5-16-2016

A Two Study Structural Modeling Based Approach for Ensuring Retention of Empirical Structures and Optimizing Short Form Development.

Meredith Kathleen Ginley

A TWO STUDY STRUCTURAL MODELING BASED APPROACH FOR ENSURING RETENTION OF EMPIRICAL STRUCTURES AND OPTIMIZING SHORT FORM DEVELOPMENT

by

Meredith Kathleen Ginley

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Psychology

The University of Memphis

August 2016

Abstract

Ginley, Meredith Kathleen. Ph. D. The University of Memphis. August, 2016. A Two Study Structural Modeling Based Approach for Ensuring Retention of Empirical Structures and Optimizing Short Form Development. Major Professors: James P. Whelan, Ph. D. and Andrew W. Meyers, Ph. D.

Retaining an empirically supported model while reducing assessment parameters becomes challenging in short form measurement development. In 2012 Larwin and Harvey proposed a systematic item reduction approach using structural equation modeling (SIR-SEM). The SIR-SEM permits retention of a strong connection to an empirically supported model while reducing some of the challenges of working with a large measurement battery. The application of the SIR-SEM strategy to reduce the number of items needed to assess an empirically supported multidimensional model of impulsivity (Ginley, Whelan, Meyers, Relyea, & Pearlson, 2014) is presented using a two-study procedure. To complete the item reduction, a SAS/Stat version of the SIR-SEM was developed and model fit statistics with extensive empirical support were adopted. In Study 1, the SIR-SEM approach successfully eliminated 84% of the items while retaining 33 items to assess three impulsivity dimensions: behavioral activation, preference for stimulation, and inhibition control. Study 2 tested the resulting 33-item impulsivity measure, the Memphis Impulsivity Measure (MIM), in an independent sample of participants. This second study confirmed model fit. Each of the three MIM dimensions had similar moderate levels of internal consistency. The Pearson correlations for each dimension score indicated good two-week test-retest reliability. The MIM was also found to be largely demographically invariant and to have a significant relation with target risk behaviors including: gambling frequency, symptomology, and classification, alcohol use problems, and alcohol use classification, and drug use involvement, and complexity of involvement.

Table of Contents

Section	Page
1. Introduction	1
2. Method: Study 1	9
3. Results: Study 1	14
4. Method: Study 2	19
5. Results: Study 2	25
6. Discussion	34
References	40

Introduction

In order to better understand constructs, improve model parsimony, and create shortened forms of measures, there is often a need to eliminate items or data points. Shortened measures take less time to administer and eliminating items can allow for identification of the most parsimonious measurement model (Epstein, 1984). However, item removal can come at a cost to construct and measurement validity (Smith, McCarthy, & Anderson, 2000). Innovative statistical techniques, such as Larwin and Harvey's (2012) systematic item reduction using structural equation modeling (SIR-SEM) can be used to reduce the number of variables without losing the connection between individual variables and the existing theoretical or empirical model. In this paper two studies were used to demonstrate SIR-SEM as an alternative approach for retaining an empirically supported model while reducing assessment parameters.

This application specifically used SIR-SEM to derive and test an efficient and effective short-scale measure for the assessment of a multidimensional model of impulsivity. Study 1 presented the SIR-SEM methodology for an empirically confirmed structural model from a battery of impulsivity measures (Ginley, Whelan, Meyers, Relyea, & Pearlson, 2014; Meda et al., 2009). Additionally, an essential step for short-from development was to determine the validity and reliability of the new alternative measure in an independent sample (i.e., Smith et al., 2000). Study 2 showed the performance of the measure created by the SIR-SEM in a new sample of participants to determine if the resulting short-form was actually a valid alternative to the initial comprehensive battery (Smith et al., 2000).

When data is taken out of context in order to enable the fit of mathematical models, the separation between the data, the sample, and the literature at large inhibits accurate interpretation of constructs being explored. As Platt (1964) argued, a field of study advances more effectively when researchers adopt scientific methods that allow for inductive inferences. Adopting the SIR-SEM approach allows scientists to actively ask questions that build upon existing theories and empirical evidence rather than using methods that lessen ability to move systematically. Additionally, retaining a connection with literature-driven models decreases the risk of misidentifying factor constructs.

Parsimony also increases the ease of interpretation and raises confidence that the measured phenomenon contains the fewest assumptions possible without interference from factors that do not meaningfully contribute to a theory (Epstein, 1984; Larwin & Harvey, 2012). Factor analysis can be used to create parsimony through the identification of items that best represent the measured constructs, pulling patterns from within a large multivariate expanse of responses. Exploratory factor analysis (EFA) based strategies eliminate items that do not meaningfully contribute to specific variable patterns (e.g., Clark & Goldsmith, 2006; Whiteside & Lyman, 2001). Items are deleted in an EFA based on quantitative and qualitative guidelines, such as poor item loading after factor rotation or belonging to a factor that is unlikely to replicate (e.g., Gorsuch, 1983). However, the true representativeness of any given sample in relation to the population from which it is drawn presents a major limitation to the accuracy of the correlation matrix. Accounting for the measurement model that accompanies the validated measure with which the sample was collected provides a useful comparison point to determine how representative the current sample may be (Brown, 2006). However, this appraisal is not possible within an exploratory factor analysis framework.

Alternatively, confirmatory factor analysis (CFA) allows for comparison between the present data set and a previously validated structural measurement model. Comparisons between validated models and experimental models (i.e., new models extracted from a current data set) can reveal more parsimonious findings (Browne & Cudeck, 1993). Experimental models are often created using respecifications, or the modification and deletion of parameters. These respecifications are post hoc and follow initial comparison between the new sample's data structure and a theoretical model, or can be hypothesized prior to analysis. When parameter modification is completed, there is a reliance on either a hypothesis-driven or a chance-based approach of changing paths based on systematic procedures (Tabachnick & Fidell, 2013). The procedure often takes the form of testing nested models to determine if the elimination of a certain path improves the fit, then deciding if a more complex model is a meaningful improvement (Santor et al., 2011). Optimally, these "guesses" are related to interpretations based on the researchers' understanding of the theory or literature that informed the initial data model. Furthermore, the testing of nested models is predicated on the assumption of normality of the items. Complicating the tests even more is that many behavioral scientists use Likert scales that are rarely normally distributed.

Modification statistics can provide some recommendations for empirically driven respecification based on the observed relationships of residuals (Tabachnick & Fidell, 2013). These statistics give information about the covariance between items and underlying model structure patterns. However, parameter adjustments conducted based on modification indices should be done with caution as they can lead to over-fitting a model, factor loadings that are difficult to explain, or adding parameters unlikely to replicate (Hatcher, 1994).

Larwin and Harvey (2012) proposed a systematic item reduction procedure as an alternative to the traditional post-hoc respecification procedures of CFA. Their procedure uses structural equation modeling to retain a measure's underlying, empirically confirmed factor structure while systematically removing items that do not improve model fit. This innovative method allows the user to determine maximum model parsimony and select the fewest items needed, while ensuring consideration of the original measurement constructs. Using a jackknifing procedure, one item is removed at a time to identify and delete specific variables from the dataset that do not significantly contribute to model fit (Larwin & Harvey, 2012; Rensvold & Cheung, 1999). Comparing each new jackknifed model to the original model is done using a comprehensive set of fit indices. This procedure does not require items to be normally distributed, which is advantageous for psychology researchers.

Larwin and Harvey's procedure is appropriate for the removal of items with an empirically determined underlying measurement model. By attending to the statistical influence of each item, model integrity can be accounted for and maintained. Model fit is addressed along with continuous testing of measurement invariance and structural invariance. The reduction procedure continues to run based on predetermined fit recommendations until the optimal number of items has been identified. These steps are completed automatically by the program based on chosen fit indices and stop rules without relying on repeated efforts to respecify models or continuous determinations by the researcher related to changes in model fit.

Despite the advantages to using the item reduction strategy proposed by Larwin and Harvey (2012), this novel approach has not have received wider attention for several reasons. First, this approach conflicts with the historical reliance on factor analysis within psychology for strategies of item reduction. Second, there is an absence of information about a direct interdisciplinary application of this strategy. Third, Larwin and Harvey's use of FORTRAN (FORmula TRANslation) programming, which is considered an "industrial strength programming language" (Michaelson, 2015) and requires compliers many researchers do not have easy access to, has made SIR-SEM beyond the computational ability of many. Finally, the initial SIR-SEM approach relies on several indices of model fit that are infrequently used and correspondingly have few guidelines for interpretation, potentially making interpretation and application less accessible.

