Supplemental Material
Table of Contents

Note: This Supplemental Material may not exactly replicate the final version of the Supplemental Material published in the APA journal. It is not the copy of record. Please use the DOI link on my website to access the PDF through your institution, allowing full access to the published type-set article.

To obtain a copy of the MHSAS and/or access the published article, please visit:
http://drjosephhammer.com/research/mental-help-seeking-attitudes-scale-mhsas/

APA-Style Citation for the Paper this Supplemental Material accompanies:

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

Mental Help Seeking Attitudes Scale (MHSAS)

INSTRUCTIONS: For the purposes of this survey, “mental health professionals” include psychologists, psychiatrists, clinical social workers, and counselors. Likewise, “mental health concerns” include issues ranging from personal difficulties (e.g., loss of a loved one) to mental illness (e.g., anxiety, depression).

Please mark the circle that best represents your opinion. For example, if you feel that your seeking help would be extremely useless, you would mark the circle closest to "useless." If you are undecided, you would mark the "0" circle. If you feel that your seeking help would be slightly useful, you would mark the "1" circle that is closer to "useful."

If I had a mental health concern, seeking help from a mental health professional would be...

	3	2	1	0	1	2	3	
Useless	◯	◯	◯	◯	◯	◯	◯	Useful
Important	◯	◯	◯	◯	◯	◯	◯	Unimportant
Unhealthy	◯	◯	◯	◯	◯	◯	◯	Healthy
Ineffective	◯	◯	◯	◯	◯	◯	◯	Effective
Good	◯	◯	◯	◯	◯	◯	◯	Bad
Healing	◯	◯	◯	◯	◯	◯	◯	Hurting
Disempowering	◯	◯	◯	◯	◯	◯	◯	Empowering
Satisfying	◯	◯	◯	◯	◯	◯	◯	Unsatisfying
Desirable	◯	◯	◯	◯	◯	◯	◯	Undesirable

Scoring Key

The MHSAS contains nine items which produce a single mean score. The MHSAS uses a seven-point semantic differential scale. Please note that the scale labels (3, 2, 1, 0, 1, 2, 3) are only provided to assist participants, and are not to be used in scoring the MHSAS. To counteract possible response sets, the valence of the item anchors was counterbalanced across the nine items. For example, the “useless – useful” item had the positively-valenced term (i.e., useful) on the right side of the scale, whereas the “important – unimportant” item had the positively-valenced term (i.e., important) on the left side of the scale. In order to properly calculate the MHSAS mean score, where a higher mean score indicates more favorable attitudes, it is necessary to reverse-code items 2, 5, 6, 8, and 9. After reverse coding, a score of “1” (the circle to the farthest left of the seven-point scale) on a given item should indicate an unfavorable attitude, a score of “4” (the middle circle of the seven-point scale) on a given item should indicate a neutral attitude, and a score of “7” (the circle to the farthest right side of the seven-point scale) on a given item should indicate a favorable attitude. Once reverse-coding is complete, calculate the MHSAS mean score by adding the item scores together and dividing by the total number of answered items. The resulting mean score should range from a low of 1 to a high of 7. For example, if someone answers 9 of the 9 items, the mean score is produced by adding together the 9 answered items and dividing by 9. Likewise, if someone answers 8 of the 9 items, the total score is produced by adding together the 8 answered items and dividing by 8. Per Parent’s 20% recommendation (2014; DOI: 10.1177/0011000012445176), a mean score should only be calculated for those respondents who answered at least 8 of the items. For more information about the MHSAS, please visit: http://DrJosephHammer.com

Supplemental Material for Anticipated Effect Sizes for Hypotheses 1 through 10

This section of the supplemental material provides citations for past empirical studies in which the relationship between help seeking attitudes and other constructs was analyzed. The effect sizes for these statistical results informed the hypothesized strength of effects that we anticipated finding when testing the relationship between help seeking attitudes, as measured by the MHSAS, and the various theoretically-related constructs embedded in the nomological network of help seeking attitudes. The expected effect size for each effect is presented in the same order as the hypotheses are presented in the manuscript. For brevity, each hypothesis is listed with two supporting citations. Please see http://DrJosephHammer.com/es1 for the full list of citations supporting each hypothesis.

