Using Item Response Theory to Develop Ultra-Brief (SSOSH-3) and Revised (SSOSH-7) Self-Stigma of Seeking Help Scales

Rachel E. Brenner[1], Kimberly F. Colvin[1], Joseph H. Hammer[2], and David L. Vogel[3]
[1] Department of Educational & Counseling Psychology, University at Albany, State University of New York
[2] Department of Educational, School, and Counseling Psychology, University of Kentucky,
[3] Department of Psychology, Iowa State University

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article is available via its DOI: 10.1177/1073191120958496

Author Note
Rachel E. Brenner https://orcid.org/0000-0003-1069-9591
Kimberley F. Colvin https://orcid.org/0000-0002-6173-7108
Joseph H. Hammer https://orcid.org/0000-0002-4028-5878
Rachel E. Brenner is now at the Department of Psychology, Colorado State University.
Correspondence concerning this article should be addressed to Rachel E. Brenner, Department of Psychology, Colorado State University, Fort Collins, CO 80523; e-mail: rachel.brenner@colostate.edu

Abstract

The current research developed ultra-brief (SSOSH-3) and revised (SSOSH-7) versions of the Self-Stigma of Seeking Help scale. Item-Response Theory was used to examine the amount of information each item provided across the latent variable scale and test whether items functioned differently across women and men. In a sample of 857 community adults, results supported removal of three reverse-scored items to create the SSOSH-7. The three most informative items were retained to create the SSOSH-3. Differential item functioning testing supported the use of both versions across women and men. Results replicated in an undergraduate student sample (N = 661). In both samples, the SSOSH-3 (α's = .82-.87) and SSOSH-7 (α's = .87-.89) demonstrated evidence of internal consistency. The SSOSH-3 (r's $\geq$ .89) and SSOSH-7 (r's $\geq$ .97) were highly correlated with the original SSOSH across samples and demonstrated significant correlations with help-seeking constructs and in similar magnitude to the original SSOSH.

Keywords: self-stigma, Self-Stigma of Seeking Help scale, help seeking, item response theory, mental health stigma

Using Item Response Theory to Develop Ultra-Brief (SSOSH-3) and Revised (SSOSH-7) Self-Stigma of Seeking Help Scales

Nearly one in five adults in the U.S. has a diagnosable mental illness, yet less than 10% received therapy in the past year (Substance Abuse and Mental Health Services Administration, 2017). A major psychological barrier that prevents individuals from seeking treatment is *self-stigma* of seeking psychological help, or the negative stereotypes/shame one would feel towards themselves for seeking psychological help (Vogel, Wade, & Haake, 2006). Self-stigma has been linked to worse help-seeking attitudes (Vogel et al., 2017), lower intention to seek counseling (e.g., Brenner, Cornish, Heath, Lannin, & Losby, 2020; Pattyn, Verhaeghe, Sercu, & Bracke, 2014), lower rates of accessing online information about mental health and counseling services (Lannin, Vogel, Brenner, Abraham, & Heath, 2016), and lesser use of services over a two-year period (Seidman, Wade, Vogel, & Armistead-Jehle, 2019). As such, researchers increasingly use self-report measures of self-stigma to identify factors that contribute to its development (e.g., Pederson & Vogel, 2007; Vogel, Bitman, Hammer, & Wade, 2013) and test interventions to reduce it (e.g., Cornish, Brenner, Vogel, & Wade, 2019; Lannin, Guyll, Vogel, & Madon, 2013).

Self-stigma research predominantly uses the Self-Stigma of Seeking Help scale (SSOSH; Vogel et al., 2006). PsycINFO indicates the SSOSH has been administered in over 150 studies and 12 countries worldwide (e.g., Vogel, Armstrong et al., 2013; Vogel et al., 2017). Although the SSOSH has received psychometric support (Vogel et al., 2006), the field would benefit from developing revised and ultra-brief versions of the SSOSH through examining individual item functioning. Reverse-scored items can be problematic (Dalal & Carter, 2015) and some SSOSH items may demonstrate issues with content validity. Revised and ultra-brief versions would allow for use in large-scale intervention studies as well as medical and other settings in which the full set of 10 items is not feasible to administer. Therefore, the current research aims to use Item Response Theory (IRT) to create revised and ultra-brief versions of the SSOSH that best reflect and efficiently measure the construct.

Revising Self-Stigma of Seeking Help Scales

One reason to examine the functioning of SSOSH items centers on its use of reverse-scored items. Half of the items are reversed scored. Although reverse scored items have the benefit of discouraging acquiescent responding (DeVellis, 2016), researchers underscore the limitations of reverse-scored items such as challenges to validity and increased systematic and random error (Dalal & Carter, 2015) and suggest that the disadvantages to reverse-scored items often outweigh any benefits (DeVellis, 2016). Importantly, reverse scored items in the SSOSH may inadvertently reflect attitudes or expectations regarding the efficacy of therapy rather than direct self-stigmatization for seeking help. This could threaten the content validity of the scale and its relations with other outcomes. Consider SSOSH item 4, "My self-esteem would increase if I talked to a therapist." This might reflect the extent to which a person believes their self-esteem would increase as an outcome of therapy rather than their anticipated decrease in self-esteem for deciding to seek help. A similar issue presents with "My self-confidence would remain the same if I sought professional help for a problem I could not solve." Moreover, in a measurement invariance study of the SSOSH, item 7 ("I would feel okay about myself if I made the choice to seek professional help") "was the most variant" (Vogel, Armstrong, et al., 2013, p. 307). Specifically, for item 7, Vogel, Armstrong et al. (2013, Table 5) freed the factor loadings in three of the five comparison countries before finding support for metric invariance and freed the intercepts in four of the five comparison countries before finding support for scalar invariance. Researchers have eliminated reverse-score items to create briefer versions of other existing measures (e.g., Berle et al., 2011). Thus, it is important to further examine SSOSH items and remove potentially problematic items to create a scale that more purely reflects self-stigma.

From a research standpoint, briefer SSOSH measures might increase its inclusion in large-scale research such as epidemiological medical studies and nationally representative public health assessments. These studies tend to use brief versions of existing measures or choose items from existing scales (McDermott et al., 2019). The Healthy Minds Study (Healthy Minds Network, 2019), for example, uses three items that assess for perceived societal stigma and two items that assess person's stigma towards help seekers, though no self-stigma items are included.

Creating a brief measure could allow researchers to examine the impact of self-stigma on a grander scale and increase the likelihood that large-scale studies include this construct.

In addition, briefer versions of the SSOSH might help address current limitations in help-seeking research. Although self-stigma has been cross-sectionally linked to help-seeking behavior (Lannin et al., 2016), few studies have examined the impact of self-stigma longitudinally (see Seidman, Wade, et al., 2019, for an exception). Inclusion of the SSOSH in longitudinal studies may be more feasible with briefer versions, as shorter surveys typically experience less participant attrition (Rolstad, Adler, & Rydén, 2011). Briefer versions of the SSOSH may also facilitate more inclusive research. As response rates are higher for shorter surveys (Nakash, Hutton, Jørstad-Stein, Gates, & Lamb, 2006), shorter surveys can facilitate collecting data from small or difficult-to-access populations (Sheldon, Graham, Pothecary, & Rasul, 2007). Examining the differential impact of self-stigma on health disparities in access to care across privileged and marginalized populations may be more feasible with briefer versions.

Developing revised and ultra-brief versions of the SSOSH could also ease and expand the consideration of self-stigma in non-traditional settings. Medical practitioners, for example, are an important referral source for mental health concerns. Most clients seek help for depression from their primary care physician first rather than a psychologist (Druss et al., 2008; Reust, Thomlinson, & Lattie, 1999). Those with mental health concerns might already have a trusting relationship with their medical provider, whereas they might feel hesitant to disclose psychological concerns to an unfamiliar therapist (Mitchell & Coyne 2007). The rise of integrated care, which includes greater coordination between medical and mental health services, further increases the importance of medical providers in this process. Unsurprisingly, medical intake forms increasingly include mental health screens for depression or anxiety (Mitchell, 2019), which allow medical practitioners to make referrals for psychological help.