The Present Systematic Item Reduction Demonstration

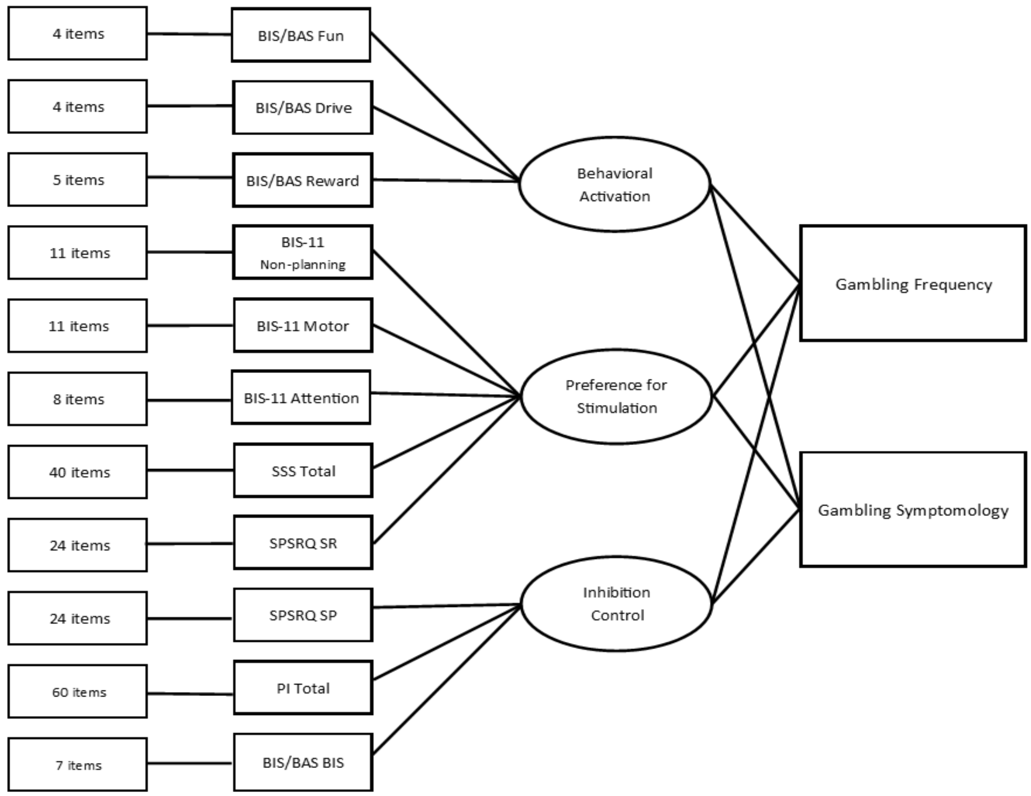
This paper first tested the SIR-SEM in a large dataset of impulsivity measures for extraction of a more parsimonious assessment model that retains an empirically supported factor structure. A second study explored an independent administration of the short form and its resulting factor structure and measurement validity. To understand the true utility of the measure, correspondence between the new short form and target behavior were also included. The new measure was evaluated to determine if it corresponded with participant behavior and was independent of demographic covariates (e.g., Smith et al., 2000). Together the two studies provided a concrete example of the SIR-SEM procedure in an application that would be of significant benefit to the psychology literature.

An Applied Example

There is a current need within the addictions literature to efficiently and validly measure the complex risk factor of impulsivity. Impaired control over engagement in a risk behavior has long been identified as a hallmark of addictive behavior and substance use disorder (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001; Potenza, 2006). Impulsivity has also been identified as a risk factor for the development of gambling problems (Ginley et al., 2014; Langewisch & Frisch, 1998; Petry, 2001). High scores on self-report measures of impulsivity for those with either a gambling disorder or a substance use disorder have been hypothesized to correspond to an underlying mechanism for addiction risk. Identification of markers such as impulsivity have helped inform the reconceptualization of the addictive disorders classification in the DSM 5 to include both substance and behavioral addictions (American Psychiatric Association, 2013; Potenza, 2006).

A multidimensional approach to measuring impulsivity in addictive disorders has received theoretical and empirical support (e.g., Ginley et al., 2014; Meda et al., 2009; Reynolds, Ortengren, Richards, & de Wit, 2006; Whiteside & Lynam, 2001). When measured with a comprehensive, multiple measure battery comprised of over 200 items, a three dimensional model of self-reported impulsivity significantly corresponded to gambling (Ginley et al., 2014) and alcohol and substance use (Dager et al., 2014; Figee et al., 2015; Hyatt et al, 2012; Meda et al., 2009; Yarosh et al., 2014). The measures of impulsivity from this empirical model each have their own empirically confirmed factor structures and were selected for their established ability to connect subsets of behaviors related to impulsivity to aspects of substance abuse. Measures were chosen based on their use in the addiction literature and their relevance to theoretically unique impulsivity related dimensions (see Ginley et al., 2014; Meda et al., 2009). The general model appears in Figure 1. The first dimension, behavioral activation, measures the delay between desired behavior and behavior engagement. Behavioral activation is based on the behavioral approach system of Gray's reinforcement sensitivity theory (Gray, 1970). Low risk and symptomatic gamblers scored significantly higher on behavioral activation than non-gamblers suggesting gamblers take less time to consider risk behavior engagement then non-gamblers, however, those with drug use disorders or at risk for an alcohol use disorder were found to score no different than healthy controls on this dimension (Meda et al., 2009). The second dimension, preference for stimulation captures an individual's tendency to engage in risk behaviors. This dimension consists of the overlap between general sensation seeking, which is conceptualized as the desire to seek novel, varied, and complex sensations and experiences

(Breen & Zuckerman, 1999) and Barrett's model of impulsivity as a combination of anxiety and psychomotor agitation (Patton, Stanford, & Barratt, 1995). Those with higher addictive behavior symptomatology scored higher on preference for stimulation. The third factor, inhibition control, corresponds to how individuals process high-risk behaviors and experience anxiety about potential poor outcomes. This dimension is a combination of the behavioral inhibition system from Gray's reinforcement sensitivity theory (Gray, 1970) and of obsession/compulsivity that has been related to reward approach or avoidance. As with preference for stimulation, those with higher addictive behavior symptomology had significantly more difficulty with inhibition control than those without.



Note. Solid lines represent empirically supported paths between variables. Barratt Impulsiveness Scale: 11th version (BIS-11; Patton et al., 1995), Behavioral Inhibition/ Behavioral Activation Scales (BIS/BAS; Carver & White, 1994, Sensation Seeking Scale: Form V (SSS Form V; Zuckerman, Eysneck & Eysenck, 1978), Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ; Torrubia et al., 2001) Padua Inventory (PI; Sanavio, 1988).

Figure 1. The model of multidimensional impulsivity as related to gambling behavior from Ginley, Whelan, Meyers, Relyea, and Pearlson (2014).

Despite its power at detecting differences between those with behavioral and substance use addictions and those without, the comprehensiveness of the nearly 200-item impulsivity battery utilized by Ginley and colleagues (2014) and Meda and colleagues (2009) presented a burden to participants and a problem for those interested in further testing this model. A more efficient assessment method would facilitate its use in future research. Our challenge, therefore, was to reduce the number of items necessary to measure these three dimensions of impulsivity. In Study 1 of this paper, we showcase the SIR-SEM procedure for reducing the measurement model. We decided to abandon Larwin and Harvey's use of FORTRAN and LISREL. Instead, we adopted their metacode, suggested criteria for factor model estimation, and guidelines for program iteration termination. We then developed a program in SAS/STAT 9.3 Proc Calis (SAS Institute, 2011). (A copy of this program is available upon request.) In Study 2 a cross validation was completed to establish the reliability and validity of these dimensions once they have been reduced and turned into a brief impulsivity measure, the Memphis Impulsivity Measure (MIM), through a replication study. This was done using a separate diverse sample of college students. Estimated validity coefficients were obtained through comparisons between the empirically supported factors of behavioral activation, preference for stimulation, and inhibition control, and outcome variables of addictive behaviors including gambling, alcohol and drug use.

Study 1

Method

Participants

A sample of 1,623 college students consented to complete an assessment battery that consisted of demographics questions, a battery of self-report impulsivity measures, a measure of gambling frequency, and a measure of gambling symptomology. Participants were 67% female

($n = 1091$), 50% college freshmen ($n = 805$), and had a mean age of 20.55 ($SD = 4.44$). The participants identified themselves as follows: 47.9% Caucasian, 38.9% African American, 3.6% Hispanic, 2.8% Asian, 0.4% American Indian, 0.1% Native Hawaiian or other Pacific Islander, and 6.3% other or did not report. Demographic information appears in Table 1.

Table 1
Demographics of the Study 1 and Study 2 Participant Samples

	Study 1 Sample Characteristics (*n* = 1623)	Study 2 Sample Characteristics (*n* = 530)
Age, *M* (*SD*)	20.6 (4.44)	19.5 (1.69)
Female, *n* (%)	1091 (67%)	277 (52.3%)
Race/Ethnicity, (%)		
African American	38.9%	31.5%
Caucasian	47.9%	55.3%
Hispanic	3.6%	4.5%
Other or multiple ethnicities reported	6.8%	5.5%
Asian	2.8%	3.2%
Risk Behavior Involvement, *n* (%)		
Gambling	Past year participation	312 (58.9%)
	Without gambling problems	404 (76.2%)
	Gamblers with some adverse symptoms	114 (21.5%)
	Probable gambling disorder	12 (2.3%)
Alcohol Use	Past year participation	304 (57.4%)
	Non-drinkers	226 (42.6%)
	Social Drinkers	229 (43.2%)
	Alcohol Use Disorder	75 (14.2%)
Drug Use	Past two-week participation	143 (27%)

Note. Not all participants responded to all demographics questions. Percentages are calculated with non-responders coded as missing. Risk behavior involvement only reported for Study 2 participants.