1. Hypothesis 1: MHSAS should demonstrate a moderate to large association with subjective norms (Mo & Mak, 2009; Vogel, Wester, Wei, & Boysen, 2005).
2. Hypothesis 2: MHSAS should demonstrate a moderate to large association with perceived behavioral control (Hess & Tracey, 2013; Mak & Davis, 2014).
3. Hypothesis 3: MHSAS should demonstrate a moderate to large association with intention (Codd & Cohen, 2003; Mak & Davis, 2014).
4. Hypothesis 4: Those who have previously sought mental health services should report higher MHSAS score (small to large effect) than those who have never sought mental health services (Elhai, Schweinle, & Anderson, 2008; Masuda, Anderson, & Edmonds, 2012).
5. Hypothesis 5: MHSAS should demonstrate an inverse, moderate to large association with public stigma of seeking help (Elhai et al., 2008; Komiya, Good, & Sherrod, 2000).
6. Hypothesis 6: MHSAS should demonstrate an inverse, moderate to large association with self-stigma of seeking help (Hackler, Vogel, & Wade, 2010; Pederson & Vogel, 2007).

7. Hypothesis 7: Women should report a higher MHSAS score (small to medium effect) than men (Elhai et al., 2008; Masuda et al., 2012).
8. Hypothesis 8: MHSAS should demonstrate an inverse, moderate association with anticipated risks (Shaffer, Vogel, & Wei, 2006; Vogel & Wester, 2003)
9. Hypothesis 9: MHSAS should demonstrate an inverse, large association with anticipated utility (Shaffer et al., 2006; Vogel & Wester, 2003).
10. Hypothesis 10: MHSAS should demonstrate a large association with the ATSPPH-SF and IASMHS-PO. This anticipated effect size is not based on prior literature given we could find no published study that examined the correlation between the scores of these two instruments. Rather, it is based on the premise that two instruments measuring the same construct should correlate strongly (DeVellis, 2016).

Supplemental Material for the Study 1 Scale Development Section Regarding Measuring Evaluation and not the Potency or Activity Dimensions of Attitudes

This section of the supplemental material provides a rationale for including sixteen items measuring potency or activity in the initial item pool.

Extant empirical research has found that the items from semantic differential scales tend to fall into one of three dimensions: evaluation (e.g., good vs bad), potency (e.g., powerful vs. powerless), and activity (e.g., fast vs. slow; Heise, 1970). Measuring attitudes in the context of TPB requires that the items selected for the scale be evaluative in nature, and should not primarily tap the potency or activity dimensions (Ajzen, 2006). Ajzen recommends that items that typify the other two dimensions be included in the item pool so that factor analysis can be used to separate out items that load on the evaluation factor (eligible to be included in the MHSAS) from those that load on the other two dimensions (not eligible). This is particularly important when developing a new semantic differential scale, due to the issue of semantic instability, which is tendency for certain items to have different meanings in different contexts. For example, the terms "hot" and "cold" may mean one thing when evaluating a person but a completely different thing when evaluating a physical object. It is important that items selected for the MHSAS are demonstrably evaluative in the context of seeking professional mental health services. Given this, we added eight potency items (i.e., big/little, powerless/powerful, deep/shallow, high/low, long/short, heavy/light, thick/thin, hard/soft) and eight activity items (i.e., fast/slow, alive/dead, noisy/quiet, active/passive, burning/freezing, lazy/industrious, calm/excited, hot/cold) to the item pool. These 16 items were drawn from published lists of items that tend to load strongly on the potency and activity dimensions (e.g., Heise, 1970). The final pool consisted of 65 items, which were administered to Study 1 participants.

Supplemental Material for the Study 1 Analysis Plan Section Regarding Data Cleaning Procedures

This section of the supplemental material describes the data cleaning procedures used prior to conducting Study 1 analyses.

The initial, pre-cleaned dataset originally contained 982 individuals. Cases with significant (> 20%; Parent, 2013) missingness on items for any scale (n = 117) excluding demographic items, or more than one wrong response to the attention check items (n = 8) were deleted. Parent (2013) provided both real world and simulation study evidence that available item analysis and multiple imputation produce similar results across low levels of missing data when a) data is not missing not at random, (b) less than 10% of all data on each subscale is missing, (c) sample size is not small (i.e., significantly larger than 50 participants), (d) subscales demonstrate adequate internal reliability, and (e) subscales with any missing data contain more than four items. Therefore, available item analysis was used for all analyses conducted in SPSS and full information maximum likelihood (FIML) was used for all analyses conducted in M*plus*. Importantly, when analyses were re-run using the uncleaned dataset that included the 117 cases that had more than 20% missing data on any given subscale, results did not differ substantively (e.g., correlation between MHSAS and Intention went from .682 to .678, correlation between MHSAS and anticipated risks went from -.404 to -.393, and correlation between MHSAS and IASMHS-PO went from .504 to .502). No variables exceeded the cutoffs of 3 and 10 for high skewness index and kurtosis index values, respectively (Weston & Gore, 2006). In the retained sample (n = 857), missing data ranged from a low of 0% for many items to a high of 0.7% for four of the MHSAS items. Little's missing completely at random (MCAR) test was found to be non-significant (p = .85), indicating the missing cases were not significantly different from the non-missing cases. Subsequently, the entire Study 1 sample was used for convergent, incremental, and ME/I analyses.

