & Croudace, 2006). Despite the PGSI’s promise (Neal, Delfabbro, & O’Neill, 2004), there are several factors that limit its use.

First, few studies beyond those by the PGSI developers have been conducted to investigate its psychometric properties (Brooker, Clara, & Cox, 2009; Holtgraves, 2009). Second, although the PGSI has been investigated in samples from Canada (Ferris & Wynne, 2001), Australia (McMillen & Wenzel, 2006), Great Britain (Orford, Wardle, Griffiths, Sproston, & Erens, 2010), China (Loo, Oei, & Raylu, 2010), and Singapore (Arthur et al., 2008), it has been examined in only one developing or poor country (China). There are several indices to determine whether a country falls in the “developing” category. According to the

¹University of Houston, Houston, TX, USA

²University of Cape Town, Rondebosch, South Africa

³Georgia State University, Atlanta, GA, USA

⁴University of Alabama at Birmingham, Birmingham, AL, USA

Corresponding Author:

Carla Sharp, Department of Psychology, University of Houston, Houston, TX 77024, USA

Email: csharp2@uh.edu

National Institutes of Health, developing countries are defined as the low- and middle-income economies. Using World Bank classifications, which in turn rely on gross national income (GNI) per capita, the GNI for South Africa is \$5,760, compared with \$41,980 for Canada, \$41,370 for the United Kingdom, \$43,770 for Australia, and \$37,220 for Singapore. The GNI for China is \$3,650, but this should be interpreted in the context of China as an emerging super economy. Moreover, although both South Africa and China may be considered developing countries, significant differences, both economic and cultural, exist that may lead to different prevalence in gambling.

Overreliance on data collected from wealthy countries impedes the use of the PGSI to detect problem gambling among population sectors that are largely absent in such countries. The poor have been shown to spend higher proportions of their income on gambling than their wealthier counterparts (Schissel, 2001). Given the explicit aim of the PGSI to be sensitive to risk and harm associated with problem gambling, it is evident that its psychometric properties merit investigation in a country, such as South Africa, where far higher proportions of people are much more vulnerable in the face of personal financial losses. Moreover, variation in the availability and accessibility of gambling across different countries may affect the significance of problem gambling as a public health problem, further warranting gambling research in non-Western countries. South Africa operates a modern regulated environment for legal gambling comparable to that found in most advanced countries. There are private casinos, both stand-alone and resort style, operating according to well-monitored and well-enforced license conditions in all major population centers. Most are accessible only by private automobile and do not attract significant patronage except from wealthy and upper-middle-class citizens and tourists. There is a national lottery, recently suffering from sharply declining participation, and scratch cards for draws in running lotteries with smaller prizes on sale in small shops everywhere. In all these respects, South Africa's gambling environment would seem familiar to a resident of a typical U.S. state. However, South Africa also has a large network of illegal and informal small gambling venues concentrated in poorer communities. Almost all gambling by the 70% of South Africans who live below the U.S. poverty line occurs in these venues or in the context of widespread street lotteries and card and dice games run by organized criminal interests.

Another limitation of prior psychometric studies of the PGSI is that most have relied on classical test theory (CTT) approaches to data analyses in lieu of latent trait approaches. Although CTT has served scale development well over the last half century, there are several advantages of using latent trait approaches to determine the internal construct validity of a measure. Comprehensive reviews of these advantages exist and readers are referred to these (e.g., Embretson &

Reise, 2000). Here, we will briefly discuss some of the advantages of latent trait approaches such as full information confirmatory factor analysis (CFA) and item response theory (IRT), as they pertain to the evaluation of the internal factor structure of population screening instruments developed for clinical epidemiological research, such as the PGSI. First, CTT approaches, such as factor analysis or principal components analysis, assume that item responses are on a continuous metric. However, item responses using Likert-type response scale options are categorical. For population screens of clinical symptoms, such as the PGSI, the modal response is usually zero, representing the endorsement of the "never" response category (Sharp et al., 2006); as a result, population screens tend to produce skewed data. The magnitude of Pearson correlations, the basis of CTT approaches, is influenced by endorsement rates (often referred to as difficulty factors). Full-information CFA and IRT analyze the categorical item responses (rather than, e.g., Pearson or polychoric correlations, which are referred to as limited information). In addition, IRT provides information useful for evaluating the performance of each item; items that underperform may then be removed and tests may be shortened. This is of particular importance for clinical population screens such as the PGSI, where it is often desirable to have brief instruments to reduce the burden on respondents.