Measuring self-stigma could provide meaningful context for physicians when interpreting mental health results and discussing next steps with clients. In one study, distressed participants with high self-stigma were half as likely to seek counseling information as distressed participants

with low self-stigma (Lannin et al., 2016). Including a screening measure of self-stigma, therefore, could help medical practitioners identify and tailor their discussion of psychological services for patients with high self-stigma who may be resistant to seeking help. This could also lead to basic trainings regarding how to discuss seeking help with distressed patients with high self-stigma. Providers in these settings often prefer briefer screening measures due to limited time and patient fatigue (McDermott, Levant, Hammer, Borgogna, & McKelvey, 2019). To address these constraints, researchers have developed very brief mental health measures, or screening tools comprised of 1-4 items (Mitchell & Coyne 2007), from existing brief measures. For example, the Patient Health Questionnaire-2 (PHQ-2; Kroenke, Spitzer, & Williams, 2003) is the ultra-brief version of the of the PHQ-9 (Kroenke, Spitzer, & Williams, 2001).

Brief measures of self-stigma would also ease the assessment of self-stigma in applied clinical settings such as University Counseling Centers. Self-stigma inhibits comfort self-disclosing to a counselor (Seidman, Lannin, Heath, & Vogel, 2019), deters treatment adherence, and contributes to early therapy dropout (Wade Post, Cornish, Vogel, & Tucker, 2011). Understanding a client's self-stigma could thus help practitioners identify and address this barrier in therapy. As Berle et al. (2011) note, "Clinicians also need to balance the demands of multiple measures that achieve a suitable breadth of coverage of disorders and clinical phenomena, while not burdening clients with hours of self-report questionnaires," (p. 345). Developing an ultra-brief self-stigma screener could thus help clinicians address self-stigma without burdening clients.

The Present Study

The SSOSH is the most widely used measure of self-stigma of seeking psychological help. However, reverse-scored items present potential validity concerns. Development of briefer versions of the SSOSH could address these concerns by removing problematic or uninformative items and also benefit researchers and practitioners. Revised and ultra-brief versions of the SSOSH could expand the field's understanding of self-stigma by increasing its accessibility for researchers conducting large-scale studies, longitudinal research, and/or research

with underrepresented populations. From a clinical standpoint, an ultra-brief, screener version of the SSOSH could allow medical physicians referring patients to seeking psychological help to consider self-stigma and allow clinicians to deter client dropout by addressing self-stigma in therapy. Therefore, in the current study, we aim to develop revised and ultra-brief versions of the original SSOSH.

To accomplish this, we conducted Item Response Theory and correlational analyses across community (Sample 1) and undergraduate (Sample 2) samples. Item response theory (IRT) allows for the placement of items and people on the same scale, while accounting for the difficulty of each item and the individual's responses to each item, rather than a total score (Hambleton, Swaminathan, & Rogers, 1991). It is then possible to determine where items provide the most information along the latent trait scale. With this information, researchers can select a subset of items to create a short form of the original scale that provides similar scoring of individuals as the original scale. Therefore, to address the current concerns, the current study used IRT to develop revised short and revised ultra-brief versions of the SSOSH.

Using IRT, we selected items based on responses to the original SSOSH in Sample 1 and used responses to SSOSH items in Sample 2 to cross-validate the selection of items. We then examined bivariate correlations of the full SSOSH-10 and the revised and ultra-brief SSOSH scales with each other and with external constructs in the nomological help-seeking net. This included public stigma of seeking help, attitudes toward seeking help, and intention to seek help. To further bolster our examination of validity, we used different measures of attitudes and intention across samples. Based on psychometric literature (e.g., Dalal & Carter, 2015; DeVellis, 2015), we hypothesized that the reverse-scored items will yield the least information and, therefore, be more likely removed as part of the development of the new versions of the scale (Hypothesis 1). Consistent with the extant help-seeking literature (e.g., Brenner et al., 2020; Tucker et al. 2013) we hypothesized that the new versions of the SSOSH would demonstrate positive correlations with public stigma and negative correlations with attitudes towards seeking help and intention to seek help (Hypothesis 2).

Method

Participants

Per recommendations when conducting an IRT analysis with the graded response model (Reise & Yu, 1990), we determined 500 participants as the minimum size of each sample.

Sample 1. Community adults (N = 857; Age, *M* = 43.6*, SD* =16.1*, Range* =18-85; 69.1% women) were recruited from ResearchMatch, a national health volunteer registry created by several academic institutions and supported by the U.S. National Institutes of Health as part of the Clinical Translational Science Award program. Approximately 67% of the sample identified as White, 11% as African-American/Black, 7% as Latino/a, 7% as Multiracial, 6% as Asian American/Pacific Islander, 1% Other, and 1% Native American/Alaskan Native. Approximately 1% reported having less than a high school diploma, 4% earned a high school diploma or GED, 9% earned an associate's degree or attended vocational school, 14% had some college, 35% earned a bachelor's, 38% earned a graduate/professional degree, and 1% preferred not to answer. Approximately 27% reported they had never sought help from a mental health professional, while 72% reported they had.

Sample 2. A total of 661 undergraduate students (62.3% female) at a large Midwestern university were recruited from the psychology department subject pool (Age, *M* = 22.0, *SD* = 2.2, *Range* = 18-47). The sample included students in their first (53.6%), second (25.0%), third (13.4%), fourth (7.9%), and other (0.2%) year. Participants reported they were European American White (83.5%), Asian American/Pacific Islander (5.9%), International student (5.1%), Latino-American (1.8%), Multi-racial American (1.7%), Black or African American (1.1%), and other (0.9%). These demographics are similar to the demographics of the university as a whole. Of these participants, 20.7% reported previously seeking psychological help.

Procedure

Review and approval for the study procedures were obtained from the University Institutional Review Board. Sample 1 participants were contacted via the ResearchMatch registry messaging system regarding a study about therapy. Interested participants were directed

to an online survey that began with an informed consent page, continued with the survey items, and ended with a conclusion page. Participants had the option of entering a drawing for one of several $25 Amazon.com gift cards. The SSOSH item data used in this study was derived from a larger dataset focused on a developing a non-stigma construct related to seeking psychological help (masked for blind peer review). Sample 2 students signed up for the study via the subject pool website. Interested participants were directed to an online survey that began with an informed consent page, continued with the survey items, and ended with a conclusion page. Participants earned research credit in their beginning psychology or communications course.

Measures

   Self-stigma of seeking help (Samples 1 and 2). The Self-Stigma of Seeking Help scale (SSOSH; Vogel et al., 2006) is a 10-item measure of self-stigma of seeking psychological help. Participants rate items on a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The SSOSH includes items such as "I would feel inadequate if I went to a therapist for help," (Vogel et al. 2006, p. 328) as well as reversed-scored items such as "My self-confidence would NOT be threatened if I sought professional help," (Vogel et al. 2006, p. 328). Previous studies indicate positive correlations of the SSOSH scores with other forms of stigma (e.g., societal stigma, self-stigma of mental illness; Tucker et al., 2013) and negative correlations with help seeking attitudes (Tucker et al., 2013), intentions (Brenner et al., 2020), and behavior (Lannin et al., 2016). SSOSH scores demonstrate internal consistency among community ($\alpha = .92$) samples and college ($\alpha = .90$) samples (Tucker et al., 2013), as well as 2-month test-retest reliability (.72) among college students (Vogel et al., 2006). In this study, SSOSH scores demonstrated internal consistency estimates of .90 in Sample 1 and .86 Sample 2.