Measures

Demographic Questionnaire. This questionnaire asked about participant age, gender, race, and ethnicity.

Barratt Impulsiveness Scale: 11th version (BIS-11). The BIS-11 (Patton et al., 1995) was developed to assess biological and behavioral correlates of impulsiveness. Respondents ranked 30-items on a 4-point scale anchored to responses of "Rarely/Never," "Occasionally," "Often," and "Almost Always." There are three subscales (Stanford et al., 2009): attentional impulsiveness, motor impulsiveness, and nonplanning impulsiveness. Higher scores on any subscale indicate higher impulsivity (Patton et al., 1995). When tested in an adult sample of participants ages 17-45, the three second order factors had Cronbach's alphas ranging from 0.59 to 0.74 (Stanford et al., 2009).

Behavioral Inhibition/ Behavioral Activation Scales (BIS/BAS). The BIS/BAS (Carver & White 1994) assesses the two components of Gray's reinforcement sensitivity theory (Gray, 1970). Participants rate 24 questions on a 4-point scale ("Very true for me" to "Very false for me"). Four of the items are filler items that are not included in any of the scales. The BIS is used to assess the behavior inhibition system, and high BIS predicts feelings of anxiety and withdrawal behavior when placed in a new situation (Carver & White, 1994). The BAS assesses the behavioral approach system. High BAS predicts greater brain activation when presented with positive events and a strengthened drive to behave in a way that produces approach behavior for both conditioned and unconditioned stimuli (Carver & White, 1994; Smillie, Pickering & Jackson, 2006). In a parametric analysis with college students, Cronbach's alphas were 0.73 for Reward Responsiveness, 0.65 for Drive, .72 for Fun-Seeking, and 0.82 for the BIS subscale (Caseras, Avila, & Torrubia, 2002).

Sensation Seeking Scale: Form V (SSS). This 40-item self-report measure captures a person's affinity for or against a variety of activities that are considered risky behaviors or high sensation activities (Zuckerman, Eysenck, & Eysenck, 1978). The SSS yields the total Sensation Seeking Score (Zuckerman, 1996). In an analysis of reliability and validity with college students, the SSS total score showed moderate reliability with a Cronbach's alpha of 0.75 (Ridgeway & Russell, 1980).

Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ). The SPSRQ (Torrubia, Avila, Molto, & Caseras, 2001) was also designed to assess BIS and BAS (Dawe & Loxton, 2004). The 48 yes-no questions assess two dimensions. The first, Sensitivity to Punishment (SP), assesses the inability to stop potential behavior when made aware of potential punishment, and the second, Sensitivity to Reward (SR), is the tendency to engage in goal-focused behavior in situations associated with reward (Torrubia et al., 2001). With a sample of college students, the Cronbach's alphas for SP and SR were .83 and .76, respectively (Caseras et al., 2003).

Padua Inventory (PI). The PI (Sanavio, 1988) has been used to assess obsessionality and compulsivity in community samples. The measure was devised using statements made by individuals meeting criteria for obsessive compulsive disorders and then reduced through factor and item analysis to its present 60 items (Sanavio, 1988). The measure uses a five-point severity inventory (0 = not at all, 1 = a little, 2 = quite a lot, 3 = a lot, 4 = very much). A total score is obtained by summing all responses. Cronbach's alpha with college students ranged from 0.77 to 0.89 (Sternberger & Burns, 1990).

Procedure

The Institutional Review Board reviewed and approved the protocol. Participants were recruited from an undergraduate subject pool. All participants were provided with informed consent materials that emphasized the voluntary nature of participation, a participant's right to withdraw, and the protection of confidentiality. Those providing consent were then administered the assessment packet via an online survey delivery system. They completed the survey questionnaires during a single data collection session and were awarded course credit as compensation.

Results

Analytic Plan

Participant responses were analyzed with the SIR-SEM procedure that used a SAS-based program written for this project. All analyses were conducted in either SPSS Version 20 or SAS Version 9.3.

Unanswered responses were determined to be missing at random. Missing responses for the impulsivity items were uncommon; every item was completed by at least 90% of respondents. For any missing items in the impulsivity measures, an individual's item score was imputed using the subscore average from the completed items.

Systematic Item Reduction Procedure

A three-factor measurement model was used to capture the specific contribution of each item from the impulsivity battery on the three impulsivity dimensions. The subscales modeled were: behavioral activation which contained four items from BIS/BAS Fun Seeking, four items from BIS/BAS Drive, and five items from BIS/BAS Reward Responsiveness; preference for stimulation which contained the 11 items from BIS-11 Nonplanning Impulsiveness, 11 items from BIS-11 Motor Impulsiveness, eight items from BIS-11 Attentional Impulsiveness, 40 items from the SSS total score, and 24 items from the SPSRQ Sensitivity to Reward scale; and inhibition control which contained 24 items from the SPSRQ Sensitivity to Punishment Scale, 60 items from the PI total score, and seven items the BIS/BAS BIS scale.

Model fit was evaluated using the fit statistic recommendations of Kline (2012)[1]. The chi-square statistic was used as it provided a test of the null hypothesis when the reproduced covariance matrix has a specific model structure. Given the large influence of sample size on the chi square statistic, the ratio of the model chi square to its degrees of freedom was also assessed, with a ratio of five or smaller considered to be acceptable model fit (as recommended by Bollen, 1989, p. 278). The Steiger-Lind root mean square error of approximation (RMSEA; Steiger, 1990) has a 90% confidence interval, and a correction for model complexity that takes sample size into account. An RMSEA less than or equal to .05 indicates close approximate fit, and values between .05 and .08 suggest reasonable error of approximation (Browne & Cudeck, 1993). The Bentler comparative fit index (CFI, Bentler, 1990) is an incremental fit index that compares the hypothesized model to a null model that assumed zero population covariance among the observed variables. CFI values greater than .90 indicate reasonably good fit (Hu & Bentler, 1999). Finally, the model was evaluated using the standardized root mean square residual (SRMR), which is based on transforming both the sample covariance matrix and the predicted covariance matrix into correlation matrixes. This allows for a measure of the overall difference between the observed and predicted correlations. SRMR values less than .10 are considered favorable and values less than .08 are considered good fit (Hu & Bentler, 1999).

---

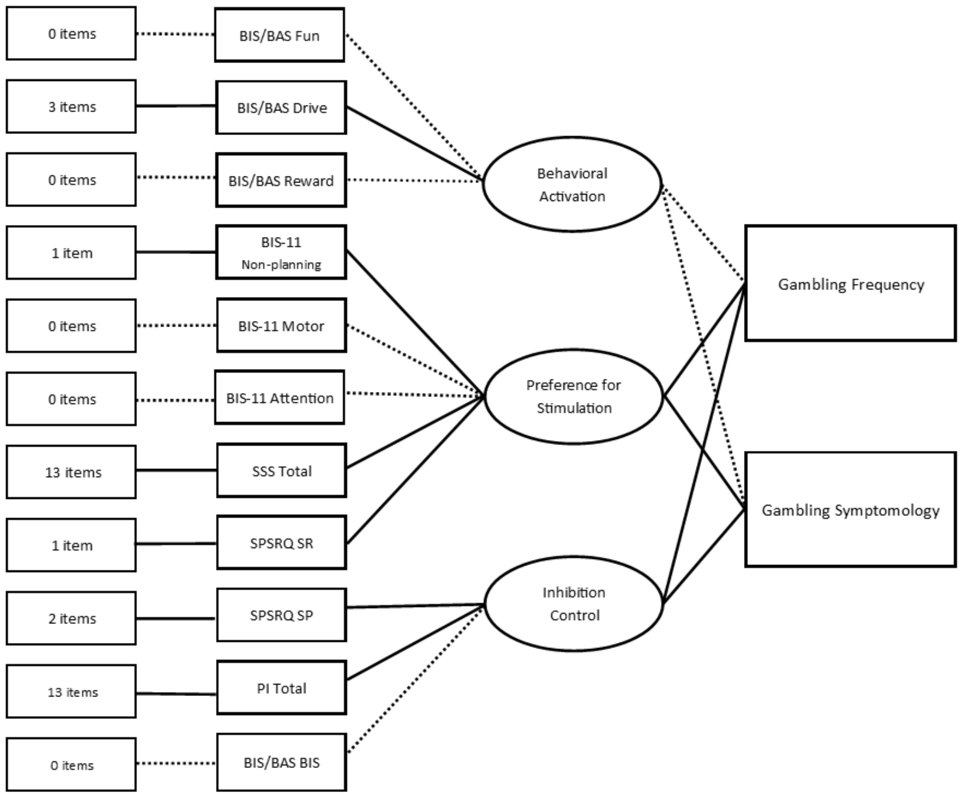
[1]Larwin and Harvey (2012) also incorporated several alternative fit statistics in their evaluation of model fit. These include the Sentorra-Bentler scale correction (SB χ^2), which is used to account for possible kurtosis within items (Schermelleh-Engel, Moosbrugger, & Muller, 2003); the expected value of

the cross validation index (ECVI), which shows how well the model fit compares with the population covariance matrix (Brown & Cudeck, 1993); and the Consistent Akaike Information Criterion (CAIC), which looks at model fit in relation to changes in model complexity (Akaike, 1978). However, given the infrequency with which these fit indices are reported in the psychology literature, they were not evaluated as fit statistics for this demonstration.

