Assessing the Factor Structure of the Canadian Problem Gambling Index: Does Qualitative Stability allow Quantitative Comparison?

Scott B. Maitland, Ph.D.
Gerald R. Adams, Ph.D.

University of Guelph
Guelph, Ontario, Canada
N1G 2W1
EXECUTIVE SUMMARY ....................................................................................................................................................... 4
GENERAL INTRODUCTION ................................................................................................................................................. 7
SOME CRITICAL QUESTIONS ............................................................................................................................................ 8
LIMITATIONS OF PRIOR ANALYSES ............................................................................................................................... 9
CONFIRMATORY FACTOR ANALYSIS ............................................................................................................................................... 10
ANALYTIC STRATEGIES .................................................................................................................................................... 12
  STATISTICAL PROCEDURES .................................................................................................................................................. 12
    Comparative fit indexes ................................................................................................................................................ 12
STUDY RATIONALE IN RELATION TO THE EMPIRICAL AND THEORETICAL LITERATURE ..................... 14
SAMPLING AND RECRUITMENT ..................................................................................................................................... 15
ANALYSES .............................................................................................................................................................................. 15
RESULTS ................................................................................................................................................................................. 16
  MODEL FIT AND COMPARISONS ........................................................................................................................................... 19
    Reliability ........................................................................................................................................................................ 19
    One-Factor versus Two-Factor Model ........................................................................................................................................ 19
  MODEL FIT FOR SMALLER SUBSAMPLES .......................................................................................................................... 22
    Models for 10% Random Sample ................................................................................................................................ 22
    Models for 20% Random Sample ................................................................................................................................ 24
  SEX INVARIANCE OF ONE-FACTOR CPGI MODEL ........................................................................................................... 27
  SEX INVARIANCE OF THE TWO-FACTOR CPGI MODEL ..................................................................................................... 30
DISCUSSION AND IMPLICATIONS .................................................................................................................................. 33
REFERENCES ........................................................................................................................................................................ 36
Executive Summary

In gambling research, The South Oaks Gambling Screen has been widely used in the United States as a focal measure of problem gambling. However, this instrument has had limited research on its psychometric properties. The Canadian Problem Gambling Index (CPGI) was developed for use with Canadian populations with ongoing evaluation of the psychometric properties of reliability, validity and generalizability. The CPGI is comprised of nine items that are often summed to derive a single additive problem gambling score. This technique implies that a single factor or dimension underlies the items of the instrument. However, it has been suggested that the construct validity of this instrument may be based on a two factor structure of problem gambling behaviors and the consequences of those behaviors.

The CPGI has been previously subjected to exploratory factor analytic methods. However, the non-normality of the frequency distributions of the 9 items presents considerable problems when using EFA techniques. Therefore, a more sophisticated research technique, confirmatory factor analysis, was applied to the data to assess whether a single or two factor solution best accounts for the measurement properties of construct validity.

Establishing construct validity for any measurement device, using either exploratory or confirmatory factor analysis (CFA), requires establishing whether the factor or factors can be replicated and determination of whether measurement equivalence across samples and comparison groups (e.g., gender) exists. Two gold standards for valid measurement are (1) to provide evidence of a factor structure that is replicable, and (2) a demonstration of measurement invariance or equivalence between comparison groups.

The primary purpose of this investigation was to assist the investigators who constructed the CPGI in determining the factor structure of the instrument and the measurement equivalence/invariance between the two most often compared groups in gambling research; males and females. The latter effort is important in that gambling researchers make comparisons between genders, often reporting a
substantial difference. However, this difference may be due to inappropriately making direct comparisons when the same model does not exist between groups. It is possible that observed gender differences are an artifact of methodological shortcomings versus true differences between the groups compared.

This investigation utilized extant data provided by the Ontario Problem Gambling Research Centre and a variety of multivariate analyses were undertaken using confirmatory factor analysis. CFA techniques allow the investigator to determine if there is a good fit between the proposed models and provide a variety of statistics that can be used to determine if a single or two-factor model has the best fit. CFA can also be used to determine if the best fit model has measurement equivalence between samples or group comparisons. The best fitting model is judged to be the most acceptable, valid, and generalizable.

