Title: Internal Structure and Reliability of the Internalized Stigma of Mental Illness Scale (ISMI-29) and Brief Versions (ISMI-10, ISMI-9) among Americans with Depression

Key words: Internalized stigma of mental illness; Self-stigma; Bifactor analysis; Model-based reliability; scale development

Hammer, J. H., & Toland, M. D. (2017). Internal structure and reliability of the Internalized Stigma of Mental Illness Scale (ISMI-29) and brief versions (ISMI-10, ISMI-9) among Americans with depression. *Stigma and Health, 2*, 159-174. doi: 10.1037/sah0000049

Joseph H. Hammer
Department of Educational, School, and Counseling Psychology
University of Kentucky
243 Dickey Hall
Lexington, KY 40506
jhhammer@gmail.com

Michael D. Toland, Ph.D.
251C Dickey Hall
University of Kentucky
Lexington, KY 40506
(858) 257-3395
Toland.md@uky.edu

Abstract

The Internalized Stigma of Mental Illness (ISMI) Scale is designed to measure self-stigma among persons with psychiatric disorders. Research provides varying degrees of evidence supporting the reliability and validity of the original English language version of the ISMI-29 as well as an array of translated and/or modified versions for particular disorders or populations used with samples from around the world. The present paper was the first to use confirmatory factor analysis to investigate the internal structure of the English ISMI-29, ISMI-24 (a version that excludes the Stigma Resistance subscale), and ISMI-10 short-form (ISMI-10) using a sample of 758 community-dwelling adults who identified as depressed. Across all three versions, results of bifactor analyses and model-based reliability estimates provided evidence for the calculation and interpretation of the total score as a measure of the general internalized stigma of mental illness construct, but did not support the use of individual subscales. Results also supported the inclusion of the Stigma Resistance subscale items when calculating the ISMI total scores. Also, a new nine-item unidimensional short form of the ISMI (ISMI-9) was developed and evaluated. Results suggested that the ISMI-9 may offer certain psychometric advantages over the ISMI-10. Researchers are strongly encouraged to conduct further internal structure analyses of the ISMI and its various alternate versions to determine the idiosyncrasy versus generalizability of the present findings.

Keywords: Internalized stigma of mental illness; Self-stigma; Bifactor analysis; Model-based reliability; scale development

Internal Structure and Reliability of the Internalized Stigma of Mental Illness Scale (ISMI-29) and Brief Versions (ISMI-10, ISMI-9) among Americans with Depression

The stigma of mental illness is the prejudice and discrimination that results from endorsing negative stereotypes about people with mental illness (Corrigan & Watson, 2002). Stigma has been called the major obstacle to healthcare in America (President's New Freedom Commission on Mental Health, 2003; U.S. Department of Health and Human Services, 1999). Internalized stigma of mental illness is the harmful psychological impact that results from internalizing this prejudice and directing it toward oneself. Internalized stigma of mental illness has been empirically linked with reduced help seeking, lower treatment adherence, and impaired self-efficacy and self-esteem, among other detrimental outcomes (Corrigan, 2004; Corrigan, Watson, & Barr, 2006; Link, 1987; Link, Struening, Neese-Todd, Asmussen, & Phelan, 2001). If mental health professionals wish to systematically study the nature, causes, and consequences of internalized stigma of mental illness and verify the efficacy of stigma reduction interventions, instruments that measure this form of stigma in a valid and reliable way are necessary. One of the most widely used measures of internalized stigma is the Internalized Stigma of Mental Illness Scale (ISMI; Ritsher, Otilingam, & Grajales, 2003). The ISMI-29 is a 29-item self-report instrument with each item rated on a 1 (*strongly disagree*) to 4 (*strongly agree*) Likert-type scale. Higher scores reflect higher levels of reported internalized stigma of mental illness. The items assume that respondents self-identify as having a mental illness (e.g., "Because I have a mental illness, I need others to make most decisions for me") and thus are most appropriately used with clinical populations.

Published manuscripts provide varying degrees of evidence supporting the reliability and validity of the original English language version of the ISMI-29 as well as an array of translated and/or modified versions for particular disorders or populations used with samples from around the world (for a review, see Boyd, Adler, Otilingam, & Peters, 2014a). In their multinational review, Boyd and colleagues reported internal consistency of reliability coefficient (α) estimates for the ISMI-29 total score ranging from .80 to .92 (Boyd et al., 2014a). They also noted the internal consistency estimates for the five subscales scores ranged from .47 (Stigma Resistance) to .91 (Alienation), with all subscales but Stigma Resistance typically demonstrating estimates above .75. Test-retest reliability coefficient estimates for the ISMI-29 total score were also reported to range from .90 to .92. In addition, the review suggested that the ISMI-29 demonstrated acceptable convergent validity correlations with external criterion variables (e.g., self-esteem, depression).

In regard to the ISMI-29's internal structure, the Boyd et al. (2014a) review cites the foundational paper describing the development and exploratory factor analysis (EFA) of the ISMI-29 (Ritsher et al., 2003). However, no published studies have subjected the original English version of the ISMI-29 to confirmatory factor analysis (CFA). It is important to use an independent sample to confirm an instrument's internal structure (Worthington & Whittaker, 2006) because it is possible for an initial dataset to be idiosyncratic, leading to the identification of an initial factor structure that does not generalize well to subsequent samples (i.e., a lack of structural generalizability). Therefore, a CFA on this version of the ISMI-29 is an important next step. The present paper is the first to document the results of such a CFA to examine the internal structure of the ISMI-29.

Furthermore, despite the documented presence of strongly correlated subscale scores (Chang, Wu, Chen, Wang, & Lin, 2014), the ISMI-29 has never been subjected to a bifactor CFA. A bifactor model would specify a single general (primary) "internalized stigma of mental illness" factor reflecting the common variance of all items as well as five narrower group (specific) factors reflecting the variance of their assigned items. Thus, each item simultaneously loads on the general factor and its assigned group factor. These six factors are all first-order factors set orthogonal to each other. The general factor reflects the primary factor independent of the group factors, whereas each specific factor reflects the unique factor of interest after controlling (partialling out) the general factor. A bifactor CFA and follow-up model-based reliability estimates allow researchers to answer key questions about multidimensional instruments (Reise, Bonifay, & Haviland, 2013). For instance, is it appropriate to calculate and interpret the total score of the overall instrument? And, is it reasonable to calculate and interpret subscale scores for each of the purported factors?

The ISMI-29 total score and subscale scores are currently being used to investigate important research questions and make clinical decisions across the United States and beyond (e.g., Boyd et al, 2014a; Boyd, Otilingam, & DeForge, 2014b). Thus, it is crucial to use these best practice techniques to confirm that the total and subscale scores are appropriate measures of internalized stigma of mental illness and its five purported subdimensions (subscores), respectively. According to the *Standards for Educational and Psychological Testing,* it is appropriate to interpret or report subscores when meaningful rationales and evidence are provided for such purposes (*The Standards*; Standard 1.14; American Education Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 27). Moreover, *The Standards* indicate that when more than one score is provided by an instrument, "the distinctiveness and reliability of those scores should be demonstrated" (AERA et al., 2014, p. 27). This is the second important contribution of the present paper.

Boyd et al. (2014b) recently developed a 10-item version of the ISMI (ISMI-10), which is designed to be scored as a unidimensional measure. However, the internal structure of the ISMI-10 (short form) has not yet been examined. The present study addressed this third gap in our psychometric understanding of the ISMI. The following literature review will distill what is known about the development and psychometric properties of the ISMI, thereby providing context for this study's hypotheses and results.