   Public stigma of seeking help (Samples 1 and 2). The Stigma Scale for Receiving Professional Psychological Help (SSRPH; Komiya et al., 2000) is a 5-item measure of perceived societal stigma toward people who seek psychological help. Participants rate items on a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). A higher score indicates greater public stigma. The SSRPH score has demonstrated validity through positive links with measures

of public toward mental illness and self-stigma of seeking help (e.g., Tucker, et al., 2013), and inverse relationship with attitudes toward seeking help (Tucker et al., 2013). The SSRPH scores demonstrated internal consistency among community adults (α = .82) and college students (α = .76; Tucker et al., 2013), similar to the current study Sample 1 (α = .80) and Sample 2 (α = .70).

Attitudes toward seeking help (Sample 1). The Mental Help Seeking Attitudes Scale (MHSAS; Hammer, Parent, & Spiker, 2018) is a nine-item measure of participants' evaluation (unfavorable vs. unfavorable) of seeking psychological help. Using a 7-point Likert scale, participants rate semantic differential scale items anchored by bipolar adjectives on either end (e.g., Useless to Useful). A higher score indicates a more positive attitude toward seeking help. The MHSAS items demonstrated internal consistency across multiple samples (α's > .92) and temporal stability (intraclass correlation coefficient [2,k] = .86) over three weeks (Hammer et al,. 2018). The MHSAS also demonstrated convergent/concurrent validity through positive associations with other measures of help-seeking attitudes and with help seeking subjective norms, perceived behavioral control, and anticipated utility of seeking help. Known-group evidence of validity was provided when women and those who had previously sought mental health services reported more favorable attitudes than did their demographic counterparts. In Sample 1 of the present study, the MHSAS scores demonstrated internal consistency (α = .93).

Intentions to seek help (Sample 1). The Mental Help Seeking Intention Scale (MHSIS; Hammer & Spiker, 2018) is a three-item measure of respondents' intention to seek help from a mental health professional if they had a mental health concern. A higher score indicates greater intention to seek help. The MHSIS demonstrated predictive evidence of validity by predicting (with 69.7% accuracy) the future help-seeking behavior of community adults with a current mental health concern. The MHSIS also demonstrated internal consistency (α = .94; Hammer & Spiker, 2018), similar to the current study (α = .95).

Attitudes toward seeking help (Sample 2). The Attitudes Toward Seeking Professional Psychological Help–Short Form (ATSPPH–SF; Fischer & Farina, 1995) is a 10-item measure of attitudes toward seeking help. Participant rate their level of agreement with items (e.g.,

"Psychotherapy would not have value for me") on a 4-point Likert-scale from 0 (*disagree*) to 3 (*agree*). Five items are reversed scored such that higher scores reflect more positive attitudes toward seeking psychological help. The ATSPPH-SF scores demonstrates validity through strong correlations with the original ATSPPH ($r = .87$; Fischer & Farina, 1995), positive associations with intentions to seek help, and inverse associations with self-stigma of seeking help, self-stigma of mental illness, and public stigma of seeking help (Tucker et al., 2013). The scale has demonstrated internal consistency in community ($\alpha = .84$) and college student ($\alpha = .79$) samples (Tucker et al., 2013). In Sample 2 of the present study, the ATSPPH–SF scores demonstrated internal consistency e ($\alpha = .86$).

Intentions to seek help (Sample 2)*.* Consistent with help-seeking research (e.g., Brenner et al., 2020), the 10-item *Psychological and Interpersonal Concerns* subscale of the 17-item *Intentions to Seek Counseling Inventory* (ISCI; Cepedo-Benito & Short, 1998) was used to capture intentions to seek psychotherapy. Participants rate the likelihood that they would seek help if they were experiencing each problem (e.g., loneliness, depression) on a 4-point scale from $1 = $ *very unlikely* to $4 = $ *very likely*. Higher scores indicate greater likelihood of seeking psychological help. This subscale has demonstrated positive associations with attitudes toward seeking psychotherapy and previous help seeking (Vogel et al., 2006). Scores have demonstrated internal consistency in college student samples ($\alpha = .89$; Brenner et al., 2020), similar to the current study ($\alpha = .84$).

Data Analysis Plan

Fitting the IRT model (Sample 1). We fit the polytomous IRT model, Samejima's graded response model (GRM; 1997), to Sample 1 responses to the ten SSOSH items. The GRM estimates the probability of a response in category *x* or higher, given a participant's underlying trait, θ_i:

$$P\left(u \geq x \mid \theta_i\right) = \frac{e^{a_j\left(\theta_i - b_{jx}\right)}}{1 + e^{a_j\left(\theta_i - b_{jx}\right)}}, \tag{1}$$

where i refers to the person, a_j is item discrimination and b_{jx} is the location of one of the item thresholds. Adjacent probabilities are subtracted to calculate the probability of a specific response. The mirt package (Chalmers, 2012) for the R environment (R Core Team, 2013) was used to fit the model. Estimation used the fixed-quadrature point expectation-maximization (EM) algorithm, with 61 quadrature points. IRT-based person scores were calculated using the expected *a-posteriori* (EAP) estimator.

Unidimensionality is an assumption of the GRM. To meet this assumption, the first factor should account for at least 20% of the variance (Hattie, 1985). As such, our first step was to conduct an exploratory factor analysis of the 10 items using principal axis factoring. After fitting the IRT model, we used Yen's *Q1* (Yen, 1981) to evaluate item fit, where the null hypothesis is that the data fit model exactly; the *Q1* statistic follows a χ^2 distribution. To avoid over-identification of misfit due to chi-square sensitivity to sample size, standardized residuals were inspected for any items whose *Q1* statistic indicated possible misfit. Finally, we looked at the standardized residuals across all items, which should be unimodal and centered around 0.

We also conducted a differential item functioning (DIF) analysis, using ordinal logistic regression, to consider possible violations of unidimensionality related to gender. That is, we conducted the DIF analysis to determine whether an individual's response to an item was a function of not just the underlying latent trait, but of gender as well. To conduct the DIF analysis, we fit three ordinal logistic regression models, predicting item response, for each item. In the first model, the latent trait, as estimated by the IRT-based score, is the only coefficient; in the second model, we added a variable indicating gender; the third model included the latent trait, gender, as well as the interaction of the latent trait and gender. An item was considered to have exhibited DIF if the coefficient for gender in the second model or the coefficient for the interaction between gender and the latent trait in the third model was statistically significant at an alpha of .01.

Item selection (Sample 1). After fitting the GRM, we inspected the resulting item response curves and information functions. Per Hypothesis 1, we considered the item content

and the resulting psychometric information, such as item response curves, item information, and standardized residuals, to identify items for removal from or inclusion in possible short forms. We selected items for a revised version and a subset of those for an ultra-brief version. The standard error of the IRT-based estimates for both versions of the scale were also computed:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}, \tag{2}$$

where $I(\theta)$ is the information function. We then compared the standard errors for the short forms to those from the original 10-item scale. The correlations between the total scores from the newly created short forms were calculated and compared to the full 10-item scale scores.

Cross-validation of the new scales (Sample 2). In Sample 2, we fit the GRM to responses to the 10 SSOSH items to evaluate whether the items functioned similarly across both samples with respect to relative item difficulty and item information. We then computed scores for the revised and ultra brief versions of the SSOSH for the Sample 2 participants; the correlations between each of the three SSOSH versions and external measures were examined.

Criterion-related validity (Samples 1 and 2). We computed scores for the revised, ultra-brief, and original SSOSH and examined correlations between each of the three SSOSH versions with existing measures in the help-seeking stigma nomological net.

Results

Fitting the IRT model

The EFA on the community sample, Sample 1, indicated that the data were unidimensional enough for IRT, with the first factor explaining 48.7% of the variance, exceeding the 20% minimum. We then fit the GRM to the 10 SSOSH items (see Table S1 for the item parameters and fit statistics). When the GRM was estimated, the average IRT-based score for the given sample was fixed to be 0. Six items were flagged for misfit with Yen's $Q1$ statistic (See Supplementary Table 1). As described in the next section, graphs of their standardized residuals were then inspected to avoid over-identification of misfit due to chi-square sensitivity

to sample size. As an example, in Figure 1 we display the category response curves for item 1, which indicate that a person with an average IRT-based score would be most likely to respond *Disagree* to item 1. Specifically, about 70% of individuals with average self-stigma scores are likely to respond *Disagree*, 20% *Strongly disagree*, 8% *Agree & disagree equally*, 2% *Agree*, and 0% (<.001%) would be predicted to respond with *Strongly agree*. The individual item information curves are provided in Figure 2, which demonstrates that item 1 is the most informative for latent scores greater than about one standard deviation below the mean. For latent scores smaller than one standard deviation below the mean none of the items provided much statistical information.