The methods of IRT can also be applied to the detection of differential item functioning (DIF). The investigation of DIF is particularly important for cross-cultural adaptation of clinical screens. DIF analysis evaluates whether items function differently for certain cultural, racial, or language groups after taking into account any group mean difference. In the context of the present study, we apply these methods to investigate DIF between four of the language groups that responded to the PGSI.

We report here on the first study to use the PGSI in a large representative sample drawn from a developing country, South Africa. It is also the first study to apply IRT to investigate the underlying factor structure and individual item functioning of the PGSI. Because of the fact that South Africa has 11 official languages, administering the PGSI in a representative sample posed unique challenges in terms of translation and back-translation of the measure. Therefore, in addition to the above, we conducted a DIF analysis prior to the main IRT analysis to ensure equivalence of item functioning across the four most commonly spoken language groups (English, IsiZulu, Sesotho, and Afrikaans).

Method

Participants

The PGSI was administered to a representative sample of the South African metropolises consistent with the screen's

purpose of measuring prevalence of problem gambling in the general population. The survey was targeted at the country's large urban environments, where opportunities for gambling in illegal venues are concentrated, because this is the only form of gambling other than scratch card draws conveniently accessible to most South Africans. South Africa has four cities with population more than 1 million: Johannesburg, Cape Town, Durban, and Tshwane/Pretoria. We drew representative samples from the greater metropolitan area of each of them to form a sample of $N = 3,000$.

Given the relatively low penetration of household telephone lines in the country, it was deemed crucial to conduct face-to-face interviews with respondents. Telephonic interviews, which are common in the field of gambling studies, were not deemed appropriate in the South African context. To ensure adequate coverage of the metropolitan areas, probability proportional to size cluster sampling was employed. The sampling methodology was organized into three stages. In Stage 1, the 2001 South African Census was used as the sampling frame to select sample points. Only noninstitutional, nonrecreational, and nonindustrial enumeration areas (EAs) were included in the sample design. The sampling frame was explicitly stratified by the metropolitan areas included in the study. Prior to drawing the sample points, the EAs in each stratum were arranged according to main place code, subplace code, and EA number. This was done to ensure the best possible coverage of the metropolitan areas. Three hundred sample points were used in total, and the allocated number of sample points in a stratum (Cape Town, Durban, Pretoria/Tshwane, and Johannesburg) was drawn systematically with probability proportional to size, with the number of households serving as the measure of size. Ten households were drawn in all sample points within a stratum to ensure that each household in the area had an equal selection probability.

In Stage 2, a geographical information system was used to select a random starting point in an EA where the selection of dwellings would take place. In instances where the street data did not provide a street name, interviewers oriented themselves using other street names and prominent features—such as schools, police stations, churches, and rivers—on the map. Once the starting point was located, interviewers worked systematically to select every n th household in the EA. If a household refused to participate in the study, another household was selected in its place from the same sample point. Importantly, interviewers were dispatched during work hours, after hours, and on weekends to ensure that there was an adequate representation of individuals in a sample point.

Finally, in Stage 3, all eligible members of the household—those aged 18 years and older¹—were listed on a Kish Grid. A household member was then randomly selected from this grid. Crucially, once the respondent had been selected, only this person could be interviewed in the household. The

substitution of households only took place after three unsuccessful attempts to contact the respondent. These visits were made on two different days and at different times.

The response rate for the study was 60.5%, which is slightly lower than prevalence studies conducted in other parts of the world (see, e.g., Volberg, 1994; Welte, Barnes, Wieczorek, Tidwell, & Parker, 2002), but is not surprising given the vagaries of life in South Africa. Finally, the data set was weighted according to gender, age, race, and community size using the All Media and Products Survey² to account for any skews in the data. Thus, the NUPSGB was carefully designed to provide a representative sample of South Africans in the four major metropolitan areas of the country. However, for the current study, we use unweighted data given the IRT analysis approach. In addition, 56.7% of the full sample reported never having gambled, which precluded administration of the PGSI.

The full sample included $N = 3000$ adult (aged 18 years and older) individuals (51.2% male; mean age = 39.34 years; $SD = 15.77$) with whom face-to-face individual surveys were conducted by trained fieldworkers. The sample consisted of 65.3% Black, 11.8% Colored, 5% Indian, and 19.7% White. "Colored" is used in South Africa to refer to a distinctive community, living mainly in the Western Cape province, who are descended primarily from Indonesians brought as slaves to the area in the 17th century, or from the indigenous Khoikhoi people and from pairings of early European settlers with members of the non-Bantu indigenous population of the Cape who are now extinct as a distinct ethnicity. Afrikaans is the standard first language of members of the Colored community. Other racial groups include Asians, who are mainly people of Indian descent, and Blacks, who are descendants of African peoples who migrated in a southerly direction from central Africa (see <http://www.statssa.gov.za/>). The racial breakdown is representative of the demographics of the large cities, in which Black people are underrepresented when compared with South Africa as a whole.