While the programming for the current project was conducted in SAS in order to increase accessibility, the exact jackknife procedure detailed by Larwin and Harvey (2012) (e.g., Rensvold & Cheung, 1999) was replicated. First, the model structure and fit statistics was calculated for the full, hypothesized model based on the findings of Ginley and colleagues (2014) and detailed above (also see Figure 1). Then, the model was re-estimated 198 times with one item removed, a different single item removed each time. Next, the program was used to rank models in relation to the original full model based on RMSEA and CFI fit statistics, with ranking preference going to the CFI in the event of discrepancy between the fits. Upon selection of the best fitting new model with only 197 items, a second item was deleted, with a different second item removed each run and re-estimated in comparison to the model with one item already removed.

Removal of items and re-estimation continued for as long as the reduced model met three criteria: the model that had been created by elimination of items retained a Pearson correlation of at least $r \geq .95$ with the original model (Byrne, Shavelson, & Muthen, 1989; Newcomb & Bentler, 1988), each impulsivity factor continued to contain at least three items (Bagozzi, 1981; van der Sluis, Dolan, & Stoel, 2005), and finally the structural integrity of the model was retained, and the reduced model still demonstrated good model fit (Bollen, 1989). The jackknifing process was considered complete based on all three criteria when only 33 items remained in the model. For the final reduced three-factor model the chi-square test was significant, $\chi^2(492, N = 1623) = 1147.39$, $p < 0.001$, and the ratio of the model chi-square to its degrees of freedom indicated a reasonable model fit. The RMSEA indicated good fit (RMSEA = .029, *95% CI*: .027, .031). The CFI indicated a good fit (CFI = .94). The SRMR value was considered a good fit (SRMR = .03).

At this point three items corresponded with the behavioral activation factor, 15 items with preference for stimulation, and 15 items with inhibition control (see Figure 2). Per Larwin and Harvey's (2012) termination rules, program iteration was discontinued. No further items were deleted, as the removal of any additional items would have resulted in fewer than three items corresponding with behavioral activation.



Note. Solid lines represent statistically significant paths between variables. Dotted lines indicate no statistically significant relationship. Barratt Impulsiveness Scale: 11th version (BIS-11; Patton et al., 1995), Behavioral Inhibition/ Behavioral Activation Scales (BIS/BAS; Carver & White, 1994, Sensation Seeking Scale: Form V (SSS Form V; Zuckerman et al., 1978), Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ; Torrubia et al., 2001) Padua Inventory (PI; Sanavio, 1988).

Figure 2. The multidimensional impulsivity model from the systematic item reduction (SIR-SEM) approach as related to the risk behavior of gambling

Study 2

Method

Participants

A sample of 599 participants between the ages of 18 and 25 years old were recruited from a large urban university. To obtain a representative sample, participants were sampled purposefully (Shadish, Cook, & Campbell, 2002) and recruitment occured via a variety of methods: from a Psychology subject pool, from other undergraduate classes, and from established student groups. Participants from the subject pool received course credit for research partication. Others received no compensation.

Sixty-nine participants were removed from the final analyses as they did not complete at least half of the study questionniares. The mean age of the remaining 530 participants was 19.5 (*SD* = 1.69) Participants were 52.3% female (*n* = 277) and they placed themselves in ethnic and racial categories, as follows: 55.3% Caucasian, 31.5% African American, 4.5% Hispanic, 3.2% Asian, and 5.5% Other or multiple races reported. Demographic information appears in Table 1.

Participants reported engagement in a range of risk behaviors. Fifty-nine percent (*n* = 312) of participants had gambled in the past year. The majority of participants gambled a few times per year and engaged in a variety of gambling activities with lottery ticket purchases being the most popular activity (34.3%, *n* = 182). Participants' past year gambling symptomology scores classified 76.2% (*n* = 404) as without gambling problems, 21.5% (*n* = 114) as gamblers with some adverse symptoms, and 2.3% (*n* = 12) as probable pathological gamblers.

Fifty-seven percent (*n* = 304) of participants reported having consumed alcohol in the past. The majority of those who had drunk alcohol in the past (48.3%, *n* = 141) indicated that they drank on average monthly or less. Eight percent (*n* = 42) reported binge drinking as classified by having six or more drinks on one occasion. Participants' AUDIT scores classified 42.6% (*n* = 226) as non-drinkers, 43.2% (*n* = 229) as social drinkers or as having minimal adverse symptoms, and 14.2% (*n* = 75) as meeting criteria for an alcohol use disorder.

Twenty-seven percent (*n* = 143) of participants reported having used an illicit drug or a medication in a way it had not been prescribed within the past two weeks. Of those who had used drugs in the past two weeks, the majority (33.6%, *n* = 48) indicated that they only used a single substance on only one or two days. Marijuana was the most commonly used drug (18.5%, *n* = 98), followed by medications used in ways other than as prescribed including stimulants (7.7%, *n* = 41), painkillers, (7.2%, *n* = 38), and sedatives, (6.2%, *n* = 33).

Measures

Demographics. These questions assessed age, gender, race/ethnicity, parent's level of educational attainment, monthly income, and whether they are a first generation college student.

Memphis Impulsivity Measure (MIM). The MIM included 33-items resulting from the SIR-SEM reduction procedure described in Study 1. As the original impulsivity battery measures used different response formats, response options for all questions were modified to be on a scale from 1 to 4 (1 = very true for me, 2 = somewhat true for me, 3 = somewhat false for me, 4 = very false for me), as this was the most common format used by the original scales. An independent investigator then checked and verified items were free of lexical or grammatical errors. This investigator was familiar with all the impulsivity items but was not working directly on the present project. All items in the final measure, except item 18, are reverse scored. Items 1-3 contribute to the behavioral activation dimension, items 4-18 to preference for stimulation, and items 19-33 to inhibition control. Completion of the MIM takes participants approximately 10 min.

Measures to Establish Validity

Beck Depression Inventory (BDI-II). The BDI-II (Beck, Steer, & Brown, 1996) is one of the most widely used instruments for assessing intensity of depression and for detecting depression in the general population. It contains 21 items and total score range from 0 to 63. Within a college student sample, Cronbach's alpha scores range from .91 (Dozois, Dobson, & Ahnberg, 1998) to .93 (Beck et al., 1996). BDI score was hypothesized to not be strongly correlated with impulsivity.

Risky Families Assessment Questionnaire (RFQ). The Risky Families Assessment Questionnaire (Taylor, Eisenberger, Saxbe, Lehman, & Leiberman, 2006) is a 13-item measure of childhood family environment. The total score of this measure captures adverse childhood experiences. This measure has been linked to neural response to threat detection and emotional stimuli. Adverse childhood experiences were found to have significant correspondence with later life health risk events. Participant's total level of risky family environment was hypothesized to correlate with inhibition control.

Life Events Scale for Students (LESS). This 36-item measure (Clements & Turpin, 1996) asks participants to indicate if they have experienced a number of events over the past year. In a sample of college students, the consistency of events reported ranged from 61% at one month follow up to 54% at six month follow up, which was found to be an adequate decay of memory over time (Clements & Turpin, 1996). A total score of stressful life events is captured by this measure, with increased numbers of stressful life events hypothesized to correlate with preference for stimulation.

The Positive and Negative Affect Schedule (PANAS). This 20-item measure (Watson, Clark, & Tellegan, 1988) consists of words that describe feelings and emotions. The words are divided into two scales; the positive affect scale and the negative affect scale. In a sample of college students, the internal consistency of positive affect scale was shown to range from .86 to .90, and the negative affect scale ranged from .84 to .87 within the same sample. The positive affect scale is hypothesized to correlate with behavioral activation. The negative affect scale is hypothesized to correlate with inhibition control.

DSM 5 Level 1 Cross Cutting Symptoms Measure (Narrow et al., 2013). Is a 22-item self-rated measure that assesses 13 psychiatric domains that are important across several psychiatric diagnoses. Each item asks how much a particular symptom has bothered an individual during the past two weeks. Items are rated on a 5-point scale from 0 (none or not at all) to 4 (severe or nearly every day). A total score on this measure captures level of distress caused by psychological symptomology. As general psychological distress may be significantly correlated with engagement in risk behaviors, we will use this measure to control for psychological distress when evaluating the relations between the MIM dimensions and risk behaviors.

Measures to Establish Correspondence with Risk Behaviors

South Oaks Gambling Screen (SOGS): (Lesieur & Blume, 1987). The past year gambling behavior form of this measure is a 20-item measure of gambling pathology that is frequently used with college student and adult samples. A major benefit of this measure is it is very sensitive to low levels of gambling difficulty. The SOGS is correlated with DSM-IV diagnostic criteria for pathological gambling in both clinical and general population samples and had an overall good sensitivity and specificity (Stinchfield, 2002; Weinstock, Whelan, Meyers, & McCausland, 2007). SOGS based scores of gambling symptomology were captured using both the SOGS total score measure, and the recommended cut points of 0 for no problems, 1-4 for some problems, and 5 or more for probable pathological gambler.