DIF analysis results indicated that no items functioned differentially due to gender (see Table 2). In other words, the model fit as expected in that only the level of the underlying latent trait was needed to predict a response. Gender did not affect the predicted response for any item.

Item Selection

As illustrated in Figure 2, three items (4, 5, and 9) provided almost no information. In addition to providing almost no information, these items were flagged for misfit by Yen's *Q1* statistic; that is, their standardized residuals did not reflect a random pattern, as would be expected if the model fit the data due to measurement error. Instead, their residuals showed a systematic under- or over-prediction of response category, or distinct patterns such as the observed values consistently smaller than expected in the middle of the latent trait scale and greater than expected at the end. Taking these separate issues together, the low information and lack of fit indicated that the GRM and the underlying latent trait did not adequately reflect the actual responses and that these three items do not belong in the same scale as the remaining items. Therefore, we removed these items to create the revised SSOSH-7. The GRM was fit to these seven items and we found adequate item-level and model-level fit.

Consistent with Hypothesis 1, the inadequate items are all reverse-scored. However, two reverse-scored items (2 and 7) provided adequate information across the scale. Items 1, 6, and 8 provided the most information and were retained for the ultra-brief SSOSH-3. Thus, a seven-

item (1, 2, 3, 6, 7, 8, and 10) revised and three-item (1, 6, and 8) ultra-brief forms were selected. The SSOSH-3 (α = .87) and SSOSH-7 (α = .89) were highly correlated with the SSOSH-10 (r's $\geq$ .91, p < .001) and each other (r = .94, p < .001), providing further evidence that they measure the same construct. The ultra-brief SSOSH-3 and revised SSOSH-7 are presented in Table 1.

Retained items provide good information across the middle and higher ends of the self-stigma latent trait (see Figure 3). Those with estimated scores less than -1.0 would still be estimated to be placed on the low end of the scale, even if the exact location is less precise.

Cross Validation of the SSOSH-3 and SSOSH-7

Next, we examined whether the item selection results would replicate in a new sample. We fit the GRM again to the responses from Sample 2, our undergraduate sample, to the 10 SSOSH items. Item parameters for both samples, with those of Sample 2 linked onto the scale of Sample 1, are provided in Table S1. The linking was done using the Stocking-Lord method in the R package plink (Weeks, 2010). Consistent with Sample 1, items, 4, 5, and 9 provided almost no information. We found a proxy for item location by averaging the four threshold parameters for each item. Using this information, we could see that the relative position of the items was about the same in both samples and the correlation between the item locations in the two samples was .92. In addition, the correlation between the item discrimination values was .84. These similarities indicated that the items functioned similarly for both samples and we could proceed to test the proposed short versions of the SSOSH. As with Sample 1, the SSOSH-3 (α = .82) and SSOSH-7 (α = .87) were highly correlated with the SSOSH-10 (r's $\geq$ .89) and with each other in Sample 2 (r = .93), providing further evidence that they measure the same construct.

Criterion-Related Validity of the SSOSH-3 and SSOSH-7 across Samples

We examined whether the SSOSH-3 and SSOSH-7 demonstrate similar bivariate correlations with criterion-related outcomes (public stigma, attitudes, and intention) as the original SSOSH across samples. We present these bivariate correlations in Table 3. Consistent with Hypothesis 2, the SSOSH-3 and SSOSH-7 were positively linked with public stigma and inversely linked with attitudes and intention. These associations were similar across all three

SSOSH versions (e.g., intention *r*'s = .22–.26, *p*'s < .001). Moreover, all forms of the SSOSH demonstrated similar correlations with public stigma and attitudes across samples. For example, correlations with public stigma for the SSOSH-3, SSOSH-7, and SSOSH-10, respectively, were .58, .59 .60, and .59 in Sample 1 and .50, .51, and .48 in Sample 2 (*p*'s < .001).

Discussion

The current study used item response theory to create ultra-brief (SSOSH-3) and revised (SSOSH-7) versions of the original SSOSH (SSOSH-10). To develop the new versions, we evaluated individual item performance—based on item fit and amount of information provided— in a community sample, and then replicated these findings in a college student sample. The best three and seven items were retained to create the SSOSH-3 and SSOSH-7, respectively. These items provide good information across the middle and higher ends of the self-stigma latent trait. Both new versions of the SSOSH were highly correlated with the original SSOSH-10. In support of criterion-related validity, the SSOSH-3, SSOSH-7, and SSOSH-10 demonstrated similar correlations with help-seeking constructs (i.e., public stigma, attitudes, intention) in the hypothesized directions in each sample. DIF analysis indicated that these items functioned similarly across men and women. Gender did not affect the predicted response for any item.

It is important to note that the creation of the short forms was not driven solely by the IRT parameters and item information. We first developed the hypothesis that reverse-scored items would be removed as they may in inadvertently reflect attitudes or expectations regarding the efficacy of therapy rather than direct self-stigmatization for seeking help. Therefore, item selection was driven by predictions developed prior to IRT analyses that were confirmed through IRT parameters and item information. Consistent with predictions, the three worst-performing items were reverse-scored. These items provided almost no information, justifying their removal from the scale. More broadly, results support the notion that reverse-score items tend to perform worse than regularly scored items. However, results also indicate that some reverse-scored items can perform well. Examining information curves may, therefore, be one way to determine how to include reverse-scored items for researchers who wish to create or revise a measure with reverse-

scored items due to some of the benefits they entail. Without such an examination, inclusion of reverse-scored items might create a risk of redundancy or retention of problematic items.

The SSOSH-3 and SSOSH-7 provide good information across the middle and higher ends of the self-stigma latent trait. This is a strength as more precisely placing respondents on the middle and upper ends of the scale is critical for understanding the impact of self-stigma by those more likely to be negatively influenced by it. In addition to providing good information across the middle and higher ends of the self-stigma latent trait, for almost all respondents the standard errors are very small and almost identical across all SSOSH versions. Although the SSOSH-3 and SSOSH-7 have larger standard errors on the lower end of the scale (see Figure 3), scores in this lower region of the scale were limited to a small percentage of participants; only 10.9% of both samples had scores less than -1.0 SD on the self-stigma scale and only 4.7% had scores less than -1.5 SD. More notably, it is not particularly important to place respondents with high precision on the low end of the scale, as this reflects people who are less likely to avoid seeking help due to self-stigma.

Generalizability of the SSOSH-3 and SSOSH-7

Study results support the generalizability of the SSOSH-3 and SSOSH-7 across unique samples. For example, examination of item functioning across a community sample and a college student sample supported the use of these measures with community adults as well as college students, the latter a common population of interest in the help seeking literature (e.g., Brenner et al., 2020; Lannin et al., 2016; Keum, Hill, Kivlighan, & Lu, 2018). Results provide initial support for the use of the SSOSH-3 and SSOSH-7 among those who have and have not sought help, given the majority of Sample 1 participants had sought help and the majority of Sample 2 participants had not. This is consistent with measurement invariance findings supporting the use of the SSOSH-10 across help-seeking history (Brenner, McDermott, Hammer, & Tucker, 2019), though this would need to be tested with the SSOSH-3 and SSOSH-7 items in future research. In addition, the DIF analysis supported the use of the SSOSH-3 and SSOSH-7 across men and women. Given that both samples demonstrated the same patterns of item

behavior, we can be confident that these findings are sample independent. Beyond item functioning, the SSOSH-3 and SSOSH-7 demonstrated similar relationships with public stigma and attitudes toward seeking psychological help across samples, providing further support for the instruments. The consistent findings across these groups suggest that the SSOSH-3 and SSOSH-7 may generalize to other populations, though this requires future research.