Measures

The PGSI (Ferris & Wynne, 2001) is a brief and easy-to-administer population screen that consists of nine items, four of which assess problem gambling behaviors (betting, tolerance, chasing, borrowing) and five of which assess the adverse consequences of gambling (problems with gambling, criticized by others, guilt, health problems, financial problems). Items are answered on a 4-point scale (0 = *never*; 1 = *sometimes*; 2 = *most of the time*; 3 = *almost always*). The initial validation study of the PGSI demonstrated a unidimensional factor structure, good internal consistency ($\alpha = .84$), adequate test-retest reliability ($r = .78$), and construct validity as evidenced by correlations with

gambling frequency (Ferris & Wynne, 2001). In the current study, we found Cronbach's α of .80, which is in the same band as other studies, where α was .86 (Brooker et al., 2009; Holtgraves, 2009), but slightly lower than .92 (McMillen & Wenzel, 2006), .92 (Arthur et al., 2008), and higher than .77 (Loo et al., 2010).

For the purposes of the current study, the measure was translated and back-translated into the 11 official languages of South Africa. We confine our analysis to four language groups (English, IsiZulu, Sesotho, and Afrikaans) for which there was sufficient sample sizes to pursue the detection of DIF. Therefore, $n = 1,469$ were included for the DIF and IRT analyses. To summarize, there were $n = 2,584$ participants across the four language groups, of which $n = 1,469$ endorsed ever having gambled.

Results

The IRT model fitting and the computation of the test statistics were performed using a beta version of IRTPRO (Cai, du Toit, & Thissen, in press; Thissen, 2009). Goodness of fit of the IRT models was evaluated using M_2 statistics and its associated root mean square error of approximation (RMSEA) values (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Maydeu-Olivares & Joe, 2005, 2006; Thissen, 2009). The M_2 statistic reflects "goodness of fit," and nonsignificant p values reflect adequate fit of the model to the item response data. However, this statistic, like other chi-square statistics obtained in the context of CFA, are generally unrealistic because there will be some error in any strong parametric model (Browne & Cudeck, 1993). The RMSEA is an index that may be computed for any statistic and provides a metric for model error. Following Browne and Cudeck's (1993) suggested "rules of thumb," values of RMSEA of .05 or less indicate close fit, values of .08 or less indicate reasonable fit, and values greater than .1 indicate poor fit of the model.

Before evaluating the psychometric properties of the nine gambling items using the item response data for all four languages, DIF analyses were done to investigate the equivalence of item functioning for the four most commonly spoken language groups (English, IsiZulu, Sesotho, and Afrikaans) that had adequate sample sizes for DIF analyses. In these analyses, we evaluated the similarity of item parameters (slope and threshold) estimated for the respondents who were interviewed in English (the original language of the PGSI) compared with those interviewed in IsiZulu, Sesotho, and Afrikaans. Because there were many instances of too few or no responses (fewer than 3) in categories *most of the time* and *almost always* for the smaller language groups (i.e., Afrikaans and Sesotho; ns approximately 200 for these two groups), these analyses were performed using the two-parameter logistic (2PL) binary IRT model collapsing *sometimes*, *most of the time*, and *almost always* into a single category representing endorsement.

One of the assumptions underlying the use of unidimensional IRT is that a single continuous construct accounts for the covariation among the item responses. This assumption and the fit of the IRT model were evaluated simultaneously by investigating the fit of a unidimensional 2PL model and evaluating the presence of local dependence (LD) among pairs or triplets of the gambling items. LD is a term used to describe excess covariation among item responses that is not accounted for by a unidimensional IRT model (i.e., a single factor). The detection of LD implies that the single factor model does not adequately explain item covariation. To investigate LD, the χ^2 LD statistic (Chen & Thissen, 1997) was used.

In separate analyses for each language group, the fit statistics did not indicate significant departures of fit for the 2PL unidimensional model (all M_2 statistics had p values larger than .08 with associated RMSEA values no larger than .02). The LD statistics are standardized chi-square values; values 10 or greater are considered noteworthy. None of the LD statistics were greater than 2.0.