The statistical analyses, which are presented in a technical format in this report, have resulted in the following observations:

- A comparison of a single factor model to a two factor model of behavior predicting consequences revealed that the two factor model is the best fit and most valid way of describing the construct of problem gambling using the CPGI.
- Reliability estimates, often referred to as internal consistency, were stronger (or higher) for males than females for both the behavior and consequences factors.
- In comparisons of random samples of the large data set, to control for excess power and determine if the model works for smaller samples, it was observed that a 10% (n = 463) sample was not adequate to appropriately fit either a one- or two-factor model. Modeling a random 20% sample indicated a lack of fit for a one-factor model. Although a two-factor model had a poor fit it was significantly better than the one factor model.
• The model comparisons examining one-factor and two-factor structures were at best a poor fit for the 10 and 20% samples. However, a reasonable fit can be obtained when a large enough sample is used. Due to the non-normality of gambling problem behavior, and very low prevalence of the issue, large samples are necessary to assure a good fit.

• Comparisons of the one and two-factor models for measurement invariance between males and females demonstrated a lack of invariance. This means that the instrument is measuring different constructs for males than females. This conclusion was drawn from the use of random 10%, 20% and full sample analyses.

The most apparent implications of the analyses reported in this technical report are as follows. A two-factor model was the best fit for the construct validity of the instrument with large data sets. The indicators of behaviors and consequences, within the two-factor model, indicated differential weights for the items within the two constructs for males versus females. The construct validity of the CPGI was stronger for males than females. This means that direct comparisons between males and females using this instrument can be misleading and should not be done. It remains unclear whether behaviors or consequences in the two-factor model can predict other forms of problem gambling attitudes, behaviors, or further consequences. Further examination of the classification system of the CPGI should consider whether behaviors or consequences are best used in defining problem gambling categories for males and females.
General Introduction

Gambling was once considered a sin but is now a common leisure activity. In the United States, one national survey has revealed that 82% of households surveyed have indicated that at least one person in the family has gambled in the last few months (Welte, Barnes, Wieczorke, Tidwell, & Parker, 2002). Further, these gamblers report an average of gambling 60 times in the past year. Some 23% gamble weekly and gambling includes a highly diversified activity ranging from casino table games, betting pools, off-track and track betting, lotteries, and other forms of gaming. The Committee on the Social and Economic Impact of Pathological Gambling (1999) has examined the evidence on gambling and points to the downside of widespread gambling. This report, and others like it, has stimulated an expanding interest in the study of problem gambling.

To estimate the prevalence of gambling behavior for various groups assessment devices have been created to provide both sample and population-based frequency. A variety of instruments have been constructed to estimate gambling prevalence. Among many other measures The South Oaks Gambling Screen (Winters, Stinchfield & Fulkerson, 1993), the Gamblers Anonymous Twenty Questions (Custer & Custer, 1978), and various adaptations of the American Psychiatric Association criteria for pathological gambling (Fisher, 1992) have been widely used. Derevensky and Gupta (2000) conclude that the South Oaks Gambling Screen has been the most widely used instrument to assess problem gambling prevalence. Ferris, Wynne and Single (1999) have evaluated these instruments and along with others have been critical of their construction and use (e.g., Volberg, 1994; Volberg & Steadman, 1992). Derevensky and Gupta (2000) indicate that while the South Oaks Gambling Screen was developed against the DSM-III criteria (American Psychiatric Association, 1980), important assessment of reliability and validity have not been reported.

There is an accumulating set of investigations examining the psychometric and methodological shortcomings of the SOGS (in its various forms). It has been argued that the meaning of some of the
items may not be understood (e.g., Ladouceur, Bouchard, Rhèaume, Jacques, Ferland, Leblond, & Walker, 2000), prevalence rates maybe over or under-estimated (e.g., Derevensky, Gupta & Winters, 2003), there may be many false positive classifications due to methodology (Gambino, 1997), and issues of scoring error (Jacques & Ladouceur, 2003). There have also been concerns about the use of DSM based instruments for studying problem gambling (e.g., see Fisher, 2000)

Poulin (2002), a Canadian investigator at Dalhousie University, has examined the psychometric adequacy of the South Oaks Gambling Screen Revised for Adolescents in a large study of youth from four Atlantic provinces. She reports that version of the SOGS has adequate stability and internal consistency and predictive validity with substance abuse problems. Further, it is found that marked differences occur for the proportion of male than female gamblers. Indeed, it is widely observed in the gambling literature that gambling problems are more prevalent for males than females. Poulin provides no evidence for the construct validity of the SOGS-RA. This is a common problem for most instruments that are used to assess gambling problems.

