An item response theory analysis of the Problem Gambling Severity Index in a national representative sample

Carla Sharp, Lynne Steinberg, and Ilya Yaroslavsky

University of Houston

Andre Hofmeyr and Andrew Dellis

University of Cape Town

Don Ross

University of Cape Town and Georgia State University

Harold Kincaid

University of Alabama at Birmingham

Corresponding author:

Carla Sharp, Ph.D.

Department of Psychology

University of Houston

Houston, TX, 77024, USA

Email: csharp2@uh.edu

Lines: 257 (incl references; excl title page, abstract, figures and tables)

Abstract

Increases in the availability of gambling heighten the need for a short screening measure of problem gambling. The Problem Gambling Severity Index is a brief measure that allows for the assessment of social and environmental aspects of gambling with the facility to identify levels of problem gambling. We 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 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.

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 in North America, 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 (CPGI; Ferris & Wynne, 2001) was designed to provide an alternative to the more frequently used South Oaks Problem Gambling Survey (Leisieur & Blume, 1987). The SOGS has received much criticism for taking a categorical and "medical" view of problem gambling at the expense of social and environmental aspects of problem gambling. Because the SOGS was developed specifically for use in clinical settings, it does not include less severe behavioral items and may under-identify individuals with sub-threshold problem gambling (Strong et al., 2003; Holtgraves, 2009). It also does 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 non-problem gamblers or pathological gamblers (a dichotomous 0/1 classification), the PGSI takes a more dimensional approach in that items are answered on a four-point scale (0 = never; 1 = sometimes, 2 = most of the time, 3 = almost always). The PGSI is therefore able to identify different subgroups of problem gamblers with different risk status (no, low, moderate, and high). Despite the PGSI's promise (Neal, Delfabbro, & O'Neill, 2004), there are factors that limit its use. Few studies beyond those by the PGSI developers have been conducted to investigate its psychometric properties (Brooker, Clara, & Cox, 2009; Holtgraves, 2009). Moreover, while the PGSI has been investigated in samples from Canada (Ferris & Wynne, 2001), Australia (McMillan & Wenzel, 2006), Great Britain (Ordford et al., 2010) and Singapore (Arthur et al., 2008), it has not yet been examined in a developing or poor country sample. Finally, all psychometric studies of the PGSI have relied on classical test theory approaches to data analyses in lieu of more appropriate latent trait approaches. The advantages of using latent trait approaches to determine the internal construct validity of a measure over classical test theory (e.g. principal component analysis and Cronbach's alpha) are well known and readers are referred to more comprehensive reviews (e.g., Embretson & Reise, 2000).

We report here on the first study to use the PGSI in a large representative sample of South Africans. It is also the first study to apply item response theory (IRT) to investigate the underlying factor structure and individual item functioning of the PGSI. Due to 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 differential item functioning (DIF) analysis prior to the main IRT analysis to ensure equivalence of item functioning across different language groups.

Method

The PGSI was administered to a representative sample of the South African metropoles consistent with the screen's purpose of measuring prevalence of problem gambling in the general population. A face-to-face individual survey of $N = 3000$ adult (+18 years of age) individuals (51.2% male; mean age = 39.34; $SD = 15.77$) in the Cape Town, Durban, Johannesburg and Tshwane metropoles of South Africa was conducted by trained fieldworkers. The sample consisted of 65.3% Black, 11.8% Coloured, 5% Indian, and 19.7% White. This breakdown is representative of the demographics of the large cities, in which Black people are under-represented by comparison with South Africa as a whole. Enumeration Areas (EAs), defined according to the 2001 national census, were the primary sampling units used in the study and the data was adjusted for clustering at this level. The data was also stratified according to metropolitan area and was weighted to account for oversampling. The weighted data are representative of the civilian population of the sampled metropolitan areas of South Africa on a variety of socioeconomic variables including region, age, race/ethnicity, and sex, based on the All Media and Products Survey (AMPS[1]). 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 the administration of the PGSI.

The PGSI (Ferris & Wynne, 2001) is a brief and easy-to-administer population screen that consists of 9 items, 4 of which assess problem gambling behaviors (betting, tolerance, chasing, borrowing) and 5 of which assess the adverse consequences of gambling (problems with gambling, criticized by others, guilt, health problems, financial problems). 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. 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, 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.

[1] The AMPS is conducted annually and is representative of the metropolitan areas of South Africa. The AMPS was used to weight the data because it more accurately reflects the demographic profile in South African metropolitan areas than the now outdated most recent national census.

Results

The IRT model fitting and the computation of the test statistics were performed using a beta version of IRTPRO (Thissen, 2009; Cai, du Toit, & Thissen, forthcoming). Goodness of fit of the IRT models was evaluated using the M_2 statistics and its associated RMSEA values (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Maydeu-Olivares & Joe, 2005; Maydeu-Olivares & Joe, 2006, Thissen, 2009).