DIF detection involved comparing the 2PL item parameters (one slope and one threshold) for each item estimated separately for each group, after using all nine items with equal parameters to estimate the population mean and variance for the focal group. DIF detection was done with Wald tests (Langer, 2008). An overall chi-square test evaluates the hypothesis of item parameter differences overall; this chi-square is partitioned into that attributable to the (a) slope (discrimination) parameter (indicating group differences in item discrimination) and to the (b) threshold (difficulty) parameter (indicating group differences in item endorsement rates). With the exception of one slope parameter comparison, none of the item parameters show significant DIF. The one exception involves the slope parameter estimated for Item 8 for those interviewed in Afrikaans. The slope parameter is estimated as 59.9 (which is effectively infinite, as an IRT slope value) as a consequence of a zero cell in the cross-tabulation table involving response to Item 1 ("Bet more than you could afford to lose") and Item 8 ("Gambling caused financial problems for you or your household"); specifically, all respondents who answered "never" to Item 1 also answered "never" to Item 8, leaving no respondents in one cell of the cross-tabulation. As a consequence, DIF detection cannot be done for this item (comparing English and Afrikaans). Overall, there was no evidence of DIF between respondents interviewed in English and Afrikaans, Sesotho, or IsiZulu, respectively. The remaining analyses were therefore conducted using the combined sample for the four language groups.

Next, we investigated the psychometric properties of the nine items for the combined language groups ($n = 1,469$). An analysis of the frequencies for each of the four categorical responses showed that between 78% and 95% of the sample answered *never* for each of the gambling items. The next question addressed before selecting an appropriate item

Table 1. Nominal Model Slope Parameters, Standard Errors, Scoring Function Values, and Intercept Parameters

Item Summary	Slope	SE	Scoring Function Value				Intercepts			
			0	1	2	3	0	1	2	3
1. Bet more than you could afford	1.16	0.18	0.00	1.84	2.24	3.00	0.00	-3.38	-5.81	-7.69
2. Needed to gamble . . . feeling	1.52	0.23	0.00	1.70	2.40	3.00	0.00	-3.97	-6.53	-9.29
3. Try to win back money	1.07	0.12	0.00	1.59	2.38	3.00	0.00	-2.25	-4.08	-5.68
4. Borrowed money	1.29	0.28	0.00	1.73	2.21	3.00	0.00	-4.84	-7.49	-9.87
5. Problem with gambling	2.36	0.54	0.00	1.08	1.36	3.00	0.00	-4.34	-6.83	-15.12
6. Health problems, stress or anxiety	1.56	0.27	0.00	2.05	2.45	3.00	0.00	-5.46	-7.98	-10.15
7. Criticized your betting	1.39	0.23	0.00	1.73	2.11	3.00	0.00	-4.50	-6.03	-9.19
8. Financial problems	3.60	0.87	0.00	0.92	1.35	3.00	0.00	-5.65	-9.26	-22.74
9. Felt guilty	1.70	0.28	0.00	1.47	1.80	3.00	0.00	-4.10	-6.23	-10.42

Note. The scoring function values and intercept parameters are listed for each of the four response alternatives.

response model was whether responses in the remaining categories (*sometimes, most of the time, almost always*) were meaningfully ordered. For each item, the score based on the eight remaining items, for each categorical response, was calculated. For all items, the score on the eight remaining items was monotonically increasing as the number of the response category increased. Thus, it appears that a multiple category response (as opposed to a binary model) may be useful, and the responses lie on a continuum in the anticipated order, if perhaps unequally spaced. Because of the pattern of item responses, heavily concentrated in the *never* category, the recently revised version of the nominal categories IRT model (Thissen, Cai, & Bock, 2010) was selected for analysis due to its facility to detect differences in the steepness of the slope parameter across the four response alternatives.

The unidimensional IRT nominal model showed satisfactory fit, $M_2(297) = 365.44, p = .01, RMSEA = 0.01$, with no indication of LD among the nine gambling items. Table 1 presents the abbreviated item content, the slope parameters, associated standard errors, the intercept parameters, and the scoring function values for the nine items. To illustrate the functions of the nominal model, Figure 1 shows the traces lines for two of the items, graphing the probability of a response in a category as a function of the value of the underlying construct: Item 6 (“Has gambling caused you any health problems, including stress, or anxiety”) and Item 5 (“Have your felt that you might have a problem with gambling”).