The SOGS measure also has a separate frequency table where participants rate the average frequency of engagement in several specific gambling activities over the past year. The original frequency table was expanded to request that for each of the nine specified gambling activities, participants indicate whether they gambled, “Not at all,” “A few times a year,” “About once a month,” “About once a week,” “A few times per week,” and “Almost daily.” This modification allowed for a more precise estimate of gambling frequency. Gambling frequencies for each gambling activity and the total gambling frequency were calculated. Participants who did not report an activity frequency data point were scored a 0 for that gambling activity.

Alcohol Use Disorders Identification Test (AUDIT). (Allen, Litten, Fertig, & Barbor, 1997). The AUDIT is an efficient measure for early stage screening for alcohol use related problems (Reinert & Allen, 2002). The measure assesses the three domains of intake, dependence, and adverse consequences. The AUDIT has performed well in college student populations with high sensitivity and specificity (Aertgeerts et al., 2000; Clements, 1998). A total score can be calculated to gain an understanding of level of alcohol related consequences. In adult populations a cut point of 8 can be used to indicate alcohol use disorder (Reinert & Allen, 2002).

DSM 5 Level 2-Substance Use-Adult-Assessment Measure (DSM 5 Drug Use) (American Psychiatric Association, 2016). This measure is an adaptation from the NIDA-modified ASSIST (National Institute on Drug Abuse, 2016). It is an “emerging measure” from the American Psychiatric Association for adults 18 years of age and older. This measure was developed for administration during an initial patient intake evaluation. It has been shown to have good test-retest reliability ranging from .73 to .78 (Narrow et al., 2013). It is also recommended for use in research to enhance clinical decision-making. While the measure is not fully diagnostic on its own, it does capture engagement in use and frequency of use over the past two weeks for 10 different drugs and medications. Participants were instructed that medications were only to be endorsed if they were taken "on your own" that is, either without a prescription or not as prescribed. Individual items can be interpreted independently. Ratings on multiple items at scores greater than 0, or the total measure score, indicates greater severity and complexity of substance use but is not a direct proxy for total frequency of use.

Procedure

The Institutional Review Board from the participating university reviewed and approved the protocol. Potential participants were provided with an online link to the study materials. Accessing the link took the participants to an informed consent document that emphasized the voluntary nature of participation, the right to withdraw at any time, and protection of privacy and confidentiality. Those agreeing to participate were then invited to complete the assessment survey. The online administration took approximately one hour to complete. At the end of the survey, participants were given contact information for questions about the study or if they desired information related to treatment.

To assess test-retest reliability, 34 participants from undergraduate psychology classes who had not participated in the initial survey were invited to take the MIM at two different administration time points spaced two weeks apart. The students participated as a class activity on a voluntary basis with no compensation provided.

Results

Analytic Plan

Preliminary analyses include a description of the sample's demographics and risk behavior engagement. The data set was also evaluated for missing data and whether missing data was random. The factor structure of the MIM was confirmed to verify that items were appropriate for inclusion. Internal consistency and temporal stability for each of the confirmed dimensions was calculated.

To assess validity, we completed a series of correlations between the impulsivity dimensions and other variables of interest. The MIM total score and dimension scores were compared to outcome variables of gambling frequency, gambling symptomology, alcohol use involvement, and drug use to determine the predictive ability of these factors for capturing risk behaviors.

The MIM dimensions were also considered in relation to demographics, risk behavior history, and mental health symptoms. First the influence of these variables on MIM response patterns was assessed. A series of regression equations then explored possible control variables. If covariates of interest were identified, those were then controlled for in the analyses. These steps were taken to evaluate if the the MIM was independent of demographic variables, but retained a relation with risk behaviors of gambling, alcohol use, or drug use.

Missing Data

For those who had completed more than half of the study measures, unanswered responses were determined to be missing at random (Brown, 2006; Downey & King, 1998). Missing responses for the impulsivity items were uncommon; every item was completed by at least 99% of respondents. No retained respondent skipped more than 3 items. For any missing items in the impulsivity measures, an item score was imputed using the item neutral value. Missing data on the other dependent variable items were also uncommon (< 2%). Nonresponses on these measures were not added into individual sum scores. The data imputation allowed for 530 subjects to be included in the analyses.

Confirmatory Factor Analysis

A CFA of the MIM was conducted using the SAS Proc Calis (SAS Institute, 2011) procedure and a generalized least squares method of parameter estimation. This procedure required the explicit identification of all relevant item pattern loadings on their factors. Additionally, this method assumed a nonsingular correlation matrix, multivariate normality, and independence of observations. The generalized least square approach was chosen over a maximum likelihood approach as it has been shown to perform slightly better with sample sizes of approximately 500 participants (Hu, Bentler, & Karo, 1992).

Model fit was evaluated using the fit statistic recommendations of Kline (2012), including the chi square statistic and chi square ratio, RMSEA, CFI, and SRMR. See Study 1 for the complete description of these fit indices. Additionally, the Goodness of Fit index (GFI) was also considered as it adjusts estimates of model fit based on the number of parameters in the model. GFI values greater than .90 indicate acceptable model fit.

The model was determined to have an overall good fit. The chi-square test was significant, χ^2 (df = 489) = 893.68, $p < .01$. This finding is acceptable since the ratio of the model chi-square to its degrees of freedom indicated a very good model fit (Bollen, 1989) and that the factor structure largely accounted for the variability of the data. The RMSEA indicated close fit (RMSEA = .040, *95% CI*: .036, .044). The CFI indicated a poor fit (CFI = .48), but GFI was acceptable (GFI = 90.). The SRMR value was considered a favorable fit (SRMR = .08).

Additionally, an examination of the factor loadings suggested all items were loading significantly on the subscales. An examination of the residual table indicated an approximately normal distribution.

Internal Consistency

Each of the three MIM dimensions had similar levels of moderate internal consistency; behavioral activation (α = .78; 95% CI: .75-.82), preference for stimulation (α = .68; 95% CI: .64-.72), and inhibition control (α = .83; 95% CI: .81-.85). These alpha levels were not affected by dropping any single item and were greater than a recommended internal consistency level of .65 or higher (Nunnally & Bernstein, 1994). The degree of consistency between items, as evidenced by the moderate magnitude of the alpha coefficient, suggested that items capture a large and varied expanse of content. It is unlikely that any two items within a dimension contain considerable redundancy with each other.

Reliability

The 34 participants who participated in the test-retest portion of the study were demographically similar to those in the full survey administration: M_{age} = 19.5, *SD* = 1.60, 55.9% Caucasian, 32.4% African American, and 11.7% Other or multiple races reported. Participants in the test-retest sample were more likely to be female (61.8%, *n* = 21). The Pearson correlations between dimension scores on the two administrations were: behavioral activation (*r* = 0.76), preference for stimulation (*r* = 0.88), and inhibition control (*r* = 0.93).

Validity

Discriminant Validity. To ensure the MIM was measuring more than purely current physiological arousal, a Pearson correlation was obtained between several items from the PANAS and the three MIM dimensions (Table 2). There were non-significant correlations, or very small significant correlations, between self-reported current levels of excitement (behavioral activation, $r = .09$, $p < .05$; preference for stimulation, $r = -.01$, $p = .75$; inhibition control, $r = -.04$, $p = .32$), alertness (behavioral activation, $r = .10$, $p < .05$; preference for stimulation, $r = -.01$, $p = .79$; inhibition control, $r = -.08$, $p = .07$), and interest (behavioral activation, $r = .01$, $p = .76$; preference for stimulation, $r = .05$, $p = .23$; inhibition control, $r = -.10$, $p < .05$).

Table 2

Discriminant and convergent validity between impulsivity dimensions and variables of interest

		Behavioral activation	Preference for stimulation	Inhibition control
Discriminant Validity	Excitement	**0.09**	-0.01	-0.04
	Alertness	**0.10**	-0.01	-0.08
	Interest	0.01	0.05	**-0.10**
Convergent Validity	BDI	**-0.10**	**0.15**	**0.40**
	RFQ	-0.01	**0.21**	**0.22**
	LESS	0.06	**0.20**	**0.12**
	PANAS Positive	**0.17**	0.02	**-0.17**
	PANAS Negative	-0.04	**0.17**	**0.47**
	CCS	-0.04	**0.24**	**0.41**

Note. Excitement, alertness, and interest are single items selected from The Positive and Negative Affect Schedule (PANAS; Waston & Tellegan, 1988). Additional measures include: Beck Depression Inventory (BDI; Beck et al., 1996); Risky Family Questionnaire (RFQ; Taylor et al., 2006); Life Events Scale for Students (LESS; Clements & Turpin, 1996); DSM 5 Level 1 Cross Cutting Symptoms Measure (CCS; Narrow et al., 2013).