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 language groups (English, IsiZulu, Sesotho, Afrikaans). 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 to 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 separate language groups, these analyses were performed using the 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 2 PL model and evaluating the presence of local dependence (LD) among pairs or triplets of the gambling items. Local dependence 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 X^2 LD statistic (Chen and Thissen, 1997) was used.

In separate analyses for each language group, the RMSEA and LD 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 2 PL 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 χ^2 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 consequence of a zero cell in the cross-tabulation table involving response to item 1 (bet more than you could afford) and item 8 (financial problems); 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 on average, about 90% of the sample answered "never" for each of the gambling items. The next question addressed before selecting an appropriate item 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 8 remaining items, for each categorical response, was calculated. For all items, the score on the 8 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 9 items. To illustrate the functions of the nominal model, Figure 1 shows the traces lines for two of the items: item 6 ("health problems, stress, or anxiety"), and item 5 ("have your felt that you might have a problem with gambling"), graphing the probability of a response in a category as a function of the value of the underlying construct.

Table 1 and Figure 1 about here

For item 6 (“health problems, stress, or anxiety”) in the upper panel of Figure 1, notice the steeply descending trace line for the “never” (0) response category as the value of the underlying construct approaches 1.5; the trace lines for the other three response categories change more gradually as the level of the latent variable (gambling severity) increases, indicating that while differences among responses 1, 2, and 3 provide some information about the level of gambling severity, those 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 to 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 9 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 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 to 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 to the 2PL binary model. For example, using the nominal model, I = 27.7 at the construct value of 2.4 while 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.

Figure 2 about here

Discussion

The current study was the first to carry out an IRT analysis of the PGSI. It is also the first to use the PGSI in a sample representative of the South African metropoles, or indeed in any sample drawn from a developing or poor country. Several findings are of note. First, equivalence of item functioning across language groups was demonstrated. In other words, even when accounting for mean differences in gambling severity between language groups, items functioned similarly comparing English to the 3 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, Afrikaans).

Second, our results are consistent with a unidimensional factor structure for the PGSI as reported by past studies using more traditional but less sophisticated analytic techniques for categorical response data (Ferris & Wynne, 2001; Brooker, Clara, & Cox, 2009; McMillen et al., 2004; Arthur et al., 2008). The PGSI was designed to measure a single factor to facilitate its function as a population screen of the prevalence rates of problem gambling. By demonstrating a unidimensional factor structure and high discrimination parameters (slopes) for all items using appropriate data analytic techniques, we provide further evidence for its internal construct validity.

Third, and furthermore consistent with its design as a population screen (which is intended to over-identify 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.

Fourth, an investigation comparing the multiple category nominal item response model to 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). Dimensional 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 continuum view of problem gambling. Whether gambling behaviors in general are best viewed as continuous or taxonic is a separate and further question for continued research.

Taken together, 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 a national review of research on problem gambling measures (Neal et al., 2004). We furthermore 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 a developing country generally.

References

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(2)*, 109-114. doi: 10.1037/a0014841.

Cai, L., du Toit, S.H.C., & Thissen, D. (forthcoming). *IRTPRO*: *Flexible professional item response theory modeling for patient reported outcomes* [Computer software]. Chicago: 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. Stable URL: http://www.jstor.org/stable/1165285.

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: Lawrence Erlbaum Associates, Publishers.

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

Hotgraves, 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*. University of North Carolina, Chapel Hill.

Leisieur, H.R., & Blume, S.B. (1987). The South Oaks Gambling Screen (The SOGS): A new instrument for the identification of pathological gamblers. *The American Journal of Psychiatry, 144,* 1184-1188.

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., & Wenzel, M. (2006). Measuring problem gambling: Assessment of three prevalence screens. *International Gambling Studies, 6(2)*, 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.

Ordford, J., Wardle, H., Griffiths, M., Sproston, K., and 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.

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., 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.

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*. Online at www.psych.umn.edu/psylabs/CATCentral/

Table 1

Nominal model slope parameters, standard errors, scoring function values, and intercept parameters

