Replication and generalizability of the Problem Gambling Severity Index: Are results consistent and comparable across studies?

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

University of Guelph
Guelph, Ontario, Canada
N1G 2W1

Final report submitted to Ontario Problem Gambling Research Centre

March 2007

Executive Summary

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. Embedded within the CPGI is the Problem Gambling Severity Index (PGSI) comprised of nine items that are 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, theoretical development and previous work has 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. Additionally, cut-off values derived from a summation of the 9 items are used to categorize respondents into non-gambler, non-problem gambler, at-risk gambler, moderate and severe gambler categories. If two factors better fit PGSI data then the value of these cut-off points is drawn into question.

The PGSI has been primarily subjected to qualitative interpretation or exploratory factor analytic methods. However, qualitative description does little to provide definitive evidence about the construct and the non-normality of the frequency distributions of the 9 items presents considerable problems when using EFA techniques. Therefore, confirmatory factor analysis should be applied to assess whether a one or two factor solution provides best fit to the data. Additionally, these methods should be employed across multiple data sets to examine the generalizability of the PGSI.

Establishing construct validity for any measurement device requires establishing whether the factor or factors can be replicated and determination of whether measurement equivalence across samples and comparison groups (e.g., sexes) exists. The primary purpose of this investigation was to examine the factor structure of the PGSI and the measurement equivalence/invariance (ME/I) between three independent datasets. An additional consideration is whether measurement equivalence exists between men and women in each of these three data sources. The latter effort is important as gambling researchers frequently make comparisons between sexes, 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 sex differences are an artifact of comparing apples to oranges (i.e., the PGSI demonstrates differential measurement properties across sexes).

Three extant datasets were examined using CFA. These techniques allow the investigator to determine if there is a good fit between the proposed models and provides methods to compare multiple models to conclude which best represents that data (something lacking from EFA). CFA also allows for direct tests of measurement equivalence between the three samples or across sexes; again, something not attainable with exploratory methods.

The statistical analyses resulted in the following observations:

- A comparison of a one factor (PGSI score) model to a two factor (Behaviors, Consequences) model in each of the three samples revealed that the **two factor** model is the best fit.
- Reliability estimates were stronger for consequence items than behavior items across all three data sets. Comparisons across sexes showed stronger estimates for women than men.
- Previous work by Maitland and Adams (2005) suggested that a reasonable fit may be obtained when a large enough sample is used. Non-normality of gambling problem behavior as measured by PGSI and very low prevalence rates seemed to imply that large samples are necessary to assure a good fit. However, the current work shows that even large samples do not ensure adequate fit of the PGSI model.
- Random samples of approximately 500 participants were drawn from each dataset and demonstrated unacceptable fit for PGSI models. Therefore, regardless of whether a one- or two-factor solution is examined, samples larger than 500 participants are required to obtain a

minimum level of acceptable fit (*Note*. This does not assure good fit and may vary depending on sample characteristics).

- Tests of ME/I for the one- and two-factor models between sexes and across samples were rejected for all three data sets. Therefore, the PGSI is measuring different constructs for men and women and even more troubling, in each instance where it was used to collect information about problem gambling behaviors and consequences.

In summary, a two-factor model was the best fit for the construct validity of the instrument with large data sets. However, the fit was questionable for a number of models. The indicators of behaviors and consequences, within the two-factor model, indicated differential weights for men versus women raising concern about using the PGSI to make comparisons across sexes. The construct validity of the PGSI was mixed across studies. Results from the Williams and Wood data showed stronger results for men than women, whereas Ladouceur et al and the Wiebe study did not show a clear pattern in the strength of factor loadings. Also interesting was that the regression between behaviors and consequences was clearly stronger for men in the Williams and Wood and Ladouceur datasets, whereas the opposite was found for the Wiebe data. Faced with these disparate patterns of results, comparisons between sexes would be misleading. Regarding replicability and generalizability of the PGSI, results across three large independent samples found differential measurement properties for the PGSI. Therefore, faced with a lack of ME/I, comparisons between studies employing the PGSI as a screening measure for problem gambling are questionable.

General Introduction

Gambling continues to gain popularity as a 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). Prevalence rates for *problem* gamblers remain low (approx. 3-5%), whereas the pool of gamblers at risk of becoming problem gamblers is generally much higher. More research is needed about what predicts gambling behavior and the progression between gambling states (Slutske, Jackson, & Sher, 2003; Wiebe, Cox & Falkowski-Ham, 2003), however, an even more fundamental concern is to demonstrate that gambling indexes operate as described and that screening methods used for categorizing gamblers into gambling states are valid, reliable and generalizable.

Most studies of problem gambling behavior rely on a screening measure to categorize participants into distinct groups. However, questions have been raised about the ability of these measurement instruments to accurately measure prevalence of gambling behavior. A variety of instruments including The South Oaks Gambling Screen (Winters, Stinchfield & Fulkerson, 1993), the Gamblers Anonymous Twenty Questions (Custer & Custer, 1978), various adaptations of the American Psychiatric Association criteria for pathological gambling (Fisher, 1992), and the Canadian Problem Gambling Index (CPGI, Wynne, 2000) were created to measure gambling and to provide methods of classifying gamblers. Derevensky and Gupta (2000) conclude that the South Oaks Gambling Screen has been the most widely used instrument to assess problem gambling prevalence, however, it is not without its problems. Ferris, Wynne and Single (1999) evaluated these instruments and along with others were critical of their construction and use (e.g., Volberg, 1994; Volberg & Steadman, 1992). Derevensky and Gupta (2000) indicate that whereas the South Oaks Gambling Screen was developed against the DSM-III criteria (American Psychiatric Association, 1980), important assessment of reliability and validity were not 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 may be 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 have been raised (Jacques & Ladouceur, 2003). The use of DSM based instruments for studying problem gambling has also been criticized (e.g., see Fisher, 2000; Slutske et al, 2003).

In an analysis of the psychometric adequacy of the South Oaks Gambling Screen Revised for Adolescents, Poulin (2002) reported that the SOGS had adequate stability, internal consistency and predictive validity with substance abuse problems. It was also reported that marked differences occur for the proportion of male than female gamblers and this disparity between sexes has been widely observed in the gambling literature. It is quite common to report sex differences in gambling, however, to do so assumes the instruments employed function equivalently for men and women. Finally, Poulin provided no evidence for the construct validity of the SOGS-RA, thereby demonstrating another common problem encountered when assessing instruments used to study problem gambling.

Regarding the Canadian Problem Gambling Index (CPGI), a number of studies have examined psychometric properties of the scale. In creating the index, phase one of development focused on how problem gambling was conceptualized, defined and measured in the field and the development of a new concept of problem gambling and a means of measuring it (Ferris, Wynne, & Single, 1999). Of particular importance for the current study, phase two emphasized the psychometric properties of the scale (Ferris & Wynne, 2001). It is reported that the CPGI resulted from a thorough review and synthesis of the gambling literature and consensus of expert opinion from gambling researchers. 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 et al., 1999; Wiebe & Single, 2002; Wynne, 2003). Embedded within the PGSI are 9 items used to create a screening measure of problem gambling called the Problem Gambling Severity Index (PGSI items below);

How often have you:

1. Bet more than you could really afford to lose?
2. Needed to gamble with larger amounts of money to get the same feeling of excitement?
3. Gone back another day to try to win back the money you lost?
4. Borrowed money or sold anything to get money to gamble?
5. Have you felt that you might have a problem with gambling?
6. Have people criticized your betting or told you that you had a gambling problem, regardless of whether or not you thought it was true?
7. Have you felt guilty about the way you gamble or what happens when you gamble?
8. Has your gambling caused you any health problems, including stress or anxiety?
9. Has your gambling caused any financial problems for you or your household?

Each PGSI response is scored from 0 (never) to 3 (almost always) and the total score of the 9-items (range 0-27) is used to determine the problem gambling level: non-problem gambling = 0; low-risk gambling = 1 – 2; moderate risk gambling = 3 – 7; problem gambling = 8 – 27. Interestingly, even in the face of widespread criticism, items from the SOGS and DSM were used in the creation of the CPGI. Additionally, there is some conflicting information regarding the number of constructs the CPGI (and in particular, the PGSI) measures. Early reports suggest that multiple dimensions were expected and, to some degree, demonstrated, however, the standard use of the 9 PGSI items is the additive strategy described above.

The CPGI has been used in all ten Canadian provinces, Australia, Norway and Iceland since its development (McReady & Adlaf, 2006). However, widespread use of the measure does not ensure that it is functioning properly with regard to reliability, validity, and generalizability. For example, whereas the nine PGSI items are often summated to derive a problem gambling score, previous work by Maitland and Adams (2005) showed that two factors fit better than a single scale score for the nine items, thereby raising question about the best use of information from the PGSI. PGSI items 1 – 4 were used to create a behaviors factor whereas PGSI items 5 – 9 were modeled as a consequences factor.

Furthermore, this study found low to moderate reliability estimates and also demonstrated a lack of ME/I when examining the PGSI items for men and women. This implies that the construct of problem gambling as measured by the PGSI is different between sexes, thereby raising concern about the comparability of results from this index. Extrapolating these results across studies and other potential comparison groups raises serious concern as to what is being tapped by the PGSI items and how these results are being employed.

Critical Questions and Concerns

To further address the concerns raised by Maitland and Adams (2005), the current study must raise some broader questions. Are problem gambling concerns the same regardless of sampling variation (i.e., both with regard to sampling methods and the characteristics of different samples)? Is problem gambling the same across independently collected samples or across sexes? When a researcher or group employs the CPGI/PGSI, the implicit assumption is that the index will work equally well for comparisons across any samples, groups, or time points collected (e.g., see comparisons in British Columbia Problem Gambling Prevalence Study, 2003). Worded differently, researchers using the index expect that their results have a degree of comparability with other colleagues using CPGI, therefore some degree of generalizability is assumed to exist. Not only did a two factor model consistently fit better than the one factor model described by Wynne and others, but other problems were noted by

Maitland and Adams. Whereas significant factor loadings resulted when the entire sample (n=4631) was tested some non-significant factor loadings were found when smaller, random samples of 10% and 20% were examined. Furthermore, regardless of sample size, a lack of ME/I was noted in all instances, suggesting that valid comparison across groups could not be made, regardless of sample size.

Limitations of Prior Analyses

Most factor solutions reported in the literature are exploratory analyses. The inherent problem with the use of exploratory factor analysis is the ease in which results can be misused and/or misinterpreted. In order to determine how many factors are needed, competing models consisting of one-, two- and/or more-factors are needed. The decision as to how many factors exist should be based on a combination of theory and statistical results. If a study only tests a one-factor model in the absence of comparisons to competing models, one can only accept a one-factor model as best fitting. The onus is still on the researcher to demonstrate that the model fits adequately, however, the lack of models against which to compare makes this result tenuous at best. Furthermore, it is impossible to test for ME/I using exploratory techniques. The report by Maitland and Adams (2005) employed confirmatory factor analysis (CFA), thereby raising the level of rigor as well as allowing direct tests of ME/I. Even so, Maitland and Adams' earlier work was based on a single data source; a prevalence study data set collected by Wiebe (2002). Whereas this work clearly entered new territory with regard to statistical methodology, it also left some important issues unanswered. The concern whether one or two factors best define the PGSI items should be examined in additional datasets. Maitland and Adams showed that two factors better fit the PGSI however they also demonstrated that some significant results were driven by excessive power in large datasets. Previously significant factor loadings became non-significant when examined in smaller random subsets of the data. Model fit was also degraded in these smaller samples. This study suggested that the low prevalence rates of problem gambling might require larger samples to adequately fit the PGSI. The current study employed three large independent data sets to address these concerns.

The structure of the PGSI was examined in each of three datasets, as well as evaluating all three datasets simultaneously to determine whether one or two factors best fits PGSI data. Additionally, there are limited attempts to examine how well the PGSI performs across comparison groups, with the Maitland and Adams report the only previously example of tests of measurement invariance of PGSI to date. Therefore, ME/I was examined for the 9 PGSI items across the three datasets. Tests of sex invariance were examined in each dataset. A brief overview of confirmatory factor analysis and ME/I follows.

Confirmatory Factor Analysis

CFA allows one to determine the best model that represents 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. Is problem gambling as measured by the PGSI simply a single factor, thereby allowing the use of a simple additive strategy across all 9 items as is so commonly used? The work of Maitland and Adams (2005) suggested that one could not entirely rule out the use of the additive strategy, however, two factors appeared to represent the underlying structure of the PGSI better than a unitary solution. If some PGSI items measure behaviors and other items measure consequences, this might suggest that respondents cannot experience the consequences in the absence of performing the behaviors. This issue may have serious consequences for the additive model normally applied to PGSI data (i.e., totaling scores on the 9 items into one score) as well as the cut-off values used to classify gamblers based on the additive strategy. The current study brings together datasets from three independent studies that employed the PGSI as part of their measurement battery. We extend the work of Maitland and Adams (2005) by examining the structure of the PGSI in each of these three distinct datasets as well as looking at the consistency (or lack thereof) of results across these data sources. Each dataset was tested separately to determine whether a unidimensional or single factor model fits the PGSI data better than a two-factor model that separates behaviors from consequences. Results were then extended to examine this question simultaneously across the three samples.

How does CFA differ from EFA?

A number of issues differentiate exploratory factor methods from CFA. EFA relies upon the statistical software to determine the number of factors to retain (though the researcher may specify a preconceived notion beforehand), whereas in CFA, the researcher specifies the number of factors that define a model. Next, *all* variables will have a factor loading onto *all* factors in an EFA. CFA allows the researcher to specify which variables are expected to load onto which factors, thereby eliminating trivial or non-salient loadings from the model. EFA provides very little in the way of evaluating model fit whereas most software packages that perform CFA also produce an excess of fit indexes (e.g., more than 30 indexes are provided in AMOS or LISREL). Therefore, careful selection and evaluation of model fit is much better in CFA. Finally, it is impossible to directly compare results from multiple EFA models without conducting extensive hand calculations to compare coefficients. Additionally, this inability to compare models directly removes the capacity to assess ME/I. Therefore, future comparisons of the CPGI/PGSI should employ CFA to examine the underlying structure of the index, especially if the goal is evaluation of comparability of results across any groups or studies.

Why is ME/I important?