Convergent Validity with Non-Risk Behaviors. To further understand the validity of the MIM dimensions of behavioral activation, preference for stimulation, and inhibition control, each dimension was correlated with variables of interest (Table 2). Behavioral activation was found to have a small but significant, relation to the BDI (r =-.10, $p < .05$), and the PANAS positive total score (r =.17, $p < .05$). Preference for stimulation was found to be modestly correlated with the DSM cross cutting symptom inventory total score (r =.24, $p < .05$) and the RFQ total (r =.21, $p < .05$). Preference for Stimulation was significantly, but weakly correlated with the BDI (r =.15, $p < .05$), and the PANAS Negative subscale (r =.17, $p < .05$). Inhibition control was found to be correlated with the DSM cross cutting symptom inventory total score (r =.41, $p < .05$), the PANAS Negative subscale (r =.47, $p < .05$), and the BDI total score (r =.40, $p < .05$). Inhibition control was also significantly, but weakly correlated with the RFQ total (r =.22, $p < .05$), the LESS total score (r =.12, $p < .05$), and the PANAS positive total score (r =-.17, $p < .05$).

Relations between demographic variables and the MIM. To gain a better understanding of how the MIM dimensions relate to specific demographic characteristics, one-way analysis of variance or *t*-tests were completed for demographic variables of gender, race/ethnicity, martial status, monthly income, first generation college student status, and highest parental education level. No significant differences were found when the MIM dimensions were compared by race/ethnicity, grade, martial status, monthly income, first generation college student status, and highest parental education level.

A significant gender difference was found in preference for stimulation, with males (M = 39.78, SD = 5.79) having significantly higher scores than females (M = 36.72, SD = 6.28), t(528)

= 5.81, $p < .05$. No significant gender differences were found for either the inhibition control or behavioral activation factor.

Correspondence of the MIM with Risk Behaviors

Gambling behavior. The relationship between scores on the MIM dimensions and measures of problem gambling severity were examined. A positive correlation was found between preference for stimulation and gambling frequency, $r = .16$, $p < .05$ where increases in preference for stimulation was associated with increases in gambling frequency. Inhibition control and behavioral activation were not significantly correlated with gambling frequency.

Regression analyses were then completed. The overall model of the three impulsivity factors significantly predicted gambling frequency, $R^2 = .03$, $F(3,526) = 4.70$, $p < .05$. A closer examination of how the individual factors contributed to the model indicated that preference for stimulation scores, $b = .16$, $t(529) = 3.62$, $p < .05$ significantly contributed to the model, but inhibition control, $b = .01$, $t(529) = .26$, $p = .80$ and behavioral activation scores, $b = -.03$, $t(529) = -.72$, $p = .47$ did not.

A second set of correlations revealed a positive correlation between the SOGS score of gambling symptoms and preference for stimulation $r = .15$, $p < .05$ and inhibition control $r = .13$, $p < .05$. Higher scores on preference for stimulation and higher scores on inhibition control were associated with higher rates of gambling symptomatology.

Since gambling symptoms were measured as a count variable, a Poisson regression was run to predict the count of SOGS-based diagnostic symptom criteria a participant would meet based on the three dimensions. Preference for simulation, *Wald* χ^2 $(1, n = 530) = 13.76$, $p < .05$, and inhibition control, *Wald* χ^2 $(1, n = 530) = 14.06$, $p < .05$, were both found to be significant predictors of problematic gambling symptoms. Behavioral activation was not, *Wald* χ^2 (1, n =

530) = 1.13, p = .29. Significant differences were explored among the three impulsivity factors and gambling classification. Separate analyses of variance showed significant classification relations with the preference for stimulation dimension, $F(2,527) = 7.27$, $p < .05$, and inhibition control, $F(2,527) = 3.44$, $p < .05$, but not for behavioral activation, $F(2,527) = 0.54$, $p = .58$. Post hoc comparisons using the LSD test on preference for stimulation revealed that those with no gambling problems (M = 37.64, SD = 6.25) were significantly lower than gamblers with some symptoms (M = 39.71, SD = 5.98), and probable pathological gamblers (M = 41.92, SD = 4.32). Post hoc comparisons for inhibition control revealed that those with no gambling problems (M = 35.17, SD = 8.45) scored significantly lower than probable pathological gamblers (M = 40.42, SD = 3.32). Gamblers with some symptoms and those with no gambling problems were not significantly different from each other.

Alcohol Use Involvement. The relationship between scores on the MIM dimensions and measures of alcohol use were examined. A positive correlation was found between preference for stimulation and AUDIT total score, $r = .41$, $p < .05$ where increases in preference for stimulation was associated with increases in alcohol use problems. A positive correlation was also found between behavioral activation and AUDIT total score, $r = .13$, $p < .05$ where increases in behavioral activation was associated with increases in alcohol use problems. Inhibition control was not found to be significantly correlated with alcohol use problems.

Regression analyses were then completed. The overall model of the three impulsivity factors significantly predicted AUDIT total score, $R^2 = .17$, $F(3,526) = 36.11$, $p < .05$. A closer examination of how the individual factors contributed to the model indicated that preference for stimulation scores, $b = .41$, $t(529) = 9.81$, $p < .05$ significantly contributed to the model, but inhibition control, $b = -.03$, $t(529) = -.68$, $p = .50$ and behavioral activation scores, $b = .04$, $t(529) = 1.02$, $p = .31$ did not.

Significant differences were explored among the three impulsivity factors and alcohol use disorder classification. Analyses of variance showed significant relations with the preference for stimulation dimension, $F(2,527) = 55.70$, $p < .05$, and behavioral activation, $F(2,527) = 5.76$, $p < .05$, but not for inhibition control, $F(2,527) = 0.40$, $p = .67$. For preference for stimulation, post hoc comparisons using the LSD test revealed that non-drinkers ($M = 35.56$, $SD = 5.87$) were significantly lower than social drinkers ($M = 39.16$, $SD = 5.62$), and those meeting criteria for alcohol use disorders ($M = 43.09$, $SD = 5.23$). Post hoc comparisons for behavioral activation revealed that non-drinkers ($M = 45.80$, $SD = 9.29$) scored significantly lower than those meeting criteria for an alcohol use disorder ($M = 49.87$, $SD = 8.81$). Non-drinkers and social drinkers were not significantly different from each other on behavioral activation. Social drinkers and those meeting criteria for alcohol use disorder were also not significantly different from each other.

Drug Use Behavior. The relation between scores on the MIM dimensions and measures of drug use over the past two weeks were examined. A positive correlation was found between preference for stimulation and total severity and complexity of substance use, $r = .27$, $p < .05$ where increases in preference for stimulation was associated with increases in total severity and complexity of substance use. A positive correlation was also found between inhibition control and total severity and complexity of substance use, $r = .09$, $p < .05$ where increases in inhibition control was associated with increases in total severity and complexity of substance use. Behavioral activation was not found to be significantly correlated with total severity and complexity of substance use.

Regression analyses were then completed. The overall model of the three impulsivity factors significantly predicted total severity and complexity of substance use, $R^2 = .07$, $F(3,526) = 14.11$, $p < .05$. A closer examination of how the individual factors contributed to the model indicated that preference for stimulation scores, $b = .27$, $t(529) = 6.09$, $p < .05$ significantly contributed to the model, but inhibition control, $b = .04$, $t(529) = .87$, $p = .38$ and behavioral activation scores, $b = -.02$, $t(529) = -.52$, $p = .61$ did not.

Significant differences were explored among the three impulsivity factors and drug use involvement. An independent samples *t*-test showed significant relations with the preference for stimulation dimension, $t(528) = 10.10$, $p < .05$, and inhibition control, $t(528) = 3.13$, $p < .05$, but not for behavioral activation, $t(528) = 1.25$, $p = .21$. A closer examination of group difference for preference for stimulation revealed that those who had not used drugs in the past two weeks (M = 36.66, SD = 5.77) were significantly lower than past two-week drug users (M = 42.30, SD = 5.53). Similarly, those who had not used drugs in the past two weeks (M = 34.92, SD = 8.40) scored significantly lower on inhibition control than those who had used drugs in the past two weeks (M = 37.41, SD = 7.37).

Discussion

An analytic challenge when reducing items within a large measurement battery is how to best retain an empirically supported factor structure while also eliminating a large number of items. Recently, Larwin and Harvey (2012) proposed the SIR-SEM approach to provide an analytic strategy that retains a strong connection to an empirically supported model while a researcher reduces the size of the dataset. In this paper we presented an application of this SIR-SEM approach for item reduction. For this example, the goal was to reduce the number of items needed to assess an empirically supported multidimensional model of impulsivity (Ginley et al., 2014). In addition, we tested our version of the SIR-SEM approach using SAS rather than FORTRAN and LISERL. Upon completion of the reduction, the resulting measure, the MIM, was then evaluated both for model fit, validity, and correspondence with risk behaviors in a new sample of college students. Overall the MIM was found to be valid, reliable, largely demographically invariant and significantly predictive of a range of risk behaviors including gambling, alcohol use, and drug use.