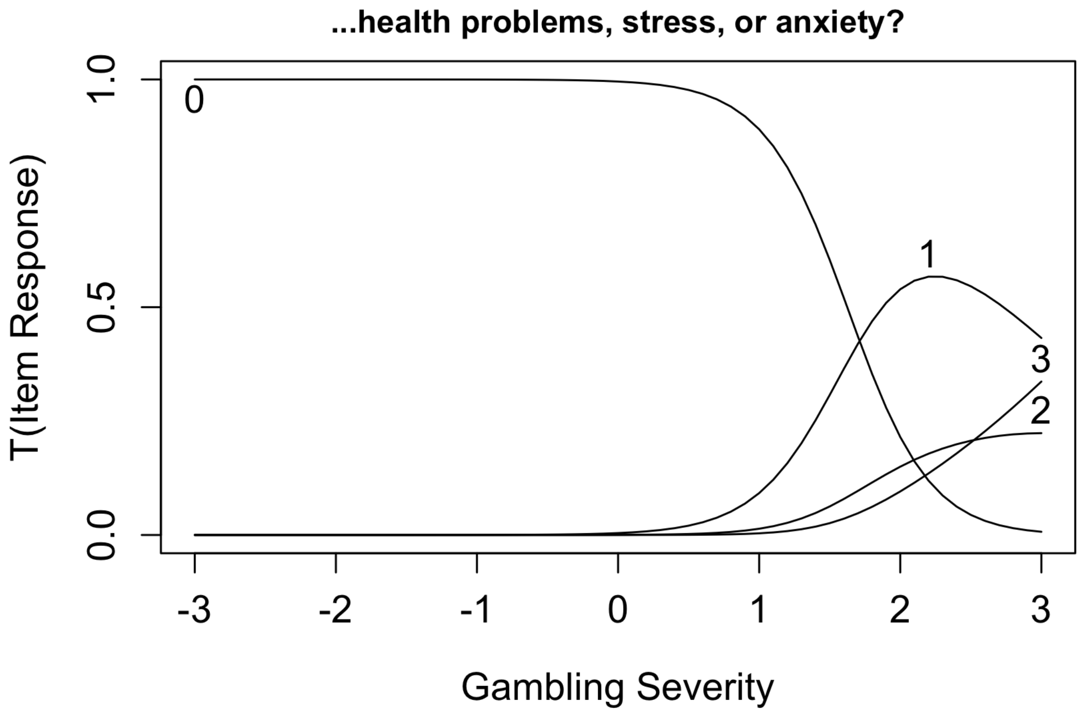
			Scoring Function Value				Intercepts			
Item summary	Slope	s.e.	0	1	2	3	0	1	2	3
1. bet more than you could afford	1.16	0.18	0.00	1.84	2.24	3.00	0.00	-3.38	-5.81	-7.69
2. needed to gamble …feeling	1.52	0.23	0.00	1.70	2.40	3.00	0.00	-3.97	-6.53	-9.29
3. try to win back money	1.07	0.12	0.00	1.59	2.38	3.00	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

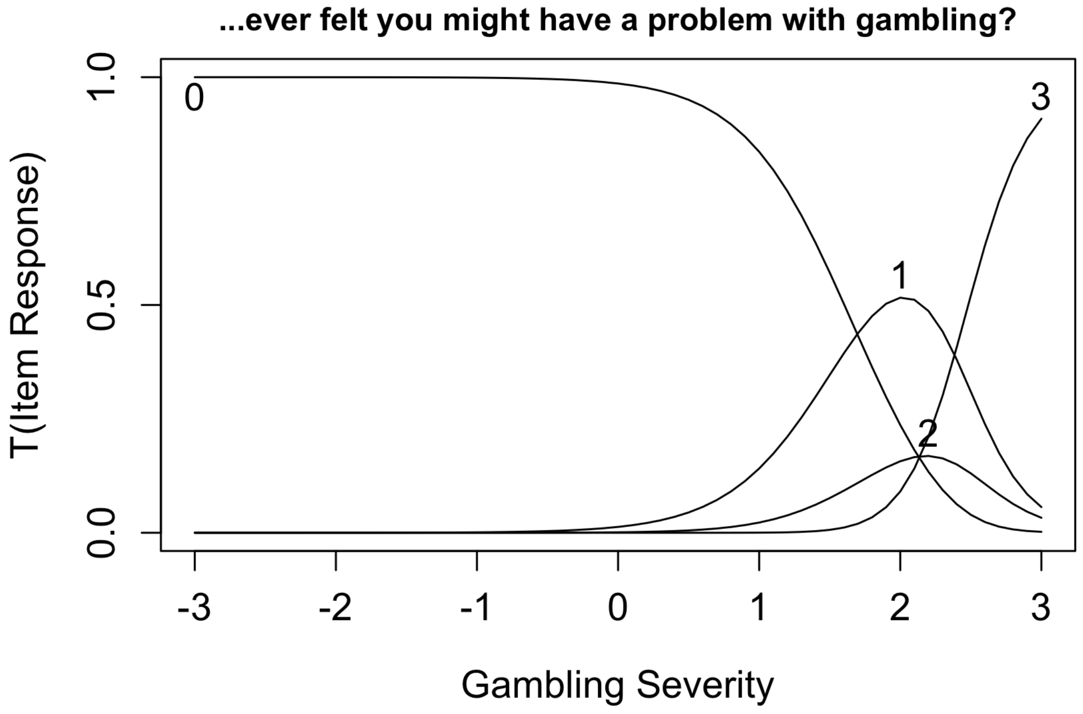
Note: The scoring function values and intercept parameters are listed for each of the four response alternatives.