In response to these and other shortcomings in assessment, a Canadian instrument has been constructed with intentions of building a psychometric sound instrument based on estimates from a Canadian population. This report focuses on further issues of measurement and conceptualization of this instrument.

**Some Critical Questions**

As with any new field of inquiry, topically-specific measurement instruments are generally required to ensure accuracy and comparability of collected data. Specifically regarding gambling research, critical questions must be raised as to what topics are of interest and what domains such an instrument should examine. For example, is problem gambling endemic and if so, are the concerns the same regardless of sampling variation (i.e., both with regard to sampling methods and the characteristics of different samples)? Another parallel question would ask what constitutes problem gambling? Is
problem gambling the same across sexes, age groups, and at a more global level, countries? Does problem gambling and how we define it remain the same over time or has the construct changed with regard to which behaviors are most salient in defining problem gambling as technology evolved? Additionally, if the construct has changed, are there standard developmental patterns or expectations as to how problem gambling evolves over time?

Addressing country specific concerns, the Canadian Problem Gambling Index (CPGI) was developed to measure gambling within Canada. This end result of a multi-year endeavor has already been widely utilized in a variety of samples and settings, with one end goal of determining risk status and prevalence rates of problem gambling.

The CPGI resulted from a thorough review and synthesis of the current gambling literature and consensus of expert opinion from gambling researchers. Additionally, this measure is partly derived from previous measures from the field of gambling research (Wynne, 2003). Reports about instrument development provide detailed information regarding operational definitions and some information about psychometric properties (i.e., reliability estimates and validity) of the constructs of the CPGI (Ferris, Wynne, & Single, 1999; Wiebe & Single, 2002; Wynne, 2003). However, whereas the nine CPGI items are often summated to derive a problem gambling score, an alternative interpretation of the CPGI items would suggest that the scale measures: (a) problem gambling behaviors; and, (b) consequences of those behavior. In this alternative model, it is clear that problem gambling behaviors must be present, and precede consequences in a temporal sequence, for participants to provide valid responses to consequence questions.

Limitations of Prior Analyses

Previous factor analytic work has used exploratory factor analytic methods to examine the data. The response scales for the 9 CPGI items, all measuring whether a participant has acted out a behavior or experienced a consequence of that behavior, result in non-normally distributed data for all items. One
expectation of exploratory factor analytic methods is that data are derived from normal distributions, thereby making scores on the 9 CPGI items inappropriate for exploratory factor analysis. Numerous studies have demonstrated that the use of non-normally distributed data can provide inaccurate results in exploratory factor analyses.

Therefore, the structure that defines the CPGI has not been reliably demonstrated using existing exploratory methodology, nor has the scale been subjected to confirmatory factor analysis (CFA) to examine replicability of factor structures across comparison groups (gender, age, ethnicity) or time ((i.e., tests of factorial invariance a.k.a., measurement invariance, measurement equivalence or measurement equivalence/invariance (ME/I)).

**Confirmatory Factor Analysis**

CFA allows one to determine the best model that represents covariations or relations found in a set of data and also allows for comparison of competing models to determine which model best represents how the items function together to create a scale or multiple subscales. Competing models raise a number of empirical questions. For example, is problem gambling as measured by the CPGI simply a single factor, thereby allowing the use of a simple additive strategy across all 9 items? Alternatively, if one believes that certain CPGI items measure behaviors, and other items measure consequence, this suggests that respondents cannot experience the consequences in the absence of performing the behaviors. This implies a clearly different model than the usual additive model applied to CPGI data (i.e., totaling scores on the 9 items into one score). Analyses were conducted to determine if a unidimensional model fits the CPGI data or whether a two-factor model, separating behaviors from consequences fitted the data better.

