Journal of MultiDisciplinary Evaluation
Volume 17, Issue 39, 2021

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Measuring Mental Health Literacy: Development of the Mental Health Awareness and Advocacy Assessment Tool

Ty B. Aller
Utah State University

Joshua R. Novak
Auburn University

Elizabeth B. Fauth
Utah State University

Sarah Schwartz
Utah State University

Background: Mental health literacy programs are a common community-based approach used to address the prevention of mental health issues on college campuses. Current assessment strategies used to evaluate the effectiveness of these programs often lack strong theoretical rational and psychometric rigor.

Purpose: The purpose of this study was twofold. First, based upon extant literature, theory, and standard clinical practice, we propose a process-based model of mental health literacy that includes three macro factors—identifying mental health issues, locating empirically based resources, and responding to mental health issues—and three micro processes of how they unfold—acquiring knowledge, building self-efficacy, and applying skills (behavior). The second aim was to test the psychometric properties of a new tool created to evaluate this process-based model—the *Mental Health Awareness and Advocacy Assessment Tool* (MHAA-AT).

Setting: Not applicable.

Intervention: Not applicable.

Research Design: A national sample of 296 college attending participants were recruited from Amazon's Mechanical Turk. Participants responded to a demographic questionnaire and the newly developed MHAA-AT. Psychometric properties were examined through item response theory, exploratory factor analyses, and bivariate correlations.

Findings: Results suggest the MHAA-AT is a sound measure and demonstrates appropriate item, person, and trait characteristics on declarative knowledge items, and single factor structures on self-efficacy and behavior items with moderate to high reliability and validity. While additional testing is need among other samples, results suggest that the MHAA-AT is a quality assessment tool.

Keywords: *College students; mental health literacy; item response theory; measurement*

Introduction

Epidemiological studies examining college students' mental health issues in the United States report that the estimated prevalence of undergraduate students experiencing depression or anxiety is 15.6% and 13% for graduate students, with 2% of all students reporting suicidal ideation in the past four weeks with similar numbers reported internationally (Eisenberg et al., 2007; Auerbach et al., 2018). The onset of mental health conditions at this age (18-25 years old) is likely the result of a convergence of many changes, including post-pubertal biological processes, new psychosocial factors, evolving health behaviors, and managing new life transitions. Since mental health issues are often associated with lower grade point averages and reduced likelihood of graduating (Eisenberg et al., 2009), colleges have a vested interest in providing cost-efficient, empirically supported, community level prevention services that target students' mental health needs (Kitzrow, 2009). Although several prevention programs exist, these approaches have yet to articulate a theoretical model of how individuals progress from knowledge to behavioral changes and how to measure that progression. The present paper highlights the rationale for expanding existing measurement to be more processed-based, and presents a new, practical, and psychometrically strong assessment tool that assesses the key components of a participants' declarative knowledge, self-efficacy, and behaviors in mental health literacy. In creating the tool, we sought to develop an easy to use, easy to interpret, assessment of mental health literacy for interventionists in higher education settings.

Mental Health Literacy

Defined, mental health literacy is knowledge and beliefs about mental disorders that aid in their recognition, management or prevention (Jorm et al., 1997). Research has identified that between 39% and 50% of participants could identify depression while only 27% of participants could identify schizophrenia (Dahlberg et al., 2008; Jorm et al., 2005). Because of these low proportions, prevention scientists have developed programs that have a positive influence on diverse populations' mental health literacy (see Jorm, 2012 for a full review). Mental health literacy programs, commonly address the following topics: a) the ability to recognize specific disorders or different types of psychological distress; b) knowledge and beliefs about risk factors and causes; c) knowledge and beliefs about self-help interventions; d) knowledge and beliefs about professional help available; e) attitudes which facilitate recognition and appropriate help-seeking; and f) knowledge of how to seek mental health information.

Measuring Outcomes in Mental Health Literacy

Existing approaches assessing declarative knowledge. Declarative knowledge refers to general facts needed to effectively identify and more comprehensively understand mental health issues, and is often assessed in two main ways—using vignettes (written by clinicians, these describe symptomology based on the Diagnostics and Statistical Manual of Mental Disorders [DSM; American Psychological Association, 2013]) or Likert scales (Jorm et al., 1997; Jorm et al., 2005; Jung et al., 2016). Although the vignette method is helpful in providing full description of the symptoms, as well as map onto a person's knowledge of the issues, they are tedious to evaluate on a large scale (O'Connor & Casey, 2015).

In contrast, the Likert-scale response approach ("Relative to the average person, how knowledgeable are you about mental illnesses (such as depression and anxiety disorders) and their treatments?" (responses range from 1 [Not at all], to 5 [extremely]; Lipson et al., [2014]), is efficient in large samples, the items included often do not fully depict the construct of 'knowledge', and are more akin to the construct of 'metacognition' (e.g., what do you *think you know* about the construct). A more effective approach to assessing declarative knowledge may include using multiple-choice questions assessing recognition of mental health symptoms, similarities and differences among mental health issues, resources to treat these disorders, and skills related to responding to these issues. Currently, there

are several studies incorporating this approach, but these measurements only commonly report the internal consistency of items, but not other important psychometric properties (Wyman et al., 2008; Quinnett, 2013). There is, however, one measure titled the *Mental Health Literacy Scale* that reports strong psychometric properties (i.e., validity and reliability estimates) and maps onto the concept of mental health literacy seamlessly (O'Connor & Casey, 2015), but does not theoretically articulate a process-based learning approach.