If replication is the key to science then methods of making valid comparisons are required to accomplish this goal. Unfortunately, some critical flaws are found when one examines the previous literature concerned with CPGI/PGSI. First, exploratory factor analyses are still being conducted even with the ease of access to confirmatory programs that would produce more rigorous results. Second, most studies employ a one-factor model and do not question whether this model fits the data appropriately (i.e., comparison of competing models is ignored). All researchers who employ the additive strategy of all 9 PGSI items are endorsing a single construct whether the decision is conscious or not. Third, a recent review found that upon interviewing researchers who have employed the CPGI in their studies, "the CPGI is considered to have provided a consistent measure of gambling and problem gambling in Canada (p. 7, McCready & Adlaf, 2006)." Even so, one researcher described the measure as state of the art but admitted that the state of the art is not very good. This supports the idea that researchers are assuming that the PGSI is working as advertised, thereby relying on the weakest form of validity (i.e., face validity). In the absence of conducted formal tests of ME/I for the PGSI across comparison groups researchers are taking a large gamble themselves.

More about ME/I

Upon determining which factor structure is the best fitting model for the PGSI in each dataset (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 within each study, as well as across all three studies simultaneously, before assuming that quantitative scores on the constructs are qualitatively comparable (Horn & McArdle, 1992; Maitland & Adams, 2005; 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). This concern is critical within each study and across the three independently collected studies as a failure to show equivalent factor structures would indicate a lack of comparability of results.

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 and one only moves to the next level of constraints if the prior level has been found to be acceptable. The framework of Meredith (1993) was employed here and includes models testing for: (a) *configural invariance* tests similarity of factor patterns across comparison groups, ignoring differences in the magnitude of factor loadings between groups; (b) *weak metric invariance of factor loadings*, is a test of no differences between the location and magnitude of factor loadings between groups [*Note*. 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)]. Additional constraints are used to test (c) *strong invariance of observed variable intercepts* where observed variable intercepts are constrained to be equivalent across sexes; and, (d) *strict invariance*, testing if uniqueness terms are invariant across sexes or studies only when found tenable. Finally, if factor loadings are found to be invariant, a test of equivalence of any structural relationships/regressions may also be conducted (e.g., the regression between gambling behaviors and gambling consequences).

Analytic Strategies

Statistical Procedures

Models were tested using AMOS 6.0 (SPSS, 2005). Response patterns for the PGSI items result in non-normally distributed data. For example, most respondents only utilize one or two response options (i.e., never, sometimes) when answering PGSI items. Therefore, PGSI data will rarely (or never) produce normally distributed outcomes and this issue requires attention. Previous exploratory studies fail to address this issue. Furthermore, use of non-normal data with exploratory methods results in unreliable estimates, therefore leading to tenuous conclusions for any results based on these data. As reported by Maitland and Adams (2005), the non-normal nature of this data was assessed 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 transformation attempts. Additionally, one of the primary assessment procedures proposed for examining the validity and generalizability of the PGSI, ME/I, is best suited to applications using covariance matrices. Methods for testing factorial invariance with non-normal data are still being developed. Therefore all analyses in the current study were conducted on covariance matrices and 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 or across studies. 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 $p < .01$[1].

Comparative fit indexes

Model fit was evaluated by examining the following fit indexes: (a) model χ^2; (b) Goodness of Fit Index (GFI; Jöreskog & Sörbom, 1984); (c) Non-Normed Fit Index (NNFI; Bentler & Bonnett, 1980); (d) Comparative Fit Index (CFI; Bentler, 1990); (e) Root Mean Square Error of Approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980).

[1] When comparing the one- and two-factor models, a significant difference in χ^2 values indicates that the second model (i.e., two-factors) makes a significant improvement in fit when compared to the one-factor model.

Model fit and model comparisons

For readers who are unfamiliar with the terminology of assessing model fit we offer the following description. To engage in model fitting 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. Unfortunately, chi-square is artificially inflated when used in large samples and therefore, will often result in a significant test. However, the chi-square test is useful when testing nested models (models within models) and is the standard measure used for this purpose. The goodness-of-fit index (GFI) compares the explained covariance to total measured covariance. The adjusted goodness-of-fit index (AGFI) is similar to the goodness-of-fit index except it adjusts for the degrees of freedom in the model and results in a more conservative estimate than GFI. Finally, the root mean square error of approximation (RMSEA) 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 (CFI) compares the tested model to a null model that has no paths that link the variables, thus making the variables independent of each other. A score of .90 or higher is considered acceptable-- the closer to 1.0 the better the fit. The normed fit index (NFI) is 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 (NNFI) is a generalized version of the Tucker and Lewis index and was created to correct some of the problems found in the NFI.

The current study provides an analysis of the PGSI across three studies, including prevalence studies from Ontario (Williams & Wood, 2004), Quebec (Ladouceur, Jacques, Chevalier, Sevigny, Hamel, & Allard, 2005), and a follow-up to a prevalence study completed in Ontario (Weibe, 2003 & 2005). Our goals included: (a) subjecting the PGSI items to Confirmatory Factor Analysis (CFA) to determine if the one-factor model proposed by Wynne (2003) provided the best fit to the data for each dataset. We compared the one-factor model to a two-factor model described by Maitland and Adams (2005) to determine if one- or two-factors best describes the PGSI items. Next, we subjecting the best fitting model to (b) tests of ME/I, to determine if the factor structure of the PGSI is stable across sexes (within each study) as well as across all three studies simultaneously. From a broader methodological (and practical) perspective, the proposed research documents the degree to which the PGSI should be used for comparative purposes, and provides evidence about the degree of generalizability and replicability of results employing this screening instrument.

Study Rationale in Relation to the Empirical and Theoretical Literature

The importance of this study cannot be overstated. The CPGI and PGSI were developed to provide a "gold standard" tool for use in measuring gambling behavior in Canada and beyond. As noted in the recent review by McCready and Adlaf (2006), this index has been used in numerous studies within Canada and is growing in popularity in other countries around the world. Remarkably, this growth has occurred in the absence of firm empirical evidence to support or contradict the underlying factor structure of the index. Limited evidence to date using confirmatory analyses shows mixed results and researchers should be cautious about using the PGSI to make group comparisons (Maitland & Adams, 2005). If measurement invariance is demonstrated for the PGSI in the current study (e.g., either between sexes or across all three samples), researchers may be reassured that they may make appropriate comparisons across different samples/studies. Positive results from this study will serve to strengthen the PGSI . Negative results (e.g., finding the proposed factor structure does not result or that measurement invariance is absent) provides further validation of the results reported by Maitland and Adams (2005) and raises questions about how the PGSI should be employed to study problem gambling.

Sampling and Recruitment

Data from three sources were used: (1) The Demographic Sources of Ontario Gaming Revenue study (Williams & Wood, 2004) included a sample of 2535 participants comprised of 1288 men and 1247 women, however missing data reduced the usable sample to 2498 (1275 men, 1223 women or 99% with PGSI data). The high PGSI rate was a result of the sampling procedure. Participants were asked how much money they had spent in a typical month in the past year on lottery, raffle or instant win tickets; playing Sports Select; playing slot machines and table games at Ontario casinos and racetracks; horse race betting; and bingo. If they responded that $9 or more was spent on such activities they completed the entire measurement battery. PGSI scores for all valid cases ranged from 0 – 27 (M = .88). (2) The Prevalence Study of Pathological Gambling in Quebec was conducted in 2002 (Ladouceur, et al., 2005). This study employed the CPGI and SOGS and comprised a sample of 8842 participants including 4352 men and 4490 women. However, only 669 or 8% of participants comprised of 415 men and 254 women had PGSI data and were available for analysis. The low portion of the sample that completed the PGSI was a result of the screening method to determine problem gambling. Respondents had to answer "yes" to one of the following criteria to be assessed for problem gambling: (1) have spent more than $520 annually on gambling; or (2) have played too much, spent too much money, or spent too much time gambling. PGSI scores for all valid cases ranged from 0 – 23 (M = 1.19). (3) The Prevalence Study of Problem Gambling in Ontario (Wiebe, 2003 & 2006) provided the final data source for this study. Wiebe employed the CPGI in a sample of 3604 participants comprised of 1531 men and 2073 women (Total: 2248; Men: 1029; Women: 1219 or 62% with PGSI data). Any person who had gambled on at least one of the 18 gambling activities in the past year were given the PGSI. PGSI scores for all valid cases ranged from 0 – 27 (M = .44).

Analyses

Assessment of psychometric properties were conducted using structural equation modeling (SEM) to examine ME/I (see Maitland & Adams, 2005; Maitland, et al., 2001; Maitland, et al., 2004; Maitland, Intrieri, Schaie & Willis, 2000; Nyberg, et al., 2003; Schaie, Maitland, Willis & Intrieri, 1998; Vandenberg & Lance, 2004). ME/I between sexes within each dataset was examined. Multi-group analyses were also conducted to examine the PGSI factor structure across all three studies simultaneously. Differences in factor structure could result at multiple levels. First, evidence could suggest that a one-factor model represents the PGSI in one study whereas a two-factor model is better in another. This type of evidence would raise serious criticism against comparing the results for the two studies; based on the disparity in measurement properties. Second, the same factor structure might result but the magnitude of the factor loadings may differ across studies. This represents *configural invariance* but does not provide enough evidence to permit comparison across the studies. Third, assuming all factor loadings are equivalent across all three studies, *weak invariance* is obtained and provides strong evidence to allow comparisons between studies. Any additional equivalence constraints that might be allowed beyond weak invariance simply provide additional strength to the argument for cross-study comparisons.

Results

Based on previous reports for the PGSI, the nine items of the scale are summed to provide an index score for respondents. This approach is shown in the conceptual model below (see Figure 1). Note that PGSI items are simply labeled V1-V9 in the same order as presented in the CPGI.

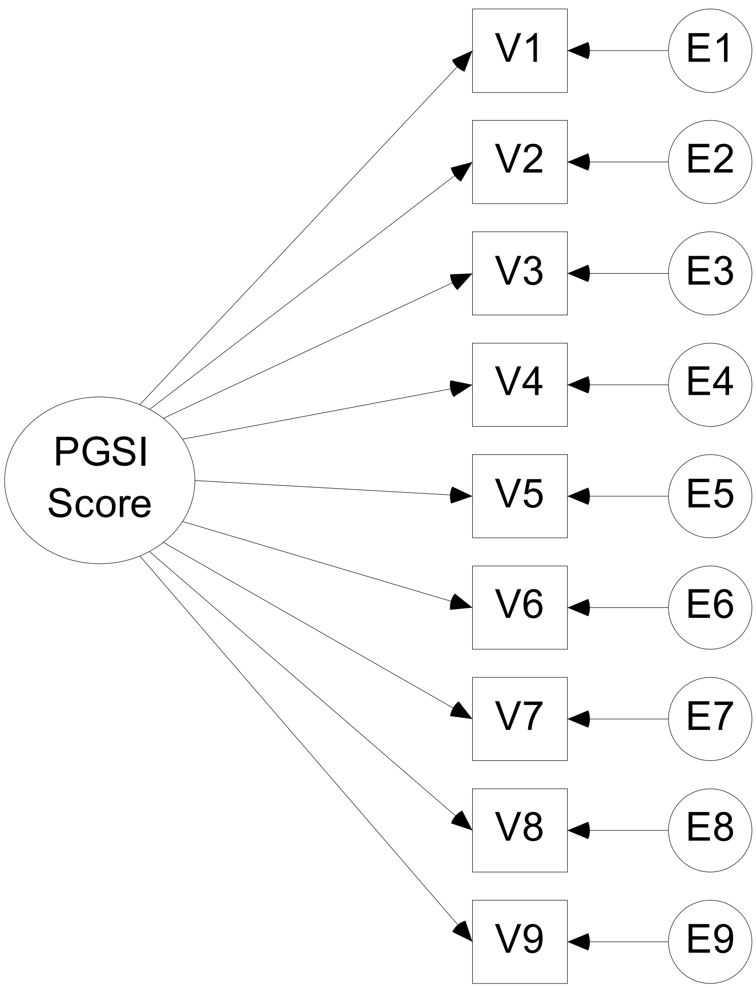

Figure 1. PGSI One Factor Model

However, examination of the nine PGSI items, the theoretical development of CPGI, and the report by Maitland and Adams (2005) suggest a competing two factor model. 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.

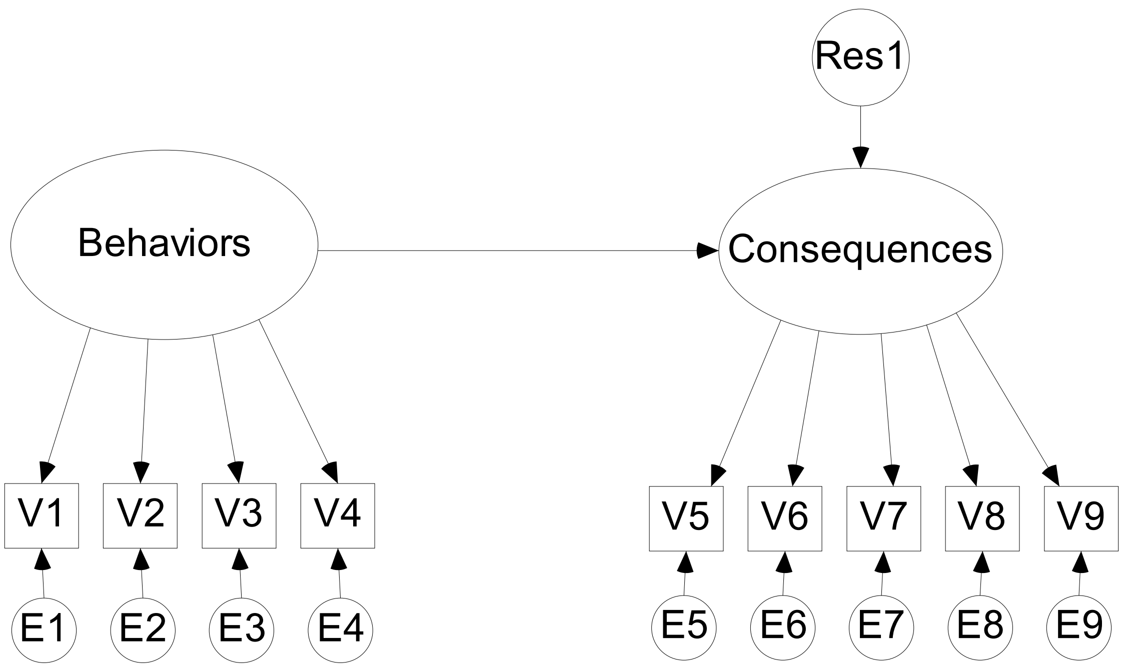

Figure 2. Two Factor PGSI Model

Model Fit and Comparisons

Information included psychometric properties and model fit/comparisons are provided for each of the three studies separately (Williams & Woods; Ladouceur et al; Wiebe), followed by simultaneous models modeling all three datasets.