Supplemental Material for the Study 1 Analysis Plan and Results Sections Regarding Exploratory Factor Analysis (EFA) Procedures

This section of the supplemental material articulates the analysis plan and results for the preliminary EFA's conducted prior to the IRT analysis. The analysis plan and results are presented together in the same narrative to ease comprehension.

We conducted a first EFA (EFA #1) with SPSS Version 23 on the 65-item pool using principal axis factor extraction and direct oblimin (oblique) rotation to confirm our assumption that the 49 evaluative items were indeed measuring the intended evaluation dimension, rather than the potency or activity dimensions. One thousand random Parallel Analysis data sets were also computed. Eigenvalues for the first four factors were higher in the actual data set (i.e., 23.80, 4.61, 2.3, 1.78, 1.49) than in the parallel analysis (i.e., 1.80, 1.71, 1.68, 1.64, 1.60). These results and the scree plot supported a four-factor solution. Thus, we re-ran the same EFA but specified the extraction of four factors (EFA #2). Examination of the pattern coefficients (see Supplemental Material Table A) revealed that nine of the 16 potency/activity items loaded most strongly on their own factor (Factor 3), whereas three loaded most strongly on Factor 1 (though they had lower primary factor loadings than most other items on that factor) and four loaded most strongly on Factor 2 (again, with lower primary factor loadings than most other items on that factor). Thus, it appeared that Factors 1, 2, and 4 were primarily evaluative in nature whereas Factor 3 was primarily assessing potency and activity. It also appeared that the potency and activity items that loaded on Factors 1 and 2 were demonstrating semantic instability (see prior description), in that they seem to be interpreted in an atypical way in this help seeking context. For example, whereas "passive-active" is typically an indicator of the activity dimension in most research contexts (Heise, 1970), when used to rate the act of seeking professional mental health help, it seems to take on an evaluative tone. In summary, given that all 49 of the items generated for the initial item pool primarily loaded on the evaluative Factors 1, 2, or 4, we concluded that these 49 items are evaluative in nature.

A third EFA (EFA #3), using the same procedures with just the 49 evaluative items, revealed the presence of a strong general evaluative help seeking attitudes factor (first eigenvalue was three times the size of the next eigenvalue; Cho et al., 2015). Therefore, we conducted a third, bifactor EFA using BI-GEOMIN rotation in M*plus* (Version 7.3; Muthén & Muthén, 1998-2012). We examined two, three, and four-factor bifactor EFA solutions. The three-factor solution demonstrated adequate model fit (S-B $\chi 2$ [1032] = 2012.80, $p < .001$; RMSEA = .044 [90% CI of .041, .047]; CFI = .938; TLI = .929; SRMR = .029), and was found to be most theoretically interpretable. Parallel Analysis (1,000 datasets) conducted in M*plus* also recommended a three-factor solution, given that eigenvalues for the first three factors were higher in the actual data set (i.e., 21.52, 3.76, 1.76, 1.29) than in the parallel analysis (i.e., 1.67, 1.60, 1.55, 1.51). Factor loadings for this three-factor bifactor EFA are presented in Supplemental Material Table B.

Factor 1 was the general evaluative help seeking attitudes factor, on which the 49 items loaded to varying degrees. In other words, the 49 items appeared to tap the general evaluative factor, while certain items also tapped narrower specific factors that are not the focus of the present investigation. Therefore, having identified that these items did indeed tap the intended evaluative dimension, the next task was to utilize IRT on the exploratory subsample to select a subset of items from the pool of 49 that would collectively provide optimal, unidimensional measurement of the general evaluative help seeking attitudes factor. See Method section of the manuscript for the Analysis Plan for this IRT analysis. It is important to note that the use of traditional IRT assuming unidimensionality was justified in this context because "smaller minor factors do not have consequential influences on estimated latent trait scores" (Toland, 2014, p. 130). When the 49 items were subjected to a confirmatory bifactor analysis (i.e., in which all 49 items were specified to load on the general evaluative factor, 13 of those 49 items were also specified to simultaneously load on a specific factor mirroring factor 2 from Supplemental Material Table B, and 6 of those 49 items were also specified to simultaneously load on a