Existing approaches assessing self-efficacy. Self-efficacy measurement approaches are arguably the most common measurement strategy implemented in studies evaluating mental health literacy, and most often use Likert-scale response options (Mitchell et al., 2013; Wyman et al., 2008). For instance, one study assessed participants' self-efficacy in their knowledge of mental health literacy by asking participants to respond to a 5-point Likert scale question, "I have a good idea of how to recognize that a student is in emotional or mental distress" (strongly agree to strongly disagree; Lipson et al., 2014). Self-efficacy is used to evaluate responding to mental health issues by asking questions such as, "I am aware of warning signs for suicide" (Wyman et al., 2008). These scales have demonstrated high internal consistency and are predictive of other health behaviors (see Sheeran et al., 2016 for a meta-analytic review). In other measures of mental health literacy, factor analyses have supported multiple factor models (single and multiple factor iterations) that include knowledge, beliefs, and resource oriented mental health literacy questions independently and combined (Jung et al., 2016).

Existing approaches assessing behavioral outcomes. Behavioral outcomes included in past studies typically assess participants' self-reported response of either their own mental health issues or an issue for someone they know well in a retrospective account (Mitchell et al., 2013; Lipson et al., 2014; Wyman et al., 2008). There are two common approaches to measuring behavioral outcomes in this domain: 1) the likelihood of responding to mental health issues and 2) responding or providing referrals to someone that is experiencing a mental health issue via a retrospective self-report. One study measured likelihood of responding to a mental health issues on a three-point Likert scale (not very likely, somewhat likely, or highly likely), based on the Question Persuade Refer (QPR) Institute's survey (Mitchell et al., 2013; Quinnett, 2013). Researchers asked participants to rate themselves on the likelihood of engaging in certain suicide prevention behaviors, such as telling a suicidal person where to get help or calling a crisis line to get help for a suicidal person—or ask participants to indicate how many times they had referred an individual experiencing suicidal thoughts to professional resources (Wyman et al., 2008). These measurement strategies assess if participants are responding to mental health issues via their self-report of their own behavior retrospectively, however the diversity of content they assess are limited to one or two issues (e.g., suicidality, seeking professional help), and typically do not assess mastery of identifying a mental health issue or locating evidenced-based resources.

Present Study

Mental Health Awareness and Advocacy Assessment Tool: A Process-Based Approach

While there are varying useful, psychometrically-sound measurement approaches to examine mental health literacy (e.g. Mental Health Literacy Scale; O'Connor & Casey, 2015), we believe current measurement approaches can be strengthened by using a process-oriented approach (defined below; Wei et al., 2017). Using the guidance of theory, research literature, past measures used to evaluate mental health literacy, and content experts in the field of mental health, we developed the *Mental Health Awareness and Advocacy* assessment tool (MHAA-AT; further described in the methods section), which is made up of three progressive domains that emphasize the process of mental health literacy: 1) the ability to identify signs and

symptoms of mental health issues (*Identifying Domain*); 2) the ability to identify and access evidence-based mental health resources (*Locating Domain*); and 3) the ability to effectively and appropriately respond to mental health issues (*Responding Domain;* See Figure 1). The MHAA-AT then examines the overall process of mental health literacy by breaking these three domains into three micro-processes: (a) acquiring knowledge (knowledge), (b) building self-efficacy (self-efficacy), and (c) applying skills (behaviors). Although these steps theoretically might operate in a linear fashion, they may occur concurrently or in a different order.

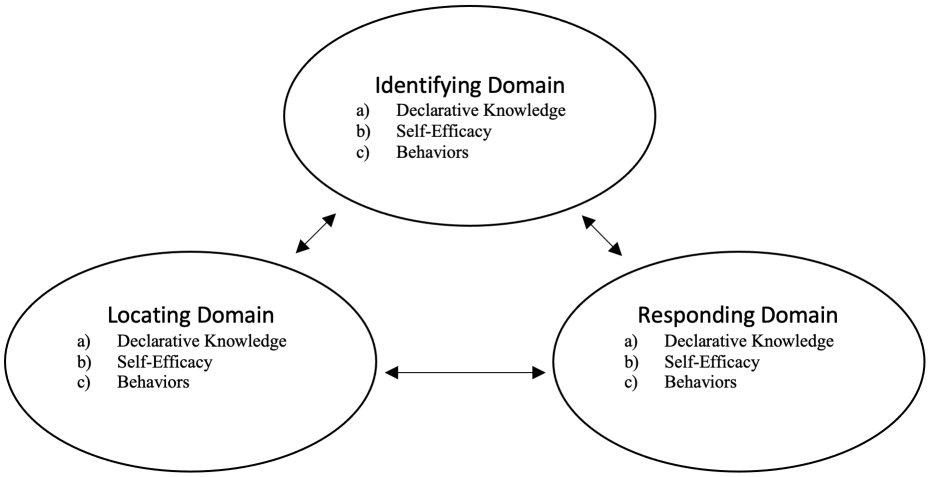

Figure 1. *Process-based model of mental health awareness and advocacy.*

Note: The circles represent the macro-processes. Micro-processes are listed within each macro-process. Declarative knowledge refers to the micro-process of acquiring knowledge; Self-efficacy refers to building self-efficacy, and behaviors refers to applying skills.