For Item 6 (“health problems, stress, or anxiety”) in the upper panel of Figure 1, a steeply descending trace line for the *never* (0) response category as the value of the underlying construct approaches 1.5 can be observed in contrast to the trace lines for the other three response categories, which change more gradually as the level of the latent variable (gambling severity) increases, indicating that although differences among responses 1, 2, and 3 provide some information about the level of gambling severity, those

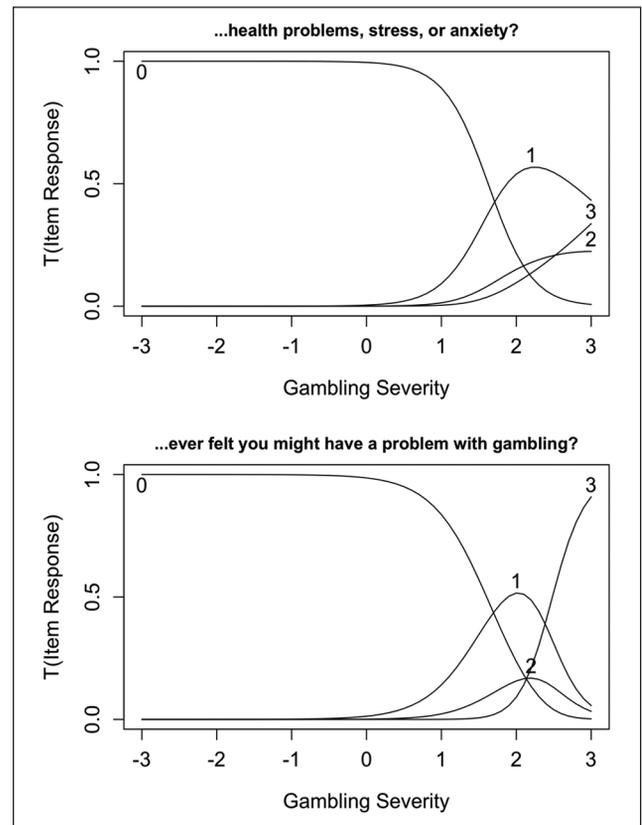


Figure 1. Trace lines that show the probability of each of the categorical responses as functions of the psychological construct for two items

differences are not as discriminating as the difference between 0 and any of the higher responses. In contrast, for Item 5 (“Have your felt that you might have a problem with gambling”), the most discriminating (steepest) curve is for Response 3, with the differences among the lower-response categories providing slightly less information.

The scoring functions (see Table 1) provide an alternate form of scoring each item. For example, for Item 5, the scoring function values are 0, 1.08, 1.36, and 3.0 for the four response alternatives, respectively. Notice that the difference between Categories 1 and 2 is much smaller than the difference between scores for Categories 2 and 3; those different differences imply that there is little psychological difference between responding *sometimes* and *most of the time* compared with the difference between *most of the time* and *almost always*. In contrast, for Item 6, the scoring function values are 0, 2.05, 2.45, and 3.0; so the difference between Categories 2 and 3 is much smaller than the difference between scores for categories 0 and 1. The scoring function values could be used in place of 0, 1, 2, and 3 for item scores. However, it is highly unlikely that practitioners would implement these scoring functions when calculating scores for the nine gambling severity items because such differences in scoring would not affect correlations with other measures, but the values describe the differential discrimination provided by the three transitions between pairs of adjacent response categories.

As previously mentioned, on average about 90% of the item responses were *never* for the nine gambling items. Such a skewed pattern of item responses may suggest that the remaining response categories individually add little to the measurement of individual differences in gambling severity. Information curves were used to evaluate whether the multiple category nominal model aids measurement compared with a simpler binary response model. Test information curves show how well the construct is measured at all levels of the underlying construct continuum. IRT information is the expected value of the inverse of the error variances for each estimated value of the underlying construct [$I(\theta) \approx 1/se^2(\theta)$]. The test information functions displayed in Figure 2 shows the varying measurement precision across the construct continuum for the nominal (solid) and 2PL IRT (dashed) models. Notice that the nominal model (solid) information curve has higher information values associated with higher values of the construct compared with the 2PL binary model. For example, using the nominal model, $I = 27.7$ at the construct value of 2.4 whereas for the binary model, $I = 6.6$ at that construct value, and is highest ($I = 17.7$) for the construct value of 1.6. Use of the nominal model, relative to the binary model, provides more information (greater measurement precision) and allows for the assessment of individual differences at higher levels of the gambling severity construct.

