Journal of Contextual Behavioral Science 5 (2016) 146–153

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

journal homepage: www.elsevier.com/locate/jcbs

Empirical Research

# How Bayesian estimation might improve CBS measure development: A case study with body-image flexibility in Hispanic students

A. Solomon Kurz [a,*], Maureen K. Flynn [b], Michael J. Bordieri [c]

[a] Department of Psychology, University of Mississippi, United States
[b] Department of Psychology, Metropolitan State University of Denver, United States
[c] Department of Psychology, Murray State University, United States

ARTICLE INFO

*Article history:*
Received 5 September 2015
Received in revised form
14 July 2016
Accepted 28 July 2016

*Keywords:*
Bayes
Body-image flexibility
BSEM
Factor analysis
Residual

ABSTRACT

The methods for examining questionnaires in psychology are steeped in conventional statistics. However, many within the social sciences have started exploring Bayesian methods as an alternative to the conventional approach. This paper highlights the usefulness of Bayesian methodology for factor analysis, using the Body Image Acceptance and Action Questionnaire (BI-AAQ) as a case study. In an all-Hispanic undergraduate sample ($n=289$), we compared techniques from Bayesian and frequentist estimation for examining the factor structure of the BI-AAQ. Results indicated Bayesian estimation was flexible and offered unique insights relative to the conventional frequentist approach. We conclude the BI-AAQ was a structurally valid measure for our all-Hispanic sample and that Bayesian methods may be fruitful for further evaluation within the contextual behavioral science community.



## 1. Introduction

Those within the emerging contextual behavioral science (CBS) tradition have endeavored to develop meaningful interventions grounded in basic and applied theory (Hayes, Barnes-Holmes, & Wilson, 2012). An important part of this strategy is developing self-report measures of key components of the psychological flexibility model (Hayes, Levin, Plumb-Vilardaga, Villatte, & Pistorello, 2013). In this regard, CBS has expanded beyond its traditional behavior analytic roots by embracing pragmatically useful mid-level functional terms and by employing statistical analyses to assess for the impact of psychological processes at the group level (Vilardaga, Hayes, Levin, & Muto, 2009). Like Vilardaga and colleagues, we argue this expansion need not result in an abandonment of the precision afforded by a functional analytic approach to human behavior as long as statistical inference is used in a manner consistent with CBS's behavioral origins.

That said, most of the CBS-based measures to date have been developed using traditional psychometric procedures consisting of exploratory and confirmatory factor analyses grounded in a classical frequentist approach to statistical probability. Notable examples include the Acceptance and Action Questionnaire (AAQ-II; Bond et al., 2011) and the Body Imagine Acceptance and Action Questionnaire (BI-AAQ; Sandoz, Wilson, Merwin, & Kellum, 2013). While this traditional measurement development strategy is not necessarily incongruent with the philosophical assumptions guiding CBS, the functional contextual foundation of CBS allows for a more pragmatic approach to measure development (see Wilson, Whiteman, & Bordieri, 2013 for an extended CBS discussion of pragmatism and truth). As Ciarrochi and colleagues note (2016), there is a fundamental tension between the two approaches as traditional psychometric theory holds stability across time and context as a hallmark indicator of quality while CBS places greater value on measures that are sensitive to context. A CBS approach to measure development places utility of the measure above traditional metrics of stability and in doing so allows for a more flexible analytic posture (Ciarrochi et al., 2016). We believe that Bayesian estimation may provide additional analytic tools consistent with CBS assumptions. In this paper we offer an overview of Bayesian estimation and provide an example of the approach applied to a CBS measure in a all-Hispanic undergraduate sample.

## 2. An introduction to Bayesian estimation

There are two major approaches to probability within the field of statistics: frequentist and Bayesian (see Fienberg, 2006). The Bayesian approach began following the posthumous publication of Reverend Thomas Bayes' famous essay on probability (Bayes & Price, 1763) and was widely adopted among statisticians up through the early twentieth century. However, following criticisms

* Correspondence to: Department of Psychology, University of Mississippi, 205 Peabody, MS 38677, United States.
*E-mail address:* ajkurz@go.olemiss.edu (A. Solomon Kurz).

http://dx.doi.org/10.1016/j.jcbs.2016.07.005
2212-1447/

from notable statisticians, such as Sir Ronald Fisher, frequentist notions of probability overshadowed the Bayesians for much of the twentieth century. Some of the criticisms were philosophical, concerning how statisticians should view probability (for an introduction, see Dienes, 2011). Other criticisms were practical; for example, Bayesian estimation is often more computationally demanding and time consuming than frequentist estimation (Krushcke, 2015). As a result, Bayesian approaches fell out of favor and frequentist approaches spread such that the vast majority of the inferential statistics reported in journals in the social sciences are frequentist.

But Bayesian estimation is more than a historic artifact and many contemporary statisticians and social scientists are reconsidering Bayesian estimation (Kruschke, 2011b). This is in part due to advances in efficient algorithms and computer speed that have made Bayesian estimation increasingly practical (Kruschke, 2015). Moreover, some contemporary statisticians have put philosophical issues aside and adopted a pragmatic statistical approach (e.g., Kass, 2011) wherein estimation procedures are chosen based on their utility for the task at hand. The reemergence of Bayesian methodology in the social sciences is evident in recent articles in methodological journals (e.g., Yuan & MacKinnon, 2009), statistical text books (e.g., Kruschke, 2015; McElreath, 2016), and at least three recent special issues in social science journals (Kruschke, 2011a; Mulder & Wagenmakers, 2016; Zyphur, Oswald, & Rupp, 2015).