Development and Factor Analyses of the English ISMI-29

The ISMI developers (Ritsher et al., 2003) generated an initial pool of 55 items, eliminated items with item-total correlations below .40, and grouped the remaining 29 items into five subscales based on the perceived themes embedded in the items: Alienation (e.g. "Having a mental illness has spoiled my life"), Stereotype Endorsement (e.g., "Mentally ill people tend to be violent"), Discrimination Experience (e.g., "People discriminate against me because I have a mental illness"), Social Withdrawal (e.g., " I don't talk about myself as much because I don't want to burden others with my mental illness"), and Stigma Resistance (e.g., "I can have a good, fulfilling life, despite my mental illness"). Stigma Resistance items are reverse-scored prior to analysis given their positive valence.

Next, the researchers then temporarily dropped the five Stigma Resistance items and used maximum likelihood EFA with varimax rotation to extract four factors from the remaining 24 items (subsequently referred to as the ISMI-24 in this paper). Thirteen of the 24 items sorted onto the expected factors (which were based on perceived themes across items by Ritsher and colleagues [2003]). In the discussion section, Ritsher and colleagues (2003) concluded that "four of the five Stigma Resistance items were poorly associated with the internalized stigma construct" (p. 46) and "we found partial support for the validity of the four other subscales as distinct facets of the internalized stigma construct… until these analyses can be replicated in a different sample, it is most parsimonious to conceptualize the ISMI as measuring a single construct" (p. 47). In summary, the original EFA of the ISMI did not provide definitive support for the purported five-factor structure of the ISMI-29 (p. 112; Stevelink, Wu, Voorend, & Brakel, 2012).

No other studies have used EFA or CFA on the English language version of the ISMI-29. However, translated versions of the ISMI-29 have been subjected to EFA and CFA. The results of such analyses on these alternative forms of the ISMI-29 cannot be solely relied on to provide evidence for the internal structure of the English version of the ISMI-29 (given that the cross-language measurement invariance of the ISMI-29 has not yet been demonstrated). However, analyses with these alternative forms can provide a glimpse into what we might expect from a CFA of the English version of the ISMI-29.

Factor Analyses of Translated Versions of the ISMI-29

The German version of the ISMI-29 was subjected to principal component analysis (PCA) with varimax (i.e., orthogonal) rotation (Sibitz, Unger, Woppmann, Zidek, & Amering, 2011). Sibitz et al. (2011) retained a 2-component solution, with all Stigma Resistance items loading primarily on the second component and the remaining 24 items loading primarily on the first component. Although scale development literature (Worthington & Whittaker, 2006) recommends against the use of PCA for most scale development applications and discourages the sole use of orthogonal rotations when the factors show intercorrelations (oblique rotations like direct oblimin should be used in such cases), these results suggested that the Stigma Resistance items function differently than the remaining 24 items. This can have important implications for whether or not it is empirically justifiable to include the five Stigma Resistance items when calculating a total instrument-wide score for the ISMI, something acknowledged by the ISMI developers (Ritsher et al., 2003). It is important to mention that the positive valence of the Stigma Resistance items (in contrast to the negative valence of the remaining items) may or may not be a key reason that the Stigma Resistance items function in a different manner.

A PCA with varimax rotation of the Amharic (official language of Ethiopia) translation of the ISMI-24 (Stigma Resistance items not included) was interpreted as reflecting the existence of the four remaining themes described by the ISMI developers (Assefa, Shibre, Asher, & Fekadu, 2012). Ociskova and colleagues (2014) conducted a maximum likelihood EFA with varimax rotation with Kaiser normalization on a Czech translation of the ISMI-29. They reported that four factors were identified and explained about 50% of the item variance, noting that the Alienation and Social Withdrawal items loaded on the same combined factor. They also indicated that 22 of the 29 items loaded on the intended factor, though the cross-loadings of these items were not described and thus cannot be judged according to standard item retention criteria.

A modified Taiwan version of the ISMI-29 was subjected to a series of CFAs (Chang et al., 2014). The authors specified a five-factor oblique first-order model (all 29 items specified to load on their intended factors and factors allowed to freely correlate with each other) as well as a second-order model (five first-order factors also specified to load on a single second-order factor thought to represent the overall construct). Both models demonstrated acceptable fit to the data, which the researchers interpreted as providing evidence of internal structure for the modified Taiwan version of the ISMI-29. There are two aspects of the findings by Chang et al. (2014) that require further explication.

First, examining the second-order model in Chang et al. (2014), the reader can see that the Stigma Resistance subscale does not significantly load on the second-order construct. These results would argue that Stigma Resistance did not function like a subdimension of the larger internalized stigma of mental illness construct; it functioned like an independent construct. If a parallel finding was found for the English version of the ISMI-29 analyzed in the present study, then this would indicate a lack of clear evidence for the use the five Stigma Resistance items when calculating the overall English ISMI-29 total score. This outcome is not unlikely given that "four of the five Stigma Resistance items were poorly associated with the internalized stigma construct" (p. 47; Ritsher et al., 2003) in the original scale development study. Importantly, this does not necessarily preclude calculating and interpreting the Stigma Resistance score as a separate score.

The second thing worth noting in Chang et al. (2014) is that the four other subscales loaded > .95 onto the second-order factor. This suggests that the four subscales' factorial independence may be at risk. In other words, we may anticipate that the present study's factor analyses of the English ISMI would reveal that these 24 items are best conceptualized as loading on a single general factor rather than four unique factors. Interestingly, this would align with the scale developer's prescient statement that "until these analyses can be replicated in a different sample, it is most parsimonious to conceptualize the ISMI as measuring a single construct" (Boyd et al., 2003, p. 47). This literature review now turns to the development of the ISMI-10 short-form (ISMI-10).

The ISMI-10

Boyd et al. (2014b) developed the 10-item version of the ISMI (ISMI-10) by re-analyzing the same dataset used to develop the ISMI-29 ($N = 127$) and cross-checking the results with a separate mental health outpatient sample ($N = 760$). The ISMI-10 contains two items from each of the five original subscales in an effort to preserve test content (i.e., coverage of the full content domain of the construct). In an effort to select the two strongest items from each subscale, the authors chose items that correlated most strongly with external criterion (e.g., public stigma) and the relevant subscale score. They also took the author-rated face-validity of each item into account. The ISMI-10 total score ($M = 2.32$, $SD = .41$, $\alpha = .75$) was found to correlate .94 with ISMI-29 total score ($M = 2.32$, $SD = .39$, $\alpha = .90$) in the validation sample; similar results arose in a larger cross-validation sample. The ISMI-10 was also found to correlate with scores from external criterion instruments to a similar degree as the ISMI-29.

Importantly, the authors stated that it was not their intention to create five two-item subscales and recommend using the total ISMI-10 score rather than dividing it into the five subscales. However, whenever any new instrument is created, it is important to factor analyze the instrument to empirically determine its internal structure. Factor analytic results that either show evidence for a unidimensional or bifactor solution (provided model-based reliability estimates support the existence of a strong, reliable general factor; Reise et al., 2013) are a prerequisite to the calculation and valid interpretation of an instrument's total score.

Therefore, the third contribution of the present study was to examine the internal structure of the ISMI-10. No known factor analyses of the ISMI-10 have been published, so this effort represents an important next step in the validation process for using ISMI-10 scores. Furthermore, the ISMI-10 developers state that a limitation of their study was that both datasets were primarily composed of male military veterans receiving VA services. Thus, the ability to examine the ISMI-10 in the context of a community-based, non-military population with a significant percentage of female respondents is another strength of the present investigation.