Content Implications

Although we used IRT to maximize information retained, it is important to consider the content implications for removing items in creating the SSOSH-7 and SSOSH-3. As illustrated in Figure 1, the SSOSH-7 and SSOSH-10 elicit similar levels of information across different levels of latent self-stigma. Moreover, the three items removed to create the SSOSH-7 (see Table 1) seemed to reflect attitudes or expectations regarding the efficacy of therapy (e.g., item 4, "My self-esteem would increase if I talked to a therapist"). Therefore, any lost information may reflect variance unwanted in the construct to begin with.

In examining Figure 1, it appears information may be lost in reducing from seven to three items. This makes sense given the ultra-brief length of the SSOSH-3. In developing the SSOSH-10 items, Vogel et al. (2006) focused on four content areas impacted as part of the self-esteem reduction inherent in self-stigma: self-regard, satisfaction with oneself and one's ability, and overall sense of worth as a person. While all of these aspects of self-esteem are interrelated, the reduction in self-confidence may be more present in the SSOSH-7 than the SSOSH-3 because it is explicitly mentioned (Item 3). Researchers interested in more fully capturing this self-confidence content area might choose to use SSOSH-7. However, the IRT results suggest that the SSOSH-3 items captures a meaningful level of information. This is consistent with the similar pattern and strength of relationships of the three SSOSH versions with outcome measures, as well as the high correlations with each other (r's $\geq$ .89). Taken together, it appears that the SSOSH-3 and SSOSH-7 items measure a similar construct.

However, a limitation of the current study is that we used participant responses to the SSOSH-10 to create the SSOSH-3, SSOSH-7, and SSOSH-10 scores, which we then used in our

validity analyses. Ideally, we would collect participant responses to each measure presented separately, even if in the same sitting. That is, participants could respond to other measures between the SSOSH-3, SSOSH-7, and SSOSH-10. This would help mitigate inflated correlations with each other due to shared error or systemic error effects (Smith et al., 2000). This would also allow for more valid testing of convergent validity (Smith et al., 2000). Although IRT might address most of the concerns that researchers face when developing a briefer measure (Smith et al., 2000), future researchers should continue to test the validity of the short forms through examining correlations between separate administrations of the three forms as well as validating the short forms in independent samples.

Use of the SSOSH-3 and Development of Benchmark Scores

The SSOSH-3 could be a useful tool for expanding the consideration of self-stigma in large-scale research such as epidemiological medical studies and nationally representative public health assessments, which tend to use brief versions of existing measures (McDermott et al., 2019). An ultra-brief measure may also help address current limitations in help-seeking research, such as the paucity of longitudinal research, lack of access to difficult to reach populations, and need for validated benchmarks in identifying persons at-risk for avoiding seeking help due to self-stigma. The SSOSH-3 might increase the feasibility of longitudinal research that examines actual use of services over time, as opposed intention to seek help as examined in the majority of the help-seeking literature. In addition, collecting data from small or difficult-to-access populations could help the field examine and address the potential differential impact of self-stigma on health disparities across privileged and marginalized populations.

Finally, identifying and validating benchmarks is an important next step in using the SSOSH-3 as a screening tool in applied settings. Self-stigma does not have known groups diagnostic categories with distinct criteria. Therefore, similar to other constructs with this limitation, researchers might develop benchmarks based on expert ratings as a starting point for future benchmark development and refinement (e.g., Modified Angoff method; see Cizek & Bunch, 2007). To validate and refine these benchmarks, researchers then might focus on

treatment seeking behavior, the outcome of interest. For example, they might assess the changes in likelihood of seeking help across scores (e.g., Stratum-Specific Likelihood Ratios; see Furukwaka, Kessler, Slade, & Andrews, 2003). Researchers should also consider the specific behavior assessed. For example, researchers might focus on help-seeking behavior after a referral rather than help-seeking behavior after a given period of time without a referral (e.g., online longitudinal survey study). Development of these benchmarks may be particularly impactful in applied settings. For example, medical providers could tailor their referral approach with distressed patients with high self-stigma, and clinicians may be able to decrease dropout rates by screening for self-stigma at intake and determining whether self-stigma should be addressed in therapy.

Additional Limitations and Future Directions for Research

While the psychometric strengths of the SSOSH-3 and SSOSH-7 were demonstrated across two unique samples drawn from different populations, the study samples are comprised mostly of white, heterosexual, cisgender female participants. Future research is needed to examine the utility of these items in more diverse groups of respondents. For example, the SSOSH-7 contains items that did not demonstrate metric and scalar configural invariance across all six countries in a 2013 study (Vogel, Armstrong et al.), yet these items were removed in the SSOSH-3. Although the SSOSH-7 captures more information, future research may indicate that the SSOSH-3 is a more appropriate measure among certain nations or cultures. Research examining men's help-seeking also indicates that self-stigma decreases as education level and income level increase, and that men living in rural areas are more prone to self-stigma (Hammer, Vogel, & Heimerdinger-Edwards, 2013). Thus, it is possible that the nature of self-stigma varies across these groups. Future researchers should conduct measurement invariance testing and other psychometric evaluation measures between the groups mentioned above as well as other groups not centrally represented in the current samples (e.g., racial minorities, sexual minorities).

As mentioned earlier, the majority of Sample 1 participants had previously sought help. Fortunately, we found the same item functioning results across samples (see Table S1).

However, additional research is needed to ensure the validity of these items across those who have and have not sought psychological help. Differences between these two groups of respondents may most likely be present in responses to the removed items that may have incidentally assessed the efficacy of therapy; the majority of Sample 1 participants could respond based on personal experience rather than solely on stereotypes and messages from other sources. Again, this should be examined in future research, including clinically distressed samples with fewer individuals who previously sought help. Likewise, the present analyses did not use samples restricted to only those reporting mental illness or clinically significant distress. Given that this is a population of keen interest to help-seeking researchers, another avenue for future research would be to verify the psychometric performance of these new versions with samples of exclusively clinically distressed respondents. Previous invariance testing supports the use of the SSOSH-10 across distress levels and help-seeking history (Brenner et al., 2020) and suggests findings will generalize for the SSOSH-3 and SSOSH-7; however, this should be examined in future work. Furthermore, both samples were drawn from English-speaking respondents residing in the United States. Given that the SSOSH-10 has been used in more than 12 countries, it would be useful to verify the utility of these two versions with international populations.

Conclusion

The current research used IRT to improve measurement of the self-stigma of seeking help. We eliminated redundant, uninformative items from the SSOSH-10 to develop a revised SSOSH-7 and retained the three most informative items to develop an ultra-brief SSOSH-3. Findings support the structural, convergent, and criterion-related validity of the SSOSH-3 and SSOSH-7. Removal of these uninformative items will improve the future efficiency of measurement of self-stigma. It may increase inclusion in research settings (e.g., longitudinal studies, large-scale studies, studies of underrepresented populations). The reduced length of the ultra-brief version could also function as a screening tool and thus facilitate increased assessment of self-stigma by medical practitioner and clinicians in applied settings.

References

Berle, D., Starcevic, V., Moses, K., Hannan, A., Milicevic, D., & Sammut, P. (2011). Preliminary validation of an ultra-brief version of the Penn State Worry Questionnaire. *Clinical Psychology & Psychotherapy*, *18*(4), 339-346. doi:10.1002/cpp.724

Brenner, R. E., McDermott, R. C., Hammer, J. H., & Tucker, J. R. (2019, August). *Measurement Invariance of the Self-Stigma of Seeking Help Scale (SSOSH)*. Poster presented at the 127th Annual convention of the American Psychological Association, Chicago, IL.