## 3. Bayes for factor analysis

Articles on self-report-based exploratory and confirmatory factor analyses (EFA and CFA, respectively) often emphasize the number of factors, factor correlations – in the case of multiple factors – and the strength of the factor. This is understandable given the primary concerns for many questionnaire developers, such as *Do the seven items in the AAQ-II form a coherent measure of psychological flexibility?* and *Does the AAQ-II correlate reasonably with known measures of mindfulness?*. Yet, other important parts of the statistical model are often underemphasized, such as residual correlations. Residual variances are the portions of the items not accounted for by the factor, and combinations of characteristics unique to the items and of measurement error (see Brown, 2015). For the simple single-factor model, the traditional presumption is that residuals are uncorrelated with each other. That is, that the correlations among the items are fully accounted for by the factor. By default, they are typically constrained to zero, which is convenient since, due to issues regarding model identification, CFAs done within a frequentist framework cannot freely estimate all possible residual correlations within a model (Muthén & Asparouhov, 2012a). Yet, it is unclear that clinical researchers should expect a factor to perfectly explain all the correlations among items. CFAs with clinically-oriented questionnaires often require residual correlations for good model fit, even in the presence of high factor loadings (Cole, Ciesla, & Steiger, 2007). For example, Bond and colleagues (2011) discovered that achieving reasonable model fit for the AAQ-II required a residual correlation for the two items including the word "painful." As MacCallum (2003) and Box (1979) reminded us in his presidential address to the Society of Multivariate Experimental Psychology, statistical models are always imperfect and are, at best, useful tools for our topics of interest. Thus, if possible, we should prefer statistical models that are flexible enough to accommodate imperfections.

This is where Bayesian estimation can help. In Bayesian inference, results are always the joint product of characteristics of the data and prior knowledge (see Zyphur & Oswald, 2015). While scale developers from both traditions apply their prior theoretical knowledge at each step of development, frequentist statistical inference only involves the characteristics of the data from the specific research sample. In contrast, Bayesians formally include prior knowledge into the statistical estimation process with what are called model priors—priors, for short. Priors may be generally characterized as either informative or uninformative. Uninformative priors have minimal influence on the analysis and are often specified for pragmatic computational purposes, to convey uncertainty about their topic area, or to yield estimates the most similar to those from frequentist estimators (Muthén & Asparouhov, 2012a). Informative priors allow researchers to influence their results with prior information, which may come from objective or subjective sources. In science, informative priors based on previous studies let researchers their aggregate knowledge, giving studies using such priors a meta-analytic flair (Zyphur & Oswald, 2015).

Muthén & Asparouhov (2012a, 2012b) recently proposed the Bayesian structural equation model (BSEM) method, wherein a special kind of informative prior, zero-mean small-variance priors, may replace the traditional exact-zero residual correlations in CFA with *approximate* zeros. In the words of Asparouhov, Muthén, and Morin (2015), "these parameters are neither completely fixed to zero nor are completely free, but are instead approximately fixed to zero" (pp. 4–5). So far, BSEM priors have substantially improved several applied factor models (e.g., Falkenström, Hatcher, & Holmqvist, 2014; Zyphur & Oswald, 2015). We propose that the approximate zeros afforded by the BSEM method will effectively accommodate the imperfections inherent in our statistical models. To examine the utility of the BSEM method for questionnaires within the CBS community, we present a case-study using a measure of body image flexibility.

## 4. Body image flexibility

Evaluating one's body negatively may be termed body image dissatisfaction (Stice & Shaw, 2002). Although body image dissatisfaction has been shown to predict eating disorder symptomatology (e.g., Brannan & Petrie, 2008; Corning, Krumm, & Smitham, 2006), it is likely linked to disordered eating through moderating variables because not everyone with body image dissatisfaction engages in problematic eating behaviors (Timko, Juarascio, Martin, Faherty, & Kalodner, 2014). Two variables shown in the literature to influence the relationship between body image dissatisfaction and disordered eating are general and body-image-related psychological flexibility (e.g., Hill, Masuda, & Latzman, 2013; Sandoz et al., 2013; Timko et al., 2014). Body image inflexibility is a specific form of psychological inflexibility that entails efforts to avoid unwanted thoughts, feelings, bodily sensations, and memories pertaining to the body, even when doing so involves actions that are incongruent with personal values. Sandoz et al. (2013) developed the first body image flexibility measure, the BI-AAQ. Results from their two principal factor analyses suggest the BI-AAQ had a single-factor structure, with standardized factor loading above .60 and item-total correlations above .52 in their samples. Ferreira, Pinto-Gouveia, and Duarte (2011) replicated their single-factor structure with responses from Portuguese participants on a translated version of the measure.

Construct validity for the BI-AAQ may be found in its negative correlations with disordered eating, body shape dissatisfaction, BMI, and internalization of the thin ideal and a positive correlation with general psychological inflexibility (Butryn et al., 2013; Sandoz et al., 2013; Timko et al., 2014). Body image flexibility has been shown to mediate the relationship between body image dissatisfaction and disordered eating (Timko et al., 2014) and the relationship between disordered eating cognitions and disordered eating pathology (Wendell, Masuda, & Le, 2012). Pre- to post-treatment improvements in BI-AAQ scores were associated with decreases in disordered eating and those with the greatest BI-AAQ increases had the largest decreases in disordered eating (Butryn et al., 2013). The BI-AAQ also demonstrated incremental validity by improving the prediction of disordered eating (Hill et al., 2013; Sandoz et al., 2013; Timko et al., 2014).

The psychometric properties of the BI-AAQ have not been examined for Hispanic persons, who constitute approximately 17% of the population (U.S. Census Bureau, 2014). Given the large number of Hispanics living in the Unites States (US), it is important they are adequately represented in clinical and psychometric research (van de Vijver, 2011). One common misconception is that body image dissatisfaction is much greater in Whites than other ethnicities. The results of a meta-analysis (Grabe & Hyde, 2006) and literature review (Ricciardelli, McCabe, Williams, & Thompson, 2007) examining ethnic differences in body image dissatisfaction dispute this belief, however. When inspecting Hispanics in particular, Hispanic women did not differ in body image dissatisfaction compared with White women (Grabe & Hyde, 2006). Likewise, evidence from the literature review suggests that body image dissatisfaction is similar for Hispanic and White men (Ricciardelli et al., 2007). Regarding disordered eating, lifetime prevalence rates of Bulimia Nervosa, Binge Eating Disorder, and any binge eating for Hispanics are within the range of prevalence estimates for Caucasian samples (Alegria et al., 2007).