The Present Study

This study had four goals and proposed five hypotheses. First, the present study used CFA to verify what internal structure was best supported by the data. Given past research suggesting that the 5-item Stigma Resistance subscale may be best treated as a separate scale (e.g., Chang et al., 2014; Sibitz et al., 2011), rather than a subdimension of the larger internalized stigma of mental illness construct, analyses of both the ISMI-29 and ISMI-24 were conducted. Given the strong correlations between the four other subscales observed in published literature (e.g., Lysaker, Roe, & Yanos, 2007), it was hypothesized (Hypothesis 1) that a bifactor structure would best account for the item covariation for both the ISMI-29 and ISMI-24. It was further hypothesized (Hypothesis 2) that Explained Common Variance estimates would suggest that the ISMI-29 and ISMI-24 should be viewed as primarily unidimensional.

Second, assuming the ISMI-29 conforms to a bifactor structure, model-based reliability estimates would then be used to determine whether or not it is advisable to calculate and interpret the ISMI total score and/or the subscale scores for each of the five factors. Past research has often found that well-respected multidimensional instruments with significant interfactor correlations should only be scored as a univocal measure and their subscales abandoned (Reise et al., 2013). Therefore, it was hypothesized that (Hypothesis 3) model-based reliability estimates would support the use of the ISMI-29 (and ISMI-24) total scores to represent the general internalized stigma of mental illness construct. In addition, it was hypothesized that (Hypothesis 4) model-based reliability estimates would not support the use of any of the subscale scores to represent the narrower purported subdomains (e.g., Alienation) of internalized stigma of mental illness.

Third, the factor structure of the ISMI-10 was examined via CFA. Given the likely presence of a strong general factor that runs throughout the 29 items of the ISMI, it was hypothesized (Hypothesis 5) that a unidimensional factor structure would best account for the item covariation of the ISMI-10. Fourth, the present study sought to use the bifactor solution as a means to develop and evaluate an alternative unidimensional short form of the ISMI, in the event that the ISMI-10 fails to demonstrate a unidimensional factor structure.

Method

Participants were 758 (645 women, 107 men, 6 other gender identity) community-dwelling adults living in the United States who identified as (1) having a mental illness and (2) dealing with some form of depression. Recruitment for the study was done via ResearchMatch, a national health volunteer registry that was created by several academic institutions and supported by the U.S. National Institutes of Health as part of the Clinical Translational Science Award (CTSA) program. ResearchMatch has a large population of volunteers who have consented to be contacted by researchers about health studies for which they may be eligible. Registry participants were contacted via the registry messaging system regarding the study, which was advertised as a mental illness attitudes questionnaire. Interested participants were then directed to an online survey that began with an informed consent page, followed by the ISMI-29 and demographic items, and ended with a debriefing page. Of note, extant research suggests that survey data derived from online measures are consistent with results from paper and pencil measures (Gosling, Vazire, Srivastava, & John, 2004).

Participants ranged in age from 18 to 83 years old ($M = 41.52$, $SD = 14.11$, $Mdn = 40$). Approximately 86% of the sample identified as White, 3% as Latino/a, 3% as African American/Black, 3% multiracial, 2% other race/ethnicity, 2% Asian American or Pacific

Islander, 1% American Indian/Native American, and 1% preferred not to answer. Approximately 50% reported being married or in a committed relationship or civil union, 31% single, 17% separated or divorced, 2% widowed, and 1% preferred not to answer. Approximately 1% reported having less than a high school education, 4% earned a high school diploma or GED, 9% earned a two-year degree, 19% had some college experience, 36% earned a four-year college degree, 32% earned a graduate or professional degree, and 1% preferred not to answer. Regarding U.S. state residence, approximately 20% reported living in East North Central, 19% in South Atlantic, 12% in Pacific, 12% in East South Central, 10% in Middle Atlantic, 10% in West North Central, 7% in Mountain, 6% in West South Central, and 3% in New England.

Results

Data Cleaning

The initial dataset contained 1,011 individuals. Cases with any missing data (n = 79), or wrong responses to attention check items (n = 43) were deleted. Because the ISMI items assume that respondents self-identify as having a mental illness, participants who indicated that they do not self-identify as having a mental illness (n = 130) were also deleted, resulting in a final sample of N = 758. No variables exceeded the cutoffs of 3 and 10 for high univariate skewness and kurtosis values, respectively (Weston & Gore, 2006). In addition, we used the MLR estimator to estimate the Model $\chi 2$ and associated fit indices that use it to protect against deviations from multivariate normality.

Classic Descriptive Analyses

According to the "4-category method" of ISMI-29 scoring (Lysaker et al., 2007), the mean score on the ISMI-29 of 2.02 (SD = .48) and on the ISMI-10 of 1.93 (SD = .50) indicates the average respondent in the present sample straddles the cutoff between “minimal to no internalized stigma” and “mild internalized stigma”. Specifically, about 50% of the sample had “minimal to no internalized stigma,” 35% had “mild internalized stigma,” 14% had “moderate internalized stigma,” and 2% had “severe internalized stigma.” The internal consistency estimates (α) were as follows: ISMI-29 = .93, 95% CI [.92, .94], ISMI-24 = .93, 95% CI [.93, .94], ISMI-10 = .82, 95% CI [.80, .84], Alienation = .84, 95% CI [.82, .86], Stereotype Endorsement = .75, 95% CI [.73, .78], Discrimination Experience = .82, 95% CI [.81, .85], Social Withdrawal = .85, 95% CI [.83, .86], and Stigma Resistance = .53, 95% CI [.48, .58]. The means (standard deviations) for the subscales are as follows: Alienation = 2.45 (.69), Stereotype Endorsement = 1.54 (.43), Discrimination Experience = 1.99 (.66), Social Withdrawal = 2.16 (.67), and Stigma Resistance = 2.00 (.47).

Evidence of Internal Structure of the ISMI-29 and ISMI-24

The internal structure of the ISMI-29 and ISMI-24 was tested via a series of CFAs with M*plus* version 6.11 (Muthén & Muthén, 1998-2012). Specifically, four competing measurement models (i.e., unidimensional, five-factor oblique, second-order, bifactor) were examined. M*plus*’ MLR option for maximum likelihood estimation was used, which calculates the Satorra and Bentler (1988) corrected/scaled chi-square test statistic (S-Bχ^2). Model fit was evaluated using the S-Bχ^2 statistic, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Standard Root Mean Square Residual (SRMR). The following fit criteria were used: RMSEA $\leq$ .06, CFI $\geq$ .95, TLI $\geq$ .95, SRMR $<$ .08 for good fit and RMSEA $\leq$ .10, CFI $\geq$ .90, TLI $\geq$ .90, SRMR $<$ .10 for acceptable fit (Weston & Gore, 2006). The unidimensional model and second-order model were both nested within the bifactor model. The five-factor oblique model was not nested within the bifactor model because the model contained more than three latent variables. Thus, corrected/scaled chi-square difference tests ($\Delta\chi^2$), Akaike's information criterion (AIC), and Bayesian information criterion (BIC) were used to compare the fit of the unidimensional, second-order, and bifactor models, whereas only the AIC and BIC were used to compare the fit of the five-factor oblique and bifactor models. Only models that achieved adequate model fit were compared via these indices. Burnham and Anderson (2002) state that an AIC value difference exceeding 6 and especially 10 provides evidence of model fit difference (as cited in Symonds & Moussalis, 2011, p. 17). A BIC value difference exceeding 10 provides strong evidence of model fit difference (Kass & Raftery, 1995). The model with the lower AIC and BIC value is considered to have superior model fit. All analyses were done at the 5% significant level.