The present study tests the MHAA-AT in a college population and addresses the following research questions: RQ1: What are the item and respondent characteristics of the declarative knowledge items of the MHAA-AT? RQ2: What is the underlying factor structure of the self-efficacy and behavior items of the MHAA-AT? RQ3: Does the MHAA-AT demonstrate strong reliability and validity?

Method

Procedure

We wanted a sample from a wide range of colleges outside of our own institution and geographic/cultural region, thus we recruited via Amazon's Mechanical Turk (MTurk; Buhrmester et al., 2011), and only accepted those participants that indicated that they self-identified as a college student and were proficient in English. Based on Costello and Osborne's (2005) recommendations, a minimum of three participants per item were collected. Participants were included in the study if they were over the age of 17 and under the age of 26, as the ages of 18-25 are commonly reflect the "traditional" college student.

Participants were routed to a survey on Qualtrics.com after selecting the mental health awareness and advocacy assessment tool study on the MTurk system. The survey contained an informed consent and the assessment tool. After reading the informed consent, participants who chose to continue completed the assessment tool. Participants

failing to meet the age requirement (18-25 years old) and educational requirement (attending college) were excluded based on Institutional Review Board approved inclusion criteria. Participants qualifying for the study received $1 for participating in the study, which is in line with MTurk time/payment standards. Previous research has suggested that while MTurk can provide quick data in a cost-efficient manner, this data can be of lower quality at times (Buhrmester et al., 2011; Paolacci et al., 2010). In response, quality insurance safeguards were embedded in the current study, and included several Instructional Manipulation Checks (IMCs), including using "captcha" or "reverse-turing test" questions, questions that have verifiable answers, and attention checks (Mason & Suri, 2012). Lastly, we blocked repeated Internet Protocol Addresses and MTurk worker identification numbers to prevent duplicate responses.

Sample

The final measurement sample included 296 college-attending 18-25 year olds (M_{age} = 22.67, *SD* =1.79*;* see Table 1 for key sample characteristics). The sample averaged in the mild depression range on the PHQ-9 (*M = 7.82, S.D.* = 6.8) and averaged in the mild anxiety range on the GAD-7 (*M = 6.62, S.D.* = 5.85). About one-third (*n* = 109, 36.8%) of the participants reported they had been diagnosed with a mental health issue, 168 (56.8%) reported they were emotionally close with someone that had experienced a mental health issue, 63 (21.3%) reported they had experienced suicidal thoughts in the past six months, 105 (35.5%) reported they had known someone that had experienced suicidal thoughts in the past six months, and 56 (18.9%) reported they had received therapy in the past six months.

Table 1
Key Sample Characteristics

	N	% of sample
Year in School	296	
Freshman	25	8.40
Sophomore	56	18.90
Junior	82	27.70
Senior	108	36.50
Graduate Professional	25	8.40
Gender Identity	296	
Female	156	52.70
Transgender Female	1	0.30
Male	132	44.60
Transgender Male	2	0.70
Gender-Questioning	2	0.70
Two-Spirit	2	0.70
Other	1	0.30
Race/Ethnicity	294	
Black or African/American	35	11.80

	N	% of sample
White/European American	185	62.50
American Indian	7	2.40
Asian	33	11.10
Hispanic or Latino	26	8.80
Bi-Racial	8	2.70
Mother's Level of Education	296	
Some High School	15	5.10
High School Graduate	60	20.30
Some College	63	21.30
Associate Degree	25	8.40
Bachelor's Degree	86	29.10
Master's Degree	33	11.10
Doctorate Level Degree	10	3.40
Father's Level of Education	280	
Some High School	16	5.40
High School Graduate	76	25.70
Some College	47	15.90
Associate Degree	22	7.40
Bachelor's Degree	73	24.70
Master's Degree	37	12.50
Doctorate Level Degree	9	3.00
Financial Stress Growing Up	295	
Not at all concerned	78	26.40
Somewhat concerned	156	52.70
Very Concerned	61	20.60

Measurement

We used a three-step process to create items included in the measure: 1) initial item development and editing; 2) item review by a panel of content experts; 3) a review by a bachelor-level student panel to increase plain language usage. First, we conducted a thorough literature review to examine studies evaluating programs covering the concept of mental health literacy. We drafted items within the declarative knowledge, self-efficacy, and behavioral outcomes section based on the guidance of previous measures in the research literature (Lipson et al., 2008; Quinnett, 2013; Wyman et al., 2008). We used these items as a benchmark to guide content development but did not use the items verbatim. Next, an extensive review of factors that hinder or facilitate help-seeking behaviors in college populations (e.g., Eisenberg et al., 2007a), correlates of mental health issues in college populations (e.g., Eisenberg et al., 2007b), and

information regarding effective responses to mental health issues (e.g., Quinnett, 2013) was completed to generate additional items. Behavioral outcomes included in the measure were generated based on the process-based model of mental health awareness and advocacy described previously (e.g., knowing about depression makes you more confident to talk to someone about depression, which leads to the help seeking behaviors).