Discussion

Despite the increase in use of IRT in educational settings, its use in personality and psychopathology measurement has lagged behind that of other areas (Reise & Waller, 2009). The current study adds to the growing number of studies that apply IRT methods to examine the psychomet-

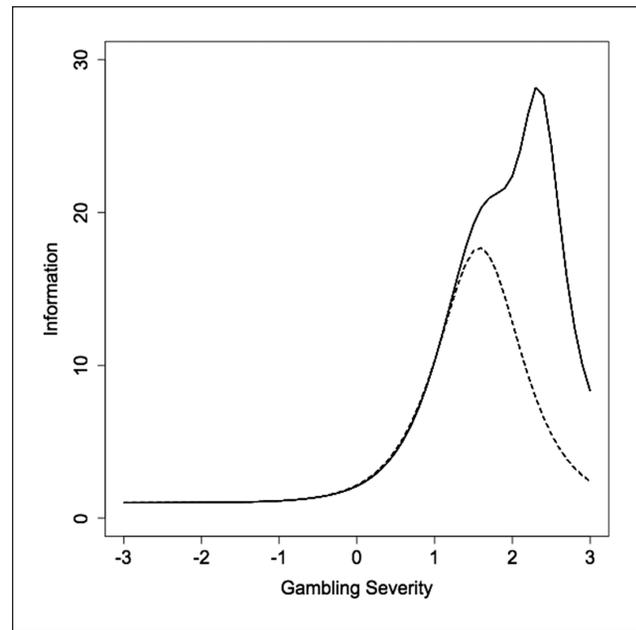


Figure 2. Test information curves showing how well the construct is measured at all levels of the underlying construct continuum

Note. The solid line represents the test information curve for the nominal categories response model. The dashed line represents the test information curve for the binary 2PL model.

ric properties of clinical measures. This study was the first to carry out an IRT analysis of the PGSI and the first to evaluate whether the use of the nominal categories model of the PGSI provides additional information at higher values of the underlying construct relative to a simpler binary model. It is also the first to use the PGSI in a developing country in Africa, more specifically, a sample representative of the South African metropolises. Several findings are of note. First, the finding that 56.7% of subjects did not engage in gambling behavior is in line with Ferris and Wynne (2001), who found 55% in this group, but slightly higher than China (42.7%; Loo et al., 2010) and lower than Singapore (61.7%; Arthur et al., 2008).

Second, DIF was not detected across language groups. In other words, even when accounting for mean differences in gambling severity between language groups, items functioned similarly comparing English with the three other language groups. This provides support for the translated versions of the PGSI into the four most often spoken official languages of urban South Africa (English, IsiZulu, Sesotho, and Afrikaans), making it the second study, in addition to Loo et al. (2010), to support the feasibility of translating the PGSI into other languages for international and cross-cultural comparison.

Third, our results are consistent with a unidimensional factor structure for the PGSI as reported by past studies using more traditional analytic techniques (Arthur et al.,

2008; Brooker, et al., 2009; Ferris & Wynne, 2001; McMillen et al., 2004) as well as more recent studies taking latent trait approaches to examining the psychometric properties of the PGSI (Loo, et al., 2010). The PGSI was designed to measure a single factor to facilitate its function as a population screen of the prevalence rates of problem gambling. High homogeneity or unidimensionality of a screening measure maximizes the ability to discriminate between diagnostically distinct groups (Costello & Angold, 1988). In this sense, and in accordance with Loo et al. (2010), who used CFA, the current results suggest that the PGSI is fit for its intended purpose. In addition, we demonstrated high discrimination parameters (slopes) for all items, providing further evidence for its internal construct validity and suitability for use as a population-based screening instrument.

Fourth, and furthermore consistent with its design as a population screen (which is intended to overidentify false positives), the endorsement of items on the PGSI showed a highly skewed pattern, with an average 90% of the item responses in the *never* category. Item discrimination parameters further confirmed the appropriateness of the PGSI for screening purposes, given the steeply descending slope associated with the *never* category, with increased probability for the other three categories with higher values of the construct. This is consistent with IRT findings of other clinical population screens, for instance, depression (e.g., Aggen et al., 2005; Sharp, Goodyer & Croudace, 2006), in which information is provided in the narrow portion of the trait continuum indicating that the items function well at the severe end of the trait.

Fifth, an investigation comparing the multiple category nominal item response model with a simpler binary response model demonstrated that the construct is optimally measured using the four-category response scale. Scores may be calculated using the scoring functions, which provide unequally spaced transitions between pairs of adjacent response categories; alternatively, traditional summed scores may be used.