In regard to the ISMI-29, both the bifactor solution and the five-factor oblique solution demonstrated adequate model fit (see Table 1 for goodness of fit statistics for all tested models). The unidimensional and second-order solutions had inadequate model fit. Relative model fit comparisons revealed that the bifactor solution fit better than the five-factor oblique solution: ΔAIC = 250.93 and ΔBIC = 162.94. In summary, the CFA results suggest that the ISMI-29 best conforms to a bifactor structure rather than a five-factor oblique, second-order, or unidimensional structure. It should also be noted that the five subscales of the five-factor oblique model correlated with each other between .65 and .89 and that the loading of the first-order Stigma Resistance factor on the second-order factor was .87 for the second order model, the significance of which will be articulated in the discussion.

In regard to the ISMI-24, the bifactor solution demonstrated good model fit. The four-factor oblique and second-order solutions demonstrated adequate model fit. The unidimensional solution demonstrated inadequate model fit. Fit comparisons revealed that the bifactor solution fit better than the four-factor oblique solution, ΔAIC = 296.34, ΔBIC = 212.98, and the second-order solution, ΔS-Bχ^2 (20) = 244.05, p < .001, ΔAIC = 303.20, ΔBIC = 210.58. In summary, similar to the ISMI-29 results, the CFA findings suggest that the ISMI-24 best conforms to a bifactor structure rather than a four-factor oblique, second-order, or unidimensional structure. Taken together, these CFA results provide evidence for a bifactor structure to best account for the item covariation for both the ISMI-29 and ISMI-24, which is consistent with Hypothesis 1.

We also calculated the percent of Explained Common Variance (ECV; Reise, Moore, & Haviland, 2010), an index of unidimensionality, attributable to the general factor and each of the five group factors. When Percent of Uncontaminated Correlations (PUC) values are higher than .80, general ECV values are less important in predicting bias; when PUC values are lower than .80, general ECV values greater than .60 and Coefficient Omega Hierarchical values greater than .70 suggest that the presence of some multidimensionality is not severe enough to disqualify the interpretation of the instrument as primarily unidimensional (p. 22, Reise, Scheines, Widaman, & Haviland, 2013). In turn, group factor ECVs establish the uniqueness of each factor, with a low group ECV indicating little unique variability due to that subscale factor.

Table 2 summarizes the factor loadings for the unidimensional and bifactor solutions for the ISMI-29 and ISMI-24. Three findings inform Hypothesis 2. First, for both the ISMI-29 and ISMI-24, most general factor item loadings were larger than the group factor loadings and similar to the item loadings from the unidimensional solution. Second, the PUC (.83, .78) and ECV (.77, .78) coefficients for the ISMI-29 and ISMI-24, respectively, were high. Third, 82% of the 29 items had Individual Explained Common Variance (IECV) coefficients above .50, which indicates that most items were better measures of the general factor than their respective group factors. Overall, these results suggest that the ISMI-29 and ISMI-24 are best conceptualized as primarily unidimensional instruments, which is consistent with Hypothesis 2.

Model-Based Reliability Estimates of the ISMI-29 and ISMI-24 Total and Subscale Scores

When an instrument's items conform to a bifactor structure, it becomes necessary to provide model-based reliability evidence that the instrument's total and subscale scores truly represent the target constructs of interest. In the absence of such evidence, researchers risk misinterpreting the meaning and significance of the instrument's total and subscale scores. Given that both the ISMI-29 and ISMI-24 conformed to a bifactor structure, the next step was to calculate a series of model-based reliability coefficients. Coefficient Omega (ω) measures the proportion of total score variance that can be attributed to all common factors (i.e., true score variance, which excludes error variance). It can also be adapted to measure the proportion of subscale score variance that can be attributed to all common factors. Coefficient Omega Hierarchical (ωH; McDonald, 1999) measures the proportion of total score variance that can be attributed to a single general factor after accounting for group (i.e., subscale) factors. Coefficient Omega Subscale (ωS) is a version of ωH that measures the proportion of subscale score variance that is uniquely due to that group (i.e., subscale) factor after controlling for the general factor.

While no definitive benchmarks for evaluating ωH and ωS exist at the time of this writing, Reise, Bonifay, and Haviland (2013) state that "tentatively, we can propose that a minimum would be greater than .50, and values closer to .75 would be much preferred" (p.137). Thus, $\omega H > .75$ would indicate that the ISMI's total score predominantly reflects a single general factor despite the presence of multidimensionality across items, which in turn would permit researchers to interpret the total score as a sufficiently reliable and appropriate measure of the general construct of internalized stigma of mental illness.

Likewise, $\omega S < .50$ would indicate that the majority of that subscale's variance is due to the general factor and that negligible unique variance is due to that group factor. In other words, that subscale score's reliability is overwhelmingly inflated (i.e., confounded) by the general factor and does not actually reliably measure the narrower subdomain construct that the subscale was purported to measure. In short, to interpret such a subscale as capturing something unique could be misleading.

Table 3 summarizes the ω, ωH, ωS, reliability of ω, and ECV coefficients for the bifactor solution for the ISMI-29 and ISMI-24. Results for both the ISMI-29 and ISMI-24 indicated that $\omega H > .75$ for the general factor and $\omega S < .50$ for all subscales. In summary, as predicted by Hypothesis 3, model-based reliability estimates provided evidence for the use of the ISMI-29 and ISMI-24 total scores to represent the general internalized stigma of mental illness construct. As predicted by Hypothesis 4, model-based reliability estimates did not provide evidence for the use of any of the five subscale scores to represent the narrower purported subdomains (e.g., Alienation) of internalized stigma of mental illness.

Evidence of Internal Structure and Model-Based Reliability Estimates of the ISMI-10 Total Score

Because the ISMI-10 is composed of only two items from each of the five purported subscales, a traditional bifactor model of the ISMI-10 was not able to converge. Therefore, only the five-factor oblique, second-order, and unidimensional models were compared, using the same procedures outlined above. Contrary to predictions, the five-factor oblique solution demonstrated good model fit. The second-order model demonstrated adequate fit. The unidimensional model demonstrated inadequate fit. Fit comparisons revealed that the five-factor oblique model fit better than the second-order model, $\Delta S\text{-}B\chi^2\ (5) = 67.53$, $p < .001$, $\Delta AIC =$

76.74, ΔBIC = 53.58. In summary, the CFA results suggested that the ISMI-10 best conforms to a five-factor oblique factor structure rather than a second-order or unidimensional structure. As was the case for the ISMI-29, the five subscales of the five-factor oblique model correlated with each other between .47 and .89, suggesting the possible existence of a strong general factor. It is not unexpected that a five factor solution will fit better than a unidimensional solution when the inter-factor correlations are high, which can result in large residuals for the unidimensional solution and misfit.

Despite the inability for a standard bifactor model using MLR to converge for the ISMI-10, it was still possible to fit an alternative bifactor model for the purpose of calculating the ECV and model-based reliability coefficients for the ISMI-10. This was achieved by setting the variance to one in lieu of using a reference item for each subscale, using the robust weighted least squares mean and variance (WLSMV) estimator, setting the parameterization to theta, and requesting 1,000 bootstrap samples.

Table 2 summarizes the factor loadings for the unidimensional and bifactor solutions for the ISMI-10 and Table 3 summarizes the ω, ωH, ωS, reliability of ω, and ECV coefficients for the bifactor solution for the ISMI-10. Three findings inform Hypothesis 5. First, many of the general item loadings were substantively larger than the group factor loadings and similar to the item loadings from the unidimensional solution. Second, the PUC (.89) and ECV (.71) coefficients for the ISMI-10 were high. Third, 80% of the 10 items had IECV values were above .50, which indicates that most items were better measures of the general factor than their respective group factors. These three findings suggest that the ISMI-10 may be best conceptualized as a primarily unidimensional instrument, which is consistent with Hypothesis 5.