The first and second author reviewed and revised the initial items to identify any potential syntax errors, content holes, and other logistical problems. Next, a panel of five content experts working in the mental health field (e.g., clinical faculty, researchers, teachers) reviewed items for face validity with particular attention to identifying content holes within the three domains. Three iterations of this process were completed, followed by presenting items to an informal focus group of four individuals with a bachelor's level education to review the plain language and note any confusion in items. The informal focus group then provided feedback they deemed appropriate related to the accessibility of the language used in the measure. In all, this process created 66 items that were included in the seminal evaluation of the MHAA-AT.

Declarative knowledge. There were 30 knowledge items; ten items assessing each of the three content areas (i.e., Identifying mental health issues, Locating empirically based resources, and Responding to mental health issues). Knowledge items were selected for inclusion if the panel agreed the items had unambiguous "right" and "wrong" answers, based on consistent findings or evidence, and included topics that should be addressed in education on that specific domain. All items in the knowledge domain were assessed using a five-answer multiple-choice test. Sample multiple choice items assessing knowledge included: "*Individuals are more likely to experience symptoms of depression when they are between the ages of: a) 6-17 years old, b) 18-29 years old, c) 30-41 years old, d) 41-52 years old, e) I don't know the answer*" and "*Which of the following has been identified by research as an effective treatment for severe major depressive disorder?*: *a) Talk Therapy, b) Journaling, c) Herbal Supplements, d) Exercise, e) I don't know the answer*". Items were coded as a one if they were answered correctly and a 0 if they were answered incorrectly. The items were then scored zero to ten with the raw score then being converted using a logarithmic function based on the non-linearity of item difficulty.

Self-efficacy. There were 20 self-efficacy items included that assessed each of the three content areas. All self-efficacy items were assessed using a 6-point Likert scale (0 = Not at all confident; 5 = Completely confident). Sample items assessing self-efficacy included: *"I can identify each of the diagnostic criteria for major depressive disorder"*, and *"I can talk to someone about accessing mental health resources for depression or anxiety issues in a kind and empathetic manner."* The 20 self-efficacy items were then averaged resulting in a score that ranged from zero to seven for the self-efficacy domain.

Behavioral outcomes. There were 15 items included in the behavioral outcomes section, all using a frequency count (N/A; No one I know has mental health issues, 0 times, 1 time, 2 times, 3 times, 4-5 times, 6+ times). Sample items assessing behavioral outcomes included: *"How often in the past three months have you recognized that someone's mental state (e.g., sadness, nervousness, uneasiness) could be indicative of a diagnosable mental health issue?"*, and *"How often in the past three months have you asked someone who showed signs/symptoms of a mental health issue if they are doing 'okay' or if they needed help?"* The 15 behavior items were then averaged giving each participant a score that ranged from zero to seven for the self-efficacy domain.

Analytic Approach

For research question 1, A 1-parameter (Rasch-type) dichotomous Item Response Theory (IRT; Bond & Fox, 2015) model was fit to each set of 10 declarative knowledge items from each domain (i.e., Identifying, Locating, Responding) data using the *ltm* package version 1.1-1(Rizopoulos, 2006) in the R software version 3.5.2 (R core team, 2018). IRT evaluates and scores response data by

simultaneously modelling item and respondent characteristics, and has measurement advantages over classical test theory (Ostini & Nering, 2006). The mathematical foundation of IRT models the probability of a correct response to each item given the respondent's trait level (e.g. amount of declarative knowledge in a specific domain) using logistic regression. It simultaneously and interpedently estimates each respondents' trait level and each items difficulty level on the same latent dimension (Ostini & Nering, 2006). The relative appropriateness of 1-parameter model in each of the domains was evaluated by examining item fit statistics, item parameter estimates standard errors, and person item maps. Respondent knowledge scores were then estimated for each subset of items separately. Descriptive characteristics for the three knowledge score distributions were calculated. Lastly, analyses were conducted to provide validity information on the declarative knowledge items.

For research question 2, the Kaiser-Meyere-Olkin (KMO) measure of sampling adequacy (values closer to 1.0 indicate appropriateness for factor analysis) and the Bartlett's test of Sphericity (p values less than .05 indicate appropriateness for factor analysis; Cerny & Kaiser, 1977) was used to determine if the underlying assumptions of principal axis factor analysis were met. Next, to identify the underlying factor structure of the self-efficacy and behavior items of the MHAA-AT, a principal axis factor analysis with oblique rotation was selected due to the non-normal distribution of data, smaller sample size, the need to account for shared variance, and to avoid any inflation of estimates of variance accounted for (Costello & Osbourne, 2005). A scree plot test (Catell, 1966) identified breaking points of factors. Factors with eigenvalues of one or higher were retained. Appropriateness of factor analysis in regard to sample size was tested using SPSS Version 25. Lastly, bivariate correlations were used to examine reliability and construct validity of the MHAA-AT.