specific factor mirroring factor 3 from Supplemental Material Table B), and the resulting standardized item factor loadings were plugged into Dueber's (2016) Bifactor Indices Calculator, the resulting ECV (.81) and OmegaH (.93) values suggested that the 49 item pool was indeed dominated by a strong general evaluative help seeking attitudes factor (Rodriguez, Reise, & Haviland, 2016). Thus, the item pool was essentially unidimensional, and thus appropriate for unidimensional IRT.

Supplemental Material for the Study 1 Analysis Plan Section Regarding Item Response Theory (IRT) Analyses

This section of the supplemental material provides a description of the advantages of using IRT to inform item selection.

A significant practical advantage of using IRT over classical test theory to inform item selection is in the utility of IRT in selecting items that range across different responses at these various levels of the construct of interest. In classical test theory, typically one conducts an EFA and selects the top-loading items as the final form of the measure. Practically, this may often result in measure composed of items that are highly correlated and generate high internal consistency values, but that are repetitive or overlapping rather than items that capture a range of participant responses. IRT aims to disperse item selection across the range of individuals at various levels of the underlying construct, resulting in selection of different items that, compared to an EFA, capture a wider range of responses. Item selection using IRT may result in better-fitting measurement models, because EFA-informed item selection may result in an item set with worse fit due to constraining of substantial inter-item covariances to zero.

Supplemental Material for the Study 1 Analysis Plan Section Regarding Measurement Equivalence/Invariance Testing

This section of the supplemental material provides information about the different forms of measurement equivalence/invariance (ME/I) and the significance of each form

Regarding ME/I: Configural invariance is present when the specified measurement model structure (allowing parameter values to vary freely) is shown to fit adequately for each group. Metric invariance is present when each item loads on its corresponding factor to a similar degree in both groups. If metric invariance is supported, weak ME/I can be concluded (Dimitrov, 2010), and this would tentatively suggest the MHSAS measures the same construct among both groups of respondents, and thus correlations between help seeking attitudes and external variables can validly be compared across these groups. Scalar invariance is present when each item intercept has a similar magnitude in both groups. If both metric and scalar invariance are present, strong ME/I can be concluded and mean differences in the MHSAS can be validly compared across groups.

Supplemental Material for the Study 1 and Study 2 Measures Sections of the Manuscript Regarding Reliability, Validity, and Purpose of Administration

This section of the supplemental material provides reliability and validity information for each instrument used in the present investigation, and designates which instruments were administered to which Sample(s), and for what purposes the instruments were used.

Demographics. Gender was used for both validity testing and measurement equivalence/invariance (ME/I) testing. Age, race, and education were reported only to characterize the nature of the sample. Demographics were administered in the surveys for Study 1 and Study 1 Time 1.

Previous Help Seeking. This item was used for both validity testing and ME/I testing. This item was administered in the surveys for Study 1 and Study 1 Time 1.

The **SSRPH** has demonstrated convergent evidence of validity through correlations with the Attitudes Towards Seeking Professional Psychological Help Scale (r = -.40; Komiya et al., 2000). The SSRPH has demonstrated internal consistency (α = .71). This instrument was used for validity testing and administered in the survey for Study 1.

The **SSOSH** has demonstrated convergent evidence of validity through correlations with attitudes toward counseling (r = -.63), intentions to seek counseling (r = -.38), and the public stigma of seeking help (r = .48; Vogel et al., 2006). The SSOSH has demonstrated test-retest reliability over a period of 2 months (α = .72) and internal consistency (α = .89). This instrument was used for validity testing and administered in the survey for Study 1.

The **ATSPPH-SF** has demonstrated convergent evidence of validity through associations with intentions to seek help (r = .50; Vogel, Wade, & Hackler, 2007) and past psychological help seeking (r = .39; Fischer & Farina, 1995). The ATSPPH-SF has demonstrated internal

consistency (α = .79 to .82; Fischer & Farina, 1995). This instrument was used for validity testing and administered in the survey for Study 1.

The **IASMHS-PO** has demonstrated convergent evidence of validity through associations with intentions to seek help (r = .24) and past psychological help seeking (r = .34; Mackenzie, Gekoski, & Knox, 2006). The IASMHS-PO has demonstrated internal consistency (α = .82; Mackenzie, Knox, Gekoski, & Macaulay, 2004). This instrument was used for validity testing and administered in the survey for Study 1.