Upon determining which factor structure is the best fitting model for the CPGI (as demonstrated by a consensus of criteria and fit indexes), and assuming any model fits well enough to be deemed acceptable, the validity and generalizability of a factor model will be subjected to tests of ME/I across
genders before assuming that quantitative scores on the constructs are qualitatively comparable (Horn & McArdle, 1992; Meredith, 1993). Evidence that a model is well-fitting for one group may not be representative of the underlying factor structure for all groups to be compared (Maitland, Dixon, Hultsch, & Hertzog, 2001; Maitland, Herlitz, Nyberg, Bäckman, & Nilsson, 2004; Nyberg et al., 2003). If the underlying structure of a factor model differs across groups, the qualitative interpretation of the model differs (i.e., what each factor represents) and therefore, any subsequent quantitative comparisons (i.e., tests of mean differences with factor or scale scores) resulting from these differences are suspect. However, if a factor structure demonstrates acceptable levels of ME/I, one is assured that comparisons of outcomes result from measuring the same concepts across groups (or within groups over time).

The next question is what constitutes an acceptable level of ME/I to allow quantitative comparisons? A brief review of the standard approach to testing ME/I provides an answer. Hertzog and Nesselroade (2003) provided a conceptual overview of ME/I. A hierarchy of restrictions is placed on factor models to test for ME/I. The widely accepted framework of Meredith (1993) was employed here: (a) configural invariance examined similarity of factor patterns across comparison groups, ignoring differences in the magnitude of factor loadings between groups; (b) weak metric invariance of factor loadings, provided a test of no differences between the location and magnitude of factor loadings between groups [weak (metric) invariance is considered the absolute minimum level of ME/I to allows for meaningful quantitative comparisons of groups by establishing comparable measurement units for the variables and factors (Cunningham, 1982, 1991; Horn, et al., 1983; Horn & McArdle, 1992; Maitland, et al., 2001)]; (c) strong invariance of observed variable intercepts where observed variable intercepts are constrained to be equivalent across sexes. Additional tests for (d) strict invariance, testing if uniqueness terms are invariant across sexes, were also examined but were not supported and are, therefore, not included in result tables.
Analytic Strategies

Statistical Procedures

Models were tested using LISREL 8.51 (Jöreskog & Sörbom, 2001) and AMOS 5.0 (Arbuckle, 2003). The non-normal nature of this data was assessed using LISREL and attempts to transform the data to approximate normal distributions were conducted. Numerous data transformations including log-linear, exponential, and cosine procedures were examined, however, data do not approximate normality regardless of which transformation was used. Additionally, one of the primary assessment procedures proposed for examining the validity and generalizability of the CPGI, measurement invariance, requires models be examined using covariance matrices. Methods for testing factorial invariance are being developed to be used for non-normal data, however, these methods are still preliminary. Therefore, all analyses were conducted on covariance matrices and mean vectors with results of the final models reported as standardized estimates for ease of interpretation. Factor scaling was accomplished by fixing one item for each factor to a value of 1.0 in the pattern matrix and the same item was used to scale factors between sexes. The chi-square difference test ($\Delta \chi^2$; Jöreskog & Sörbom, 1989) was used to compare nested models. The critical value used for all comparisons was $p < .01$.

Comparative fit indexes. Model fit was evaluated by examining the following fit indexes: (a) model $\chi^2$; (b) Non-Normed Fit Index (NNFI; Bentler & Bonnett, 1980); (c) Comparative Fit Index (CFI; Bentler, 1990); (d) Root Mean Square Error of Approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980).

For the reader that is unfamiliar with the terminology of assessing model fit we offer the following description. To engage in a model fit it is necessary to examine several indices when evaluating a model and to never rely solely on a single fit index. There are, in actuality, two kinds of fit indices. The absolute fit compares observed versus expected variances and covariances. The
comparative fit indices assess and compare the fit of a target model to some baseline model.

Absolute fit indices include several types. The chi-square compares the observed covariance matrix with the expected covariance matrix, similar to any chi-square analysis you might be familiar with except this technique uses the covariance rather than the frequency or percentage of some behavior. The chi-square is zero when there is no difference between the two matrices. This is good because it indicates no difference which means a perfect fit. In comparison, a significant chi-square indicates that the model predicts relations that are significantly different from the relations or associations observed in the sample – thus, the model should be rejected because it is a poor fit. Often the chi-square test is useful when testing nested models (models within models). The goodness-of-fit index compares the explained covariance to total measured covariance. The adjusted goodness-of-fit index is similar to the goodness-of-fit index except it adjusts for the degrees of freedom in the model. This adjustment is considered to establish parsimony of the model. Some investigators use the centrality index and it ranges from 0.0 to 1.0, with values closer to 1.0 being preferable. Often researchers suggest that a value greater than .90 is interpreted as being an acceptable model fit. The standardized root mean square residual is the average discrepancy (difference) between the observed and the expected correlations across all the parameters that are estimated in the model. Often it is suggested that the RMR is only interpretable in regard to standardized variables. Finally, the root mean square error of approximation adjusts for parsimony in the model and is less influenced (relative insensitivity) to the size of the sample used to test the model fit. A perfect fit will yield a score of 0.0, and some investigators consider a score of less than .08 as adequate, with scores of less than .05 considered good.