To answer research question 3, each domain of the measure was correlated with scores from measures of similar constructs used in the research literature. These measures included (1) the knowledge subscale from the Question, Persuade, Refer (QPR) institute (Quinnett, 2013; Wyman et al., 2008)—a 14-question measure used to assess knowledge related to suicide prevention and is commonly used to assess the knowledge gained by participating in QPR training; (2) a 7-item measure self-efficacy subscale used to assess QPR gatekeeping training (Wyman et al., 2008), (3) and general measures assessing mental health of an individual including the Patient Health Questionnaire-9 (PHQ-9; Löwe et al., 2004)—a 9-item Likert questionnaire assessing depressive symptoms, and the Generalized Anxiety Scale 7 (GAD-7; Spitzer et al., 2006)—a 7-item Likert questionnaire that assesses anxiety symptoms.

Results

RQ 1: Reliability and Dimensionality Analyses

For the purpose of data analyses, responses to the declarative knowledge items were coded in a binary fashion (correct or incorrect) with "I don't know" responses recoded as incorrect. Due to the process-based nature of the MHAA-AT, the 30 declarative items were divided into the three domains (i.e., Identifying, Locating, and Responding domains).

In the IRT framework, a 1-parameter Rasch Model was applied to the data. Mean square fit statistics (Mean Squared Error, MSW infit and outfit; see Table 2) suggested adequate unidimensionality of each of the domains (Bond & Fox, 2015). Reliability statistics of each subdomain indicate fair internal consistency. IRT simultaneously estimated both item difficulties (eta) and person-specific knowledge levels (theta) by maximum likelihood. Figure 2 contains the Person-item maps which present the overall spread of difficulty on items. Lastly, the raw to scaled-score conversions were calculated for each domain (see Table 3). In the following sections, each of the aforementioned domain specific statistics are described.

Table 2
Three IRT Analyses: Item Fit Characteristics (MSQ) for MHAA-AT

Item	Domain					
	Identifying		Locating		Responding	
	Outfit	Infit	Outfit	Infit	Outfit	Infit
1	0.90	0.86	**1.22**	1.10	0.98	0.97
2	0.92	0.92	0.87	0.89	1.18	1.00
3	**1.41**	**1.22**	0.79	0.87	0.85	0.89
4	0.94	0.97	0.94	0.97	0.99	0.90
5	1.07	1.09	1.05	1.00	1.04	0.98
6	1.14	1.16	**1.91**	**1.22**	**1.40**	0.99
7	0.73	0.77	0.84	0.87	0.96	0.99
8	0.83	0.85	0.88	0.91	0.95	0.94
9	0.70	0.80	1.07	1.00	0.73	0.83
10	**1.45**	0.86	0.74	0.83	0.88	0.83

Note: Bolded values denote MSQ-values outside the range of +/- 1.2 which may indicate inappropriate fit for the selected item in the selected domain (Bond & Fox, 2015).

Table 3
Three IRT Analyses: Item Difficulty Estimates (Eta), Conditional Probabilities (prob.), and Raw to Scaled Scores Conversion for the Declarative Knowledge items of the MHAA-AT

	Domain					
	Identifying		Locating		Responding	
Item	Eta (prob.)	Est (SE)	Eta (prob.)	Est (SE)	Eta (prob.)	Est (SE)
0	–	-3.67 (–)	–	-3.39 (–)	–	-3.69 (–)
1	1.25 (.22)	-2.70 (1.10)	0.61 (.34)	-2.45 (1.08)	-0.06 (.52)	-2.70 (1.11)
2	1.01 (.26)	-1.79 (.85)	0.37 (.40)	-1.58 (.83)	1.43 (.20)	-1.77 (.87)
3	-0.55 (.64)	-1.15 (.76)	0.46 (.38)	-.98 (.73)	-0.66 (.66)	-1.10 (.78)
4	-.30 (.58)	-.60 (72)	-0.04 (.51)	-.48 (.69)	2.01 (.12)	-.53 (.74)
5	1.35 (.20)	-.09 (71)	1.34 (.19)	-.01 (.68)	0.59 (.36)	.01 (.73)
6	1.29 (.21)	.42 (.73)	1.63 (.15)	.46 (.70)	2.43 (.08)	.54 (.74)
7	.98 (.27)	.99 (.79)	-0.61 (.66)	.97 (.74)	0.76 (.32)	1.11 (.77)
8	1.11 (.24)	1.73 (.94)	0.21 (.44)	1.58 (.83)	1.71 (.16)	1.77 (.86)
9	-0.53 (.63)	–	-0.91 (.73)	1.58 (.83)	-1.15 (.76)	2.69 (1.10)
10	4.10 (.01)	3.28 (–)	-0.45 (.62)	–	-0.76 (.68)	–

Note: Estimates are on the logit scale. Items that require more knowledge in order to answer correctly have higher values and items that discriminate at a lower level of knowledge will have smaller values. The probability is the chance of correctly responding to each item, conditional on having a knowledge level of zero. The Est. denotes the estimated score for each sub-domain given a particular raw score. For example, a raw score of 6 on the identifying domain equates to a converted score of .42.