Williams and Wood study

Reliability

Reliability for all 9 PGSI items, using Cronbach's coefficient alpha was .77 for the overall sample. Examining Behaviors and Consequences for the entire sample resulted in .57 and .75 respectively. Alpha for men was .69 (Behaviors $\alpha = .47$; Consequences $\alpha = .66$) and .83 (Behaviors $\alpha = .66$; Consequences $\alpha = .82$) was noted for women, clearly showing stronger estimates for women and for consequences when compared to behaviors. Reliability information for all datasets is presented in Appendix A.

One-Factor versus Two-Factor Model

Results from AMOS 6 showed that the one-factor model had statistically significant factor loadings (range: .46 - .67) however questionable model fit was shown by consensus of fit indexes: $\chi^2 = 438.05$, $df = 27$, $p < .001$, GFI = .962, NNFI = .888, CFI = .816, RMSEA = .078. It should be noted that the GFI is quite good, whereas *none* of the other values reach a level of acceptability. Squared multiple correlations (SMCs) for all variables range from .21 (21% of variance in item 1 accounted for by the factor) to .45 or 45% of the variance in item 9 explained by the factor (see Figure 3).

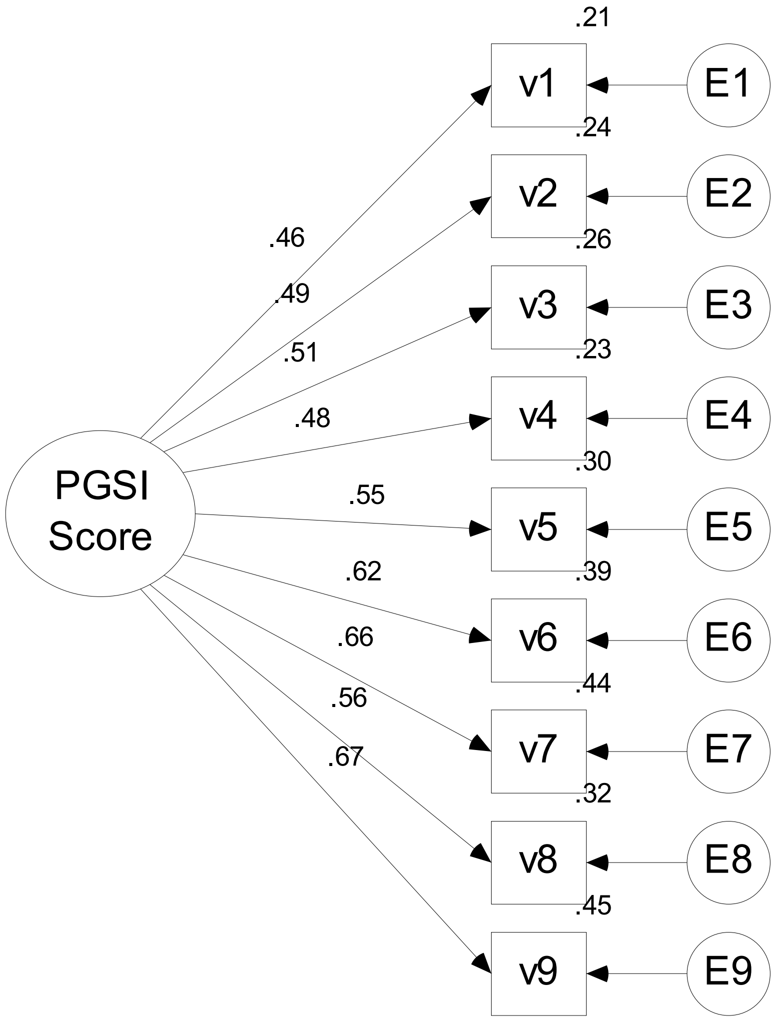

Figure 3. Williams and Woods Study – One Factor Model

The two-factor model resulted in statistically significant factor loadings ranging from .48 - .52 for Behaviors and from .55 to .67 for Consequences. Model fit was improved over the one-factor model but still questionable for some measures: χ^2 = 432.20, *df* = 26, *p* < .001, GFI = .963, NNFI =.888, CFI = .919, RMSEA =.078. SMCs for all variables ranged from .23 (23% of variance in item 1 accounted for by the Behavior factor) to .46 or 46% of the variance in item 9 explained by the Consequences factor. Comparison of the one factor and two factor models reveals $\Delta\chi^2_{M2\text{-}M1}$ = 5.85, *df* = 1, *p* < .001. A value larger than 3.84 (*p* = .05) or 6.64 (*p* = .01) indicates that the second model fits better than the first. Therefore, the two factor model shows statistically significant though marginal improvement in model fit as demonstrated by the statistically significant difference between models. Fit indexes remained similar for both 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 $\beta = .93$, $p < .001$, demonstrating that the latter are dependent on the former as expected and that 87% of variance in consequences are attributed to behaviors (see Figure 4).

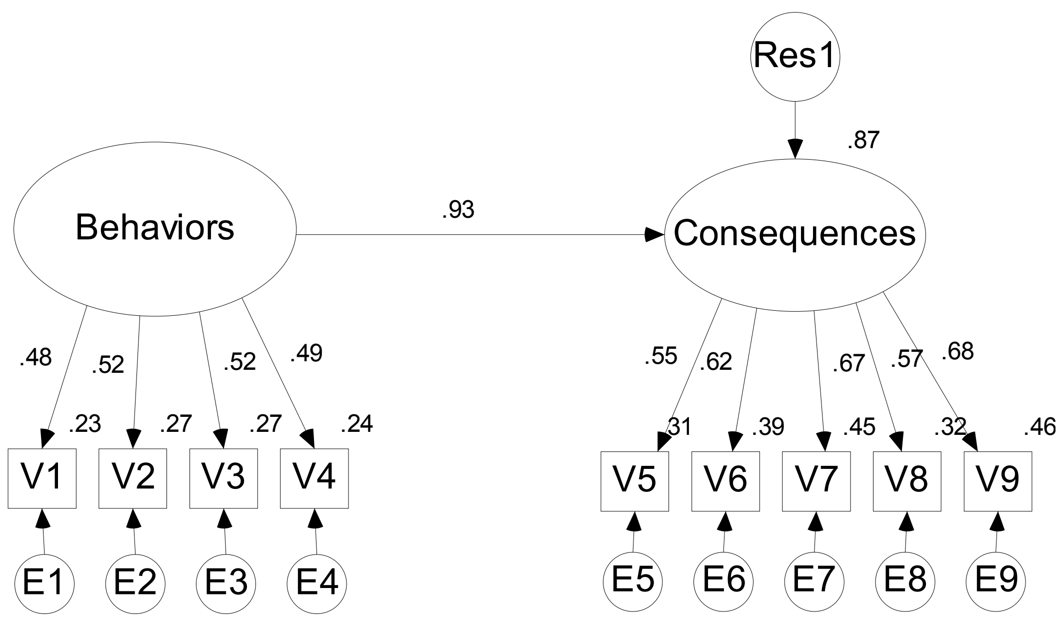

Figure 4. Williams and Woods Study – Two Factor Model

Sex Invariance of One-Factor PGSI Model

To test the hypothesis of sex invariance of the PGSI models, we expanded our strategy into a multi-group, structural model examining men and women. This initial multi-group model examined configural invariance of the one-factor PGSI models in men and women for the Williams and Wood prevalence sample (Configural Invariance Model: $\chi^2 = 935.39$, $df = 54$, $p < .001$, GFI = .923, NNFI = .798, CFI = .849, RMSEA =

.081). Results showed that the one-factor model results in statistically significant factor loadings (ranging from .35 - .69 for men; .47 - . 83 for women). Loadings for women were stronger overall. SMCs for men ranged from .12 (12% of variance in item 3 accounted for by the factor) to .47 or 47% of the variance in item 9 explained by the PGSI factor. SMCs for women ranged from .22 (22% of variance in item 2 accounted for by the factor) to.69 or 69% of the variance in item 6 explained by the factor. Standardized results for men are detailed in Figure 5.

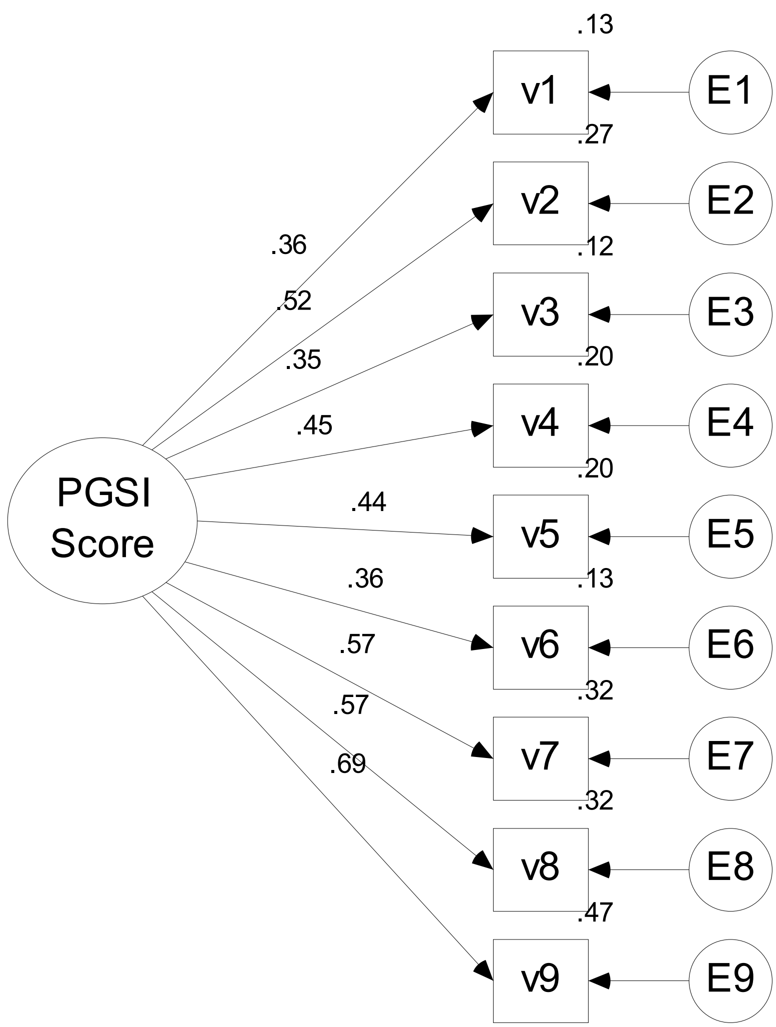

Figure 5. Williams and Woods Study – One Factor Model for Men

Standardized results for women are depicted in Figure 6.

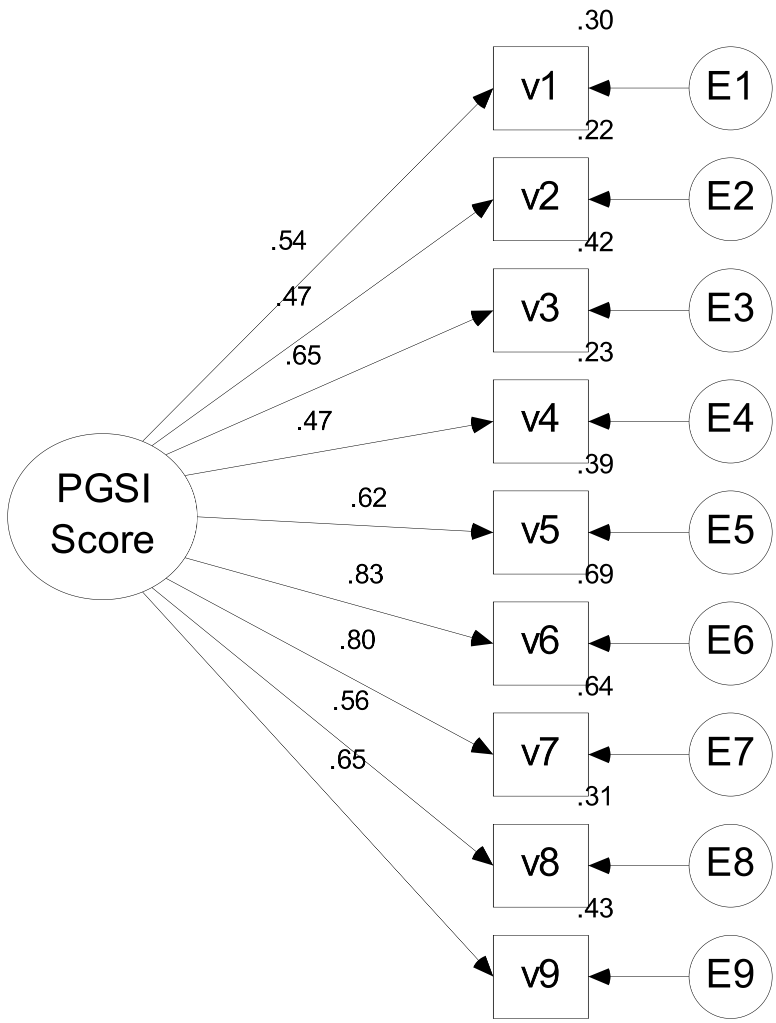

Figure 6. Williams and Woods Study – One Factor Model for Women

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: χ^2 = 1204.25, *df* = 62, *p* < .001, GFI = .901, NNFI = .772, CFI = .804, RMSEA = .086, showed worsened fit and model fit indexes that were not acceptable. Delta chi-square for 8 degrees of freedom was 268.860 chi-square points, clearly demonstrated a lack of invariance of factor loadings across genders[2]. No further invariance tests were conducted for the one-factor PGSI model in the Williams and Wood dataset.