Results for the ISMI-10 indicated that ωH > .75 for the general factor and ωS < .50 for all subscales. Thus, these model-based reliability estimates provided evidence for the use of the ISMI-10 total scores to represent the general internalized stigma of mental illness construct. Estimates did not provide clear evidence for the use of any subscale scores, congruent with the ISMI-10 developers' request that only a total ISMI-10 score be used.

Development, Evidence of Internal Structure, and Model-Based Reliability Estimates of the ISMI-9 Total Score

Because the items that compose the ISMI-10 contained too much multidimensionality to allow the ISMI-10 to demonstrate a clean CFA unidimensional factor structure, the IECV coefficients for the ISMI-29 were used to develop the ISMI-9. "For each item the IECV provides an indication of the strength of the bifactor loadings on the general factor [internalized stigma of mental illness] relative to the strength of the specific factor. Thus, it is desirable to identify items with high IECV values, which provides initial evidence of item-level unidimensionality… selecting additional items in this manner allows for the development of a unidimensional instrument that is strongly associated with the general factor without being overly influenced by the group factors." (p. 517, Stucky, Edelen, Vaughan, Tucker, & Butler, 2014). To ensure unidimensionality and retain content domain coverage of the construct (as defined by the five "topic areas" delineated by Ritsher et al., 2003, p. 34), the two items from each group factor with the highest IECV values (minimum value of .80; Stucky & Edelen, 2014) were selected for the ISMI-9. Because the Discrimination Experience group factor only contained one item with an IECV above .80, only one item from this group factor was retained. All nine retained items had an IECV ≥ .84. This process resulted in the ISMI-9, which contains items that (a) were drawn from the five "topic areas" (i.e., Alienation, Stereotype Endorsement, Discrimination Experience, Social Withdrawal, Stigma Resistance) and (b) were the cleanest measures of the general internalized stigma of mental illness factor—the factor that short forms of the ISMI seek to measure.

The unidimensional model of the ISMI-9 demonstrated good fit, S-B$\chi^2(27) = 70.35$, $p < .001$, RMSEA = .046 [90% CI of .033, .059], CFI = .976, TLI = .967, SRMR = .027. Whereas the ISMI-10 is a primarily unidimensional instrument (ECV = .71) with enough multidimensionality to preclude the good fit of a unidimensional factor solution, the ISMI-9 is a unidimensional instrument (ECV = .87, PUC = .89) that demonstrated good fit with a unidimensional factor solution. Furthermore, the ISMI-9 total score ($\alpha = .86$, 95% CI [.85, .88], $M = 2.13$, $SD = .58$) demonstrated slightly stronger internal consistency estimates than did the ISMI-10 total score ($\alpha = .82$, 95% CI [.80, .84]). In addition, the ISMI-9 total score was a slightly cleaner measure of the general internalized stigma of mental illness construct (ωH = .89) than was the ISMI-10 total score (ωH = .84). Consistent with this, 100% of the 9 items had IECV values above .80 compared with 30% for the ISMI-10. The total scores of the ISMI-9 and ISMI-10 ($r = .88$) were found to correlate .95 and .94, respectively, with the total score of the ISMI-29. In summary, the ISMI-9 total score demonstrated a slightly cleaner unidimensional structure and stronger reliability than the ISMI-10 total score.

Discussion

The present study investigated (a) the internal structure of the English language version of the ISMI-29 and ISMI-24, (b) whether or not model-based reliability estimates support the calculation and interpretation of the ISMI-29 (and ISMI-24) total and/or subscale scores, (c) the internal structure of the ISMI-10, and (d) the psychometric properties of the new ISMI-9.

Internal Structure, Scoring, and Interpretation of the ISMI-29 and ISMI-24

CFA results supported Hypothesis 1: the ISMI-29 and ISMI-24 both conformed to a bifactor structure. This suggests that the covariation among the ISMI items may be best accounted for by (a) a single general internalized stigma of mental illness factor that reflects the common variance across all items plus (b) group factors (i.e., the five subscales) that capture some additional (i.e., unique) common variance among clusters of items (p. 688, Reise, 2012). In contrast, the results did not suggest that the ISMI-29 (and ISMI-24) is a purely unidimensional instrument, a purely multidimensional instrument, nor a multidimensional instrument defined by narrower lower-order factors whose substantial intercorrelation is accounted for by a higher second-order factor. In line with Hypothesis 2, ECV analyses suggested that the ISMI-29 (and ISMI-24), despite containing some multidimensionality, may be best conceptualized as a primarily unidimensional instrument.

While the good fit of the bifactor model suggests that the internal structure of the ISMI involves both a larger general factor and narrower group factors, model-based reliability estimates were needed to determine the utility of the ISMI total and subscale scores. In other words, the finding that an instrument conforms to a bifactor structure does not, by itself, provide a compelling rationale for the calculation and interpretation of the total or subscale scores of that instrument. Results suggested that about 92% of the total score variance modeled is due to the general internalized stigma of mental illness factor, whereas only 5% of the total score variance is due to the five (or four, in the case of the ISMI-24) subscale group factors. Furthermore, 95% (i.e., ω_H of .92 divided by ω of .96) of the reliable variance in the ISMI total score was due to the general factor, which means that the general internalized stigma of mental illness factor is the only meaningful influence on total score variation. In contrast, results demonstrated that the majority of each ISMI subscale's true score variance was accounted for by the general internalized stigma of mental illness factor (70% to 94%) rather than the specific group factor (6% to 30%). Put simply, each ISMI subscale score appears to be primarily driven by a respondent's general (or overall) degree of internalized stigma of mental illness rather than the respondent's specific degree of perceived Alienation, Social Withdrawal, etc.

Taken as a whole, these findings suggest that the English ISMI, regardless of whether or not the Stigma Resistance subscale items are included in the calculation of the total score, may best be scored as a univocal measure. In other words, results supported the use of the ISMI-29 and ISMI-24 total score, but did not provide clear evidence in support of the use of subscale scores, congruent with Hypotheses 3 and 4. This also aligns with the scale developer's initial recommendation that the ISMI be conceptualized as measuring a single construct (p. 47; Ritsher et al., 2003). This also dovetails with the CFA analysis of the modified Taiwan version of the ISMI-29, which found that all subscales except Stigma Resistance intercorrelated strongly, hinting at the possibility of a strong general factor that runs through all items (Ociskova et al., 2014). It is pertinent to mention here that the finding of one interpretable general factor with narrower subfactors that do not warrant interpretation is a common result of studies that subject instruments to bifactor modeling (e.g., Brouwer, Meijer, Zevalkink, 2012; Gignac & Watkins, 2013) and has been shown to apply to well-validated instruments (e.g., Beck Depression Inventory-II; Wechsler Adult Intelligence Scale-IV).

Reconciling the Stigma Resistance Subscale

As noted in the introduction, researchers using the English ISMI-29 or its various modifications and translations have sometimes found that the Stigma Resistance subscale correlates strongly enough with the other subscales to warrant incorporating its items into the

ISMI-29 total score calculation (Assefa et al., 2012). Other times, the Stigma Resistance subscale has demonstrated a weak relationship with the other subscales, leading some to speculate Stigma Resistance (at least as it is operationalized in the ISMI-29) is best treated as a separate stigma construct rather than a subdimension of the larger internalized stigma of mental illness construct (e.g., Lysaker et al., 2007). In contrast to the CFA of the modified Taiwanese version of the ISMI-29 (Chang et al., 2014), the present CFA results indicate the Stigma Resistance subscale correlated strongly with the four other factors and loaded to a similar degree onto the second-order internalized stigma of mental illness factor. These results suggest that the construct of Stigma Resistance appears to relate to other purported subdimensions of internalized stigma of mental illness in a manner that supports the inclusion of Stigma Resistance as a subdimension of the larger internalized stigma of mental illness construct. However, given the documented inconsistency of the Stigma Resistance subscale's covariation with the other subscales across studies, researchers using the instrument are encouraged to routinely perform a preliminary factor analysis prior to making decisions regarding the use of the Stigma Resistance items in the calculation of the total score.