Table 4
Descriptive Statistics and Correlations among the MHAA-AT Domains and Other Key Measures

	Items			Descriptives				Correlations					
	#	Type	Score	Range	M (SD)	% Correct	α	(1)	(2)	(3)	(4)	(5)	(6)
MHAA-AT Domains													
(1) Declarative Knowledge	30	0/1	IRT										
Identifying Subdomain	*10*	*0/1*	*Count*	*0-10*	*3.44 (2.09)*	*34.4*	.62						
Locating Subdomain	*10*	*0/1*	*Count*	*0-10*	*4.52 (2.40)*	*45.2*	.68						
Responding Subdomain	*10*	*0/1*	*Count*	*0-10*	*3.95 (2.05)*	*39.5*	.60						
(2) Self-Efficacy	20	0-5	Mean	1 - 6	4.20 (0.66)	–	.95	.31**					
(3) Behavior	15	0-5	Mean	0 - 5	0.86 (0.87)	–	.95	.10	.43**				
Other Key Measures													
(4) QPR Knowledge	14	0/1	Count	0 - 12	8.64 (2.12)	70.2		.44**	-.01	-.01			
(5) Self-Efficacy (Wyman)	*7*	1-7	Mean	1 - 7	4.21 (0.66)	–	.78	.13*	.51**	.26**	-.02		
(6) PHQ-9	9	0-3	Sum	0 - 27	7.83 (6.80)	–	.93	.02	.26**	.49**	.41	.13*	
(7) GAD-7	*7*	0-3	Sum	0 - 21	6.62 (5.85)	–	.93	.06	.27**	.46**	.03	.09	.82**

Notes: α = Cronbach alpha reliability estimates, PHQ-9 = Patient Health Questionnaire, GAD-7 = Generalized Anxiety Disorder), correlations are Pearson's correlation coefficients * $p < .05$ (2-tailed) ** $p < .01$ (2-tailed).

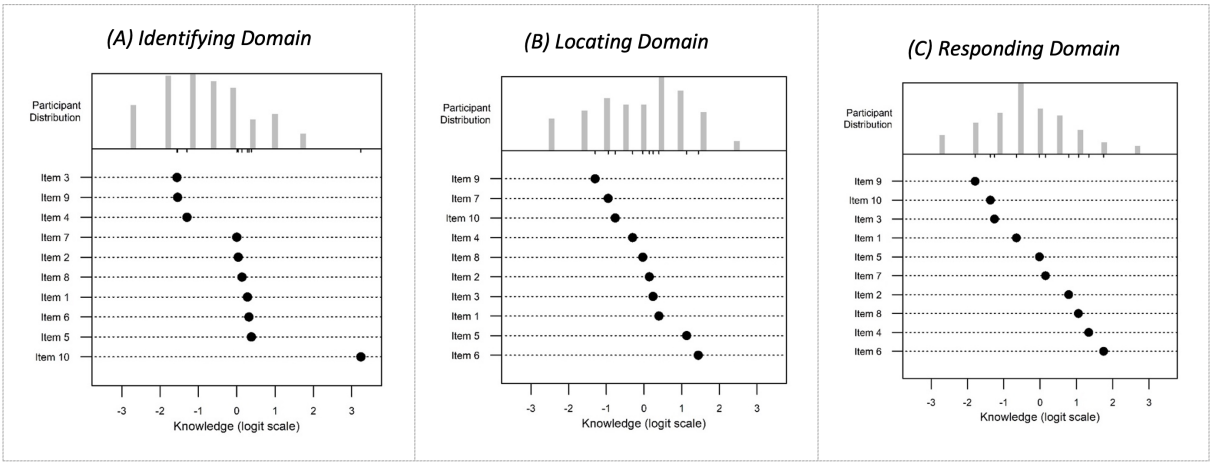

Figure 2. *Process-based model of mental health awareness and advocacy.*

Note: Estimates are on the logit scale. Items that require more knowledge in order to answer correctly have higher values (3) and items that discriminate at a lower level of knowledge will have smaller values (-3). The image provides a representation of participant spread (participant distribution) on each item (item number) and their corresponding difficulty (logit scale).

Identifying domain. The identifying domain is sufficiently unidimensional (MSQ's < 1.5; see Table 2) with the exception of item 1 and item 10. Due to the nature of these items (e.g., symptoms of depression and age of onset of anxiety disorders) having face validity with the identifying domain, the authors opted to keep these items. Internal consistency of the scale (α = .62) was adequate and was not highly influenced by the dropping of any particular item. The person item map for the identifying domain (see Panel A of Figure 2) depicts the spread of the data across the latent dimension of 'identifying mental health issues.' As is seen in this figure, the questions tend to fall within +/- 1 on the logit scale suggesting there is need for easier and more difficult questions to increase the variability of difficulty of the items on the subscale. Lastly, due to the relative non-linear shape of the slope of difficulty of items it is suggest that raw scores be converted to weighted scores in interpretation (see Table 3).

Locating domain. The Locating domain fit indices suggest the domain is sufficiently unidimensional (see Table 2). Items 1 and item 6 are slightly outside of the range of acceptable MSQ, but were kept due to the MSQ guidelines proposed by Bond and Fox (2015) being highly influenced by sample size and our sample size being moderate. Internal consistency of the locating domain (α = .68) was good and was not highly influenced by the dropping of any particular item. The person item map of the Locating domain (See Panel (B) of Figure 2) suggests more spread in difficulty of items compared to the Identifying domain, but there is still need for more questions that cover the poles of difficulty. The Locating domain also depicted a non-linear shape of the slope on difficulty of items suggesting that raw scores should be converted to weighted scores in interpretation (see Table 3).