Sex Invariance of the Two-Factor PGSI Model

Sex invariance of the two-factor PGSI model was examined next: Configural Invariance Model: $\chi^2 = 897.66$, $df = 52$, $p < .001$, GFI = .926, NNFI = .799, CFI = .855, RMSEA = .081. The two-factor model resulted in statistically significant factor loadings ranging from .35 - .53 for men's behaviors and .36 - .69 for men's consequences. SMCs for men ranged from .13 (13% of variance in item 3 accounted for by the factor) to .48 or 48% of the variance in item 9 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for men was $\beta = .96$, $p < .001$, accounting for 91% of the variance in Consequences.

Standardized factor loadings were .48 - .68 for women's behaviors and .54 - .85 for women's consequences. SMCs for women ranged from .23 (23% of variance in item 4 accounted for by the factor) to .72 or 72% of the variance in item 6 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for women was $\beta = .89$, $p < .001$, accounting for 80% of the variance in Consequences (compared to $\beta = .93$ for men). Overall model fit showed CFI, NNFI and RMSEA values were unacceptable, suggesting questionable fit to the data.

Standardized results for the configural two-factor PGSI model for the male sample are shown in Figure 7:

[2] Determination of ME/I employs $\Delta\chi^2$ to compare competing models. However, the logic for decision making differs from the earlier comparison of the one- versus two-factor model. For the hypothesis of ME/I to be considered acceptable, a non-significant $\Delta\chi^2$ is desired, indicating that the additional constraints placed on the model (e.g., constraining all factor loadings to be equivalent across groups) did not result in a statistically significant decrement of model fit.

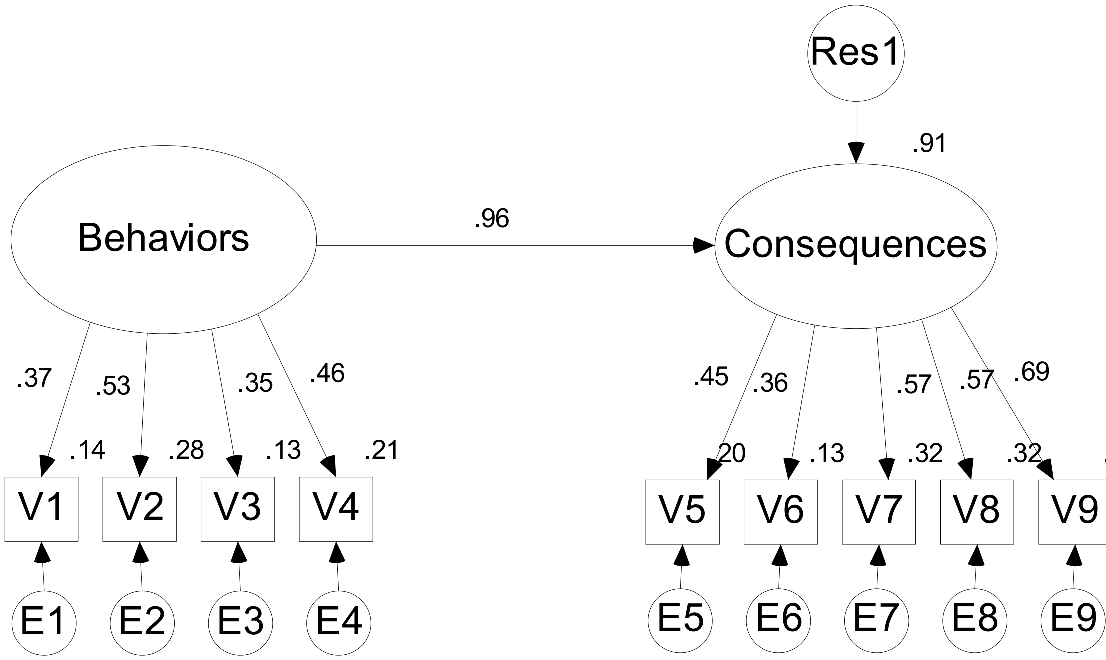

Figure 7. Williams and Woods Study – Two-Factor Model for Men

Standardized results for the configural two-factor PGSI model for all women are found in Figure 8:

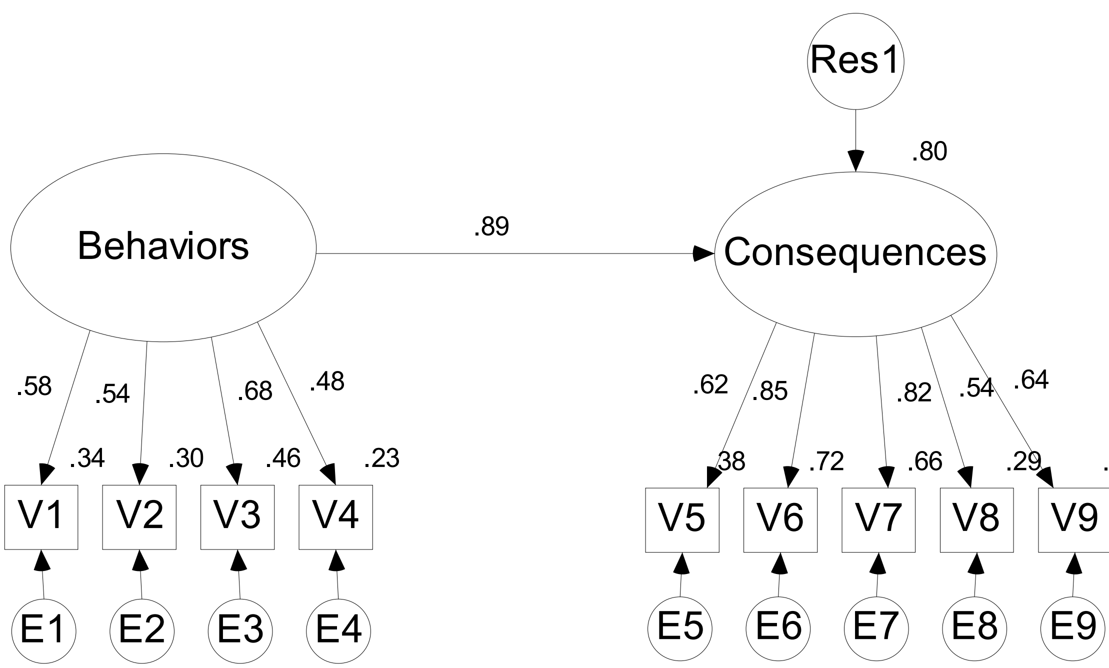

Figure 8. Williams and Woods Study – Two Factor Model for Women

Next, tests constrained all factor loadings across men and women for the two-factor model. Model results: $\chi^2 = 1162.02$, $df = 59$, $p < .001$, GFI = .904, NNFI = .769, CFI = .811, RMSEA = .087, showed worsened fit and model fit indexes that are poor. Delta chi-square for 7 degrees of freedom was 264.36 chi-square points, clearly demonstrated a lack of invariant factor loadings across genders. No further invariance tests were conducted for the two-factor PGSI model.

Therefore, a lack of sex invariance was noted for both the one- and two-factor PGSI models when examined across the data from the Williams and Wood prevalence study.

Ladouceur et al., Study

Reliability

Reliability for all 9 PGSI items, using Cronbach's coefficient alpha was .89 for the overall sample. Examining Behaviors and Consequences overall resulted in .62 and .83 respectively. Values for men were .90 (Behaviors $\alpha = .78$; Consequences $\alpha = .84$) and .88 (Behaviors $\alpha = .81$; Consequences $\alpha = .82$) for women respectively. No clear advantage was noted between groups however, alphas for consequences were higher than those found for gambling behaviors. Additionally, the Ladouceur et al.values were significantly higher than those found for the Williams and Wood study (see Appendix A).

One-Factor versus Two-Factor Model

Results showed that the one-factor model resulted in statistically significant factor loadings (ranging from .55 - .80) and reasonable model fit shown by consensus of fit indexes: M1: $\chi^2 = 231.64$, $df = 27$, $p < .001$, GFI = .925, NNFI = .906, CFI = .929, RMSEA = .107. The GFI, CFI and NNFI are good, whereas the RMSEA is not acceptable. SMCs for all variables range from .30 (30% of variance in item 4 accounted for by the factor) to .64 or 64% of the variance in item 5 explained by the factor (see Figure 9).

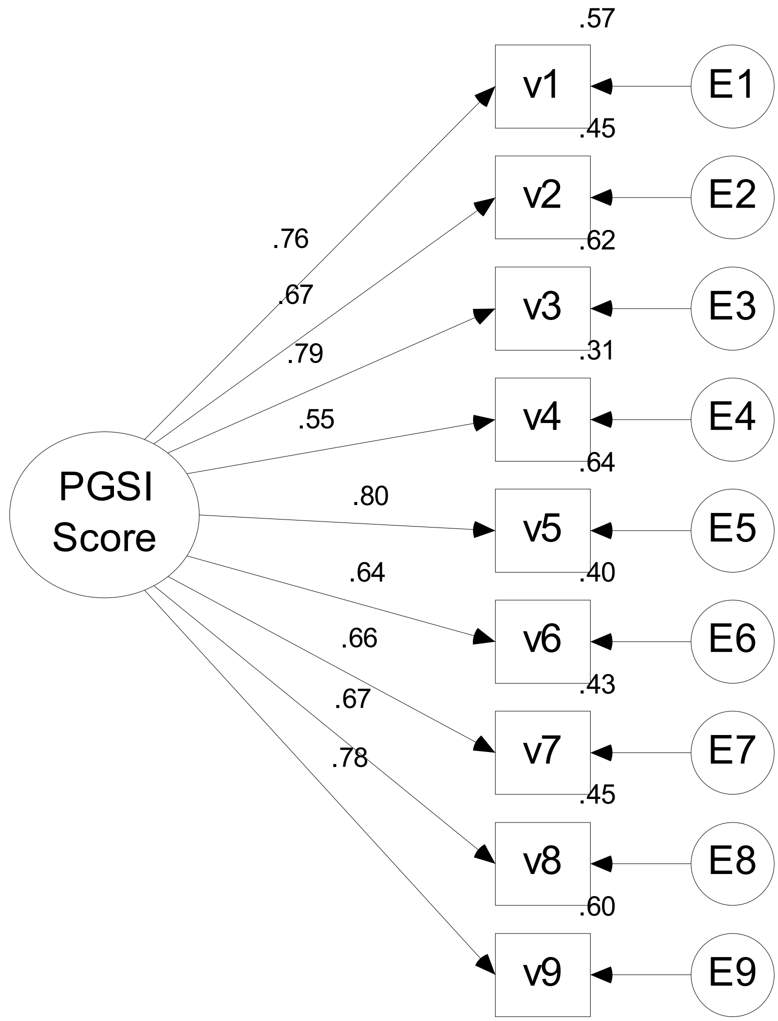

Figure 9. Ladouceur et al Study – One Factor Model

The two-factor model resulted in statistically significant factor loadings ranging from .58 - .81 for Behaviors and from .65 to .83 for Consequences. Model fit was good: $\chi^2 = 165.70$, $df = 26$, $p < .001$, GFI = .945, NNFI = .933, CFI = .952, RMSEA = .090. SMCs for all variables range from .34 (34% of variance in item 4 accounted for by the Behavior factor) to .70 or 70% of the variance in item 5 explained by the Consequences factor. Comparison of the one factor and two factor models reveals $\Delta\chi^2_{M2-M1} = 65.94$, *df* = 1, *p* < .001. Therefore, the two factor model clearly shows significant improvement in model fit as demonstrated by the statistically significant difference between models. Additionally, all fit indexes showed marked improvement for the two-factor model. As with the Williams and Wood data, this finding supported the second model as best fitting, however, the evidence is much stronger in this data. Additionally, the standardized regression coefficient between Gambling Behaviors and Consequences was $\beta = .89$, $p < .001$, demonstrating that the latter are dependent on the former as expected and that 80% of variance in consequences are attributed to behaviors (see Figure 10).

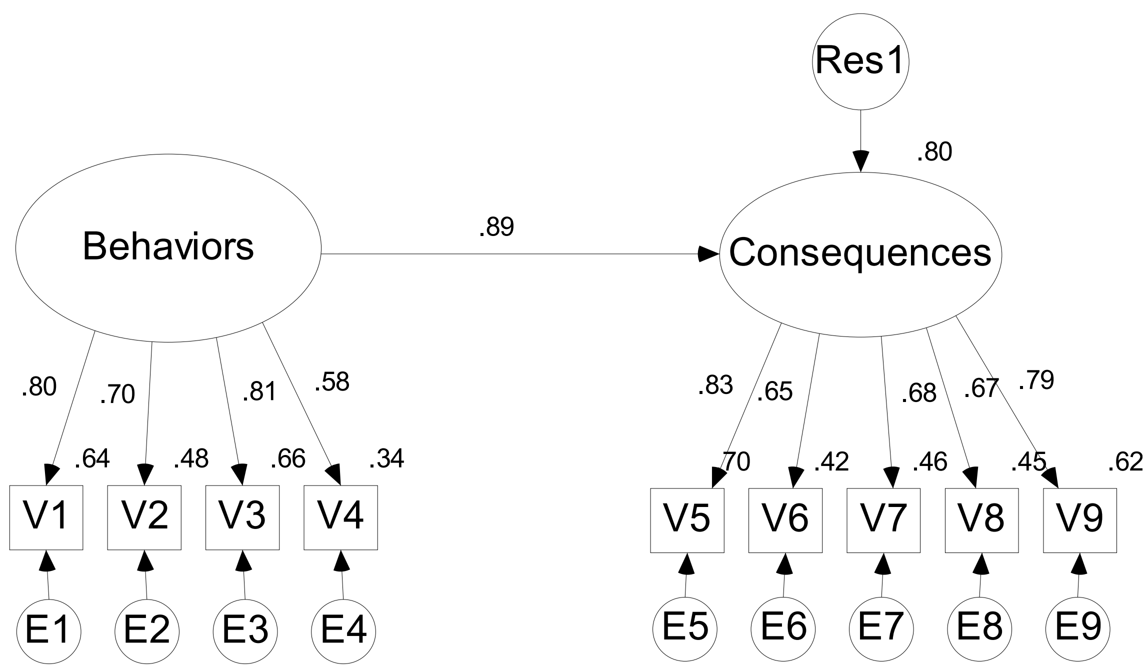

Figure 10. Ladouceur et al, Study – Two Factor Model

Sex Invariance of One-Factor PGSI Model

To test the hypothesis of sex invariance of the PGSI models in the Quebec data, we expanded our strategy into a multi-group, simultaneous structural model examining men and women. This initial multi-group model examined configural invariance of the one-factor PGSI models in men and women for the Ladouceur et al. sample (Configural