The **anticipated utility** and **anticipated risk** subscale have demonstrated convergent evidence of validity (Vogel & Wester, 2003) through associations with attitudes toward seeking psychological help (utility r = .36, risk r = -.24). The two subscales have demonstrated internal consistency (utility α = .83, risk α = .74). These subscales were used for validity testing and administered in the survey for Study 1.

The evidence of reliability and validity for the MHSAS is reported in the Results section of the manuscript. The 65-item pool for the MHSAS was administered in the survey for Study 1. The final 9-item version of the MHSAS was administered in the Study 2 (Time 1 and Time 2) surveys.

The **K6** has demonstrated validity through its ability to discriminate between clinical and nonclinical populations (Kessler et al., 2002). The K6 has demonstrated internal consistency (α = .84; Kessler et al., 2002). K6 scores below 5 indicate the presence of low psychological distress whereas scores above 5 indicate the presence of moderate (or greater) psychological distress (Prochaska, Sung, Max, Shi, & Ong, 2012). This instrument was used for validity testing and ME/I testing and administered in the survey for Study 1.

Help-seeking intention instruments that follow Azjen's practices have previously demonstrated evidence of reliability ($\alpha \geq .97$; Mo & Mak, 2009; Hammer & Vogel, 2013) and validity (e.g., significant positive associations between intention and both attitudes and subjective norms around seeking professional psychological help; Bayer & Peay, 1997; Hammer & Vogel, 2013; Mo & Mak, 2009; Schomerus, Matschinger, & Angermeyer, 2009). This instrument was used for validity testing and administered in the survey for Study 1.

Help-seeking subjective norms instruments that follow Azjen's guidelines have previously demonstrated evidence of reliability ($\alpha \geq .85$; Hammer & Vogel, 2013; Mo & Mak, 2009) and validity (e.g., significant positive association between subjective norms and intention to seek help; e.g., Bayer & Peay, 1997; Mo & Mak, 2009; Schomerus et al., 2009). This instrument was used for validity testing and administered in the survey for Study 1.

Help-seeking perceived behavioral control instruments that follow Azjen's guidelines have previously demonstrated evidence of reliability ($\alpha \geq .69$; Hess & Tracey, 2013; Mo & Mak, 2009) and validity (e.g., significant positive association between perceived behavioral control and intention to seek help; e.g., Hess & Tracey, 2013; Mo & Mak, 2009). This instrument was used for validity testing and administered in the survey for Study 1.

Supplemental Material for the Study 2 Analysis Plan Section Regarding Data Cleaning Procedures

The initial, pre-cleaned, Time 1 dataset originally contained 291 individuals. Cases with significant (> 20%; Parent, 2013) missingness on items for any scale (n = 5) excluding demographic items, or a wrong response to the attention check item (n = 3) were deleted. In the retained Time 1 sample (n = 285), there was no missing data for any of the MHSAS items. The Time 2 dataset contained 207 individuals. One case contained missing data (i.e., the respondent had missing data only on item 5). For both Time 1 and Time 2 variables, no variables exceeded the cutoffs of 3 and 10 for high skewness index and kurtosis index values, respectively (Weston & Gore, 2006).

Supplemental Material Detailing Psychometric Comparison of MHSAS, ATSPPH-SF, and IASMHS-PO Using the Study 1 Sample

When multiple instruments measuring the same construct exist, researchers and clinicians can compare the strengths and limitations of each measure in order to select the one best suited to their goals. Table 1 within the manuscript and information within this section of the Supplemental Material summarize the psychometric properties of each instrument based on data from the total sample. Regarding Table 1, readers will see that the MHSAS ($\alpha = .93$) demonstrated an internal consistency estimate whose 95% CI did not overlap with the 95% CI of the ATSSPH-SF ($\alpha = .86$) and IASMHS-PO ($\alpha = .83$), suggesting that the MHSAS total score is more internally consistent than the total scores of these other instruments.

We examined the global and local model fit for each of the three instruments, by testing a unidimensional solution in M*plus* 6.11 using the MLR estimator for each instrument. In terms of global fit, the MHSAS demonstrated good fit (S-B χ^2 [27] = 87.11, $p < .001$, RMSEA = .051 [90% CI of .039, .063], CFI = .978, TLI = .971, SRMR = .022), while the ATSPPH-SF (S-B χ^2 [27] = 248.89, $p < .001$, RMSEA = .084 [90% CI of .075, .094], CFI = .907, TLI = .881, SRMR = .046) and PO (S-B χ^2 [27] = 193.713, $p < .001$, RMSEA = .101 [90% CI of .088, .114], CFI = .895, TLI = .854, SRMR = .058) both demonstrated marginal to adequate fit.