There are a variety of comparative fit indices too. The comparative fit index compares the tested model to a null model that has no paths that link the variables, thus making the variables totally independent of each other. This index is most stable with small samples. A score of .90 or higher are considered acceptable-- the closer to 1.0 the better the fit. Delta or incremental fit index also ranges
from 0.0 to 1.0 with higher values indicating a better model fit. The normed fit index is somewhat sensitive to sample size and does not perform as consistently with smaller samples. This index tends to under-estimate when the data are not normally distributed. That is, the values may be unacceptably small for good-fitting models. Finally, the non-normed fit index is a generalized version of the Tucker and Lewis index.

We originally proposed a reanalysis of the CPGI prevalence data. Our goals included: (a) examining the data using Exploratory Factor Analytic (EFA) techniques to establish a baseline; and, (b) subjecting the resulting model to Confirmatory Factor Analysis (CFA) to determine if the subscales proposed by Wynne (2003) result from the data. Next, we proposed subjecting the best fitting model to (c) tests of measurement invariance, to determine if the factor structure of the CPGI is stable across sexes. Additional tests for the stability of the factor structure across time could be completed when longitudinal (i.e., repeated measures) data become available; however, such data was not made available for these analyses. From a broader methodological (and practical) perspective, the proposed research documents the degree to which the CPGI should be used for comparative purposes, and provides evidence about the degree of generalizability and replicability of results employing this measurement instrument.

**Study Rationale in Relation to the Empirical and Theoretical Literature**

The importance of this study cannot be overstated. The limited research in the field of problem gambling has been primarily descriptive. The CPGI was developed to provide a “gold standard” tool for use in measuring gambling behavior in Canada and beyond. However, the evidence to date suggests almost no information is available about the meaning and significance of this instrument other than reports written by the creators of the scale or their collaborators. Furthermore, the absence of confirmatory analyses to support the claims made during the scale development phase means researchers using the CPGI should be skeptical about their resulting claims about group comparisons. Moreover,
assuming that measurement invariance is demonstrated with the CPGI, authors may rest assured that comparisons across different samples are legitimate (i.e., they are not trying to equate apples and oranges). Hence, positive results from this study will serve to strengthen claims made with the CPGI and allow for wider dissemination or impact of this tool. Potential negative results (e.g., finding the factor structure does not work or that measurement invariance is absent) will serve to motivate additional analyses to determine the best fitting model to measure problem gambling.

The research conducted in this study serves as the basis for future studies of ranging impact. For example:

- Refinement, validation, and determination of the extent of generalizability of a measurement instrument for assessing problem gambling;
- Examination of the evidence for measurement invariance (i.e., comparability of the subscales that are part of the CPGI) to ensure users are making meaningful comparisons, thereby encouraging others to employ the CPGI in their own research;
- Determining and testing the relevance of the CPGI for use in a provincial and/or national study of problem gambling within Canada.

**Sampling and Recruitment**

Pre-existing data provided by the Ontario Problem Gambling Research Centre was used for this study. Data from the Prevalence Study employing the CPGI were obtained from Dr. Jamie Wiebe and comprise a sample of 4631 participants comprised of 2256 men and 2375 women.

**Analyses**

Assessment of psychometric properties were conducted using multiple methods including exploratory factor analytic work and structural equation modeling (SEM), to examine measurement equivalence/invariance (ME/I) (see Maitland, et al., 2001; Maitland, et al., 2004; Maitland, Intrieri, Schaie & Willis, 2000; Nyberg, et al., 2003; and Schaie, Maitland, Willis & Intrieri, 1998). ME/I
between sexes was examined. Additionally, the sample size for the Prevalence Study was quite large (n=4631), presenting a unique problem regarding sample size and power. Normally one is concerned with having adequate sample size to find significant results if present. A large sample as presented here creates the opposite problem whereby even trivial results with very small impact or effect, can be statistically significant. Therefore, we conducted analyses on randomly selected samples, comprised of 10 and 20% of the entire Prevalence Study sample, and report these results, as well as results for the entire study sample. This is critically important, especially when combined with the response format for the nine CPGI items, where respondents are endorsing the presence or absence of a behavior. This response format creates an additional problem that will be discussed later in the report. For now, suffice it to say that a low prevalence for any behavior including gambling, may result in the need for larger samples to obtain sufficient cases to observe the desired outcome.