Hispanics have been largely unrepresented in prior BI-AAQ studies of US samples. For example, Sandoz et al. (2013) and Timko et al. (2014) reported their samples were less than 6% Hispanic. Although there is accumulating evidence supporting the broad cultural applicability of ACT processes (Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Hayes, Muto, & Masuda, 2011), evidence is still needed to demonstrate suitability in Hispanic populations in particular. We predict the BI-AAQ will function similarly with Hispanic college students because of the universal assertion of the model, evidence supporting the applicability of ACT processes in other cultures, and similar rates of body image dissatisfaction and disordered eating in Hispanic populations.

## 5. The present study

For the present study, we examined the factor structure of the BI-AAQ in an all-Hispanic undergraduate sample. The purpose was twofold: (a) to determine whether the BI-AAQ is a useful measure of body image flexibility in Hispanic college students and (b) to compare the utility of frequentist and Bayesian-based analytic strategies. Based on prior research, we expected the BI-AAQ to yield a single-factor structure, high factor loadings, and predictive validity in our sample. For the analytic strategy, we were particularly interested in two lines of comparison :1) whether the two methods lead to similar conclusions in a simple single-factor model with no residual correlations and 2) in the event that either or both indicated poor fit for the initial model, how the frequentist and Bayesian methods compared for accommodating residual correlations.

## 6. Method

### 6.1. Participants

Undergraduate students ($N$=382) enrolled in psychology courses at a public southwestern US university volunteered for the study in exchange for course credit.

### 6.2. Measures

#### 6.2.1. Demographic questionnaire

Sex, age, country of origin, and weight were assessed via a self-report questionnaire.

#### 6.2.2. Body Image-Acceptance and Action Questionnaire

The BI-AAQ (Sandoz et al., 2013) is 12-item measure designed to measure body-image-related psychological flexibility. Users rate items such as “I shut down when I feel bad about my body shape or weight” on a 7-point Likert-type scale ranging from 1 (*Never true*) to 7 (*Always true*), with lower scores suggesting greater body-image flexibility. Sandoz et al. (2013) also Ferreira et al. (2011) found support for a single-factor solution.

#### 6.2.3. Eating Attitudes Test-26

We assessed disordered eating behavior using the EAT-26 (Garner, Olmsted, Bohr, & Garfinkel, 1982), a 26-item self-report questionnaire. With the EAT-26, users rate items on a 6-point Likert-type scale ranging from 5 (*Always*) to 0 (*Never*). Garner and colleagues suggested the EAT-26 should have three theoretically-derived subscales: Dieting, Food Preoccupation, and Oral Control. Subsequent psychometric studies have not supported the proposed 3-factor structure and the factor structures proposed across these studies have conflicted (Maïano, Morin, Lanfranchi, & Therme, 2013). However, the EAT-26 total score is a widely used and recommended indicator of disordered eating behavior (Anderson, Lundgren, Shapiro, & Paulosky, 2004). The EAT-26 total score is often used as a screening tool (e.g., Orbitello et al., 2006), can accurately separate those diagnosed with eating disorder form controls (e.g., Kiezebrink, Campbell, Mann, & Blundell, 2009; Mintz & O'Halloran, 2000), and was used as a criterion variable in previous studies with the BI-AAQ (Hill et al., 2013; Sandoz et al., 2013). In the present study, we estimated a general EAT-26 score.

#### 6.2.4. Attention check items

In order to control for careless responding (Meade & Craig, 2012), we included two attention check items in the survey battery. The item “Please fill in the ‘somewhat likely’ option for this item” was embedded within the items of a questionnaire two questionnaire slots before the BI-AAQ. Another item reading “Please click the ‘rarely’ option for this item” was embedded in a questionnaire three slots after the BI-AAQ.

### 6.3. Procedure

Approval was obtained from the university's Institutional Review Board. All participants provided informed consent and were administered a questionnaire battery using university owned survey software. Participants completed these measures in a computer lab with one or two researchers present and then received extra course credit.

### 6.4. Statistical analysis strategy

#### 6.4.1. Data preparation

Prior to primary study analyses, we assessed BI-AAQ item distributions with histograms, skew (Mdn=.39, SD=.32) and kurtosis (Mdn=−1.03, SD=.35) statistics. We checked for response quality with attention check items, invariant response patterns, and multivariate outliers. A small percentage of the BI-AAQ item values (0.1%) were missing. All analyses were conducted using M*plus* 7.3.

#### 6.4.2. Frequentist CFA

To account for the missing data and nonnormal item distributions, we used the full information maximum likelihood estimator with nonnormal robust standard errors (MLR; Yuan & Bentler, 2000; also, Enders, 2010). These models were evaluated using the model $\chi^2$, the root mean square error of approximation (RMSEA), and the Comparative Fit Index (CFI). Though nonsignificant $\chi^2$ values are frequently dismissed in the literature because of their well-known tendency to over-penalize trivial misfit in large samples, our modest sample size warranted considering the $\chi^2$ seriously (Kline, 2011). For interpreting the RMSEA and CFI, we used the following rules of thumb: RMSEA values $<.08$ and $<.05$ suggested acceptable and good fit, respectively; CFI values $>.90$ and $>.95$ suggested acceptable and good fit, respectively (Brown, 2015). We also present the Bayesian information criterion (BIC) for nested and nonnested model comparisons. The BIC has an unbounded range and models with lower relative values are generally preferred (Vrieze, 2012).

#### 6.4.3. *Bayesian CFA and BSEM*

Full information Bayesian estimation also accommodates data with missing values and nonnormal distributions (e.g., Yuan & MacKinnon, 2009). We used a Markov chain Monte Carlo algorithm, the Gibbs sampler, which takes many draws from the posterior distribution of the model parameters until convergence criteria are met. Herein, we specified three Markov chains of 200,000 iterations each and assessed their convergence with the potential scale reduction factor (PSR) and the Kolmogorov-Smirnov test (K-S). PSR values are ratios of between-chains variance over within-chains variance and they indicate convergence when that ratio stably approaches one (Gelman & Rubin, 1992). The K-S test determines whether the distributions for each Markov chain pair significantly differed from each other using *p*-value cutoffs (Muthén & Asparouhov, 2012b). Bayesian estimation requires the specification of prior distributions for all estimated parameters. Unless otherwise specified, we used M*plus* default uninformative priors, which was true for all factor loadings and intercepts. For BSEM models, we specified zero-mean small-variance priors for residual correlations using the inverse-Wishart distribution (Muthén & Asparouhov, 2012a). For a detailed discussion of the priors used in this manuscript and of BSEM priors in general, see the supplemental material.