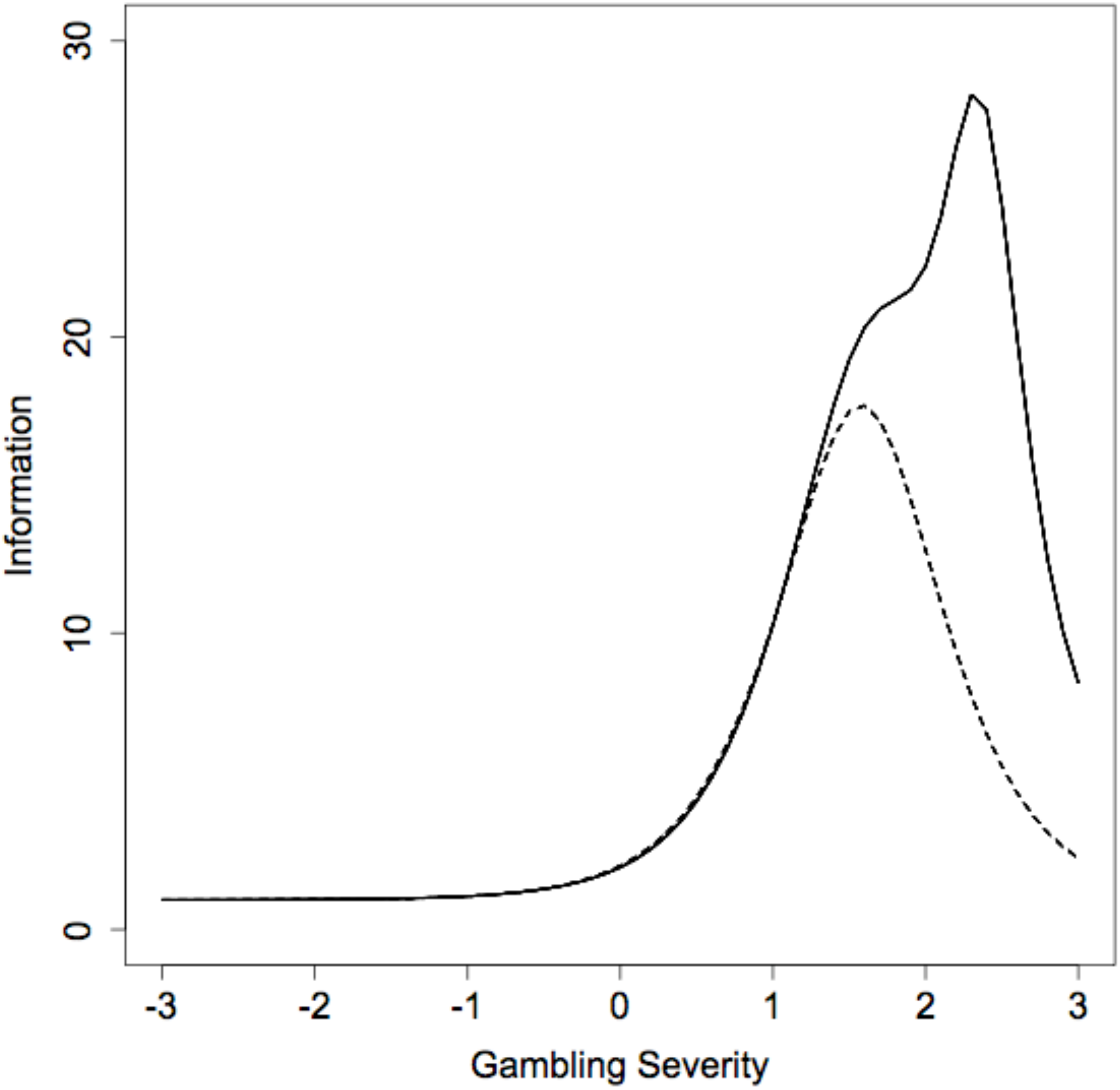
Figure Captions

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

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


...health problems, stress, or anxiety?
T(Item Response)
1.0
0.5
0.0
0
1
2
3
-3
-2
-1
0
1
2
3
Gambling Severity


...ever felt you might have a problem with gambling?
T(Item Response)
1.0
0.5
0.0
0
1
2
3
-3
-2
-1
0
1
2
3
Gambling Severity


30
20
10
0
Information
-3
-2
-1
0
1
2
3
Gambling Severity