Invariance Model: $\chi^2 = 505.79$, $df = 54$, $p < .001$, GFI = .860, NNFI = .811, CFI = .859, RMSEA = .112). The one-factor model resulted in statistically significant factor loadings ranging from .58 - .80 for men; and .46 - .82 for women. SMCs for men ranged from .33 (33% of variance in item 4 accounted for by the factor) to .64 or 64% of the variance in item 5 explained by the PGSI factor. SMCs for women ranged from .21 (21% of variance in item 7 accounted for by the factor) to.67 or 67% of the variance in item 2 explained by the factor. Standardized results for men (see Figure 11):

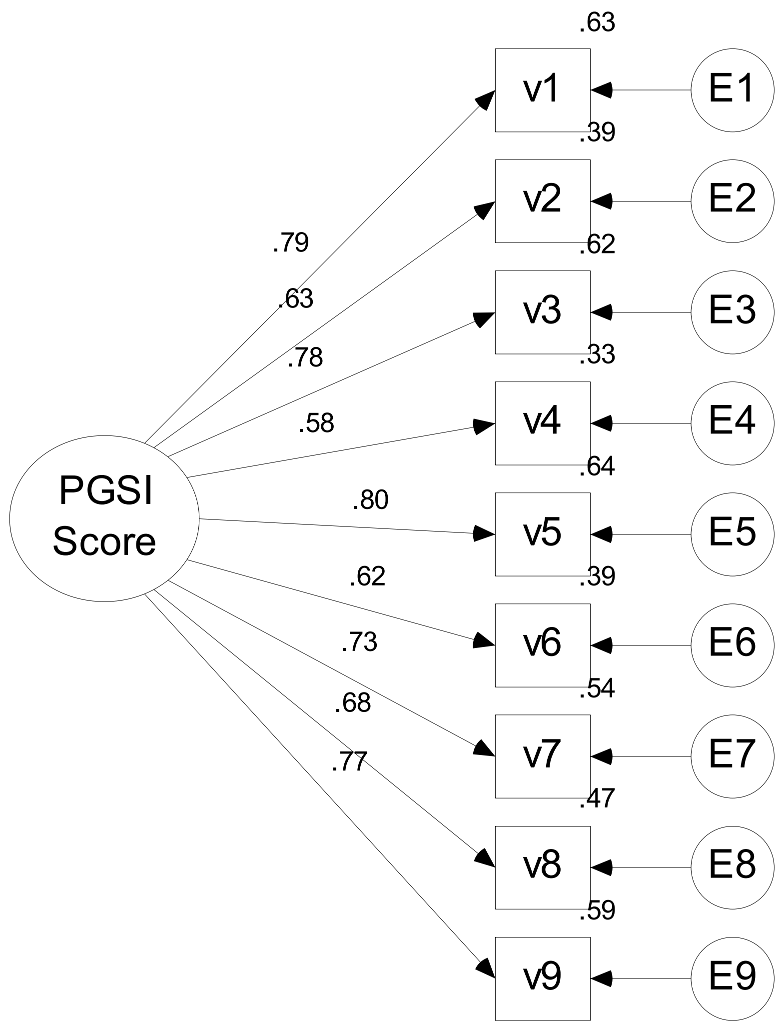

Figure 11. Ladouceur et al, Study – One Factor Model for Men

Standardized results for women are found in Figure 12:

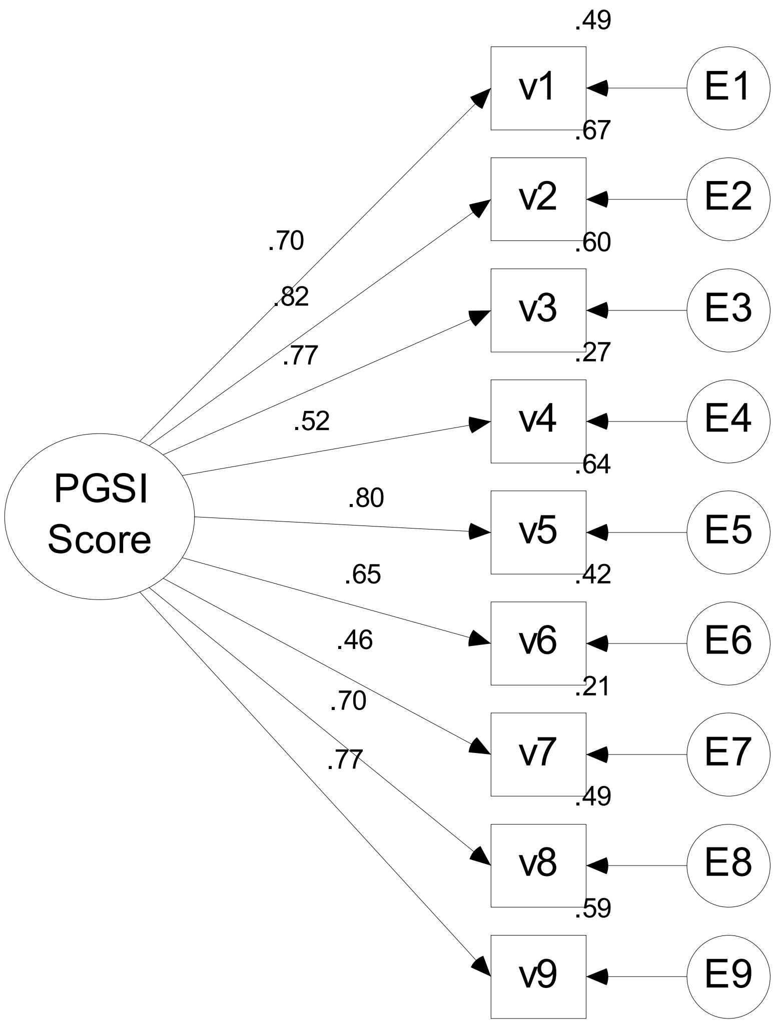

Figure 12. Ladouceur Study – One Factor Model for Women

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: $\chi^2 = 561.01$, $df = 62$, $p < .001$, GFI = .851, NNFI = .819, CFI = .844, RMSEA = .110, showed worsened fit and model fit indexes that were degraded from the configural model. Delta chi-square for 8 degrees of freedom was 55.23 chi-square points, demonstrating a lack of invariance of factor loadings across sexes. No further invariance tests were conducted for the one-factor PGSI model in the Ladouceur et al, dataset.

Sex Invariance of the Two-Factor PGSI Model

Sex invariance of the two-factor PGSI model was examined next: Configural Invariance Model: $\chi^2 = 452.44$, $df = 52$, $p < .001$, GFI = .885, NNFI = .826, CFI = .875, RMSEA = .107. The two-factor model resulted in statistically significant factor loadings ranging from .60 - .82 for men's behaviors and .63 - .82 for men's consequences; and, .55 - .82 for women's behaviors and .56 - .87 for women's consequences. SMCs for men ranged from .36 (36% of variance in item 4 accounted for by the Behavior factor) to .67 or 67% of the variance in item 1 explained by the Consequences factor. The standardized regression coefficient between Gambling Behaviors and Consequences for men was $\beta = .93$, $p < .001$, accounting for 86% of the variance in Consequences.

SMCs for women ranged from .30 (30% of variance in item 4 accounted for by the factor) to .76 or 76% of the variance in item 5 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for women was $\beta = .86$, $p < .001$, accounting for 73% of the variance in Consequences. Standardized results for the configural two-factor PGSI model for the male sample (see Figure 13):

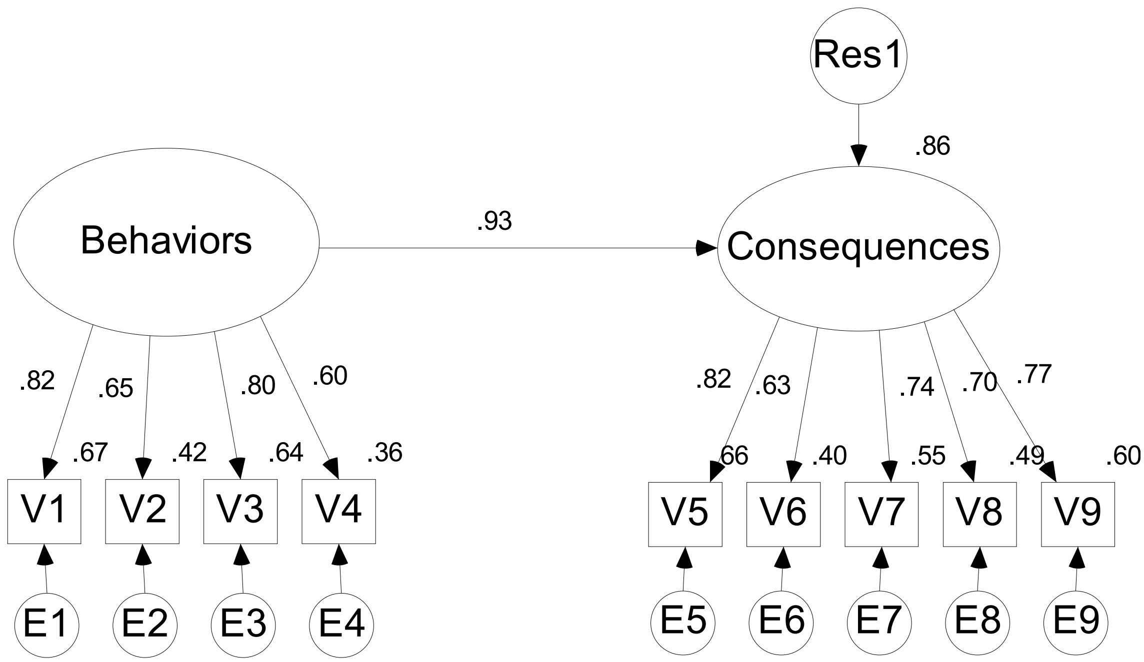

Figure 13. Ladouceur et al, Study – Two Factor Model for Men

Standardized results for the configural two-factor PGSI model for all women in the Ladouceur et al, sample are found in Figure 14:

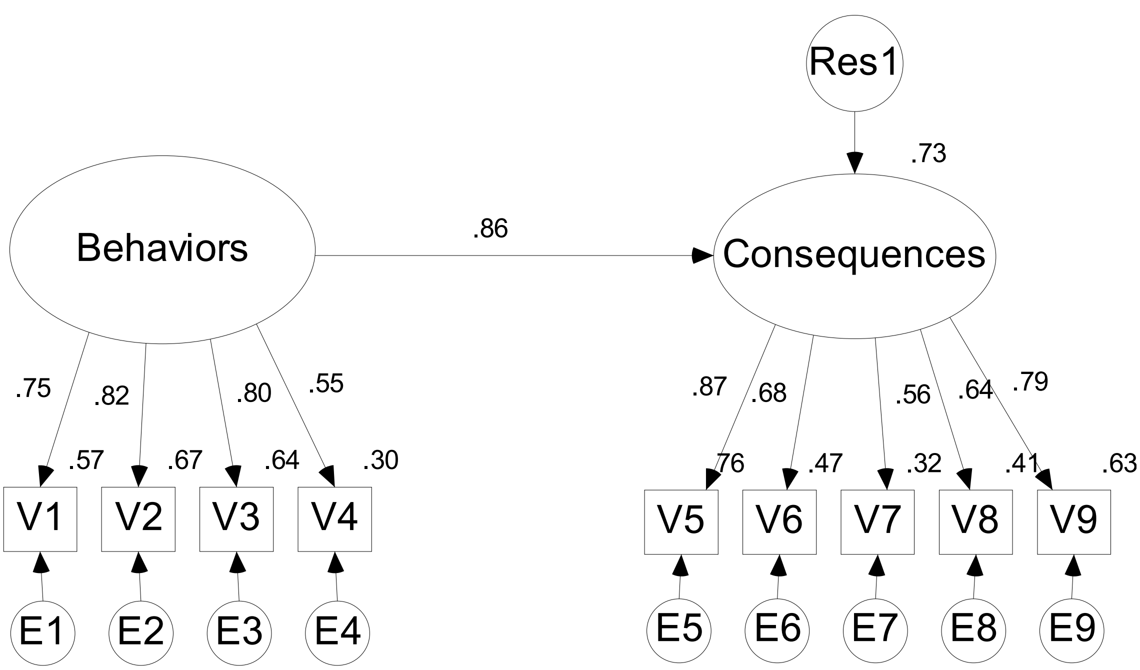

Figure 14. Ladouceur Study – Two Factor Model for Women

Tests constrained all factor loadings across men and women for the two-factor model. Model results: $\chi^2 = 514.27$, $df = 59$, $p < .001$, GFI = .870, NNFI = .826, CFI = .857, RMSEA = .108, showed worsened and unacceptable fit. Delta chi-square for 7 degrees of freedom was 61.83 chi-square points, clearly demonstrating a lack of invariant factor loadings across sexes. No further invariance tests were conducted for the two-factor PGSI model. In summary, a lack of sex invariance was noted for both the one- and two-factor PGSI models for the Ladouceur et al, study.

Wiebe study

Reliability

Reliability for all 9 PGSI items, using Cronbach's coefficient alpha was .85 for the overall sample. Examining Behaviors and Consequences overall resulted in .67 and .82 respectively. Alpha for men was .85 (Behaviors α = .70; Consequences α = .70) and .85 (Behaviors α = .61; Consequences α = .82) for women, showing the same estimates across sexes, whereas behaviors had a higher value for men and consequences was higher for women (see Appendix A).

One-Factor versus Two-Factor Model

The one-factor model resulted in statistically significant factor loadings (ranging from .58 - .77) however, questionable model fit was shown by consensus of fit indexes: M1: $\chi^2 = 782.24$, $df = 27$, $p < .001$, GFI = .923, NNFI = .870, CFI = .902, RMSEA = .112. The GFI and CFI values reached a level of acceptability whereas the NNFI and RMSEA indicated questionable fit. SMCs for all variables ranged from .34 (34% of variance in item 1 accounted for by the factor) to .61 or 61% of the variance in item 8 explained by the factor (see Figure 15).