Responding domain. The fit indices of the Responding domain are also within normal ranges and suggest the items as being unidimensional (see Table 2). Internal consistency of the locating domain (α = .60) was adequate. The internal consistency ranges do drop below ranges of acceptability suggesting that more work is needed on the scale to identify areas of 'lumpiness' within the single factor. The person item map (see Panel C of Figure 2) of the Responding domain shows the most spread in difficulty of questions.

RQ2: Self-Efficacy and Behavior Items

For the self-efficacy items, the KMO (= .95) was above the recommended level and Bartlett's test of sphericity was significant (X^2 = 3849.33, df = 190, p < .001), indicating that the self-efficacy items were suitable for factor analysis. The anti-image correlation matrices were all greater than .5, supporting the inclusion of each item in the factor analysis (Field, 2005). Initial outcomes from the self-efficacy items without a fixed number of factors to extract, extracted 3 factors with eigenvalues higher than 1. A scree plot test (Cattell, 1966) showed the breaking point after three factors. To add clarity in a single factor structure, multiple manual factor extractions from 1 to 3 were performed. Based on recommendations from Costello and Osborne (2005; item loadings above .30, no or few cross loadings, and no factors with fewer than three items, p. 3), clarity of a single-factor remained clear. The one factor structure of the self-efficacy items explained 50.58% of the variance in the MHAA-AT self-efficacy items.

For the behavior items, the KMO (= .92) was above the recommended level and Bartlett's test of sphericity was significant (X^2 = 3840.04, df = 105, p < .001), indicating that the data was suitable for factor analysis. The diagonals of the anti-image correlation matrices for the behavior items were greater than .5, supporting that the inclusion of each item in the factor analysis (Field, 2005). Initial outcomes from the self-efficacy items without a fixed number of factors to extract, extracted 2 factors with eigenvalues higher than 1. A scree plot test (Cattell, 1966) showed the breaking point after two factors. To add clarity in a single factor structure, multiple manual factor extractions from 1 to 2 were performed. Based Costello and Osborne recommendations (2005), the items from the single factor remained clear. The one factor structure of the behavior items explained 56.96% of the variance in the MHAA-AT behavior items.

RQ3: Does the MHAA-AT Demonstrate Strong Reliability and Validity?

Reliability statistics for the MHAA-AT was assessed in multiple ways. First, the internal consistency of the declarative knowledge items was assessed by breaking the thirty items into each of the three domains (see IRT section). The Identifying domain, Locating domain, and Responding domain each demonstrated adequate internal consistency (α = .62, .68, and .60 respectively). The underlying factor-structure of the self-efficacy and behavior questions of the MHAA-AT suggested that the items should not be separated into the three distinct domains and should instead be interpreted as one factor each (i.e., self-efficacy items and behavior items). The internal consistency of the self-efficacy and behavior items was good (Self-efficacy α = .95; Behavior items α = .95).

Construct validity of the MHAA-AT was assessed by completing bivariate correlations between the micro-processes (declarative knowledge, self-efficacy, and behavior items) of the MHAA-AT and psychometrically sound measures commonly used to evaluate mental health awareness and advocacy (see Table 5). The declarative knowledge items were positively correlated with the QPR Knowledge subscale (r = .44, p < .01) and the Wyman and colleagues (2008) self-efficacy subscale (r = .13, p < .05). Additionally, the MHAA-AT self-efficacy subscale was positively correlated with the Wyman et al. (2008) self-efficacy subscale (r = .51, p < .01). Lastly, the MHAA-AT subscales were also correlated with one another (declarative knowledge positively correlated with self-efficacy; self-efficacy positively correlated with behaviors), as with the PHQ-9, and GAD-7.

Discussion

Concluding the current study we determined the MHAA-AT is a reliable and valid assessment tool for assessing college students' declarative knowledge, self-efficacy, and behaviors in identifying mental health issues, locating evidence-based resources, and responding to mental health issues. For RQ1, IRT analyses of Knowledge items indicated

that the item difficulty appropriately covers the range of knowledge exhibited by the sampled population, but with room for general improvement. For instance, in the Identifying domain, item difficulty scores range from -2 to 3 on the logit scale (see Figure 2), indicating that we may need to consider developing questions that are less difficult (closer to -3), of average difficulty (between -1 and 0) and of moderate difficulty (between 1 and 3). The Locating domain, while more spread across the logit scale on item difficulty, might benefit from questions that are toward the two poles of difficulty (closer to -3 and 3 on the logit scale). The Responding domain has the most spread in item difficulty, but might benefit from questions that are deemed toward the two poles of difficulty. Internal consistency coefficients ranged from acceptable to good. These findings are notable given the inherent challenges to analyzing binary response choice measures. Accordingly, the MHAA-AT declarative knowledge items should be scored and interpreted using the number of correct responses on each domain and then converted using the theta score adjustments found in Table 3.

For RQ2, the self-efficacy items and behavior items fit a one factor model. Each item was retained with an appropriate factor loading and demonstrated high internal consistency. This finding was slightly surprising due to the proposal of the three domains being three separate micro-processes within mental health awareness and advocacy (see Figure 1). That being said, the overall macro-process (e.g., knowledge leading to self-efficacy and self-efficacy leading to behaviors) proposed via theory was initially supported by this study.