In terms of local fit, the MHSAS evidenced higher average standardized item factor loadings (.79 vs. 63 and .62, for the MHSAS, ATSPPH-SF, and IASMHS-PO, respectively) and thus lower average standardized item factor loading residuals (.39 vs. .62 and .59, for the MHSAS, ATSPPH-SF, and IASMHS-PO, respectively) than the other attitudes measures. In other words, the MHSAS items functioned as purer measures of the attitudes factor than the items of the other two measures. That the ATSPPH-SF and PO items evidenced higher residuals

supports our contention, discussed in the introduction, that the items of these two instruments may be at greater risk of measuring construct-irrelevant variance (AERA et al., 2014, p. 12).

We next conducted three separate EFAs on the scales using principal axis factor extraction and direct oblimin (oblique) rotation. Parallel analysis results for each scale (not shown) indicated that the ATSPPH-SF and PO scales (but not the MHSAS) came close to requiring the use of a second factor to adequately represent the dimensionality of those instruments, which would deviate from the TPB's recommendation that attitudes be measured as unidimensional. The uncertain dimensionality of the ATSPPH-SF and IASMHS-PO instruments has been previously documented (see introduction), which further highlights the potential advantages of the stronger unidimensionality of the MHSAS. Regarding Factor Determinacy (MHSAS = .97; ATSPPH-SF = .94; IASMHS-PO = .93) and the *H* index (MHSAS = .94; ATSPPH-SF = .88; IASMHS-PO = .86), all three scales had satisfactory estimates, with the MHSAS demonstrating the strongest estimates. Thus, the MHSAS score may offer similar or better construct replicability than the other two instruments' scores.

Supplemental Material References

Ajzen, I. (2006). Constructing a theory of planned behavior questionnaire. Available online: http://people.umass.edu/aizen/pdf/tpb.measurement.pdf

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.

Bayer, J. K., & Peay, M. Y. (1997). Predicting intentions to seek help from professional mental health services. *Australian and New Zealand Journal of Psychiatry*, *31*(4), 504-513.

Cho, S. J., Wilmer, J., Herzmann, G., McGugin, R. W., Fiset, D., Van Gulick, A. E., ... & Gauthier, I. (2015). Item response theory analyses of the Cambridge Face Memory Test (CFMT). *Psychological Assessment*, *27*(2), 552-566.

Codd, R. T., & Cohen, B. N. (2003). Predicting college student intention to seek help for alcohol abuse. *Journal of Social and Clinical Psychology*, *22*(2), 168-191.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage publications.

Dueber, D. M. (2016). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. Retrieved from http://sites.education.uky.edu/apslab/resources

Elhai, J. D., Schweinle, W., & Anderson, S. M. (2008). Reliability and validity of the Attitudes Toward Seeking Professional Psychological Help Scale-Short Form. *Psychiatry Research*, *159*(3), 320-329.

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.

Hammer, J. H., & Vogel, D. L. (2013). Assessing the utility of the willingness/prototype model in predicting help-seeking decisions. *Journal of Counseling Psychology, 60*, 83-97.

Heise, D. (1970). The semantic differential and attitude research. In G. F. Summer. *Attitude Measurement* (235-253). Chicago, IL: Rand McNally.

Hess, T. R., & Tracey, T. J. (2013). Psychological help-seeking intention among college students across three problem areas. *Journal of Counseling & Development, 91*, 321-330.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalence and trends in non-specific psychological distress. *Psychological Medicine*, *32*(06), 959-976.

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.

Mackenzie, C. S., Gekoski, W. L., & Knox, V. J. (2006). Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. *Aging and Mental Health*, *10*(6), 574-582.

Mackenzie, C. S., Knox, V. J., Gekoski, W. L., & Macaulay, H. L. (2004). An adaptation and extension of the attitudes toward seeking professional psychological help scale. *Journal of Applied Social Psychology*, *34*(11), 2410-2435.

Mak, H., & Davis, J. (2014). The application of the theory of planned behavior to help-seeking intention in a Chinese society. *Social Psychiatry & Psychiatric Epidemiology, 49*(9), 1501-1515.