Factor Structure, Scoring, and Interpretation of the ISMI-10

Contrary to Hypothesis 5, a unidimensional structure did not best account for the item covariation of the ISMI-10. Rather, a five-factor oblique structure provided the best fit to the data. However, model-based reliability coefficients suggested that 84% of the total score variance modeled is reliably due to the general internalized stigma of mental illness factor, whereas only 6% of the total score variance is due to the two-item subscale group factors. Furthermore, 92% (i.e., ω_H of .84 divided by ω of .91) of the reliable variance in the ISMI total score was due to the general factor, which means that the general internalized stigma of mental illness factor was the only meaningful influence on total score variation. These findings suggest that it may be permissible to score the ISMI-10 as a univocal measure, congruent with the ECV analyses finding that the ISMI-10 is a primarily unidimensional instrument. This aligns with the recommendations of the scale developers, who suggest researchers conceptualize the ISMI-10 as a unidimensional instrument for which a total score can be calculated (Boyd et al., 2014b). However, even though the ISMI-10 is a primarily unidimensional instrument, there is enough multidimensionality across the items that some degree of bias may be introduced when calculating a raw total score (Reise et al., 2013). Therefore, IECV coefficients based on the bifactor solution from the ISMI-29 were used to develop an alternative short form of the ISMI that would demonstrate an unequivocally unidimensional factor structure: the ISMI-9.

Factor Structure, Scoring, and Interpretation of the ISMI-9

Regarding evidence concerning internal structure (Standard 1.13; AERA et al., 2014), the ISMI-9 demonstrated a clear unidimensional factor structure. This supports the use of the ISMI-9 total score as a measure of the overall internalized stigma of mental illness. Regarding evidence of reliability/precision (Standard 2.3), the ISMI-9 total score demonstrated slightly stronger internal consistency and cleaner measurement of the general internalized stigma of mental illness construct than the ISMI-10 total score. Regarding content-oriented evidence of the validity of the ISMI-9 total score (Standard 1.1), like the ISMI-10, the ISMI-9 contains items from each of the five purported factors of the ISMI-29. However, future research is needed to examine convergent evidence for the validity of the ISMI-9 total score (Standard 1.16). Such evidence has already begun accumulating for the ISMI-10 (Boyd et al., 2014b).

Given that the ISMI-29 and ISMI-24 appear to be dominated by a single general factor, there may be less value in using these longer instruments to obtain a total score when an abbreviated instrument like the ISMI-10 or ISMI-9 may accomplish the same task with less participant burden. Given the possible psychometric advantages of the ISMI-9 total score demonstrated with the present data, we encourage researchers to further investigate the comparative validity of the ISMI-9 and ISMI-10 total scores, with particular attention to convergent evidence.

Cautions, Limitations, and Future Directions

It goes without saying that the present findings are most relevant to the versions of the ISMI (i.e., English versions of the ISMI-29, ISMI-24, ISMI-10, and ISMI-9) that were tested in the present study. There are many modifications and translations of the ISMI and it is not certain to what degree the pattern of results we observed with the versions we tested would replicate themselves with other versions used in other contexts. Future research using bifactor and model-based reliability analysis on alternate versions of the ISMI would help determine the degree to which our findings are idiosyncratic versus universally applicable to these alternative versions, which in turn influences how scholars and practitioners can use these alternative versions.

In addition, the present study's findings, like all studies' findings, were influenced by the nature of the sample. Our sample was composed of community-dwelling adults living in the United States who self-identified as having a mental illness—depression, in this case. The majority of the sample was also female, white, and reported minimal to mild internalized stigma. Importantly, this lower sample mean level of stigma contrasts with the higher mean levels of stigma reported in the 14 studies reviewed by Boyd et al. (2014a). Furthermore, our sample was drawn from ResearchMatch.org registry members who accepted our invitation to participate in a study described as a "mental illness attitudes questionnaire." ResearchMatch.org registry participants are not necessarily representative of the larger US population and the study description may have engendered a degree of participant self-selection. Therefore, it is possible that our findings are unique and only apply to this specific intersectional population. Thus, generalizability of these results should not be assumed; this should be tested directly in future research. For example, perhaps the ISMI demonstrates a different factor structure with interpretable subscale factors among inpatient, severely mentally ill, heavily self-stigmatized, racial/ethnic minority, and/or verified DSM-diagnosed populations. Thus, the present findings can serve as grist for empirically- and theoretically-informed dialogue, rather than a definitive statement about the "true nature" of the ISMI. This is particularly true given the different factor analytic outcomes that have been reported across variants of the ISMI (Boyd et al., 2014a). We also recommend that the incremental validity of the group factors (i.e., their ability to predict variance in relevant criteria beyond the variance accounted for by the general factor) be examined in future research, as this is a key criterion for judging the utility of subscale scores. This data, in tandem with model-based reliability estimates, could further inform the utility of using ISMI-29 subscale scores.

Conclusions

This paper investigated the internal structure and model-based reliability of an electronically-administered, original English version of the ISMI-29, ISMI-24, and ISMI-10 among community-dwelling adults living in the United States who identified as having a mental illness, specifically depression. Results indicated that the ISMI-29, ISMI-24, and ISMI-10 should all be scored as univocal measures given the strong general factor that runs through all the items. Researchers and clinicians who wish to calculate and interpret the ISMI subscales of the ISMI-29 or ISMI-24 are encouraged to consider the present findings prior to making that decision. Researchers are strongly encouraged to conduct further internal structure analyses of the ISMI and its various alternate versions. In the absence of such research, the generalizability of the present findings remains uncertain. Finally, this paper introduced an alternative short form of the ISMI—the ISMI-9—that offers some potential psychometric advantages over the ISMI-10. Future research on the reliability and validity of both short forms is recommended.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Assefa, D., Shibre, T., Asher, L, & Fekadu, A. (2012). Internalized stigma among patients with schizophrenia in Ethiopia: A cross-sectional facility-based study. *BMC Psychiatry, 12*, 1-10.

Boyd, J. E., Adler, E. P., Otilingam, P. G., & Peters, T. (2014a). Internalized stigma of mental illness (I) scale: A multinational review. *Comprehensive Psychiatry, 55,* 221-231.

Boyd, J. E., Oltingam, P. G., & DeForge, B. R. (2014b). Brief version of the internalized stigma of mental illness (ISMI) scale: Psychometric properties and relationship to depression, self esteem, recovery orientation, empowerment, and perceived devaluation and discrimination. *Psychiatric Rehabilitation Journal, 37,* 17-23.

Brouwer, D., Meijer, R. R., & Zevalkink, J. (2013). On the factor structure of the Beck Depression Inventory-II: G is the key. *Psychological Assessment, 25*, 136–145.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference, 2nd ed*. New York, NY: Springer.

Chang, C. C., Wu, T. H., Chen, C. Y., Wang, J. D., & Lin, C. Y. (2014). Psychometric evaluation of the internalized stigma of mental illness scale for patients with mental illness: Measurement invariance across time. *Clinical Psychology: Science and Practice, 9,* 35-53.

Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist, 59,* 614-625.

Corrigan, P. W. & Watson, A. C. (2002). The paradox of self-stigma and mental illness. *Clinical Psychology: Science and Practice, 9,* 35–53.