Brenner, R. E., Cornish, M. A., Heath, P. J., Lannin, D. G., & Losby, M. (2020). Seeking help despite the stigma: Experiential avoidance as a moderated mediator. *Journal of Counseling Psychology, 67*(1), 132-140. doi: 10.1037/cou0000365

Cepeda-Benito, A., & Short, P. (1998). Self-concealment, avoidance of psychological services, and perceived likelihood of seeking professional help. *Journal of Counseling Psychology, 45*, 1-7. doi: 10.1037/0022-0167.45.1.5

Chalmers, R., P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29. doi: 10.18637/jss.v048.i06

Cizek, G., & Bunch, M. (2007). The Angoff method and Angoff variations. Standard Setting. *Thousand Oaks, California: SAGE Publications, Inc*, 81-96.

Cornish, M. A., Brenner, R. E., Vogel, D. L., & Wade, N. G. (2019). Evaluation of an online help seeking stigma-reduction intervention for military personnel. *Stigma and Health*, *4*(4), 480–486. doi: 10.1037/sah0000167

Dalal, D. K., & Carter, N. T. (2015). Negatively-worded items negatively impact survey research. In C. E. Lance, & R. J. Vandenberg (Eds.), *More Statistical and Methodological Myths and Urban Legends* (pp. 112-132). New York, NY: Routledge.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Fischer, E. H., & Farina, A. (1995). Attitudes toward seeking professional psychological help: A
     shortened form and considerations for research. *Journal of College Student
     Development, 36*(4), 368-373. doi: 10.1037/t05375-000

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6
     and K10 screening scales for psychological distress in the Australian National Survey of
     Mental Health and Well-Being. *Psychological medicine*, *33*(2), 357-362.

Hambleton, R. K., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response
     Theory*. Newbury Park: Sage Publications.

Hammer, J. H., & Parent, M. C., & Spiker, D. A. (2018). Mental Help Seeking Attitudes Scale
     (MHSAS): Development, reliability, validity, and comparison with the ATSSPH-SF and
     IASMHS-PO. *Journal of Counseling Psychology, 65*, 74-85. doi: 10.1037/cou0000248

Hammer, J. H., & Spiker, D. A. (2018). Dimensionality, Reliability, and Predictive Evidence of
     Validity for Three Help Seeking Intention Instruments: ISCI, GHSQ, and MHSIS.
     Journal of Counseling Psychology, 65, 394-401. doi: 10.1037/cou0000256

Hammer, J. H., Vogel, D. L., & Heimerdinger-Edwards, S. R. (2013). Men's help seeking:
     Examination of differences across community size, education, and income. Psychology
     of Men & Masculinity, 14(1), 65-75.

Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of a set of test items.
     Applied Psychological Measurement, *9*(2), 139-164.

Healthy Minds Network. (2019). *Research*. Retrieved from the Healthy Minds Network website:
     http://healthymindsnetwork.org/research/hms.

Kahn, J. H. (2006). Factor Analysis in Counseling Psychology Research, Training, and Practice:
     Principles, Advances, and Applications. *The Counseling Psychologist*, *34*(5). 684–718.

Keum, B. T., Hill, C. E., Kivlighan Jr, D. M., & Lu, Y. (2018). Group-and individual-level self-
     stigma reductions in promoting psychological help-seeking attitudes among college
     students in helping skills courses. *Journal of Counseling Psychology*, *65*(5), 661-668.

Komiya, N., Good, G. E., & Sherrod, N. B. (2000). Emotional openness as a predictor of college

      students' attitudes toward seeking psychological help. *Journal of Counseling Psychology,*

      47, 138–143. doi: 10.1037//0022-0167.47.1.138

Kroenke, K., Spitzer, R. L., Williams, J. B. (2001). The PHQ-9. Validity of a brief depression

      severity measure. *Journal of General Internal Medicine, 16*, 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B. (2003). The Patient Health Questionnaire- 2: Validity

      of a two-item depression screener. *Medical Care ,41,* 1284–1292.

Lannin, D. G., Guyll, M., Vogel, D. L., & Madon, S. (2013). Reducing the stigma associated

      with seeking psychotherapy through self-affirmation. *Journal of Counseling*

      Psychology, *60*, 508-519. doi:10.1037/a0033789

Lannin, D. G., Vogel, D. L., Brenner, R. E., Abraham, T. W., & Heath, P. J. (2016). Does self-

      stigma reduce the probability of seeking mental health information? *Journal of*

      Counseling Psychology, *63*, 351-358. doi: 10.1037/cou0000108

McDermott, R. C., Levant, R. F., Hammer, J. H., Borgogna, N. C., & McKelvey, D. K. (2019).

      Development and validation of a five-item male role norms inventory using bifactor

      modeling. *Psychology of Men & Masculinity, 20*(4), 467-477. doi:10.1037/men0000178

Mitchell, A. J., & Coyne, J. C. (2007). Do ultra-short screening instruments accurately detect

      depression in primary care? A pooled analysis and meta-analysis of 22 studies. *The*

      British Journal of General Practice: The Journal of the Royal College of General

      Practitioners, 57(535), 144-151.

Nakash, R.A., Hutton, J.L, Jørstad-Stein, E.C., Gates, S., Lamb, S.E. (2006) Maximising

      response to postal questionnaires – A systematic review of randomised trials in health

      research. *BMC Medical Research Methodology*, *6*(1). doi:10.1186/1471-2288-6-5

Substance Abuse and Mental Health Services Administration (2017). *National Survey on Drug*

      Use and Health. Retrieved from https://www.samhsa.gov/data/sites/default/files/cbhsq-

      reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.htm#tab8-39B

Pattyn, E., Verhaeghe, M., Sercu, C., & Bracke, P. (2014). Public stigma and self-stigma: Differential association with attitudes toward formal and informal help seeking. *Psychiatric Services*, *65*(2), 232-238. doi:10.1176/appi.ps.201200561

Pederson, E. L., & Vogel, D. L. (2007). Male gender role conflict and willingness to seek counseling: Testing a mediation model on college-aged men. *Journal of Counseling Psychology*, *54*(4), 373–384. doi: 10.1037/0022-0167.54.4.373

Reust, C. E., Thomlinson, R. P., & Lattie, D. (1999). Keeping or missing the initial behavioral health appointment: A qualitative study of referrals in a primary care setting. *Families, Systems, & Health*, *17*(4), 399–411. doi: 10.1037/h0089892

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rolstand, S., Adler, J., Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101-1108.

Seidman, A. J., Lannin, D. L., Heath, P. J., & Vogel, D. L. (2019). Setting the Stage: The Effect of Affirming Personal Values Before Psychotherapy Intake Screenings on Perceptions of Self-Stigma and Self-Disclosure. *Stigma and Heath, 4*(3), 256-259.

Seidman, A. J., Wade, N. G., Vogel, D. L., & Armistead-Jehle, P. (2019). The impact of stigma on behavioral health care utilization among active duty service members. *Military Psychology*, *31*(1), 11-17. doi:10.1080/08995605.2018.1522927

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 85-100). New York: Springer.

Sheldon, H., Graham, C., Pothecary, N., & Rasul, F. (2007). Increasing response rates amongst black and minority ethnic and seldom heard groups. *Europe: Picker Institute*. Retrieved from: http://www.repoa.or.tz/documents_storage/Publications/Reports/06.3_Kessy_and_Urio.pdf

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. Psychological Assessment, 12, 102-111.

Tucker, J. R., Hammer, J. H., Vogel, D. L., Bitman, R. L., Wade, N. G., & Maier, E. J. (2013). Disentangling self-stigma: Are mental illness and help-seeking self-stigmas different? *Journal of Counseling Psychology*, 60(4), 520-531. doi:10.1037/e631712012-001

Vogel, D. L., Armstrong, P. I., Tsai, P. C., Wade, N. G., Hammer, J. H., Efstathiou, G., Holtham, E., Kouvaraki, E., Liao, H. Y., Shechtman, Z., & Topkaya, N. (2013). Cross-cultural validity of the Self-Stigma of Seeking Help (SSOSH) scale: Examination across six nations. *Journal of Counseling Psychology*, *60*(2), 303-310. doi:10.1037/a0032055

Vogel, D. L., Bitman, R. L., Hammer, J. H., & Wade, N. G. (2013). Is stigma internalized? The longitudinal impact of public stigma on self-stigma. *Journal of Counseling Psychology*, *60*(2), 311–316. doi: 10.1037/a0031889

Vogel, D. L., Strass, H. A., Heath, P. J., Al-Darmaki, F. R., Armstrong, P. I., Baptista, M. N., Brenner, R. E., Gonçalves, M., Lannin, D. G., Liao, H-Y, Mackenzie, C. S., Mak, W. W. S., Rubin, M., Topkaya, N., Wade, N. G., Wang, Y. F., & Zlati, A. (2017). Stigma of seeking psychological services: Examining college students across ten countries/regions. *The Counseling Psychologist, 45,* 170-192.