Masuda, A., Anderson, P. L., & Edmonds, J. (2012). Help-seeking attitudes, mental health stigma, and self-concealment among African American college students. *Journal of Black Studies, 43*(7), 773-786.

Mo, P. H., & Mak, W. S. (2009). Help-seeking for mental health problems among Chinese: The application and extension of the theory of planned behavior. *Social Psychiatry and Psychiatric Epidemiology*, *44*(8), 675-684.

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

Parent, M. C. (2013). Handling item-level missing data: Simpler is just as good. *The Counseling Psychologist*, *41*, 568–600.

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.

Prochaska, J. J., Sung, H. Y., Max, W., Shi, Y., & Ong, M. (2012). Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. *International Journal of Methods in Psychiatric Research*, *21*(2), 88-97.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150.

Schomerus, G., Matschinger, H., & Angermeyer, M. C. (2009). The stigma of psychiatric treatment and help-seeking intentions for depression. *European Archives of Psychiatry and Clinical Neuroscience*, *259*(5), 298-306.

Shaffer, P. A., Vogel, D. L., & Wei, M. (2006). The mediating roles of anticipated risks, anticipated benefits, and attitudes on the decision to seek professional help: An attachment perspective. *Journal of Counseling Psychology*, *53*(4), 442-452.

Vogel, D. L., Wade, N. G., & Haake, S. (2006). Measuring the self-stigma associated with seeking psychological help. *Journal of Counseling Psychology*, *53*(3), 325.

Vogel, D. L., & Wester, S. R. (2003). To seek help or not to seek help: The risks of self-disclosure. *Journal of Counseling Psychology*, *50*, 351-361.

Vogel, D. L., Wester, S. R., Wei, M., & Boysen, G. A. (2005). The role of outcome expectations and attitudes on decisions to seek professional help. *Journal of Counseling Psychology*, *52*(4), 459-470.

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

Supplemental Material Table A

Factor Loadings for Four-Factor Exploratory Factor Analysis

	Factor			
	1	2	3	4
Unhelpful - Helpful	.91	.00	.08	-.05
Useless - Useful	.88	-.02	.05	-.02
Unproductive - Productive	.84	-.01	.05	-.06
Unimportant - Important	.84	.04	.03	-.08
Harmful - Beneficial	.80	-.03	.03	.03
Unhealthy - Healthy	.79	.06	.02	.07
Ineffective - Effective	.77	-.13	-.01	-.08
Worthless - Valuable	.77	-.04	.09	-.06
Unnecessary - Necessary	.75	.01	-.08	-.12
Wasteful - Not wasteful	.72	-.01	.05	.11
Foolish - Wise	.72	.06	.00	.12
Bad - Good	.72	.02	.00	.12
Hurting - Healing	.70	-.07	-.01	.08
Negative - Positive	.70	-.07	.00	.05
Immature - Mature	.64	.17	-.06	.12
Disempowering - Empowering	.63	-.09	-.09	.11
Discouraging - Encouraging	.63	-.20	-.07	.01
Unsatisfying - Satisfying	.63	-.28	-.14	-.06
Irresponsible - Responsible	.60	.14	-.03	.19
Wrong - Right	.59	.05	-.09	.09
Impractical - Practical	.57	-.15	-.13	.08
Unacceptable - Acceptable	.53	.01	.02	.28
Undesirable - Desirable	.53	-.22	-.05	.10
Powerless - Powerful	.52	-.19	-.13	.13
Passive - Active	.50	.11	.00	.12
Cowardly - Courageous	.49	.07	-.15	.25
Confusing - Enlightening	.46	-.21	-.08	.09
Weak - Strong	.43	-.07	-.15	.26
Lazy - Industrious	.43	.03	-.30	.11
Risky - Safe	.41	-.31	.00	.11
Boring - Interesting	.40	-.22	-.25	-.03
A luxury - A necessity	.39	-.09	.03	.05
Stressful - Relaxing	.08	-.72	-.12	.02
Intimidating - Inviting	-.03	-.68	-.13	.12
Hard - Easy	-.05	-.68	.01	.12
Unpleasant - Pleasant	.16	-.67	-.16	.01
Uncomfortable - Comfortable	.14	-.63	-.15	.11
Unenjoyable - Enjoyable	.24	-.62	-.19	-.05
Awkward - Not awkward	-.02	-.57	.01	.32