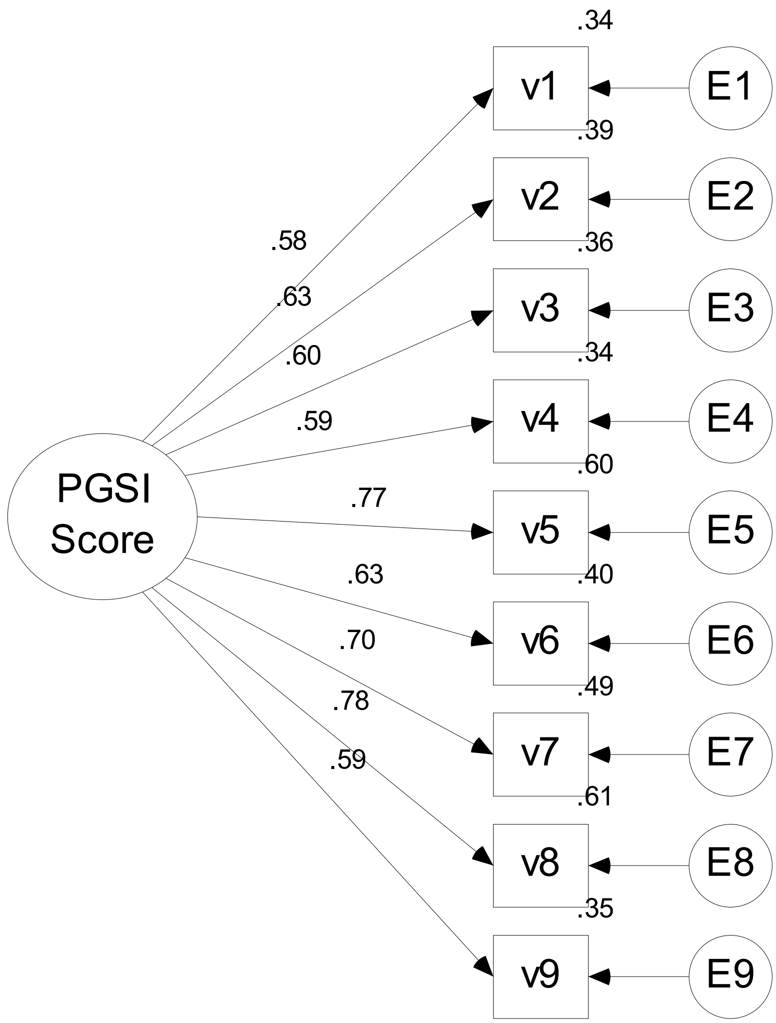

Figure 15. Wiebe Study – One Factor Model

The two-factor model resulted in statistically significant factor loadings ranging from .59 - .73 for Behaviors and from .61 to .80 for Consequences. Model fit was improved over the one-factor model: $\chi^2 = 518.19$, $df = 26$, $p < .001$, GFI = .950, NNFI = .912, CFI = .936, RMSEA = .092. SMCs for all variables ranged from .35 (35% of variance in item 1 accounted for by the Behavior factor) to .64 or 64% of the variance in item 5 explained by the Consequences factor. Comparison of the one factor and two factor models revealed $\Delta\chi^2_{M2\text{-}M1} = 264.05$, $df = 1$, $p < .001$. Some fit indexes showed marked improvement for the two-factor model (e.g., GFI, NNFI and CFI, whereas the

RMSEA remained questionable). This finding supported the second model as best fitting. Additionally, the standardized regression coefficient between Gambling Behaviors and Consequences was $\beta = .82$, $p < .001$, demonstrated that the latter are dependent on the former as expected and that 68% of variance in consequences are attributed to behaviors (see Figure 16).

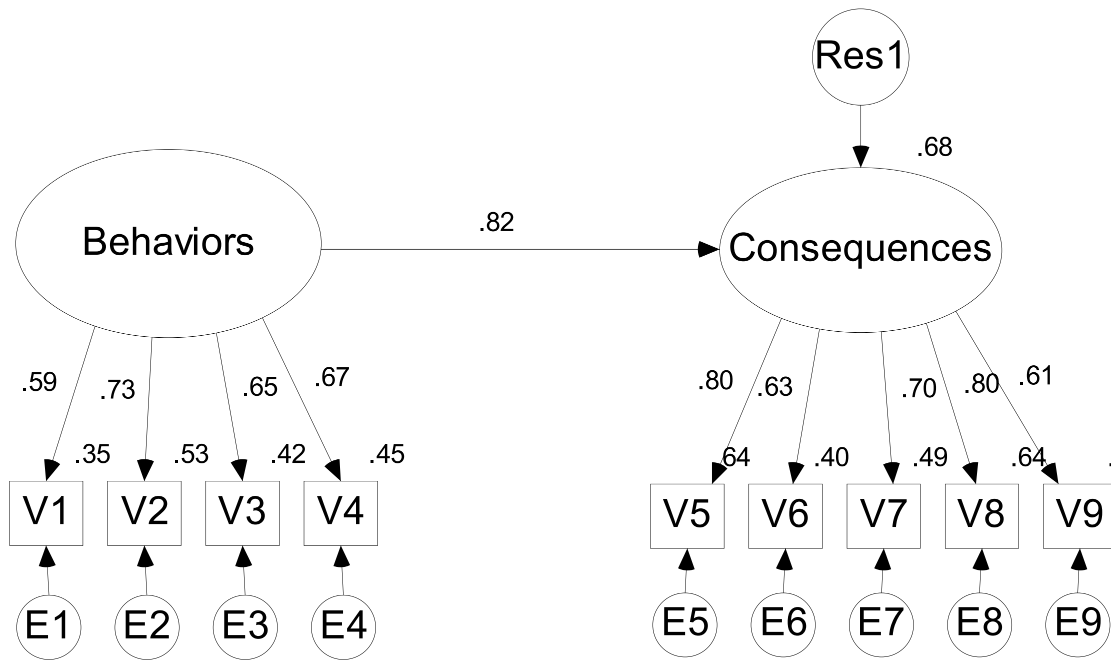

Figure 16. Wiebe Study – Two Factor Model

Sex Invariance of One-Factor PGSI Model

To test the hypothesis of sex invariance of the PGSI models in the Wiebe data, we expanded our strategy into a multi-group, structural model examining men and women simultaneously. This initial multi-group model examined configural invariance of the one-factor PGSI models in men and women (Configural Invariance Model: $\chi^2 = 949.85$, $df = 54$, $p < .001$, GFI = .909, NNFI = .855, CFI = .891, RMSEA = .086). Results showed that the one-factor model resulted in statistically significant factor loadings ranging from .55 - .81 for men; and .54 - .82 for women. SMCs for men ranged from .33

(33% of variance in item 9 accounted for by the factor) to .66 or 66% of the variance in item 5 explained by the PGSI factor. SMCs for women ranged from .29 (29% of variance in item 1 accounted for by the factor) to .68 or 68% of the variance in item 8 explained by the factor. Standardized results for men are depicted in Figure 17:

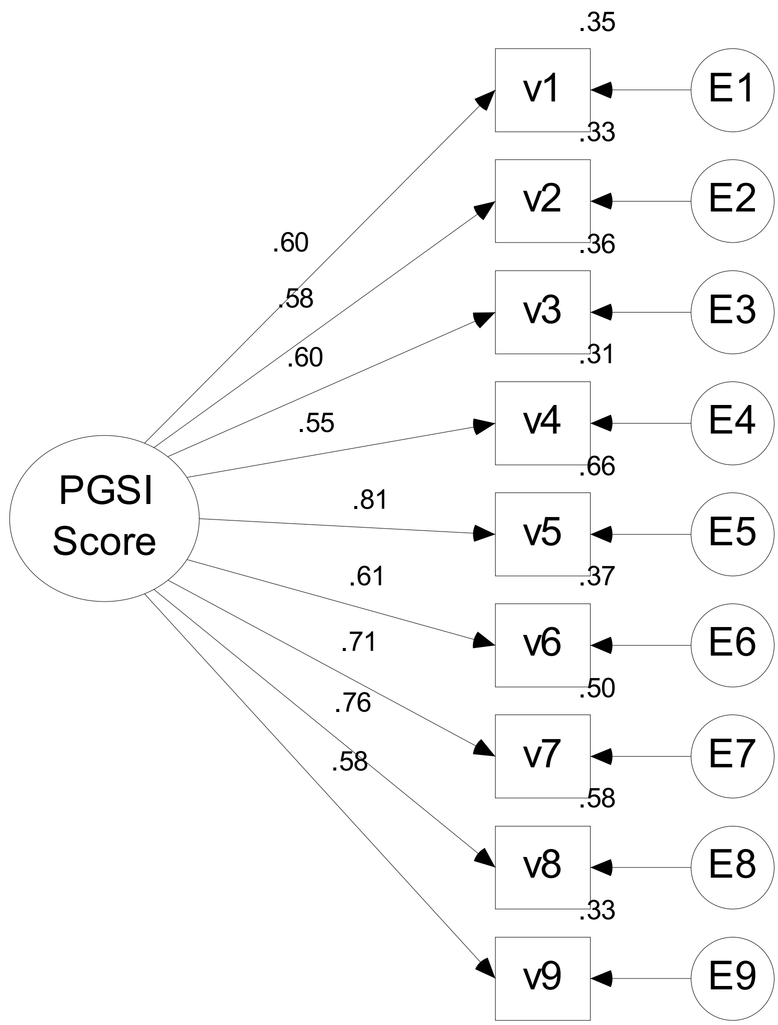

Figure 17. Wiebe Study – One Factor Model for Men

Standardized results for women are depicted in Figure 18:

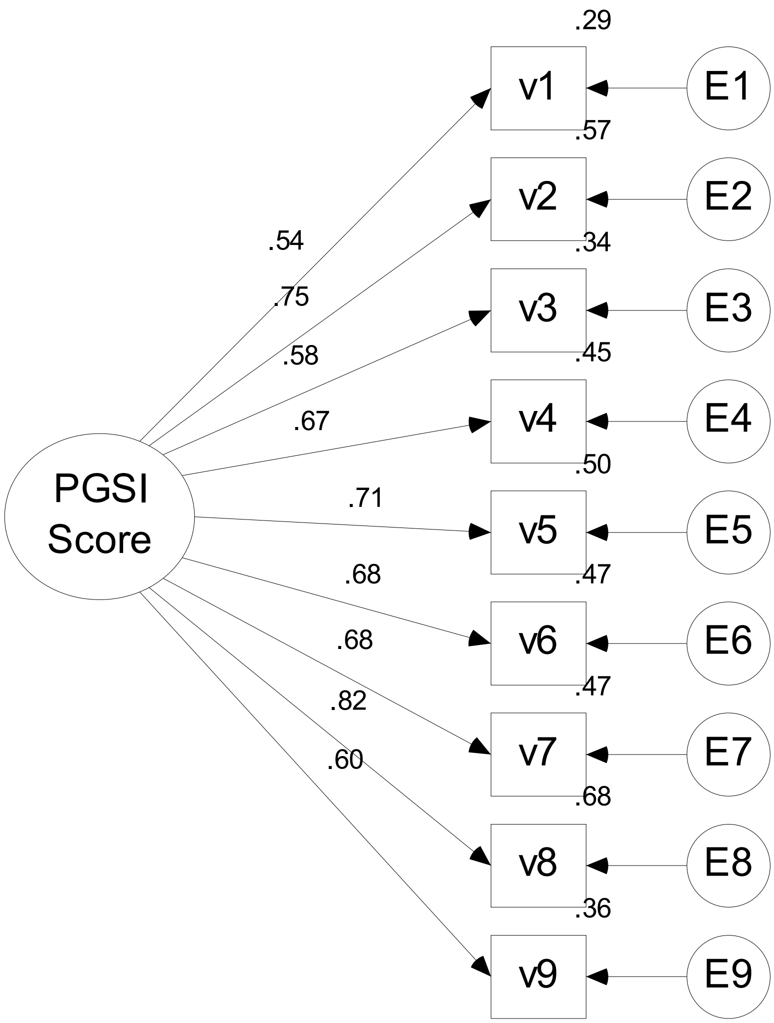

Figure 18. Wiebe Study – One Factor Model for Women

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: $\chi^2 = 1008.53$, $df = 62$, $p < .001$, GFI = .905, NNFI = .866, CFI = .885, RMSEA = .082. Delta chi-square for 8 degrees of freedom was 58.68 chi-square points, demonstrating a lack of invariance of factor loadings and poor fit across sexes.

Sex Invariance of the Two-Factor PGSI Model

Sex invariance of the two-factor PGSI model was examined next: Configural Invariance Model: $\chi^2 = 734.89$, $df = 52$, $p < .001$, GFI = .930, NNFI = 885, CFI = .917, RMSEA = .076. The two-factor model resulted in statistically significant factor loadings ranging from .65 - .70 for men's behaviors and .62 - .81 for men's consequences; and, .49 - .79 for women's behaviors and .58 - .84 for women's consequences. SMCs for men ranged from .33 (33% of variance in item 9 accounted for by the factor) to .72 or 72% of the variance in item 5 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for men was $\beta = .77$, $p < .001$, accounting for 59% of the variance in Consequences.

SMCs for women ranged from .28 (28% of variance in item 1 accounted for by the factor) to .69 or 69% of the variance in item 8 explained by the factor. The standardized regression coefficient between Gambling Behaviors and Consequences for women was $\beta = .94$, $p < .001$, accounting for 89% of the variance in Consequences.