Corrigan, P. W., Watson, A. C., & Barr, L. (2006). The self-stigma of mental illness: Implications for self-esteem and self-efficacy. *Journal of Social and Clinical Psychology, 25,* 875-884.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48*, 639-662.

Kass, R. E., & Raftery, A. E. (1995) Bayes factors. *Journal of the American Statistical Association, 90*, 773–795

Link, B. B. (1987). Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. *American Sociological Review, 52*, 96-112.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatric Services, 52*, 1621–1626.

Lysaker, P. H., Roe, D., & Yanos, P. T. (2007). Toward understanding the insight paradox: Internalized stigma moderates the association between insight and social functioning, hope and self-esteem among people with schizophrenia spectrum disorders. *Schizophrenia Bulletin, 33*, 192-199.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide (7th ed)*. Los Angeles, CA: Authors.

Ociskova, M., Prasko, J., Kamaradova, D., Latalova, K., Kurfurst, P., Dostalova, L., …, & Tichackova, A. (2014). Self-stigma in psychiatric patients – standardization of the ISMI scale. *Neuroendocrinology Letters, 35*, 624-632.

President's New Freedom Commission on Mental Health. (2003). Achieving the promise: Transforming mental health care in America—Executive summary. Retrieved from http://govinfo.library.unt.edu/mentalhealthcommission/reports/FinalReport/downloads/FinalReport.pdf

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*, 667-696.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129-140.

Reise, S. P., Moore, T. N., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92,* 544–559.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26.

Ritsher, J. B., Otilingam, P. G., & Grajales, M. (2003). Internalized stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Research, 121,* 31-49.

Sibitz, I., Unger, A., Woppman, A., Zidek, T., & Amering, M. (2011). Stigma resistance in patients with schizophrenia. *Schizophrenia Bulletin, 37,* 316-323.

Stevelink, S. A. M., Wu, I. C., Voorend, C. G. N. V., & Brakel, W. H. V. (2012). The psychometric assessment of internalized stigma instruments: A systematic review. *Stigma Research and Action, 2,* 100-118.

Stucky, B. D., Edelen, M. O., Vaughan, C. A., Tucker, J. S., & Butler, J. (2014). The Psychometric Development and Initial Validation of the DCI-A Short Form for Adolescent Therapeutic Community Treatment Process. *Journal of Substance Abuse Treatment*, *46*, 516–521.

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (183-206). London, UK: Taylor & Francis.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference, and model averaging in behavioral ecology using the Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65*, 13-21.

U.S. Department of Health and Human Services. (1999). Mental health: A report of the Surgeon General—Executive summary. Retrieved from http://www.surgeongeneral.gov/library/mentalhealth/home.html

Weston, R., & Gore, P. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34,* 719-751.

Worthington, R. L., & Whitaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806-838.

Table 1

Goodness of Fit Statistics for All Tested Measurement Models

Model	S-B χ2	df	RMSEA [90% CI]	CFI	TLI	SRMR	AIC	BIC
ISMI-29 Unidimensional	1,674.82	377	.067 [.065, .071]	.825	.811	.058	46,147.01	46,549.88
ISMI-29 Five-Factor Oblique	1,112.80	367	.052 [.048, .055]	.899	.889	.055	45,483.43	45,932.60
ISMI-29 Second Order	1,214.48	372	.055 [.051, .058]	.886	.876	.057	45,597.06	46,023.08
ISMI-29 Bifactor	882.04	348	.045 [.041, .049]	.928	.916	.044	45,232.50	45,769.66
ISMI-24 Unidimensional	1,190.69	252	.070 [.066, .074]	.851	.837	.054	37,900.52	38,233.93
ISMI-24 Four-Factor Oblique	750.17	246	.052 [.048, .056]	.920	.910	.051	37,365.65	37,726.85
ISMI-24 Second Order	758.89	248	.052 [.048, .056]	.916	.910	.051	37,372.51	37,724.45
ISMI-24 Bifactor	478.06	228	.038 [.033, .043]	.960	.952	.032	37,069.32	37,513.86
ISMI-10 Unidimensional	277.20	35	.096 [.085, .106]	.850	.807	.051	16,359.42	16,498.34
ISMI-10 Five-Factor Oblique	55.12	25	.040 [.026, .054]	.981	.966	.026	16,109.68	16,294.90
ISMI-10 Second Order	129.02	30	.066 [.055, .078]	.939	.908	.041	16,186.41	16,348.49
ISMI-10 Bifactor			Failed to converge.					

Note: All models were statistically significant at the $p < .001$ level. ISMI = Internalized Stigma of Mental Illness Scale. Statistics are based on MLR estimation. S-B χ2 = Satorra and Bentler (1988) corrected/scaled chi-square test statistic, RMSEA = Root Mean Square Error of Approximation, CI = Confidence Interval, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = Standard Root Mean Square Residual.

Table 2
Confirmatory Factor Analysis Standardized Loadings for the ISMI

	ISMI-29								ISMI-24							ISMI-10								ISMI-9	
	Uni	Bifactor							Uni	Bifactor						Uni	Bifactor							Uni	Bifactor
		Gen	F1	F2	F3	F4	F5	IECV		Gen	F1	F2	F3	F4	IECV		Gen	F1	F2	F3	F4	F5	IECV		IECV
Group Factor 1: Alienation																									
i1a	**.74**	**.76**	.05					***.99***	**.74**	**.76**	.08				.98									**.76**	.99
i2a	**.67**	**.69**	**.12**					.96	**.65**	**.65**	**.19**				.91	**.64**	**.73**	**.18**					.94		
i3a	**.49**	**.50**	.01					***.99***	**.49**	**.50**	.05				.99	**.45**	**.49**	**.45**					.54	**.51**	.66
i4a	**.64**	**.62**	**.45**					.66	**.64**	**.61**	**.45**				.65										
i5a	**.63**	**.61**	**.50**					.60	**.62**	**.59**	**.53**				.57										
i6a	**.67**	**.66**	**.35**					.76	**.67**	**.65**	**.38**				.74										
Group Factor 2: Stereotype Endorsement																									
i7se	**.52**	**.51**		**.11**				**.96**	**.51**	**.51**		**.12**			.94									**.52**	.71
i8se	**.47**	**.45**		**.24**				.88	**.47**	**.45**		**.25**			.85										
i9se	**.23**	**.21**		**.31**				.27	**.23**	**.20**		**.31**			.24	**.24**	**.32**		**.48**				.30		
i10se	**.56**	**.54**		**.30**				.86	**.55**	**.52**		**.33**			.80										
i11se	**.54**	**.52**		**.40**				.72	**.52**	**.50**		**.43**			.63										
i12se	**.42**	**.39**		**.48**				.46	**.41**	**.38**		**.47**			.42										
i13se	**.62**	**.61**		**.33**				**.89**	**.61**	**.59**		**.36**			.82	**.64**	**.81**		-.02				.99	**.61**	.99
Group Factor 3: Discrimination Experience																									
i14de	**.58**	**.56**			**.47**			.59	**.59**	**.57**			**.45**		.63										
i15de	**.58**	**.55**			**.45**			.61	**.59**	**.56**			**.43**		.64	**.59**	**.58**			**.55**			.53		
i16de	**.65**	**.61**			**.54**			.59	**.66**	**.63**			**.52**		.62	**.66**	**.67**			**.50**			.64		
i17de	**.61**	**.58**			**.47**			.64	**.62**	**.59**			**.45**		.66										
i18de	**.71**	**.74**			-.05			***.99***	**.71**	**.73**			-.07		.99									**.73**	.90
Group Factor 4: Social Withdrawal																									
i19sw	**.55**	**.54**				**.16**		.92	**.55**	**.55**				**.13**	.94										
i20sw	**.72**	**.68**				**.46**		.72	**.73**	**.71**				**.44**	.79	**.73**	**.73**				**.43**		.75		
i21sw	**.72**	**.71**				**.15**		***.99***	**.73**	**.73**				.07	.99									**.69**	.99
i22sw	**.68**	**.64**				**.39**		.80	**.69**	**.66**				**.35**	.82	**.70**	**.71**				**.44**		.73		
i23sw	**.66**	**.65**				**.15**		**.98**	**.67**	**.66**				.10	.99									**.63**	.89
i24sw	**.68**	**.66**				**.25**		.92	**.69**	**.69**				**.18**	.97										
Group Factor 5: Stigma Resistance																									
i25sr	.06	.04					.27	.02																	
i26sr	**.68**	**.69**					.13	***.99***																**.71**	.85
i27sr	**.57**	**.58**					.39	**.84**								**.56**	**.68**					**.28**	.85	**.59**	.73
i28sr	**.25**	**.22**					.35	.35								**.25**	**.36**					**.49**	.35		
i29sr	**.19**	**.18**					**.34**	.26																	