Results

Based on previous reports for the CPGI, the nine items of the scale are summed to provide an index score for respondents. This approach, diagrammed as a representation of the factor model is seen below (see Figure 1).
CPGI Conceptual Model
However, examination of the nine CPGI items reveals a competing model as previously described. This model includes four items that measure gambling behaviors as well as five items measuring consequences of those behaviors. As such, this model presents a predictive and dependent relationship between the behaviors and consequences (i.e., the consequences should not be present in the absence of the gambling behaviors). This model is presented in Figure 2.
Model Fit and Comparisons

Reliability

Reliability estimates using Cronbach’s coefficient alpha (split-half correlations:

Part 1 = items 1-5; Part 2 = items 6-9) ranged from .64 to .68 for the overall sample.

Values for men were .68 and .74 and for women .52 and .55 for parts 1 and 2
respectively, clearly showing stronger estimates for men.

One-Factor versus Two-Factor Model

Results from AMOS 5 for the entire sample show that the one-factor model
results in statistically significant factor loadings (ranging from .47 - .70) and acceptable
model fit shown by consensus of fit indexes: M1: $\chi^2 = 338.31$, df = 27, p < .001, NNFI
= .953, CFI = .972, RMSEA = .041. Squared multiple correlations for all variables range
from .22 (22% of variance in item 7 accounted for by the factor) to .50 or 50% of the
variance in item 5 explained by the factor (see Figure 3).
The two-factor model resulted in statistically significant factor loadings ranging from .49 - .63 for Behaviors and from .47 to .71 for Consequences. Model fit was also good: M2: $\chi^2 = 224.98$, df = 26, $p < .001$, NNFI = .954, CFI = .974, RMSEA = .041. Squared multiple correlations for all variables range from .22 (22% of variance in item 7 accounted for by the factor) to .50 or 50% of the variance in item 5 explained by the factor. Comparison of the one factor and two factor models reveals $\Delta \chi^2_{M2-M1} = 113.33$, df = 1, $p < .001$. Therefore, the two factor model shows significant improvement in model fit as demonstrated by the statistically significant difference between models. This
finding supported the second model as best fitting, however, results will be provided for both the one and two factor models for all analyses to follow. Additionally, the standardized regression coefficient between Gambling Behaviors and Consequences was $\beta = .96$, $p < .001$, demonstrating that the latter are dependent on the former as expected and that 92% of variance in consequences are attributed to behaviors (see Figure 4).
Model Fit for Smaller Subsamples

Models for 10% Random Sample

A random 10% sample (n = 463) was tested for both the one- and two-factor CPGI models. Results show that the one-factor model results in statistically significant factor loadings (ranging from .25 - .84) and mixed model fit results: M1: $\chi^2 = 118.34., \text{df} = 27, p < .001, \text{NNFI} = .854, \text{CFI} = .913, \text{RMSEA} = .086$. Squared multiple correlations for all variables range from .06 (6% of variance in item 9 accounted for by the factor) to .74 or 74% of the variance in item 4 explained by the factor. The significant chi-square value, NNFI index smaller than .9 and RMSEA larger than .05 indicated a lack of fit for the one-factor model in a random 10% sample (see Figure 5).
The two-factor model resulted in statistically significant factor loadings ranging from .26 - .81 for Behaviors and from .26 to .83 for Consequences. Model fit was not acceptable: $\chi^2 = 109.70$, $df = 26$, $p < .001$, NNFI = .862, CFI = .920, RMSEA = .083. The significant chi-square value, NNFI below .9 and RMSEA larger than .05 all indicate a lack of fit similar to that noted for the one-factor model. Squared multiple correlations for all variables range from .07 (7% of variance in items 2 & 9 accounted for by the factor) to .69 or 69% of the variance in item 8 explained by the factor. Comparison of the one factor and two factor models revealed: $\Delta \chi^2_{M2-M1} = 8.64$, $df = 1$, $p < .001$. 
Therefore, the two factor model showed significant improvement in model fit as demonstrated by the statistically significant difference between models, however, the questionable model fit statistics raise concern whether either model truly fits for a random 10% sub-sample of the prevalence data. The standardized regression coefficient between Gambling Behaviors and Consequences was $\beta = 1.08, p < .001$, further demonstrating problematic results as standardized regression weights should not exceed a value of 1.00 (see figure 6).