Different theories or societal conceptions of problem gambling naturally produce different screening tools, thus generating different empirical findings about the prevalence of the problem (McMillen & Wenzel, 2006). Gradient approaches to the assessment of problem gambling allow for a view of gambling as a continuum ranging from social or recreational gambling (with no adverse effects) to problem gambling (with adverse effects for the individual, family, friends, colleagues, and the community) through to pathological gambling (with severe negative consequences and meeting diagnostic criteria; Neal et al., 2004), as opposed to a simple classification in terms of meeting diagnostic criteria. Against this background, our findings are important as the PGSI (in comparison to the SOGS) was developed based on a gradient view of problem gambling.

Several limitations are of note. First, as with all self-report questionnaire-based studies, our findings may have been affected by demand characteristics placed on individuals participating in the research, recall bias, and the cross-sectional nature of the study. Longitudinal studies that allow tracking of day-to-day gambling activity to derive true severity scores and allow for the calculation of test-retest reliability are ways of addressing these limitations. In addition, future research can potentially explore the PGSI's use in other sub-Saharan African countries, where similar, if not more significant, poverty may affect gambling severity. Given the emerging evidence for the unidimensional factor structure of the PGSI across several countries (now including South Africa), it is expected that a unidimensional factor structure will be confirmed for these countries as well.

Notwithstanding these limitations, the current study contributes to the growing literature supporting the psychometric properties of the PGSI as the population screen of choice, also recently suggested by several reviews of research on problem gambling measures (Abbott & Volberg, 2006; Neal et al., 2004). With only nine items, it is short and practical to administer in large population surveys, even in developing countries such as South Africa. This is important, given that problem gambling research has been dominated by Western concepts, methodologies, and solutions (Loo et al., 2010). Here, we provide the first evidence for the internal construct validity of the PGSI for use in population-based studies of problem gambling in South Africa specifically, and in developing countries generally. As such, the current article contributes to a growing number of studies investigating the psychometric properties of easy-to-administer clinical measures in sub-Saharan Africa and South Africa in particular. In the context of the paucity of specialist psychological and psychiatric services in South Africa and other sub-Saharan African countries, there is an urgent need for psychometric evaluation and adaptation of Western psychiatric tools in developing countries (Peterson, 2004; Sharp, Skinner, Serekoane, & Ross, 2010). This necessity is further compounded by the worldwide shift toward a more community-based psychiatric service delivery approach (Botha, Koen, & Niehaus, 2006), which will require psychometrically sound epidemiological screening measures such as the PGSI.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Research results were not vetted by the NRGF, and scientific autonomy of the research is protected by the policies of the University of Cape Town, which contracted for the research with the NRGF, administered the funds, and conducted ethical oversight.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The gathering of the data for this paper was funded by the National Responsible Gambling Programme (NRGP), which is in turn funded through a levy on the South African gaming industry. Two authors (Ross, Dellis) received consulting income from the NRGP. Three other authors (Sharp, Hofmeyr, Kincaid) received research grants to their institutions for percent effort from the NRGP.

Notes

1. Eighteen years is the legal gambling age in South Africa.
2. The All Media and Products Survey is conducted annually and is representative of the metropolitan areas of South Africa. The All Media and Products Survey was used to weight the data because it more accurately reflects the demographic profile in South African metropolitan areas than the, now outdated, 2001 Census.