Note : ISMI = Internalized Stigma of Mental Illness Scale. Uni = Unidimensional Model, Gen = General Factor, F1 = Alienation Group Factor, F2 = Stereotype Endorsement Group Factor, F3 = Discrimination Experience Group Factor, F4 = Social Withdrawal Group Factor, F5 = Stigma Resistance Group Factor, IECV = Individual Explained Common Variance. Loadings for ISMI-29 and ISMI-24 are based on MLR estimation and loadings for the ISMI-10 are based on WLSMV estimation. Items are listed in the same order as Table 2 of Boyd et al. (2003).

* All bolded loadings significant at $p < .05$

Table 3

Explained Common Variance and Model-Based Reliability Estimates for the ISMI

	Gen	F1	F2	F3	F4	F5
	ISMI-29					
Omega	.96 [.96, .97]	.89 [.87, .90]	.85 [.83, .87]	.89 [.88, .91]	.89 [.88, .90]	.68 [.64, .71]
Omega Hierarchical	.92 [.91, .92]	.01 [.01, .01]	.01 [.01, .02]	.01 [.01, .01]	.01 [.00, .01]	.01 [.00, .01]
Omega Subscale		.13 [.10, .17]	.20 [.16, .25]	.24 [.20, .28]	.09 [.06, .12]	.24 [.16, .30]
Reliability of Omega	.95 [.95, .96]	.15 [.11, .19]	.24 [.19, .29]	.26 [.22, .31]	.10 [.07, .13]	.36 [.25, .44]
Explained Common Variance (ECV)	.77 [.75, .79]	.05 [.04, .06]	.04 [.04, .05]	.07 [.06, .08]	.03 [.02, .04]	.03 [.02, .04]
	ISMI-24					
Omega	.96 [.96, .97]	.89 [.87, .90]	.85 [.83, .87]	.89 [.88, .91]	.90 [.88, .90]	
Omega Hierarchical	.91 [.90, .92]	.01 [.01, .02]	.02 [.02, .02]	.01 [.01, .02]	.01 [.00, .01]	
Omega Subscale		.16 [.12, .20]	.25 [.21, .30]	.21 [.17, .26]	.05 [.01, .08]	
Reliability of Omega	.95 [.94, .95]	.18 [.13, .22]	.30 [.25, .35]	.24 [.19, .29]	.06 [.01, .09]	
Explained Common Variance (ECV)	.78 [.76, .80]	.07 [.05, .08]	.06 [.05, .07]	.07 [.06, .08]	.03 [.01, .03]	
	ISMI-10					
Omega	.91 [.89, .92]	.99 [.99, .99]	.60 [.54, .64]	.76 [.73, .78]	.83 [.81, .86]	.65 [.58, .70]
Omega Hierarchical	.84 [.82, .85]	.01 [.01, .01]	.00 [.00, .01]	.03 [.02, .03]	.02 [.01, .03]	.01 [.01, .02]
Omega Subscale		.14 [.08, .20]	.08 [.03, .14]	.31 [.25, .38]	.22 [.16, .28]	.22 [.14, .31]
Reliability of Omega	.92 [.91, .94]	.14 [.08, .20]	.13 [.05, .23]	.41 [.34, .50]	.26 [.19, .33]	.34 [.23, .46]
Explained Common Variance (ECV)	.71 [.67, .74]	.15 [.13, .17]	.12 [.10, .13]	.14 [.12, .17]	.21 [.18, .23]	.10 [.08, .12]
	ISMI-9					
Omega	.91 [.90, .92]	.99 [.99, .99]	.66 [.62, .69]	.70 [.66, .75]	.70 [.66, .73]	.75 [.70, .80]
Omega Hierarchical	.89 [.88, .90]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.00 [.00, .00]	.01 [.01, .02]
Omega Subscale		.04 [.02, .06]	.04 [.04, .08]	.07 [.06, .08]	.03 [.00, .06]	.15 [.12, .21]
Reliability of Omega	.98 [.98, .98]	.04 [.02, .06]	.06 [.06, .12]	.10 [.08, .12]	.04 [.00, .08]	.20 [.16, .28]
Explained Common Variance (ECV)	.87 [.86, .89]	.21 [.20, .24]	.15 [.13, .16]	.12 [.11, .14]	.20 [.18, .21]	.19 [.17, .21]

Note: ISMI = Internalized Stigma of Mental Illness Scale. Gen = General Factor, F1 = Alienation Group Factor, F2 = Stereotype Endorsement Group Factor, F3 = Discrimination Experience Group Factor, F4 = Social Withdrawal Group Factor, F5 = Stigma Resistance Group Factor. 95% confidence intervals for all coefficients are displayed in brackets.

Appendix

Internalized Stigma of Mental Illness Inventory – 9-item Version (ISMI-9)*

We are going to use the term "mental illness" in the rest of this questionnaire, but please think of it as whatever you feel is the best term for it.

For each question, please mark whether you strongly disagree (1), disagree (2), agree (3), or strongly agree (4).

		Strongly disagree	Disagree	Agree	Strongly agree
1.	Stereotypes about the mentally ill apply to me.	1	2	3	4
2.	In general, I am able to live life the way I want to.	1	2	3	4
3.	Negative stereotypes about mental illness keep me isolated from the 'normal' world.	1	2	3	4
4.	I feel out of place in the world because I have a mental illness.	1	2	3	4
5.	Being around people who don't have a mental illness makes me feel out of place or inadequate.	1	2	3	4
6.	People without illness could not possible understand me.	1	2	3	4
7.	Nobody would be interested in getting close to me because I have a mental illness.	1	2	3	4
8.	I can't contribute anything to society because I have a mental illness.	1	2	3	4
9.	I can have a good, fulfilling life, despite my mental illness.	1	2	3	4

Scoring Key
The ISMI-9 contains 9 items which produce a total score. Reverse-code items 2 and 9 before calculating the total score. Add the item scores together and then divide by the total number of answered items. The resulting score should range from 1-4. For example, if someone answers 8 of the 9 items, the total score is produced by adding together the 8 answered items and dividing by 8.

Interpretation of Scores
4-category method (following the method used by Lysaker et al., 2007):

1.00-2.00: minimal to no internalized stigma
2.01-2.50: mild internalized stigma
2.51-3.00: moderate internalized stigma
3.01-4.00: severe internalized stigma

2-category method (following the method used by Ritsher [Boyd] & Phelan, 2004).

1.00-2.50: does not report high internalized stigma
2.51-4.00: reports high internalized stigma

** Appendix format adapted from Boyd et al. (2014)'s ISMI-10 Appendix*