Vogel, D. L., Wade, N. G., & Haake, S. (2006). Measuring the self-stigma associated with seeking psychological help. *Journal of Counseling Psychology*, 53, 325-337.

Wade, N. G., Post, B. C., Cornish, M. A., Vogel, D. L., & Tucker, J. R. (2011). Predictors of the change in self-stigma following a single session of group counseling. *Journal of Counseling Psychology*, *58*(2), 170–182. doi: 10.1037/a0022630

Weeks, J. P. (2010). plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, *35*(12), 1-33.

Yen, W. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, *5*, 245-262. doi:10.1177/014662168100500212

Table 1

Items retained from the SSOSH-10 with Item Numbers for the SSOSH-3 and SSOSH-7

SSOSH-10	Number in SSOSH-3	Number in SSOSH-7
1. I would feel inadequate if I went to a therapist for psychological help.	1	1
2. My self-confidence would NOT be threatened if I sought professional help.	—	2
3. Seeking psychological help would make me feel less intelligent.	—	3
4. My self-esteem would increase if I talked to a therapist.	—	—
5. My view of myself would not change just because I made the choice to see a therapist.	—	—
6. It would make me feel inferior to ask a therapist for help.	2	4
7. I would feel okay about myself if I made the choice to seek professional help.	—	5
8. If I went to a therapist, I would be less satisfied with myself.	3	6
9. My self-confidence would remain the same if I sought professional help for a problem I could not solve.	—	—
10. I would feel worse about myself if I could not solve my own problems.	—	7

Note. Participants are presented with, "Directions: People at times find that they face problems that they consider seeking help for. This can bring up reactions about what seeking help would mean. Please use the 5-point scale to rate the degree to which each item describes how you might react in this situation."

Participants rate items on the following scale: 1 (*Strongly Disagree*), 2 (*Disagree*), 3 (*Agree/Disagree Equally*), 4 (*Agree*), 5 (*Strongly Agree*).

Table 2

Results of DIF Analysis Across Gender

	Model 1		Model 2			Model 3					
Item	AIC	logL	Gender (SE)	AIC	logL	Gender (SE)	Gender*θ (SE)	AIC	logL	$LR_{1:2}$	$LR_{2:3}$
1	1086.22	-538.11	-0.12 (0.18)	1087.76	-537.88	-0.22 (0.21)	0.23 (0.25)	1088.93	-537.46	0.46	1.30
2	1728.49	-859.25	0.08 (0.15)	1730.22	-859.11	0.11 (0.16)	-0.09 (0.19)	1731.98	-858.99	0.27	0.51
3	1282.71	-636.35	0.07 (0.17)	1284.55	-636.28	0.08 (0.20)	-0.02 (0.24)	1286.54	-636.27	0.16	0.17
4	2086.93	-1038.47	-0.10 (0.14)	2088.41	-1038.20	-0.09 (0.14)	-0.10 (0.15)	2089.93	-1037.97	0.52	1.00
5	2107.15	-1048.58	0.13 (0.14)	2108.28	-1048.14	0.17 (0.15)	-0.16 (0.16)	2109.28	-1047.64	0.87	1.87
6	1118.32	-554.16	0.14 (0.18)	1119.72	-553.86	0.03 (0.20)	0.26 (0.25)	1120.62	-553.31	0.60	1.69
7	1228.41	-609.20	0.08 (0.17)	1230.19	-609.10	0.21 (0.20)	-0.31 (0.22)	1230.15	-608.07	0.21	2.26
8	1228.25	-609.12	-0.28 (0.17)	1227.41	-607.70	-0.37 (0.19)	0.21 (0.21)	1228.41	-607.21	2.84	3.83
9	2119.97	-1054.98	0.28 (0.14)	2118.01	-1053.00	0.28 (0.14)	-0.02 (0.16)	2119.99	-1053.00	3.96	3.98
10	1882.42	-936.21	-0.16 (0.15)	1883.27	-935.64	-0.16 (0.15)	0.04 (0.17)	1885.21	-935.60	1.15	1.22

Note: Gender = ordinal logistic coefficient for gender in Models 2 and 3, Gender*θ = ordinal logistic coefficient for gender by latent trait interaction in Model 3, AIC = Akaike information criterion, logL = log-likelihood, *LR* = likelihood ratio test, where *LR* = -2(log-likelihood of reduced model – log-likelihood of full model). None of the Gender or Gender*θ ordinal logistic coefficients were significant at *p* < .01. None of the likelihood ratio tests were significant at *p* < .01.

Table 3

Total scores from SSOSH10 and two SSOSH short forms with external measures for both samples.

Measures	SSOSH-10	SSOSH-7	SSOSH-3	Public Stigma	Attitudes	Intent
M(*SD*)Sample 2	2.7 (.65)	2.6 (.73)	2.6 (.86)	2.2 (.45)	2.6 (.40)	2.1 (.56)
SSOSH-10	—	0.97	0.89	0.44	-0.62	-0.26
SSOSH-7	0.98	—	0.93	0.47	-0.63	-0.24
SSOSH-3	0.91	0.94	—	0.50	-0.62	-0.24
Public Stigma	0.61	0.61	0.60	—	-0.31	-0.04
Attitudes	-0.60	-0.59	-0.57	-0.47	—	0.45
Intent	-0.59	-0.57	-0.54	-0.45	0.68	—
M(*SD*)Sample 1	2.2 (.78)	2.1 (.85)	2.1 (.95)	2.3 (.61)	5.7 (1.02)	5.5 (1.49)