With Bayesian estimation, overall model fit may be assessed with the posterior predictive *p*-value (PPP; Gelman, Meng, & Stern, 1996). For structural equation models (SEMs) with continuous outcomes, in M*plus*, the PPP for model fit uses the typical model $\chi^2$ as the test statistic. This makes the PPP the proportion of times the $\chi^2$ from model-based synthetic data exceeds the $\chi^2$ for the sample data (Muthén & Asparouhov, 2012a; also, Song & Lee, 2012). The ideal PPP-value is .5 and values approaching or below .05 suggest poor fit. To compare the difference between the model $\chi^2$ for the real and model-implied synthetic data, M*plus* also produces 95% confidence intervals, which straddle zero in good fitting models (Muthén & Asparouhov, 2012a). In addition to the BIC, many within Bayesian literature recommend comparing models with the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002), which is interpreted much like the BIC; models with relatively lower values are preferred. Asparouhov and colleagues (2015) showed that when using the BSEM method, the DIC is more appropriate than the BIC because the BIC over-penalizes BSEM models with smaller sample sizes.

#### 6.4.4. *Reliability and estimate intervals*

Due to nonnormal item distributions, differences among the factor loadings, and residual correlations, quantifying internal consistency with Cronbach's α would likely produce biased estimates (Brown, 2015). Instead, we used the scale reliability coefficient ρ, which was proposed by Lord, Novick, and Birnbaum (1968) and further developed by others (e.g., Raykov, 2001, 2009; see also Brown, 2015; Kline, 2011). When estimated with modern SEM software, ρ can accommodate nonnormal distributions, varying factor loadings, and residual correlations (Raykov, 2001, 2009). Coefficient ρ may be interpreted much like α, with values approaching one indicating good scale reliability.

To be sensitive to recent criticisms of the use of *p*-values in psychology (e.g., Cumming, 2014) and to Bayesian sensibilities, we presented 95% intervals for specific point estimates. For frequentist analyses, we used confidence intervals (CI), the Bayesian counterparts for which are called *probability* intervals (PI; also often called credibility intervals). For extensive discussions contrasting CIs and PIs, see Kruschke and Liddell (2015) and Zyphur and Oswald (2015).

## 7. Results

### 7.1. *Participants and item characteristics*

The initial dataset contained responses from 382 participants. One participant with missing data on all BI-AAQ items was excluded from further analysis. We removed data from another who identified as 17 years old and from 20 more who did not identify as Hispanic. We further excluded data from 48 participants who failed one or both of the attention check items, 17 with an unvarying response pattern for all BI-AAQ items (e.g., 1 1 1…; see Meade & Craig, 2012), and seven more with multivariate outlier response patterns for the BI-AAQ items based on a Mahalanobis $D^2 p<.001$. Thus, the following analyses were based on the responses of the remaining 289 participants.

Demographically, our 289 participants were 100% Hispanic and 230 (79.6%) identified as female. Regarding country of origin, 246 (85.1%) identified the United States, 40 (13.8%) identified Mexico, 1 identified Puerto Rico, and 2 chose “Other” without further specification. Average age was 21.1years ($Mdn=20$, $SD=4.5$), average weight was 150.7 lbs. ($Mdn=143$, $SD=38.5$), and average calculated body mass index was 25.6 ($Mdn= 24.0$, $SD=5.7$).

All BI-AAQ items were positively correlated with one another, with a high average inter-item correlation of .57 ($Mdn=.62$, $SD=.13$). From the frequentist perspective, all correlations were significant with $p<.001$, with the sole exception of $r_{2,6}$, for which $p=.025$. From a more Bayesian perspective, the lower limits of the 95% probability intervals were well above zero, indicating it was very improbable that the BI-AAQ items were negatively correlated in Hispanic undergraduates. See the supplementary material for full correlation matrices and descriptive statistics for the BI-AAQ items.

### 7.2. *Confirmatory factor analyses with MLR*

First, we estimated a single-factor model with all residual correlations constrained to zero. The resulting $\chi^2$ and RMSEA indicated unacceptably poor model fit (see Table 1). However, the standardized factor loadings were fairly high ($Mdn=.81$, $SD=.12$, ranging from .48 to .86), suggesting the 12 items were reasonably reliable. We then estimated three more models by sequentially freeing the residual correlation with the largest modification index (see Table 1). Results showed model fit improved notably after each freed residual correlation. However, even the model with three residual correlations had a significant model $\chi^2$ and an RMSEA only within the adequate range. The magnitudes of the residual correlations were small (i.e., $\theta_{10,11}=.14$, $\theta_{8,9}=.14$, and $\theta_{6,9}=-.14$ in the final model) and we discerned no clear linguistic or methodological reasons why those three residual pairs should covary more so than other pairs.

**Table 1**
Fit statistics for the BI-AAQ factor analyses computed with the MLR estimator.

| | | | $\chi^2$ | | RMSEA | | | ρ | |
|---|---|---|---|---|---|---|---|---|---|
| Model by $\theta_{ij}$ | # FP | BIC | Est. | *df* | Est. | 90% CI | CFI | Est. | 95% CI |
| No $\theta_{ij}$ | 36 | 7677 | 173.167 | 54 | .087 | [.073, .102] | .934 | .943 | [.934, .952] |
| $\theta_{10,11}$ | 37 | 7642 | 142.985 | 53 | .077 | [.062, .092] | .950 | .940 | [.930, .949] |
| $\theta_{10,11}$ and $\theta_{8,9}$ | 38 | 7627 | 127.657 | 52 | .071 | [.055, .087] | .958 | .936 | [.926, .947] |
| $\theta_{10,11}$, $\theta_{8,9}$, and $\theta_{6,9}$ | 39 | 7615 | 113.309 | 51 | .065 | [.049, .081] | .966 | .939 | [.929, .950] |

*Note*. #FP=the number of free model parameters; Est.=the point estimate; CI=confidence interval; ρ=scale reliability. The model $\chi^2$ for each model had a $p<.001$.