The SIR-SEM approach in Study 1 successfully retained the empirically supported dimensions of impulsivity while also reducing the number of items. This analytic plan yielded strong initial model fit statistics, increasing confidence that the factor structure identified by Ginley and colleagues (2014) was retained in a new, larger sample of participants. Additionally, this method was successful in removing the majority of the items, ultimately removing 84% of them from the set of variables.

In Study 2 a large and diverse sample of participants completed the MIM measure that resulted from the SIR-SEM approach in an effort to cross validate Study 1 findings. This new measure had only 33 items, a universal response format, and took less than 10 mins to complete. The MIM was confirmed to have good model fit. The internal consistency of the three dimensions were also found to be adequate and with good test-retest reliability. The factors were found to be largely invariant by demographic, with the exception of preference for stimulation which was significantly related to participant gender. As hypothesized, behavioral activation had significant correspondence with the positive affect scale, however it was also found to weakly but significantly correspond with depression, a measure hypothesized to have been independent of the MIM dimensions. Preference for stimulation did not significantly correspond with life stressors as had been hypothesized, but did correspond with the depression, risky family environment, negative affect, and DSM psychopathology symptoms. The dimension of inhibition control did correspond with risky family environment as was hypothesized, and also corresponded with life stressors, depression, DSM psychopathology, and affect. The dimensions were also found to replicate the small but significant correspondence with risk dimensions that had been shown previously by the literature (Ginley et al., 2014; Meda et al., 2009). Taken together, the psychometric data obtained in this replication provided strong support for using the SIR-SEM to develop short, reliable, and valid scales from large measurement batteries (e.g., Smith et al., 2000). Interestingly, the findings of this specific multidimensional impulsivity model also appear to be less discriminant from general psychological distress then expected.

There are many benefits to reducing the multivariate space in a systematic manner. As it relates to the impulsivity and risk taking literature (e.g., MacKillop et al, 2014; Meda et al., 2009; Reynolds et al., 2006; Whiteside & Lyman, 2001; Whiteside, Lyman, Miller, & Reynolds, 2005), this item reduction provides at least two practical benefits for researchers. First, assessment of the three impulsivity dimensions can now be more efficient. Second, researchers can use the multidimensional model with confidence that it does not contain interference from less meaningful predictors (Epstein, 1984).

The reduced impulsivity model revealed by the SIR-SEM analysis did not fully support the previously reported relations between the three dimensions and the risk behaviors. Both preference for stimulation and inhibition control significantly contributed to explain target dependent variables, but behavioral activation did not. The failure of behavioral activation to correspond with any of the variables aside from alcohol use disorder classification provided an example of the benefit of using the SIR-SEM procedure. In the event items were eliminated from the impulsivity battery using only an EFA, behavioral activation would simply stop appearing.

While a researcher could note its absence by looking to the literature, he or she could generate few hypotheses about what had occurred. With the confidence that the three items identified by the SIR-SEM procedure do meaningfully represent the factor of behavioral activation, we can begin to postulate possible explanations. For example, behavioral activation might not be related to gambling variables or drug use variables due to its predictive variance being largely consumed by the preference for stimulation and/or inhibition control factors. It could also have been a sample specific finding or an artifact created by the limited number of items on this factor generating restricted variance so that only the largest group differences, those between alcohol use disorder classifications, were captured. It is possible to tie the finding to the literature and hypothesize that behavioral activation is not significantly predictive of all risk behaviors. This hypothesis would correspond with the results of Meda and colleagues (2009) who found that behavioral activation failed to significantly differentiate healthy controls from participants with addiction or at risk for addiction. It would also provide additional support for the findings of a recent meta analysis on impulsivity dimensions by Gullo, Loxton, and Dawe (2014) who identified a two factor impulsivity model based generally on approach impulsivity, what this model calls preference for stimulation, and general inhibitory processes, here called inhibition control.

By providing an example of the SIR-SEM statistical method with a psychological variable we made Larwin and Harvey's item reduction approach more accessible to behavioral science researchers and showcased the additional steps for a psychometric evaluation of the resulting short form in an independent sample (as recommended by Smith et al., 2000). Our application used PROC Calis in SAS/Stat and the Larwin and Harvey metacode to conduct the item reduction. As a commonly available analytical product used extensively in business and academia, the SAS program will allow a broad population of researchers to make use of the item reduction procedure. Another value to our application was that we decided to adopt more conventional fit statistics as recommended by Kline (2012).

However, this demonstration of a SAS based application of the Larwin and Harvey (2012) analysis has limits. First, we did not compare the SIR-SEM procedure directly with possible alternatives. For example, we could have experimentally contrasted the SIR-SEM with CFA and EFA methods. However, decisions to remove items via CFA and EFA methods, while empirically informed, can be somewhat arbitrary, making it challenging to conclude what method would be superior. Additionally, the SIR-SEM approach relied on global model fit indices as determinates of model fit. While the indices chosen were empirically informed, the decision to focus on them, instead of unmodeled cross-loadings or unmodeled error covariance when eliminating items was also a decision point that may have resulted in a model that fit the data well but did not fully consider all reliability estimates.

Additionally, while the SIR-SEM is able to do an effective job of removing items within an empirically supported model it is not able to overcome all of the limitations that go along with the development of a short form. While the MIM was found to be psychometrically valid, removal of variables necessitates less content coverage. The prediction of risk behavior by the MIM dimensions was very small, and it is possible that important risk behavior-specific impulsivity content items were removed when trying to achieve improved model fit. The multidimensional impulsivity structure identified by Ginley and colleagues (2014) was found to replicate in two new samples (i.e., the Study 1 and Study 2 samples) with good model fit, and the findings in both studies do mirror the small amount of correspondence obtained in the original studies upon which the multidimensional model is based (Ginley et al., 2014; Meda et al., 2009).

These replications on their own are a significant finding and provide considerable support for the importance of multidimensional assessment of impulsivity. However, the SIR-SEM procedure was not able to directly assess reliability and validity issues which may be stemming from including measures within the model that had only moderate validity rates to start with. Nor is a short form able to overcome how the complexity of impulsivity means it is less discriminant from other variables of interest then was hypothesized. A benefit of the shortened measure is that these validity explorations can now happen much more efficiently, moving the science along at a more rapid pace with reduced participant burden.

Conclusion

SIR-SEM provides a useful tool when researchers want to improve theoretically driven or empirically based measurement models. In this study the SIR-SEM approach was applied to a large impulsivity battery (Ginley et al., 2014; Meda et al., 2009). The SIR-SEM was found to effectively accomplish the study aims of retaining the three dimensional model of impulsivity, reducing the number of parameters needed for assessment, and producing a short impulsivity scale with a good model fit performance, validity, and correspondence with target risk behaviors when evaluated in a separate sample. In addition, this demonstration provided support for the SIR-SEM procedure in SAS and presented the use of commonly recommended fit statistics. Although further testing of this approach and model is needed, this demonstration provides social scientists with an accessible, structural modeling based approach applicable for short form measure development and testing.

References

Aertgeerts, B., Buntinx, F., Bande-Knops, J., Vanderrneulen, C., Roelants, M., Ansoms, S., & Fevery, J. (2000). The value of CAGE, CUGE, and AUDIT in screening for alcohol abuse and dependence among college freshmen. *Alcoholism: Clinical and Experimental Research*, *24*(1), 53-57.

Akaike, H. (1978) A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, *30*, 9-14.

Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, *21*(4), 613-619.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

American Psychiatric Association. DSM-5: Online Assessment measures. http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures accessed 2/2/16.

Bagozzi, R. P. (1981). Evaluating structural equation models with unobservable variables and measurement error: a comment. *Journal of Marketing Research*, *18*(3), 375-381. doi:10.2307/3150979

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the BDI-II.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246. doi: 10.1037/0033-2909.107.2.238

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

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345-370. doi: 10.1007/BF02294361

Breen, R., & Zuckerman, M. (1999). 'Chasing' in gambling behavior: Personality and cognitive determinants. *Personality and Individual Differences, 27*(6), 1097-1111.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), SAGE Focus Editions: Vol. 154. *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466. doi: 10.1037/0033-2909.105.3.456

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology*, *67*(2), 319-333. doi: 10.1037/0022-3514.67.2.319

Caseras, X., Ávila, C., & Torrubia, R. (2003). The measurement of individual differences in Behavioural Inhibition and Behavioural Activation Systems: A comparison of personality scales. *Personality and Individual Differences, 34*(6), 999-1013. doi:10.1016/S0191-8869(02)00084-3

Clark, R. A., & Goldsmith, R. E. (2006). Global innovativeness and consumer susceptibility to interpersonal influence. *The Journal of Marketing Theory and Practice*, *14*(4), 275-285. doi: 10.2753/MTP1069-6679140402

Clements, K., & Turpin, G. (1996). The Life Events Scale for Students: validation for use with British samples. *Personality and Individual Differences*, *20*(6), 747-751.

Clements, R. (1998). A Critical Evaluation of Several Alcohol Screening Instruments Using the CIDI-SAM as a Criterion Measure. *Alcoholism: Clinical and Experimental Research*, *22*(5), 985-993.