Note: Descriptive statistics, and zero-order correlations are shown below the diagonal for Sample 1 (community adults, N = 857) and above the diagonal for Sample 2 (college students, N = 661). Public stigma was measured using the SSRPH. Attitudes was measured using the HSAS in Sample 1 and ATSPPH in Sample 2. Intent was measured using the MHSIS in Sample 1 and the ISCI in Sample 2. All correlations were significant at $p < .001$.
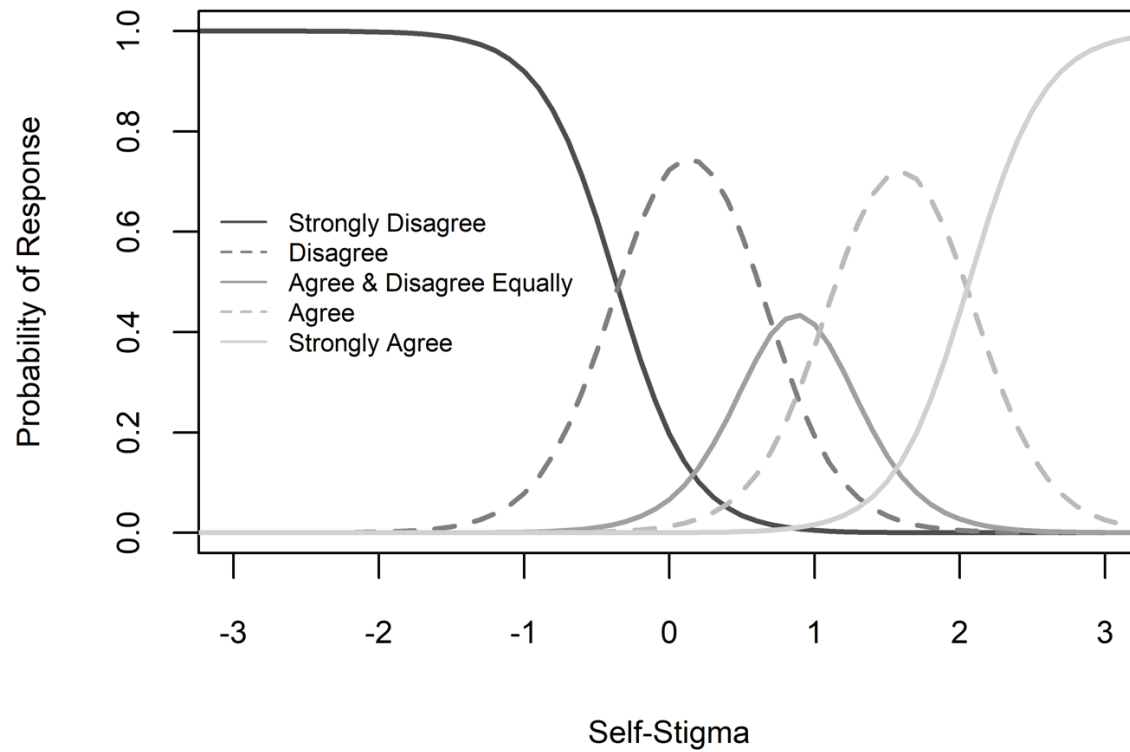
Figure 1. Item response curves for item 1 ("I would feel inadequate if I went to a therapist for psychological help") of the SSOSH
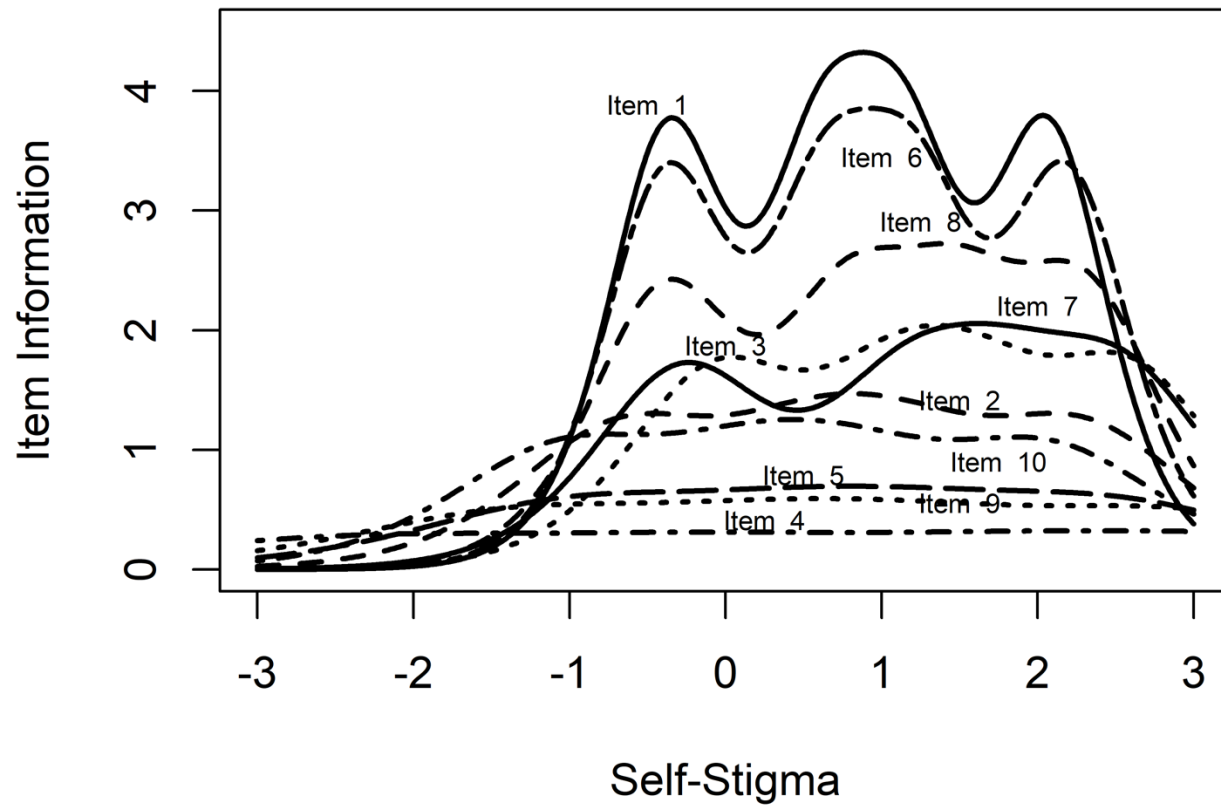
scale.
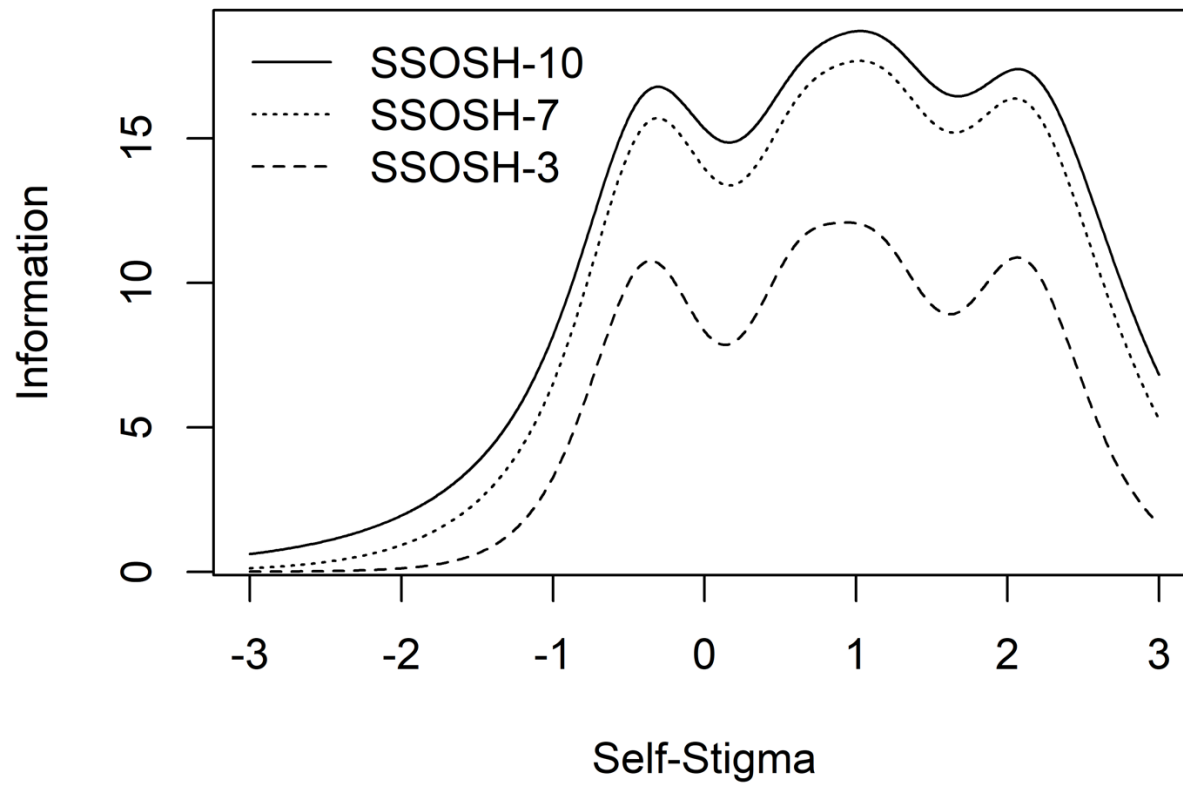
Figure 2. Item information curves for the ten SSOSH items.

Figure 3. Total test information for SSOSH-10, SSOSH-7, and SSOSH-3.