### 7.3. Confirmatory factor analyses with Bayesian estimation

All Bayesian models herein showed clear signs of convergence. For the initial model, all residual correlations were constrained to zero. As expected, this model showed poor fit, with PPP $<.001$, 95% CI [136,199] (see Table 2). The factor loadings were near identical with those from the original MLR model (*Mdn* =.83, *SD*=.12, ranging from .49 to .89). Thus, the results from the initial Bayesian model cohered with the initial frequentist one: the model should be rejected.

For the final set of models, we specified zero-mean, small-variance BSEM priors for residual correlations. Following Asparouhov et al. (2015), we specified priors with several different degrees of freedom values (*d*) in order to gauge the sensitivity of the models to the priors. For simplicity, we report results from models with the *d* value specified at 100, 200, and 300, for which 100 was the least informative and 300 was the most (see Table 2). Based on the PPP and the DIC, all three BSEM models fit the data better than all the MLR models and the Bayesian model with exact-zero residual correlations. The BSEM models for which $d=100$ and 200 clearly passed the PPP test of model fit. The BSEM factor loadings were very similar to those reported, above. The residual correlations with the greatest magnitudes were $\theta_{2,6}$, $\theta_{6,9}$, and $\theta_{8,9}$ for all BSEM models. Yet, even in the model with the weakest priors (i.e., $d=100$), which allowed them to deviate the furthest from zero, their magnitudes were small at −.09, −.09, and .09, respectively.

### 7.4. The influence of BSEM priors on relations among factors

Researchers may wonder how the BSEM priors for residual correlations influence more complex analyses, such as with multiple latent variables. Previous research and theory (Hill et al., 2013; Sandoz et al., 2013) suggest that body-image flexibility should correlate negatively with eating behavior pathology. To examine this, we used the EAT-26 as a general measure of eating behavior pathology. We estimated a general EAT-26 factor with three item parcels (Little, Rhemtulla, Gibson, & Schoemann, 2013), which were constructed by randomly assigning the items into three groups and averaging them within each parcel (see the supplementary material for details). We then estimated the correlation between the general EAT-26 factor and the BI-AAQ factor using each of the eight BI-AAQ measurement models, above, testing the question, *Does estimating residual correlation models for the BI-AAQ yield noteworthy differences in its association with the EAT-26?*

The scale reliability for the parceled EAT-26 factor was adequate: $\rho_{MLR}=.890$, 95% CI [.862, .918]; $\rho_{Bayes}=.888$, 95% PI [.863, .910]. Main results from the eight analyses are presented in Table 3. As anticipated, the correlations between the BI-AAQ and EAT-26 factors were negative and large in magnitude (Cohen, 1992). The trivial differences in the correlations and their 95% intervals among the different models suggested the differing residual correlation specifications had no meaningful impact on the correlation between the factors.

**Table 3**
Correlation between the BI-AAQ and EAT-26 Factors, by model.

| Model | *r* | 95% CI | 95% PI |
|---|---|---|---|
| MLR | | | |
| No $\theta_{ij}$ | −.580 | [−.664, −.497] | |
| $\theta_{10,11}$ | −.576 | [−.660, −.492] | |
| $\theta_{10,11}$ and $\theta_{8,9}$ | −.574 | [−.659, −.489] | |
| $\theta_{10,11}$, $\theta_{8,9}$, and $\theta_{6,9}$ | −.573 | [−.657, −.488] | |
| Bayes | | | |
| No $\theta_{ij}$ | −.577 | | [−.661, −.484] |
| $d=100$ | −.582 | | [−.666, −.489] |
| $d=200$ | −.580 | | [−.654, −.487] |
| $d=300$ | −.579 | | [−.663, −.486] |

## 8. Discussion

In this study, we examined the factor structure of the BI-AAQ with a Hispanic undergraduate sample, which is the first study to explicitly do so to our knowledge. This study is also the first to examine the factor structure of the English version of the BI-AAQ using CFA with any sample. Methodologically, we estimated the factor structure using both frequentist (MLR) and Bayesian (the Gibbs sampler) estimators to compare their relative utilities. The key point of comparison was how they handled residual correlations amongst the BI-AAQ items.

Consistent with previous studies (Sandoz et al., 2013; Ferreira et al., 2011), we found good evidence for a single-factor structure in our Hispanic undergraduate sample. However, both estimators revealed residual correlations were necessary for good model fit. With the frequentist estimator, we sequentially freed three residual correlations, which emphasized a few small-magnitude residual correlations for which we had no clear substantive or methodological interpretations. In contrast, we used the Bayesian

**Table 2**
Fit statistics for the BI-AAQ factor analyses computed with the Bayesian estimator.

| Model by $\theta_{ij}$ | | | | | PPP | | ρ | | $\theta_{ij}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Prior | # FP | *pD* | BIC | DIC | Est. | 95% CI | Est. | 95% PI | *M* | *Mdn* | *SD* |
| No $\theta_{ij}$ | 36 | 35.7 | 7545 | 7678 | <.001 | [136,199] | .944 | [.934, .953] | | | |
| $d=100$ | 102 | 68.3 | 7854 | 7413 | .439 | [−34, 40] | .939 | [.915, .956] | .003 | .002 | .035 |
| $d=200$ | 102 | 58.7 | 7878 | 7418 | .175 | [−21, 55] | .942 | [.925, .955] | .001 | .001 | .028 |
| $d=300$ | 102 | 52.8 | 7899 | 7426 | .050 | [−7, 71] | .943 | [.928, .955] | <.001 | <.001 | .023 |

*Note*. *d*=the degrees of freedom parameter, which largely dictates how informative the BSEM residual correlation priors are; *pD*=the estimated number of parameters; PPP=posterior predictive *p*-value; PI=probability interval.

estimator and Muthén and Asparouhov's BSEM method to model the residual correlations as approximate zeros. The BSEM models had good fit and factor loadings that were comparable to those of the other models. Though the residual correlations in the BSEM models were statistically and theoretically trivial in magnitude, they were large enough to degrade model fit when constrained to exactly zero. Furthermore, when we estimated the correlation between the BI-AAQ and measure of disordered eating behavior using, the EAT-26, we found the different methods of handling residual correlations resulted in trivial differences. We conclude that the BSEM models for the BI-AAQ are good representations of data. The BI-AAQ had good factorial validity and showed preliminary evidence for construct validity as a measure of body-image flexibility for our participants.