References

- Abbott, M. W., & Volberg, R. A. (2006). The measurement of adult problem gambling and pathological gambling. *International Gambling Studies, 6*, 175-200. doi:10.1080/14459790600928678
- Arthur, D., Tong, W. L., Chen, C. P., Hing, A. Y., Sagara-Rosemeyer, M., Kua, E. H., & Ignacio, J. (2008). The validity and reliability of four measures of gambling behaviour in a sample of Singapore university students. *Journal of Gambling Studies, 24*, 451-462.
- Botha, U. A., Koen, L., & Niehaus, D. J. (2006). Perceptions of a South African schizophrenia population with regards to community attitudes towards their illness. *Social Psychiatry and Psychiatric Epidemiology, 41*, 619-623.
- Brooker, I. S., Clara, I. P., & Cox, B. J. (2009). The Canadian Problem Gambling Index: Factor structure and associations with psychopathology in a nationally representative study. *Canadian Journal of Behavioral Science, 41*, 109-114. doi:10.1037/a0014841
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE.
- Cai, L., du Toit, S. H. C., & Thissen, D. (in press). *IRTPRO: Flexible professional item response theory modeling for patient reported outcomes [Computer software]*. Chicago, IL: SSI International.
- Cai, L., Maydeu-Olivares, A., Coffman, D. L., & Thissen, D. (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2p tables. *British Journal of Mathematical and Statistical Psychology, 59*, 173-194. doi:10.1348/000711005X66419
- Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265-289. Retrieved from <http://www.jstor.org/stable/1165285>
- Costello, E. J., & Angold, A. (1988). Scales to assess child and adolescent depression: Checklists, screens and nets. *Journal of the American Academy of Child and Adolescent Psychiatry, 27*, 726-737.
- Culleton, R. P. (1989). The prevalence rates of pathological gambling: A look at methods. *Journal of Gambling Behavior, 5*, 22-41. doi:10.1007/BF01022135
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London, England: Erlbaum.
- Ferris, J., & Wynne, H. (2001). *The Canadian Problem Gambling Index* (Final report). Ottawa, Ontario, Canada: Canadian Centre on Substance Abuse.
- Holtgraves, T. (2009). Evaluating the Problem Gambling Severity Index. *Journal of Gambling Studies, 25*, 105-120. doi:10.1007/s108989-008-9107-7
- Langer, M. M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation*. Chapel Hill: University of North Carolina.
- Leisieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry, 144*, 1184-1188.
- Loo, J. M., Oei, T. P., & Raylu, N. (2010). Psychometric evaluation of the Problem Gambling Severity Index—Chinese Version (PGSI-C). *Journal of Gambling Studies*. doi:10.1007/s10899-010-9221-1
- Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*, 1009-1020. doi:10.1198/016214504000002069
- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*, 713-732. doi:10.1007/s11336-005-1295-9
- McMillen, J., Marshall, D., Ahmed, E., & Wenzel, M. (2004). *2003 Victorian Longitudinal Community Attitudes Survey*. Canberra, Australia: The Centre for Gambling Research, Australian National University.
- McMillen, J., & Wenzel, M. (2006). Measuring problem gambling: Assessment of three prevalence screens. *International Gambling Studies, 6*, 147-174. doi:10.1080/14459790600927845
- Neal, P., Delfabbro, P. H., & O'Neill, M. (2004). *Problem gambling and harm: Towards a national definition* (Report prepared for the National Gambling Research Program Working Party). Melbourne, Victoria, Australia: State of Victoria, Department of Justice.
- Orford, J., Wardle, H., Griffiths, M., Sproston, K., & Erens, B. (2010). PGSI and DSM-IV in the 2007 British Gambling Prevalence Survey: Reliability, item response, factor structure and inter-scale agreement. *International Gambling Studies, 10*(1), 31-44.

- Peterson, I. (2004). Primary level psychological services in South Africa: Can a new psychological professional fill the gap? *Health Policy and Planning, 19*, 33-40.
- Reise, S. P. & Waller, N. G. (2009). Item Response Theory and Clinical Measurement. *Annual Review of Clinical Psychology, 5*, 27-48.
- Schissel, B. (2001). Betting against youth: The effects of social marginality on gambling among young people. *Youth & Society, 32*, 473-491. doi:10.1177/0044118X01032004004
- Sharp, C., Goodyer, I. M., & Croudace, T. J. (2006). The Short Mood and Feelings Questionnaire (SMFQ): A unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of Abnormal Child Psychology, 34*, 365-377. doi:10.1007/s10802-006-9027-x
- Sharp, C., Skinner, D., Serekoane, M., & Ross, M. (2010). A qualitative study of the cultural appropriateness of the Diagnostic Interview Schedule for Children (DISC-IV) in South Africa. *Social Psychiatry and Psychiatric Epidemiology*. Advance online publication. doi:10.1007/s00127-010-0241-z
- Strong, D. R., Breen, R. B., Lesieur, H. R., & Lejuez, C. W. (2003). Using the Rasch model to evaluate the South Oaks Gambling Screen for use with nonpathological gamblers. *Addictive Behaviors, 28*, 1465-1472. doi:10.1016/S0306-4603(02)00262-9
- Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system—IRT software. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved from <http://www.psych.umn.edu/psylabs/CATCentral/>
- Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 43-75). New York, NY: Routledge.
- Volberg, R. A. (1994). The prevalence and demographics of pathological gamblers: Implications for public health. *American Journal of Public Health, 84*, 237-241.
- Welte, J. W., Barnes, G. M., Wieczorek, W. F., Tidwell, M. C., & Parker, J. (2002). Gambling participation in the U.S.: Results from a national survey. *Journal of Gambling Studies, 18*, 313-37.