Dager, A. D., Anderson, B. M., Rosen, R., Khadka, S., Sawyer, B., Jiantonio-Kelly, R. E., ... Fallahi, C. R. (2014). Functional magnetic resonance imaging (fMRI) response to alcohol pictures predicts subsequent transition to heavy drinking in college students. *Addiction*, *109*(4), 585-595.

Dawe, S., & Loxton, N. (2004). The role of impulsivity in the development of substance use and eating disorders. *Neuroscience and Biobehavioral Reviews, 28*(3), 343-351. doi:10.1016/j.neubiorev.2004.03.007

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*(2), 175-191.

Dozois, D. J., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory–II. *Psychological Assessment*, *10*(2), 83.

Epstein, R. (1984). The principle of parsimony and some applications in psychology. *Journal of Mind and Behavior, 5*(2), 119-130.

Figee, M., Pattij, T., Willuhn, I., Luigjes, J., van den Brink, W., Goudriaan, A., ... Denys, D. (2015). Compulsivity in obsessive–compulsive disorder and addictions. *European Neuropsychopharmacology*. Available online: http://www.sciencedirect.com/science/article/pii/S0924977X15003867.

Ginley, M. K., Whelan, J. P., Meyers, A. W., Relyea, G. E., & Pearlson, G. D. (2014). Exploring a multidimensional approach to impulsivity in predicting college student gambling. *Journal of Gambling Studies*, *30*(2), 521-536. doi: 10.1007/s10899-013-9374-9

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gray, J.A. (1970). The psychophysiological basis of introversion-extraversion. *Behaviour Research and Therapy, 8*(3), 249-266. doi:10.1016/0005-7967(70)90069-0

Gullo, M. J., Loxton, N. J., & Dawe, S. (2014). Impulsivity: Four ways five factors are not basic to addiction. *Addictive Behaviors*, *39*(11), 1547-1556.

Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.

Hyatt, C. J., Assaf, M., Muska, C. E., Rosen, R. I., Thomas, A. D., Johnson, M. R., ... Potenza, M. N. (2012). Reward-related dorsal striatal activity differences between former and current cocaine dependent individuals during an interactive competitive game. *PLoS One*, *7*(5), e34917.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. doi: 10.1080/10705519909540118

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351.

Kline, R. B. (2012). Assumptions in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp.111-125). New York, NY: Guilford Press.

Langewisch, M. W., & Frisch, G. R. (1998). Gambling behavior and pathology in relation to impulsivity, sensation seeking, and risky behavior in male college students. *Journal of Gambling Studies*, *14*(3), 245-262. doi: 10.1023/A:1022005625498

Larwin, K., & Harvey, M. (2012). A demonstration of a systematic item-reduction approach using structural equation modeling. *Practical Assessment, Research & Evaluation*, *17*(8), 2. Available online: http://pareonline.net/getvn.asp?v=17&n=8

Evaluation, *17*(8), 2.

Lesieur, H.R., & Blume, S.B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *The American Journal of Psychiatry, 144*(9), 1184-1188.

MacKillop, J., Miller, J. D., Fortune, E., Maples, J., Lance, C. E., Campbell, W. K., & Goodie, A. S. (2014). Multidimensional examination of impulsivity in relation to disordered gambling. *Experimental and Clinical Psychopharmacology*, *22*(2), 176.

Meda, S. A., Stevens, M. C., Potenza, M. N., Pittman, B., Gueorguieva, R., Andrews, M. M., Thomas, A. D., … Pearlson, G. D. (2009). Investigating the behavioral and self-report constructs of impulsivity domains using principal component analysis. *Behavioural Pharmacology*, *20*(5-6), 390-399. doi: 10.1097/FBP.0b013e32833113a3

Michaelson, G. (2015). Teaching programming with computational and informational thinking. *Journal of Pedagogic Development*, *5*(1).

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry*, *158*(11), 1783-1793.

Narrow, W. E., Clarke, D. E., Kuramoto, S. J., Kraemer, H. C., Kupfer, D. J., Greiner, L., & Regier, D. A. (2013). DSM-5 Field Trials in the United States and Canada, part III:

development and reliability testing of a cross-cutting symptom assessment for DSM-5. *American Journal of Psychiatry*, *170*(1), 71-82.

National Institute on Drug Abuse: NIDA Quick Screen V 1.0. http://www.nida.nih.gov/nidamed/screening/nmassist.pdf, accessed 2/2/16.

Newcomb, M. D., & Bentler, P. M. (1988). *Consequences of adolescent drug use: Impact on the lives of young adults*. Thousand Oaks, CA: Sage.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory.* New York: McGraw-Hill.

Patton, J. H., Stanford, M. S., & Barratt, E.S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, *51*(6), 768-774. doi: 10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Petry, N. M. (2001). Pathological gamblers, with and without substance abuse disorders, discount delayed rewards at high rates. *Journal of Abnormal Psychology*, *110*(3), 482-487. doi: 10.1037/0021-843X.110.3.482

Platt, J. (1964). Strong inference. *Science*, *146*(3642), 347-353. doi: 10.1126/science.146.3642.347

Potenza, M. N. (2006). Should addictive disorders include non-substance-related conditions? *Addiction*, *101*(1), 142-151.

Reinert, D. F., & Allen, J. P. (2002). The alcohol use disorders identification test (AUDIT): a review of recent research. *Alcoholism: Clinical and Experimental Research*, *26*(2), 272-279.

Rensvold, R. B., & Cheung, G. W. (1999). Identification of influential cases in structural equation models using the jackknife method. *Organizational Research Methods*, *2*(3), 293-308. doi: 10.1177/109442819923005

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*(2), 305-315. doi: 10.1016/j.paid.2005.03.024

Ridgeway, D., & Russell, J.A. (1980). Reliability and validity of the Sensation-Seeking Scale: Psychometric problems in Form-V. *Journal of Consulting and Clinical Psychology, 48*(5), 662-664. doi:10.1037/0022-006X.48.5.662

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental design for generalized causal inference. Boston: Houghton-Mifflin.

Sanavio, E. (1988). Obsessions and compulsions: the Padua Inventory. *Behaviour Research and Therapy*, *26*(2), 169-177. doi: 10.1016/0005-7967(88)90116-7

Santor, D. A., Haggerty, J. L., Lévesque, J. F., Burge, F., Beaulieu, M. D., Gass, D., & Pineault, R. (2011). An overview of confirmatory factor analysis and item response analysis applied to instruments to evaluate primary healthcare. *Healthcare Policy*, *7*, 79. doi: 10.12927/hcpol.2011.22694

SAS Institute Inc. 2011. *SAS/STAT® 9.3 User's Guide*. Cary, NC: SAS Institute.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23-74.

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review*, *10*(4), 320-335. doi: 10.1207/s15327957pspr1004_3

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102.

Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L, Anderson, N.E., & Patton, J.H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*(5), 385-395. doi: 10.1016/j.paid.2009.04.008

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180. doi: 10.1207/s15327906mbr2502_4

Sternberger, L., & Burns, G. (1990). Obsessions and compulsions: Psychometric properties of the Padua Inventory with an American college population. *Behaviour Research and Therapy, 28*(4), 341-345. doi:10.1016/0005-7967(90)90087-Y

Stinchfield, R. (2002). Reliability, validity, and classification accuracy of the South Oaks Gambling Screen (SOGS). *Addictive Behaviors*, *27*(1), 1-19.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics,* (6th ed.). Boston, MA: Pearson.

Taylor, S. E., Eisenberger, N. I., Saxbe, D., Lehman, B. J., & Lieberman, M. D. (2006). Neural responses to emotional stimuli are associated with childhood family stress. *Biological Psychiatry*, *60*(3), 296-301.

Torrubia, R., Avila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences*, *31*(6), 837-862. doi: 10.1016/S0191-8869(00)00183-5

van der Sluis, S., Dolan, C. V., & Stoel, R. D. (2005). A note on testing perfect correlations in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(4), 551-577. doi: 10.1207/s15328007sem1204_3

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Weinstock, J., Whelan, J. P., Meyers, A. W., & McCausland, C. (2007). The performance of two pathological gambling screens in college students. *Assessment*, *14*(4), 399-407.

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669-689. doi: 10.1016/S0191-8869(00)00064-7

Whiteside, S. P., Lynam, D. R., Miller, J. D., & Reynolds, S. K. (2005). Validation of the UPPS impulsive behaviour scale: a four-factor model of impulsivity. *European Journal of Personality*, *19*(7), 559-574.

Yarosh, H. L., Hyatt, C. J., Meda, S. A., Jiantonio-Kelly, R., Potenza, M. N., Assaf, M., & Pearlson, G. D. (2014). Relationships between reward sensitivity, risk-taking and family history of alcoholism during an interactive competitive fMRI task. *PloS One*, *9*(2), e88188.

Zuckerman, M. (1996). The psychobiological model for impulsive unsocialized sensation seeking: A comparative approach. *Neuropsychobiology*, *34*(3), 125-129. doi: 10.1159/000119303

Zuckerman, M., Eysenck, S.B., & Eysenck, H.J. (1978). Sensation seeking in England and America: Cross-cultural, age, and sex comparisons. *Journal of Consulting and Clinical Psychology, 46*(1), 139-149. doi:10.1037/0022-006X.46.1.139