Models for 20% Random Sample

A random 20% sample ($n = 926$) was tested next. Results showed that the one-
factor model had statistically significant factor loadings (ranging from .11 - .75) and poor model fit: M1: $\chi^2 = 168.38, \text{df} = 27, p < .001$, NNFI = .669, CFI = .801, RMSEA = .075. Squared multiple correlations for all variables range from .01 (1% of variance in item 2 accounted for by the factor) to .56 or 56% of the variance in item 8 explained by the factor. The significant chi-square value, NNFI index smaller than .9 and RMSEA larger than .05 all indicated a lack of fit for the one-factor model in the random 20% sample (see Figure 7). Standardized results for the one-factor CPGI model for a random 20% sample are displayed below.
Testing the two-factor CPGI model in the random 20% sample resulted in statistically significant factor loadings ranging from .22 - .38 for Behaviors and from .32 to .77 for Consequences. Model fit was poor: M2: $\chi^2 = 158.11$, df = 26, $p < .001$, NNFI = .678, CFI = .814, RMSEA = .074. The significant chi-square, NNFI and CFI below .9 and RMSEA larger than .05 all indicate a lack of fit for the model. Squared multiple correlations for all variables ranged from .05 (5% of variance in item 2 accounted for by the factor) to .59 or 59% of the variance in item 8 explained by the factor. Comparison of the one factor and two factor models revealed $\Delta\chi^2_{M2-M1} = 8.64$, df = 1, $p < .001$. Therefore, the two factor model showed significant improvement in model fit over the one-factor CPGI model, however, model fit statistics would suggest neither model is truly acceptable. The standardized regression coefficient between Gambling Behaviors and Consequences was $\beta = .66$, $p < .001$, accounting for 44% of the variance in Consequences. The standardized solution for the two-factor CPGI model for the random 20% sample is shown below (see Figure 8).
Sex Invariance of One-Factor CPGI Model

To test the hypothesis of sex invariance of the CPGI models, we expanded our strategy into a multi-group, simultaneous structural model examining men and women in one model. This initial multi-group model examined configural invariance of the one- and two-factor CPGI models in men and women for the entire prevalence sample (Configural Invariance Model M2: $\chi^2 = 466.55$, $df = 54$, $p < .001$, NNFI = .905, CFI = .943, RMSEA = .041). Results showed that the one-factor model results in statistically significant factor loadings (ranging from .42 - .76 for men; .27 - .63 for women).
Squared multiple correlations for men range from .18 (18% of variance in item 7 accounted for by the factor) to .57 or 57% of the variance in item 8 explained by the CPGI factor. Squared multiple correlations for women range from .07 (7% of variance in item 2 accounted for by the factor) to .39 or 39% of the variance in item 7 explained by the factor. Standardized results for the one-factor CPGI model for all men in the prevalence sample (see Figure 9):

![Diagram of one-factor CPGI model for men]

Standardized results for the one-factor CPGI model for all women in the prevalence sample (see Figure 10):

![Diagram of one-factor CPGI model for women]
To test for sex invariance, the next model examined whether all factor loadings could be constrained to be identical for men and women. The model results are as follows: $\chi^2 = 853.63, \text{df} = 46, p < .001, \text{NNFI} = .841, \text{CFI} = .890, \text{RMSEA} = .053$, showed worsened fit and model fit indexes that were no longer acceptable. Delta chi-square for 8 degrees of freedom was 387.09 chi-square points, clearly demonstrated a lack of invariance of factor loadings across genders. No further invariance tests were conducted for the one-factor CPGI model.
Sex Invariance of the Two-Factor CPGI Model

Sex invariance of the two-factor CPGI model was examined next: Configural Invariance Model M2: $\chi^2 = 440.09, \text{df} = 52, p < .001, \text{NNFI} = .910, \text{CFI} = .946, \text{RMSEA} = .040$. The two-factor model resulted in statistically significant factor loadings (ranging from .55 - .75 for men’s behaviors and .42 - .76 for men’s consequences; and, .28 - .55 for women’s behaviors and .47 - .63 for women’s consequences). Squared multiple correlations for men ranged from .18 (18% of variance in item 7 accounted for by the factor) to .58 or 58% of the variance in item 8 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for men was $\beta = .95, p < .001$, accounting for 89% of the variance in Consequences.