A strength of SEM is that is allows researchers to account for sources of measurement error in their statistical models. For example, researchers have proposed models for accounting for error associated with wording effects (e.g., Weijters, Baumgartner, & Schillewaert, 2013) and hierarchical structures (e.g., De Jong, Steenkamp, & Fox, 2007). Because the residual correlations in the BSEM models, herein, hovered closely around zero, we concluded the basic single-factor model was fine for our data, but that the exact-zero assumption was overly strict. However, for instances wherein one or more BSEM residual correlations deviate substantially from zero, researchers should consider major model revisions, such as including additional factors. See Asparouhov et al. (2015) for extensive discussion on the topic.

### *8.1. Precision and scope of the BSEM method*

CBS measure development benefits from statistical models that are specific enough to faithfully summarize specific datasets, but general enough that scientists might apply those models to other datasets. Hayes et al. (2012) summarize these analytic qualities as precision and scope. In their words: "The criterion of precision means that only a limited number of analytic concepts apply to a given case; scope means a given analytic concept applies to a range of cases" (p. 2).

Some might criticize the BSEM method for yielding unnecessarily complex models, trading precision for model fit (see Stromeyer, Miller, Sriramachandramurthy, & DeMartino, 2015; nb. the rebuttal by Asparouhov et al., 2015). That is, our most complex frequentist CFA model for the BI-AAQ included 39 free parameters. The better-fitting BSEM models, in contrast, contained 102 free model parameters, the difference being 63 additional residual correlations. Yet, the BSEM models were interpretively simple: all 66 residual correlations may be summarily described as approximately zero and interpretively trivial. In this way, the BSEM models were reasonably precise and yet flexible enough that they may generalize to other samples—reasonable precision and potential scope. Contrast this with the three small residual correlations in the final frequentist model. Methodologists have long warned applied researchers against specifying residual correlations for which they have no clear theoretical justifications (e.g., MacCallum, 2003). Doing so could result in complex models that fit well in one sample but fail to replicate, failing both precision and scope. We question how likely the residual correlations for the three item pairs indicated by the frequentist estimator would generalize to other samples.

The beauty of the BSEM approximate zeros is that they may well replicate. The probability of drawing a value within a range is greater than the probability of drawing a specific value. The BSEM hypothesis of an approximate zero corresponds to a range of values and the traditional null hypothesis of an exact zero corresponds to a specific value. Thus, we expect approximate BSEM zero residual correlations are more likely to replicate in another sample than traditional exact-zero residual correlations. Furthermore, their trivial magnitudes do not draw attention from the most important part of the model, a theoretically-coherent single-factor structure composed of high factor loadings.

### *8.2. Strengths, limitations, and future directions*

Building on the work of Verhagen and Fox (2013), who used Bayesian estimation to account for measurement invariance across time in a latent growth curve model, this study further highlights the utility of Bayesian estimation for CBS analyses. Our study also represents an early comparison of frequentist and Bayesian factor analysis using the BSEM method. This study also focused on Hispanic individuals, an important and growing ethnic minority within the US that has been underrepresented in the literature. However, this study was limited in that the sample consisted of undergraduates at one US university near the Texas/Mexico border. Hispanics are diverse with respect to race, culture, ethnicity, traditions, SES, immigration history, and country of origin (Santiago-Rivera, Arredondo, & Gallardo-Cooper, 2002). This topic would benefit from future research involving more encompassing Hispanic samples or, alternatively, specific subgroups.

Interest in Bayesian estimation is spreading rapidly throughout the social sciences (Zyphur & Oswald, 2015). Herein, we presented a specific application of Bayesian statistics to factor analysis. The potentials are broader. Researchers have applied Bayesian statistics to many other areas related to factor analysis, such as item response theory (Fox, 2010) and multilevel SEM (Song & Lee, 2012). Bayesian estimation may also be applied to staple procedures, such as ANOVA and regression (Kruschke, 2015; McElreath, 2016). More generally, because Bayesian statistics are not tied to large sample theory, they may be particularly helpful for small sample sizes, especially when with informative priors (e.g., Zyphur & Oswald, 2015). And perhaps most importantly for science, Bayesian statistics offer a viable alternative to null-hypothesis significance testing. Bayesian estimates allow researchers do directly test research hypotheses, rather than settling for testing the null (Kline, 2013; Kruschke, 2011; Kruschke & Liddell, 2015; Zyphur & Oswald, 2015).

In conclusion, Sandoz and colleagues' BI-AAQ showed good factorial validity and scale reliability for our all-Hispanic sample, and also showed preliminary construct validity with a large negative association with eating behavior pathology. Thus, the BI-AAQ may be a useful measure for assessing body image flexibility and its role in disordered eating in Hispanic-American individuals. From a methodological perspective, Bayesian estimation using the BSEM method provided helpful insights into the BI-AAQ's factor structure and allowed for a more flexible measurement model than with frequentist methods. We recommend researchers examine the utility of the BSEM method for other questionnaires and consider using Bayesian estimation, more broadly.

## References

Alegria, M., Woo, M., Cao, Z., Torres, M., Meng, X. L., & Striegel-Moore, R. (2007). Prevalence and correlates of eating disorders in Latinos in the United States. *International Journal of Eating Disorders*, *4*, S15–S21. http://dx.doi.org/10.1002/eat.20406.

Anderson, D. A., Lundgren, J. D., Shapiro, J. R., & Paulosky, C. A. (2004). Assessment of eating disorders: review and recommendations for clinical use. *Behavior Modification*, *28*, 763–782. http://dx.doi.org/10.1177/0145445503259851.

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: comments on Stromeyer et al. *Journal of Management*, *41*, 1561–1577. http://dx.doi.org/10.1177/0149206315591075.

Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M.A. and F.R.S. *Philosophical Transactions of the Royal Society of*

*London, 53*, pp. 370–418. ⟨http://dx.doi.org/10.1098/rstl.1763.0053⟩.
Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., & Zettle, R. D. (2011). Preliminary psychometric properties of the acceptance and action questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*, 676–688. http://dx.doi.org/10.1016/j.beth.2011.03.007.
Box, G. E. (1979). Some problems of statistics in everyday life. *Journal of the American Statistical Association*, *74*, 1–4. http://dx.doi.org/10.1080/01621459.1979.10481600.
Brannan, M. E., & Petrie, T. A. (2008). Moderators of the body dissatisfaction-eating disorder symptomatology relationship: replication and extension. *Journal of Counseling Psychology*, *55*, 263–275. http://dx.doi.org/10.1037/0022-0167.55.2.263.
Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY, USA: Guilford Press.
Butryn, M. L., Juarascio, A., Shaw, J., Kerrigan, S. G., Clark, V., O'Planick, A., & Forman, E. M. (2013). Mindfulness and its relationship with eating disorders symptomatology in women receiving residential treatment. *Eating Behaviors*, *14*, 13–16. http://dx.doi.org/10.1016/j.eatbeh.2012.10.005.
Ciarrochi, J., Zettle, R.D., Brockman, R., Duguid, J., Parker, P., Sahdra, B., & Kashdan, T. B. (2016). Measures that make a difference: A functional contextualistic approach to optimizing psychological measurement in clinical research practice. In R.D. Zettle, S.C. Hayes, D. Barnes–Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 320–346). Malden, MA, US: Wiley Blackwell.
Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155.
Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381–398. http://dx.doi.org/10.1037/1082-989X.12.4.381.
Corning, A. F., Krumm, A. J., & Smitham, L. A. (2006). Differential social comparison processes in women with and without eating disorder symptoms. *Journal of Counseling Psychology*, *53*, 338–349. http://dx.doi.org/10.1037/0022-0167.53.3.338.
Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, *25*, 7–29. http://dx.doi.org/10.1177/0956797613504966.
De Jong, M. G., Steenkamp, J. M., & Fox, J. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, *34*, 260–278. http://dx.doi.org/10.1086/518532.
Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290. http://dx.doi.org/10.1177/1745691611406920.
Enders, C. K. (2010). *Applied missing data analysis*. New York, NY USA: Guilford Press.
Falkenström, F., Hatcher, R. L., & Holmqvist, R. (2014). Confirmatory factor analysis of the patient version of the working alliance inventory-short form revised. *Assessment Advance Online Publication*. , http://dx.doi.org/10.1177/1073191114552472.
Fienberg, S. E. (2006). When did Bayesian inference become "Bayesian". *Bayesian Analysis*, *1*, 1–40. http://dx.doi.org/10.1214/06-BA101.
Ferreira, C., Pinto-Gouveia, J., & Duarte, C. (2011). The validation of the body image acceptance and action questionnaire: ëxploring the moderator effect of acceptance on disordered eating. *International Journal of Psychology and Psychological Therapy*, *11*, 327–345. http://www.ijpsy.com/volumen11/num3/302.html.
Fox, J.-P. (2010). *Bayesian item response modeling: theory and applications*. New York: Springer.
Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitudes test: psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871–878. http://dx.doi.org/10.1017/S0033291700049163.
Gelman, A., Meng, X. L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*, 733–807. http://www.stat.columbia.edu/~gelman/research/published/A6n41.pdf.
Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472. http://dx.doi.org/10.1214/ss/1177011136.
Grabe, S., & Hyde, J. S. (2006). Ethnicity and body dissatisfaction among women in the united states: a meta-analysis. *Psychological Bulletin*, *132*, 622–640. http://dx.doi.org/10.1037/0033-2909.132.4.622.
Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavior science: creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science*, *1*, 1–16. http://dx.doi.org/10.1016/j.jcbs.2012.09.004.
Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy*, *44*, 180–189. http://dx.doi.org/10.1016/j.beth.2009.08.002.
Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behavior Research and Therapy*, *44*, 1–25. http://dx.doi.org/10.1016/j.brat.2005.06.006.
Hayes, S. C., Muto, T., & Masuda, A. (2011). Seeking cultural competence from the ground up. *Clinical Psychology: Science & Practice*, *18*, 232–237. http://dx.doi.org/10.1111/j.1468-2850.2011.01254.x.
Hill, M. L., Masuda, A., & Latzman, R. D. (2013). Body image flexibility as a protective factor against disordered eating behavior for women with lower body mass index. *Eating Behaviors*, *14*, 336–341. http://dx.doi.org/10.1016/j.eatbeh.2013.06.003.
Kass, R. E. (2011). Statistical inference: the big picture. *Statistical Science*, *26*, 1–9. http://dx.doi.org/10.1214/10-STS337.
Kiezebrink, K., Campbell, D., Mann, E., & Blundell, J. (2009). Similarities and differences between excessive exercising anorexia nervosa patients compared with DSM-IV defined anorexia nervosa subtypes. *Eating and Weight Disorders*, *14*, e199–e204. http://dx.doi.org/10.1007/BF03325117.
Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY, USA: Guilford Press.
Kline, R. B. (2013). *Beyond significance testing: statistics reform in the behavioral sciences* (2nd ed.). Washington, DC, US: American Psychological Association.
Kruschke, J. K. (2011aa). Introduction to special section on Bayesian data analysis. *Perspectives on Psychological Science*, *6*, 272–273. http://dx.doi.org/10.1177/1745691611406926.
Kruschke, J. K. (2011bb). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312. http://dx.doi.org/10.1177/1745691611406925.
Kruschke, J. K. (2015a). *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan* (second ed.). Burlington, MA, US: Academic Press/Elsevier.
Kruschke, J. K., Liddell, T. M. (2015). The Bayesian new statistics: two historical trends converge. Retrieved from: ⟨http://dx.doi.org/10.2139/ssrn.2606016⟩.
Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*, 285–300. http://dx.doi.org/10.1037/a0033266.
Lord, F., Novick, M., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley.
MacCallum, R. C. (2003). 2001 presidential address: working with imperfect models. *Multivariate Behavioral Research*, *38*, 113–139. http://dx.doi.org/10.1207/S15327906MBR3801_5.
Maïano, C., Morin, A. S., Lanfranchi, M., & Therme, P. (2013). The eating attitudes Test-26 revisited using exploratory structural equation modeling. *Journal of Abnormal Child Psychology*, *41*, 775–788. http://dx.doi.org/10.1007/s10802-013-9718-z.
McElreath, R. (2016). *Statistical rethinking: a Bayesian course with examples in R and Stan*. Boca Raton, FL: Chapman & Hall/CRC Press.
Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437–455. http://dx.doi.org/10.1037/e518362013-127.
Mintz, L. B., & O'Halloran, M. S. (2000). The eating attitudes test: validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment*, *74*, 489–503. http://dx.doi.org/10.1207/S15327752JPA7403_11.
Mulder, J., & Wagenmakers, E.–J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: practical relevance and new developments". *Journal of Mathematical Psychology*, *1–5*, 72. http://dx.doi.org/10.1016/j.jmp.2016.01.002.
Muthén, B., & Asparouhov, T. (2012aa). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. http://dx.doi.org/10.1037/a0026802.
Muthén, B., & Asparouhov, T. (2012bb). Rejoinder to MacCallum, Edwards, and Cai (2012) and Rindskopf (2012): mastering a new method. *Psychological Methods*, *17*, 346–353. http://dx.doi.org/10.1037/a0029214.
Orbitello, B., Ciano, R., Corsaro, M., Rocco, P. L., Taboga, C., Tonutti, L., & Balestrieri, M. (2006). The EAT-26 as screening instrument for clinical nutrition unit attenders. *International Journal of Obesity*, *30*, 977–981. http://dx.doi.org/10.1038/sj.ijo.0803238.
Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*, 315–323. http://dx.doi.org/10.1348/000711001159582.
Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development*, *42*, 223–232. http://dx.doi.org/10.1177/0748175609344096.
Ricciardelli, L. A., McCabe, M. P., Williams, R. J., & Thompson, J. K. (2007). The role of ethnicity and culture in body image and disordered eating among males. *Clinical Psychology Review*, *27*, 582–606. http://dx.doi.org/10.1016/j.cpr.2007.01.016.
Sandoz, E. K., Wilson, K. G., Merwin, R. M., & Kellum, K. K. (2013). Assessment of body image flexibility: the body image-acceptance and action questionnaire. *Journal of Contextual Behavioral Science*, *2*, 39–48. http://dx.doi.org/10.1016/j.jcbs.2013.03.002.
Santiago-Rivera, A. L., Arredondo, P., & Gallardo-Cooper, M. (2002). *Counseling Latinos and la familia: a practical guide*. Thousand Oaks, CA: Sage.
Song, X.-Y., & Lee, S.-Y. (2012). *Basic and advanced Bayesian structural equation modeling: with applications in the medical and behavioral sciences*. New York, NY: John Wiley & Sons.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, *64*, 583–639. http://dx.doi.org/10.1111/1467–9868.00353.
Stice, E., & Shaw, H. E. (2002). Role of body dissatisfaction in the onset and maintenance of eating pathology: a synthesis of research findings. *Journal of Psychosomatic Research*, *53*, 985–993. http://dx.doi.org/10.1016/S0022-3999(02)00488-9.
Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The prowess and pitfalls of Bayesian structural equation modeling: important considerations for management research. *Journal of Management*, *41*, 491–520.