Standardized results for the configural two-factor PGSI model for male sample are depicted in Figure 19:

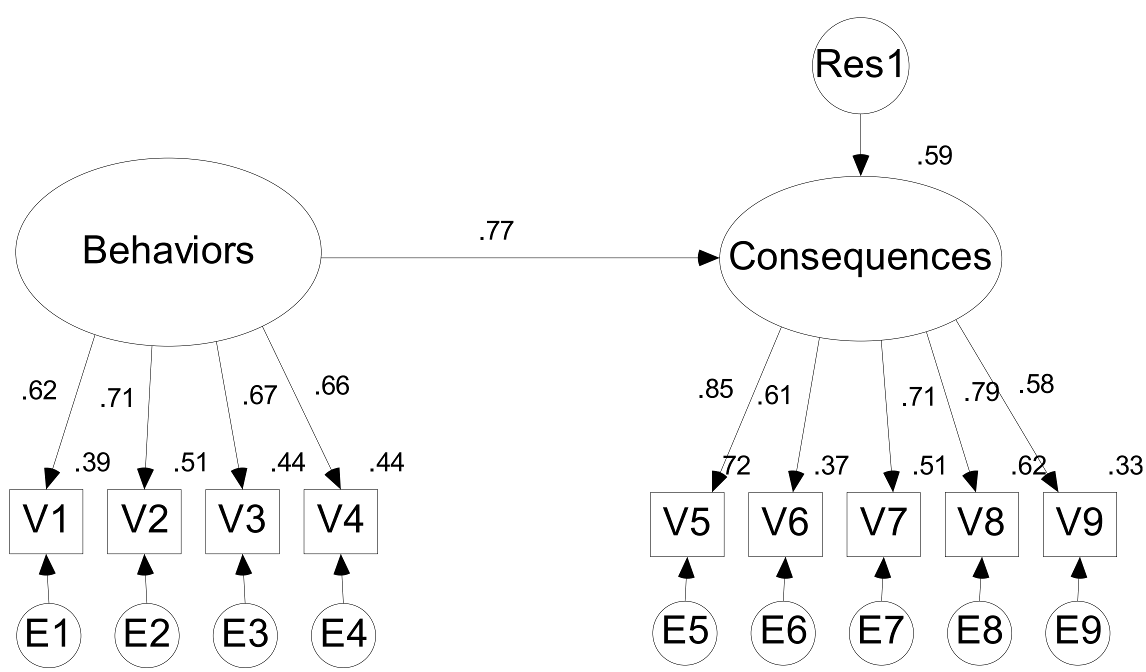

Figure 19. Wiebe Study – Two Factor Model for Men

Standardized results for the configural two-factor PGSI model for all women in the Wiebe sample are depicted in Figure 20:

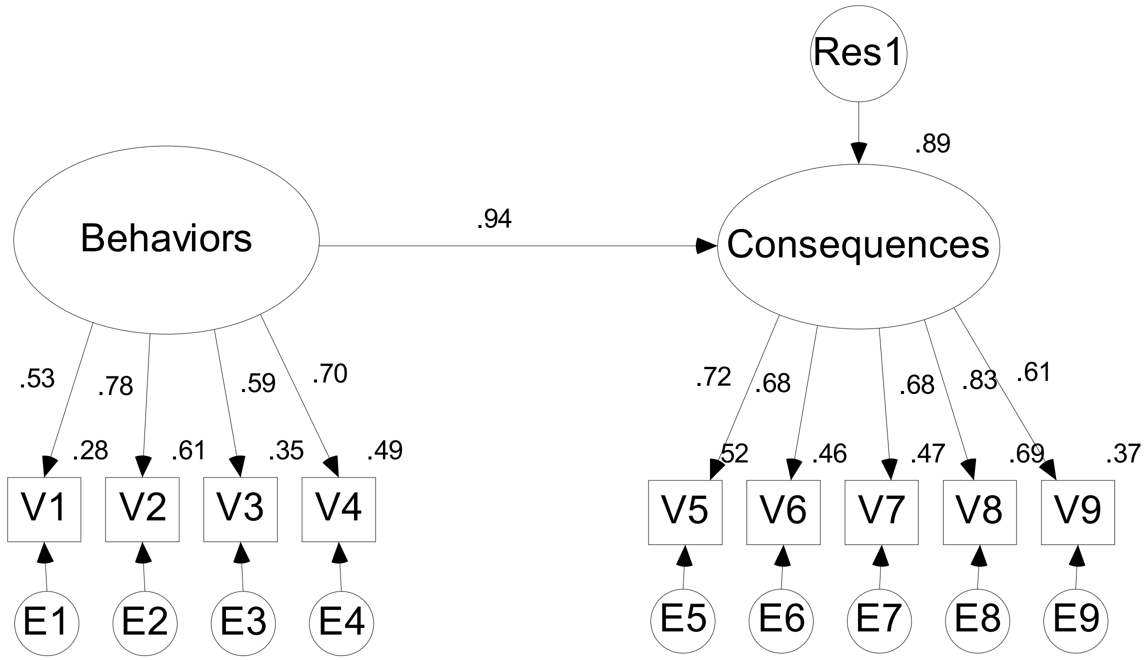

Figure 20. Wiebe Study – Two Factor Model for Women

The next model constrained all factor loadings across men and women for the two-factor model. Model results: $\chi^2 = 790.87$, $df = 59$, $p < .001$, GFI = .926, NNFI = .891, CFI = .911, RMSEA = .074. Delta chi-square for 7 degrees of freedom was 55.98 chi-square points, demonstrated a lack of invariant factor loadings across sexes. Similar to findings for the other two datasets, a lack of sex invariance was noted for both the one- and two-factor PGSI models for the Wiebe study.

Simultaneous Modeling of Three Studies

Following the same sequence of models completed independently for each data set, we now provide results showing the one- and two-factor PGSI models where all three datasets are modeled simultaneously. The overall chi-square goodness of fit test is additive, therefore, the value for the combined model reflects the total of the results from modeling each data source separately. The benefit of simultaneous model is most evident in that it allows direct tests of ME/I across the three studies. We examined the PGSI items across all three studies to determine if results are comparable.

One-Factor versus Two-Factor Model

Results show that the one-factor model resulted in statistically significant factor loadings for each group however mixed model fit was shown by consensus of fit indexes: M1: $\chi^2 = 1452.00$, $df = 81$, $p < .001$, GFI = .941, NNFI = .882, CFI = .912, RMSEA = .056. Fit indexes revealed good fit for GFI, CFI and RMSEA, whereas the NNFI was lower than desired. Factor loadings although statistically significant, and SMCs, were highly variable across the three studies (see table 1).

Table 1. Standardized configural estimates for one-factor PGSI model

	Williams & Wood		**Ladouceur et al**		**Wiebe**	
PGSI Item	**Factor Loading**	**SMC**	**Factor Loading**	**SMC**	**Factor Loading**	**SMC**
1	.46	.21	.76	.57	.58	.34
2	.49	.24	.67	.45	.63	.39
3	.51	.26	.79	.62	.60	.36
4	.48	.23	.55	.31	.59	.34
5	.55	.30	.80	.64	.77	.60
6	.62	.39	.64	.40	.63	.40
7	.66	.44	.66	.43	.70	.49
8	.56	.32	.67	.45	.78	.61
9	.67	.45	.78	.60	.59	.35

Therefore, the one-factor simultaneous model is considered to demonstrate reasonably good fit to the data. Path diagrams corresponding to these results were already presented (see Figure 3 for Williams & Wood, Figure 9 for Ladouceur, and Figure 15 for Wiebe).

The two-factor model also resulted in statistically significant factor loadings for all three studies (see table 2). Model fit was improved over the one-factor model: $\chi^2 =$

1107.13, *df* = 78, $p < .001$, GFI = .955, NNFI = .908, CFI = .934, RMSEA = .049. Comparison of the one factor and two factor models reveals $\Delta\chi^2_{M2\text{-}M1} = 344.87$, *df* = 3, $p < .001$. This result clearly supports the **two factor model as best fitting** the data across all three studies. Fit indexes improved noticeably between the one- and two-factor models. The standardized regression coefficients between Gambling Behaviors and Consequences were remarkably consistent between the studies: $\beta_{Williams} = .93$, $p < .001$, $\beta_{Ladouceur} = .89$, $p < .001$, $\beta_{Wiebe} = .82$, $p < .001$; thereby accounting for approximately 87%, 80%, and 68% of variance in consequences attributed to behaviors respectively. Factor loadings and SMCs remained stable or became stronger in the two-factor model versus results for the one-factor model.

Table 2. Standardized configural estimates for two-factor PGSI model

	Williams & Wood		**Ladouceur et al**		**Wiebe**	
PGSI Item	**Factor Loading**	**SMC**	**Factor Loading**	**SMC**	**Factor Loading**	**SMC**
Behavior Item						
1	.48	.23	.80	.64	.59	.35
2	.52	.27	.70	.48	.73	.53
3	.52	.27	.81	.66	.65	.42
4	.49	.24	.58	.34	.67	.45
Consequence Item						
5	.55	.31	.83	.70	.80	.64
6	.62	.39	.65	.42	.63	.40
7	.67	.45	.68	.46	.70	.49
8	.57	.32	.67	.45	.80	.64
9	.68	.46	.79	.62	.61	.37

Path diagrams for the two-factor configural solution for each sample were previously reported (see Figures 4, 10 & 16 for Williams & Woods, Ladouceur et al., & Wiebe respectively).

Invariance of One-Factor PGSI Model Across Three Studies

To test the hypothesis of invariance of the one-factor PGSI models across all three datasets, we constrained the factor loadings to be equal for all groups, thereby testing for weak metric invariance. These results were compared against the configural invariance results reported above. The model results: $\chi^2 = 2140.06$, $df = 97$, $p < .001$, GFI = .912, NNFI = .854, CFI = .868, RMSEA = .056. Delta chi-square for 16 degrees of freedom was 688.06 chi-square points, demonstrated a clear lack of invariance of factor loadings across the three studies for the one-factor PGSI model.

Invariance of the Two-Factor PGSI Model Across Three Studies

Invariance of the two-factor PGSI model was examined next. Factor loadings were constrained to be equivalent across all three datasets and these results were compared against the configural model reported earlier. Results for this weak metric model: $\chi^2 = 1770.95$, $df = 92$, $p < .001$, GFI = .926, NNFI = .873, CFI = .892, RMSEA = .058. Delta chi-square for 14 degrees of freedom was 663.82 chi-square points, demonstrating a lack of invariant factor loadings for the two-factor PGSI model across studies. Whereas the two-factor solution consistently fit better than did a one-factor PGSI model, it is clear that a lack of invariance resulted when employing these models to compare men and women or to examine results across multiple studies, therefore suggesting differential measurement properties exist between sexes and groups.

What if only Behaviors are invariant?

The question remains whether behaviors but not consequences might be invariant across the three datasets. This model was examined to determine if PGSI gambling behaviors might be consistent. Results of the configural model: $\chi^2 = 1107.13$, $df = 78$, $p < .001$, GFI = .955, NNFI = .908, CFI = .934, RMSEA = .049. Constraining the factor loadings for only the behavior items: $\chi^2 = 1343.43$, $df = 84$, $p < .001$, GFI = .947, NNFI = .896, CFI = .919, RMSEA = .053. Delta chi-square for 6 degrees of freedom was 236.30 chi-square points, showing reasonable fit but a lack of invariance for the behavior items of the PGSI.

What if only Consequences are invariant?

Alternatively, is it plausible that consequences might be invariant across the three datasets in the absence of invariant gambling behaviors? This model was examined to determine if PGSI consequence items might be consistent. Results of the configural were identical to those reported above for the behavior-only model. Constraining the factor loadings for only the consequence items: $\chi^2 = 1557.36$, $df = 86$, $p < .001$, GFI = .938, NNFI = .881, CFI = .905, RMSEA = .056. Delta chi-square for 8 degrees of freedom was 450.23 chi-square points, clearly demonstrating a lack of invariance for the consequence items.

Faced with clear evidence that a lack of ME/I exists in PGSI data across sexes and studies, some additional evaluations were conducted. Breaking from the traditional path of testing for invariance of all factor loadings at once, or testing all loadings on a given factor, sequential tests were conducted that examined invariance for each PGSI item individually (i.e., is PGSI item 1 invariant across the three samples; is PGSI item 2 invariant across the three studies, etc). Results revealed that only PGSI item 7 (How often have you felt guilty about the way you gamble or what happens when you gamble?) demonstrated measurement invariance.

Discussion and Implications

Results from this study both confirm and extend the findings reported by Maitland and Adams (2005). To our knowledge, that report remains one of the few applications of

CFA to PGSI data and the results showed questionable model fit and a lack of invariance between sexes. The reason that the previous and current work of Maitland and Adams is critically important is the need to provide firm evidence of the underlying structure of the PGSI employing appropriate statistical methods. Previous reports using exploratory factor methods suggested that a one-factor model was supported for the nine items comprising the index. However, most of these studies employed EFA without adequate description of what extraction methods were employed (i.e., principal components analysis or exploratory factor analysis) or the criteria that was used to determine the number of factors to retain. Additionally, previous studies have set out to demonstrate that a one-factor model fits PGSI data in the absence of another model. Without conducting formal tests of nested (i.e., competing) models, the best one can conclude is that a one-factor model results but this does not provide evidence that the model fits the data well. Furthermore, exploratory factor methods provide sparse information to aid with assessment of model fit, therefore, a high degree of subjective interpretation is introduced by the researcher. Not surprisingly, such reports started with the goal of demonstrating that one factor underlies the PGSI items and, indeed, this is what they report (e.g., Ferris & Wynne, 2001; Wynne, 2003).

The current study confirmed the finding of Maitland and Adams (2005), that two factors consistently fit PGSI data better than a one factor model. We tested nested one- and two-factor models across three large independent and evidence suggested that a two-factor model better fit the data across all three studies. Therefore, a two-factor model comprised of gambling behaviors and gambling consequences factors has now been demonstrated as the best fitting model from 4 key gambling studies conducted within Canada. Early theoretical description of the PGSI items supported the concept that PGSI includes two constructs (i.e., behaviors, consequences; Ferris & Wynne, 2001; Wynne, 2003), therefore it is somewhat ironic that a unitary dimension has become the primary structure used to define the construct.

The three datasets demonstrated mixed results, with everything from poor to good model fit for the one- and two-factor PGSI models. However, it should be noted that in all cases, fit of the two-factor model was always stronger (as demonstrated by improved fit indexes) than was found for the one-factor solution. Even so, the pattern of factor loadings across studies and across sexes was inconsistent (i.e., which loadings are strongest, are they stronger for men or women, etc). Furthermore, for the two-factor model, the strength of the regression between behaviors and consequences differed between sexes and across studies. The Williams and Wood and Ladouceur data showed a stronger relationship for men, whereas the Wiebe study demonstrated a substantially larger relationship for women, replicating the relationship found for the two-factor model in Maitland and Adams (2005). This suggests that regardless of whether one utilizes a one- or two-factor model for PGSI, it remains questionable whether the 9 PGSI items work together to form a coherent screening instrument. This has serious implications for the use of the recommended cut points to create categories of gambling based upon PGSI items.