Squared multiple correlations for women ranged from .08 (8% of variance in item 2 accounted for by the factor) to .40 or 40% of the variance in item 7 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for women was $\beta = .92, p < .001$, accounting for 84% of the variance in Consequences.

Standardized results for the configural two-factor CPGI model for male sample (see Figure 11):
Standardized results for the configural two-factor CPGI model for all women in the prevalence sample (see Figure 12):
Tests constrained all factor loadings across men and women for the two-factor model. Model results: $\chi^2 = 823.84$, $df = 59$, $p < .001$, NNFI = .838, CFI = .894, RMSEA = .053, showed worsened fit and model fit indexes that are mixed. Delta chi-square for 7 degrees of freedom was 383.74 chi-square points, clearly demonstrated a lack of invariant factor loadings across genders. No further invariance tests were conducted for the two-factor CPGI model.

A lack of sex invariance was noted for both the one- and two-factor CPGI models when examined across the entire prevalence sample. Additional tests were examined in
the random 10% and 20% samples and a lack of sex invariance was noted.

**Discussion and Implications**

Results from this study provide the first application of CFA to examine the underlying structure of the CPGI. Previous reports using exploratory factor methods suggested that a one-factor model was supported for the nine items comprising the scale. Using confirmatory factor techniques, we tested whether support was found for a one-factor model versus a two-factor model that was comprised of a gambling behaviors and a gambling consequences factor. Item content supports the two-factor model as meaningful. Therefore, the notion of behaviors preceding consequences or consequences as dependent on participation in gambling behaviors is meaningful.

The scale of measurement used for the nine gambling items results in non-normal distributions for each item, thereby violating the expectation of normality of input variables for analysis. Transformations of the data were not successful in creating normally distributed outcomes, and analyses were conducted using different matrix formats to assess the impact of this non-normality. Analyses were conducted on polychoric correlation matrices and covariance matrices with similar patterns of results. Covariance matrices are required to conduct tests of factorial invariance (ME/I), therefore, subsequent analyses were reported for results from covariance matrix analysis.

An additional concern, peculiar to the data used for these analyses but applicable to all data concerning problem gambling, is linked to the issue of prevalence. Problem gambling as defined in the literature has a relatively low prevalence. As such, sample size must be large to gather data on enough participants to have a range of variability in responses to the nine CPGI items. The current sample was 4631 adults. Another
problem results from the use of very large samples in CFA procedures. CFA requires samples large enough to ensure adequate power to obtain statistically significant results, however, samples that are too large will result in statistically significant, but potentially meaningless results. This concern was addressed by examining the entire sample of 4631 participants, and also testing models in a randomly drawn 10% sample (n=463) and a randomly drawn 20% sample (n=926). The 10% sample provides adequate power to obtain meaningful results, however, the low prevalence and non-normality of data made it prudent to examine the 20% sample as well.

Results from the overall sample suggested adequate model fit for both the one- and two-factor models for the CPGI. Tests to examine which model provides better fit suggested that the two-factor model was better fitting. This result raises question about the use of the CPGI as a single, additive scale. The strong relation demonstrated between gambling behaviors and consequences of those behaviors further supported this conclusion. The one- and two-factor models were then examined in the 10% and 20% samples. A general lack of fit was found for either model in the random 10% sample. Whereas statistically significant factor loadings were found, other indicators of acceptability of the model suggested a lack of fit. Similar results were noted for the random 20% sample, which resulted in statistically significant factor loadings but model fit indexes not obtaining acceptable levels.

Rejection, or at best, questionable fit results for the 10% and 20% samples, further demonstrated the problems previously described in fitting a model to a low-prevalence phenomenon. All factor analysis methods examine the manner in which items related to one another, not the actual level of response on that item. As most of this sample is
comprised of persons who did not endorse the gambling behaviors or consequence items, it appears that limited variability in responses was found even in a sample of 926 participants.
References