Inconvenient - Convenient	.19	-.54	-.19	.04
Frustrating - Not frustrating	.19	-.51	-.07	.11
Soft - Hard	.12	.51	-.13	-.08
Light - Heavy	.04	.48	-.12	-.04
Distressing - Comforting	.34	-.46	-.03	.11
Slow - Fast	.17	-.39	.13	-.02
Expensive - Inexpensive	.04	-.36	.14	-.01
Calm - Excited	.05	.25	.17	-.10
Dead - Alive	.25	-.09	-.49	.12
Shallow - Deep	-.28	-.09	.49	-.07
Little - Big	.24	.12	-.44	-.04
Quiet - Noisy	.08	.13	.42	-.06
Short - Long	-.10	.32	-.39	.01
Freezing - Burning	-.05	.06	-.38	-.01
Cold - Hot	.04	-.16	-.37	.00
Low - High	-.15	.23	.37	.05
Thin - Thick	.03	.09	-.34	-.03
Needy - Not needy	-.07	-.08	.03	.66
Shameful - Not shameful	.19	-.16	.07	.57
Attention seeking - Not attention seeking	.11	.08	-.03	.53
Pathetic - Not pathetic	.33	-.02	.01	.53
Desperate - Not desperate	.17	-.06	.02	.51
Humiliating - Not humiliating	.15	-.32	.10	.50
Embarrassing - Not embarrassing	.04	-.39	-.01	.44
Selfish - Unselfish	.28	.00	-.10	.34
Dependent - Independent	.17	-.22	-.17	.24

Note: Results of Exploratory Factor Analysis using principal axis factor extraction with oblique rotation (direct oblimin) when four factors were specified for extraction. $N = 490$. Bold indicates the item was sourced from Heise's (1970) potency item list or activity item list.

Supplemental Material Table B

Factor Loadings for Three-Factor Bifactor Exploratory Factor Analysis

Item	F1: General Evaluative Attitudes Factor	F2: Specific Factor	F3: Specific Factor
Unhelpful - Helpful	.84	-.11	-.16
Useless - Useful	.85	-.08	-.13
Unimportant - Important	.73	-.13	-.14
Unproductive - Productive	.77	-.08	-.15
Unhealthy - Healthy	.78	-.16	-.03
Unnecessary - Necessary	.68	-.07	-.20
Harmful - Beneficial	.81	-.08	-.08
Ineffective - Effective	.78	.03	-.18
Bad - Good	.78	-.11	.00
Worthless - Valuable	.70	-.07	-.12
Hurting - Healing	.78	-.01	-.06
Wasteful - Not wasteful	.77	-.09	.02
Immature - Mature	.64	-.21	.00
Negative - Positive	.76	-.02	-.06
Disempowering - Empowering	.76	-.01	-.01
Unsatisfying - Satisfying	.77	.19	-.17
Discouraging - Encouraging	.76	.11	-.11
Irrresponsible - Responsible	.65	-.21	.06
Wrong - Right	.65	-.11	-.01
Impractical - Practical	.73	.06	-.02
Cowardly - Courageous	.65	-.13	.08
Unacceptable - Acceptable	.70	-.09	.14
Undesirable - Desirable	.72	.12	.00
Weak - Strong	.68	-.01	.11
Boring - Interesting	.57	.17	-.13
Confusing - Enlightening	.64	.12	-.02
Risky - Safe	.64	.20	.03
A luxury - A necessity	.45	.01	.00
Selfish - Unselfish	.54	-.06	.20
Stressful - Relaxing	.50	.62	-.03
Intimidating - Inviting	.44	.57	.09
Unpleasant - Pleasant	.57	.59	-.08
Hard - Easy	.37	.55	.11
Uncomfortable - Comfortable	.58	.53	.03
Unenjoyable - Enjoyable	.58	.53	-.11
Awkward - Not awkward	.49	.44	.26
Inconvenient - Convenient	.56	.45	-.04
Frustrating - Not frustrating	.54	.41	.03

Distressing - Comforting	.66	.36	.02
Embarrassing - Not embarrassing	.54	.27	.36
Expensive - Inexpensive	.16	.29	.02
Dependent - Independent	.50	.14	.13
Needy - Not needy	.41	.00	.48
Shameful - Not shameful	.63	.04	.45
Attention seeking - Not attention seeking	.44	-.11	.34
Pathetic - Not pathetic	.69	-.07	.36
Desperate - Not desperate	.53	-.02	.33
Humiliating - Not humiliating	.61	.20	.39

Note: Results of Three-Factor Bifactor Exploratory Factor Analyses using BI-GEOMIN rotation. N = 490. Bold indicates the nine items that were ultimately selected via Item Response Theory for the final version of the Mental Help Seeking Attitudes Scale.