http://dx.doi.org/10.1177/0149206314551962.

Timko, C. A., Juarascio, A. S., Martin, L. M., Faherty, A., & Kalodner, C. (2014). Body image avoidance: an under-explored yet important factor in the relationship between body image dissatisfaction and disordered eating. *Journal of Contextual Behavioral Science*. , http://dx.doi.org/10.1016/j.jcbs.2014.01.002.

U.S. Census Bureau (July 2014), State and County QuickFacts. Retrieved from: 〈http://quickfacts.census.gov/qfd/states/00000.html〉.

Wendell, J. W., Masuda, A., & Le, J. K. (2012). The role of body image flexibility in the relationship between disordered eating cognitions and disordered eating symptoms among non-clinical college students. *Eating Behaviors*, *13*, 240–245. http://dx.doi.org/10.1016/j.eatbeh.2012.03.006.

Wilson, K. G., Whiteman, K., & Bordieri, M. J. (2013). The pragmatic truth criterion and values in contextual behavioral science In: S. Dymond, & B. Roche (Eds.), *Advances in relational frame theory & contextual behavioral science: research & application*. Oakland, CA: New Harbinger.

van de Vijver, F. J. R. (2011). Capturing bias in structural equation modeling In: P. S. E. Davidov, & J. Billiet (Eds.), *Cross-cultural analysis: methods and applications* (pp. 3–34). New York, NY, USA: Routledge.

Verhagen, J., & Fox, J.-P. (2013). Longitudinal measurement in health-related surveys. A Bayesian joint growth model for multivariate ordinal responses. *Statistics in Medicine*, *32*, 2988–3005. http://dx.doi.org/10.1002/sim.5692.

Vilardaga, R., Hayes, S. C., Levin, M. E., & Muto, T. (2009). Creating a strategy for progress: a contextual behavioral science approach. *The Behavior Analyst*, *32*, 105–133. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686981/pdf/bhan-32-01-105.pdf.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228–243. http://dx.doi.org/10.1037/a0027127.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: an integrative model. *Psychological Methods*, *18*, 320–334. http://dx.doi.org/10.1037/a0032121.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data In: M. E. Sobel, & M. P. Becker (Eds.), *Sociological methodology* (pp. 165–200). Washington, D.C.: ASA.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*, 301–322. http://dx.doi.org/10.1037/a0016972.

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: a user's guide. *Journal of Management*, *41*, 390–420. http://dx.doi.org/10.1177/0149206313501200.

Zyphur, M. J., Oswald, F. L., & Rupp, D. E. (2015). Rendezvous overdue: Bayes analysis meets organizational research. *Journal of Management*, *41*, 387–389. http://dx.doi.org/10.1177/01492063145492.