Therefore, the two-factor model was better fitting than the unidimensional model across all data sets and for men and women. Combined with the consistent evidence that PGSI failed to demonstrate measurement invariance, the use of PGSI as a single additive dimension is questioned. Results from four independent datasets (the current study plus the work in Maitland & Adams, 2005) provide compelling evidence that the PGSI may not be functioning as designed. Furthermore, the demonstration of differential measurement properties across studies and sexes provide additional evidence that the PGSI produces different answers when used to measure problem gambling by different researchers, across different samples, and across sexes. If the PGSI measures something different in each group then the validity of the cut points used to categorize gamblers with PGSI must be questioned. The current study only evaluated ME/I across sexes and data from three independent studies. Missing from these results is work addressing whether or not the PGSI measures the same construct across other comparison groups that might be examined in gambling research (e.g., age groups, different education levels, levels of problem gambling behavior, etc) as well as concern about the stability of PGSI when used to measure problem gambling longitudinally. In all examples mentioned above, a lack of invariance would certainly produce problems for any comparisons made.

Limitations and concerns

Generalizability of Ladouceur data

A number of concerns must be identified. Whereas the Williams and Wood and Wiebe studies has available PGSI data for the majority of their participants, the Ladouceur et al study only had valid PGSI items for 8% of participants. A goal of this study was to examine how PGSI compared to SOGS, however, even if participants were randomly assigned and received either PGSI or SOGS, this would mean that a maximum of approximately 16% of the 8800+ participants completed this data. For purposes of conducted the analyses proposed in the current study, the 669 valid cases provided adequate power. However, serious reservations are expressed about the representativeness of this small subset of data, and even stronger caution is warranted when trying to generalize results from these data to the population from which it was drawn or results from other studies. Additionally, the method of identifying gamblers in this study and the selective subset of participants who received PGSI items raises question about the prevalence estimates produced from these 669 participants.

Non-normality of PGSI data

The data for all nine PGSI items is non-normal, thereby violating assumptions of normality of input variables for analysis. As reported by Maitland and Adams (2005) transformations were not successful in creating normally distributed outcomes, and analyses were conducted using different matrix formats to assess the impact of this non-normality. In that study, analyses were conducted on polychoric correlation matrices and covariance matrices with similar patterns of results. Covariance matrices are required to conduct tests of ME/I, therefore, analyses for the previous and current studies were derived using covariance matrices. The issue of non-normality of PGSI is not discussed in most reports that have employed CPGI/PGSI.

Low prevalence and limited range of responses

Maitland and Adams reported a potential link to prevalence rates of problem gambling and sample size in their previous work. It is commonly known that problem gambling as currently defined has a very low prevalence rate (i.e., ~ 1% categorized as severe problem gamblers), regardless of data source. Large samples must be gathered to include participants with scores that fall into the different categories of problem gambling as defined by the nine PGSI items (e.g., at-risk, moderate or severe problem gambler).

Additionally, low endorsement and limited variability of the PGSI items (i.e., most respondents will only utilize the lower 2 response categories) makes non-normal data inevitable. Maitland and Adams (2005) demonstrated that statistically significant results found in a large sample of 4631 became non-significant when results were examined in smaller, random 10 and 20% samples drawn from their data and model fit degraded. This work suggested that it was necessary to examine PGSI in large data sets to obtain significant factor loadings and adequate model fit. The current study contradicts this statement to some degree. Significant factor loadings were found for all three datasets however, model fit remained questionable in some of these models and would certainly be worse in smaller subsets of the data[3]. Based on results of the current study samples of less than 500 participants resulted in PGSI models being rejected due to poor fit even if factor loadings were statistically significant.

Finally, these results raise concern about the replicability and generalizability of PGSI. Researchers accept the face validity of PGSI and assume the additive strategy is sound. However, even if the two-factor solution is accepted as best fitting, consistency of results is not assured. The relation between gambling behaviors and consequences (as measured by PGSI items) demonstrates a strong degree of redundancy (or at minimum, a highly dependent relationship). It is entirely possible that a smaller subset of PGSI items could be employed to obtain a similar result, thereby saving time, money, and potential frustration of the respondent.

[3] Analyses were conducted on smaller subsets for each study. The Williams & Wood and Wiebe data sets were approximately the same size and random 20% samples were drawn resulting in subsamples of 492 and 455 respectively. The Ladouceur sample was much smaller but a random 70% sample resulted in a subsample of approximately the same size (n=459). Analyses for all models in this report were examined on these smaller subsamples. Whereas factor loadings remained statistically significant in these models, consistently worsened fit was noted and would result in rejection of models as ill fitting and unacceptable. Therefore, the ability to adequately fit PGSI models in samples of less than 500 is questioned (regardless of whether a one- or two-factor solution is modeled).

References

Bentler, P. M. (1990). Fit indices, LaGrange multipliers, constraint changes, and incomplete data in structural models. *Multivariate Behavioral Research, 25*, 163-172.

Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-600.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Cunningham, W. R. (1982). Factorial invariance: A methodological issue in the study of psychological development. *Experimental Aging Research, 8*, 61-65.

Cunningham, W. R. (1991). Issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 106-113). Washington, DC: American Psychological Association.

Derevensky, J.L., & Gupta, R. (2000). Prevalence estimates of adolescent gambling: A comparison of the SOGS-RA, DSM-IV-J, and the GA 20 Questions. *Journal of Gambling Studies, 16*, 227-251.

Derevensky, J.L., Gupta, R., & Winters, K. (2003). Prevalence rates of youth gambling problems: are the current rates inflated? *Journal of Gambling Studies*, *19*, 405 -425.

Ferris, J., Wynne, H., & Single, E. (1999). *Measuring problem gambling in Canada: Final report –Phase I Inter-Provincial Task Force on Problem Gambling*. Ottawa, ON: Canadian Centre on Substance Abuse.

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

Fisher, S. (2000). Developing the DSM-IV-DSM-IV criteria to identify adolescent problem gambling in non-clinical populations. *Journal of Gambling Studies, 16*, 253-273.

Gambino, B. (1997). Method, method: who's got the method? What can we KNOW about the number of compulsive gamblers. *Journal of Gambling Studies, 13*, 291-206.

Gambino, B. (1997). The correction for bias prevalence estimation with screening tests. *Journal of Gambling Studies, 13*, 343-351.

Hertzog, C. & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging, 18*, 639-657.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist, 1*, 179-188.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18,* 117-144.

Jacques, C., & Ladouceur, R. (2003). DSM-IV-J criteria: a scoring error that may be modifying the estimates of pathological gambling among youths. *Journal of Gambling Studies, 19*, 427-431.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications*. Chicago: SPSS.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8* (Version 8.51) [Computer software]. Chicago: Scientific Software International.

Labouvie, E. W. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 493-502). Washington, DC: American Psychological Association.

Ladouceur, R., Bouchard, C., Rheaume, N., Jacques, C., Ferland, F., Leblond, J., & Walker, M. (2000). Is the SOGS an accurate measure of pathological gambling among children, adolescents and adults? *Journal of Gambling Studies*, *16*, 1-24.

Ladouceur, R., Jacques, C., Chevalier, S., Sevigny, S., & Hamel, D. (2005). Prevalence of pathological gambling in Quebec in 2002. *Canadian Journal of Psychiatry, 50,* 451-456.

Maitland, S. B. & Adams, G. R. (2005). *Assessing the factor structure of the Canadian Problem Gambling Index: Does qualitative stability allow quantitative comparison?* A report prepared for the Ontario Problem Gambling Research Centre.

Maitland, S. B., Dixon, R. A., Hultsch, D. F., & Hertzog, C. (2001). Well-being as a moving target: Measurement equivalence of the Bradburn Affect Balance scale. *Journal of Gerontology, 56B*, P69-P77.

Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., & Nilsson, L-G. (2004). Selective sex differences in declarative memory. *Memory and Cognition, 32,* 1160-1169.

Maitland S. B., Intrieri, R. C., Schaie, K. W., & Willis, S. L. (2000). Gender differences and changes in cognitive abilities across the adult life span. *Aging, Neuropsychology, and Cognition, 7,* 32-53.

McReady, J. & Adlaf, E. (2006). *Performance and enhancement of the Canadian Problem Gambling Index (CPGI): Report and recommendations*. Report for Inter-provincial Funding Partners for Research Into Problem Gambling.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525-543.

Ministry of Public Safety and Solicitor General (2003). *British Columbia problem gambling prevalence study*. Victoria, BC: Author.

Nyberg, L., Maitland, S. B., Rönnlund, M., Bäckman, L., Dixon, R. A., Wahlin, Å., & Nilsson, L-G. (2003). Selective adult age-differences in memory performance in an age-invariant multi-factor model of declarative memory. *Psychology and Aging, 18,* 149-160.

Poulin, C. (2000). Problem gambling among adolescent students in the Atlantic provinces of Canada. *Journal of Gambling Studies*, *16*, 53-78.

Poulin, C. (2002). An assessment of the validity and reliability of the SOGS-RA. *Journal of Gambling Studies*, *18*, 67-93.

Schaie, K. W., Maitland, S. B., Willis, S. L., & Intrieri, R. C. (1998). Longitudinal invariance of adult psychometric ability factor structures across 7 years. *Psychology and Aging, 13,* 8-20.

Slutske, W. S., Jackson, K. M., & Sher, Kenneth, J. (2003). The natural history of problem gambling from age 18 to 29. *Journal of Abnormal Psychology, 112,* 263 -274.

Wiebe, J. (2003). *Gambling and Problem Gambling in Ontario: A Replication Study Three Years Later*

Wiebe, J., Barnes, G. M., Wieczorek, W. F., Tidwell, M. & Parker, J. (2002). Gambling participation in the U.S. – Results from a national study. *Journal of Gambling Studies*, *18*, 313-337.

Wiebe, J., Cox, B., & Falkowski-Ham, A. (2003). *Psychological and social factors associated with problem gambling in Ontario: A one year follow-up study.* Toronto, ON: Responsible Gambling Council.

Wiebe, J., Mun, P., & Kauffman, N. (2006). *Gambling and Problem Gambling in Ontario 2005*. Report prepared for the Ontario Problem Gambling Research Centre.

Wiebe, J. & Single, E. (2002). *Gambling risk practices and consequences*. Unpublished manuscript.

Wiebe, J. W., Single, E., & Falkowski-Ham, A. (2001). *Measuring gambling and problem gambling in Ontario.* Toronto, ON: Responsible Gambling Council.

Williams, R. & Wood, R. (2004). *The Demographic Sources of Ontario Gaming Revenue Study*. Report prepared for the Ontario Problem Gambling Research Centre.

Wynne, H. J. (2003). *Introducing the Canadian Problem Gambling Index*. Unpublished manuscript.

Appendix A. Reliability information for all PGSI data/models

Items	**Williams & Wood Coefficient Alpha**	**Ladouceur et al Coefficient Alpha**	**Wiebe Coefficient Alpha**
9 PGSI Items - Overall	*.77*	*.89*	*.85*
Men	.69	.90	.85
Women	.83	.88	.85
Behavior Items – Overall	*.57*	*.62*	*.67*
Men	.47	.79	.70
Women	.66	.81	.61
Consequence Items – Overall	*.75*	*.83*	*.82*
Men	.66	.84	.70
Women	.82	.82	.82

Appendix B. Comparison of PGSI Models Williams & Wood Study

One vs. Two Factors	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	$\Delta\chi^2(df)$
One Factor Model	27	438.05	.001	.962	.078	.888	.816	
Two Factor Model	26	432.20	.001	.963	.078	.888	.819	5.85(1)
Sex Invariance Models								
Configural One Factor Model	54	935.39	.001	.923	.081	.798	.849	
Weak Invariance One Factor Model	62	1204.25	.001	.901	.086	.772	.804	268.86(8)
Configural Two Factor Model	52	897.66	.001	.926	.081	.779	.855	
Weak Invariance Two Factor Model	59	1162.02	.001	.904	.087	.769	.811	264.36(7)

Note. GFI = LISREL Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; NNFI = NonNormed Fit Index; CFI = Comparative Fit Index

Appendix C. Comparison of PGSI Models Ladouceur et al. Study

One vs. Two Factors	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	$\Delta\chi^2(df)$
One Factor Model	27	231.64	.001	.925	.107	.906	.929	
Two Factor Model	26	165.70	.001	.945	.090	.933	.952	65.94 (1)
Sex Invariance Models								
Configural One Factor Model	54	505.79	.001	.860	.112	.811	.859	
Weak Invariance One Factor Model	62	561.01	.001	.851	.110	.819	.844	55.23(8)
Configural Two Factor Model	52	452.44	.001	.885	.107	.826	.875	
Weak Invariance Two Factor Model	59	514.27	.001	.870	.108	.825	.857	61.83(7)

Note. GFI = LISREL Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; NNFI = NonNormed Fit Index; CFI = Comparative Fit Index

Appendix D. Comparison of PGSI Models Wiebe Study

One vs. Two Factors	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	$\Delta\chi^2(df)$
One Factor Model	27	782.24	.001	.925	.112	.870	.902	
Two Factor Model	26	518.19	.001	.940	.092	.912	.936	264.05 (1)
Sex Invariance Models								
Configural One Factor Model	54	949.85	.001	.909	.086	.855	.891	
Weak Invariance One Factor Model	62	1008.53	.001	.905	.082	.866	.885	58.68(8)
Configural Two Factor Model	52	734.89	.001	.930	.076	.885	.917	
Weak Invariance Two Factor Model	59	790.87	.001	.926	.074	.891	.911	55.98(7)

Note. GFI = LISREL Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; NNFI = NonNormed Fit Index; CFI = Comparative Fit Index

Appendix E. Comparison of Multigroup PGSI Models

One vs. Two Factors	*df*	χ^2	*p*	GFI	RMSEA	NNFI	CFI	$\Delta\chi^2(df)$
One Factor Model	81	1452.00	.001	.941	.056	.882	.912	
Two Factor Model	78	1107.13	.001	.955	.049	.908	.934	344.87 (3)
Sample Invariance Models								
Configural One Factor Model			Same as one factor results above					
Weak Invariance One Factor Model	97	2140.06	.001	.912	.056	.854	.868	688.06(16)
Configural Two Factor Model			Same as two factor results above					
Weak Invariance Two Factor Model	92	1770.95	.001	.926	.058	.873	.892	663.82(14)
Additional Invariance Models								
Configural Two Factor Model			Same as two factor results above					
Weak Invariance Behavior Only Model	84	1343.43	.001	.947	.053	.896	.919	236.30(6)
Weak Invariance Consequence Only Model	86	1557.36	.001	.938	.056	.881	.905	450.23(8)

Note GFI = LISREL Goodness of Fit Index; RMSEA = Root Mean Square Error of Approximation; NNFI = NonNormed Fit Index; CFI = Comparative Fit Index