For RQ3, the MHAA-AT demonstrated strong convergent validity (see Table 5). As would be expected, the MHAA-AT declarative knowledge items were significantly correlated with the QPR knowledge items, a measure commonly used in the literature base (Lipson, 2014; Mitchell et al., 2013; Reis & Cornell, 2008). The MHAA-AT self-efficacy items were also significantly correlated with the Wyman and colleagues (2008) measure of self-efficacy in gatekeeping knowledge and behaviors. Interestingly, the MHAA-AT self-efficacy items were positively associated with both depression and anxiety symptoms. These findings could be due to personal exposure to mental health symptoms, treatments, and responses based on personal experiences positively influencing more participant confidence in the material assessed on the MHAA-AT. Future research should explore this relationship more as it is possible that students receiving campus therapy resources could have more mental health literacy and therefore be a potential resource for college campuses experiencing increasing mental health issues. This strengths-based interpretation could provide interventionists with more possibilities in peer-led programming.

Of particular interest was the MHAA-AT statistics that partially support the assessment tool being process-based. Specifically, the MHAA-AT declarative knowledge items were positively associated with the MHAA-AT self-efficacy items, but not the behavior items. This provides partial support for the process-based model in that as participants' knowledge increased so did their self-efficacy, but as Bandura (2005) suggests, knowledge does not equate to action. Participants' self-efficacy was significantly correlated with their behavior. In short, the data seem to suggest that as declarative knowledge increases, as does self-efficacy, but knowledge isn't directly linked to self-reported behaviors. Future research will need to examine student demographic correlates in relation to MHAA-AT scores as it may provide more theory-based understandings of how helps students build mental health literacy and help pinpoint a specific mechanism that may be driving the change in student learning.

Limitations

First, the study used mTurk for data collection, and although this approach is more increasingly common in the social sciences, there are intrinsic limitations associated with this method of data collection, including participant inattention and compensation (self-selection bias). Although we attempted to address these limitations through the use of attention questions and fair payment, these threats cannot be fully accounted for via online data collection methods. Additional research on the MHAA-AT is needed to address

the ability to demonstrate strong psychometric properties in larger and more diverse populations, as well as using multiple data points to help identify stability of the constructs across time (test-retest reliability). Second, the sample here was higher than average in anxiety and depression (mild, for both), which—according to theory (Bandura, 1982)—might influence specific domains of the measure. Future research should investigate if this holds true in samples with lower anxiety and depression.

Conclusion

Despite this being a first study addressing the psychometric properties of the MHAA-AT, we believe that the results suggest the tool is ready for use in larger, basic/descriptive research and/or intervention-based research projects on college campuses. The MHAA-AT seems to help identify the process of mental health awareness and advocacy. For instance, if a student scores lower in particular areas of declarative knowledge (e.g., identifying mental health issues) their scores were associated with being less likely to be confident in the same area and ergo less likely to identify mental health issues in a variety of contexts. This is especially important for interventionists, especially those working on a college campus, wishing to tailor their interventions to most directly influence a specific type of outcome. Additionally, the MHAA-AT is easy to use and can be completed in less than 20 minutes and provides a robust formulation of students' specific strengths and areas of understanding related to mental health literacy. In sum, the MHAA-AT provides a theory-based, easy to implement tool, to assess students' mental health literacy regardless of the modality of intervention while also providing easy to interpret output data that can help gauge large populations in a cost-effective manner.

References

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., ... & Murray, E. (2018). WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. *Journal of Abnormal Psychology, 127*(7), 623-638. DOI: 10.1037/abn0000362

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorder* (5th ed.). Washington, DC: Author.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122-147. DOI: 10.1037/0003-066X.37.2.122

Bandura, A. (2005). The primacy of self-regulation in health promotion. *Applied Psychology*, *54*(2), 245-254. DOI: 10.1111/j.1464-0597.2005.00208.x

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3-5. DOI: 10.1177/1745691610393980.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. DOI: 10.1207/s15327906mbr0102_10

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, *12*(1), 43-47. DOI: 10.1207/s15327906mbr1201_3

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(7), 1-9.

Dahlberg, K. M., Waern, M., & Runeson, B. (2008). Mental health literacy and attitudes in a Swedish community sample–Investigating the role of personal experience of mental health care. *BMC Public Health*, *8*(1), 8-18. DOI: 10.1186/1471-2458-8-8

Eisenberg, D., Golberstein, E., & Gollust, S. E. (2007a). Help-seeking and access to mental health care in a university student population. *Medical Care*, *45*(7), 594-601. DOI: 10.1097/MLR.0b013e31803bb4c1

Eisenberg, D., Golberstein, E., & Hunt, J. B. (2009). Mental health and academic success in college. *The BE Journal of*

Economic Analysis & Policy, *9*(1). DOI:10.2202/1935-1682.2191

Eisenberg, D., Gollust, S. E., Golberstein, E., & Hefner, J. L. (2007b). Prevalence and correlates of depression, anxiety, and suicidality among university students. *American Journal of Orthopsychiatry*, *77*(4), 534-542. DOI: 10.1037/0002-9432.77.4.534

Field, A. P. (2005). Discovering Statistics using SPSS. London: Sage.

Jorm, A. F. (2012). Mental health literacy: Empowering the community to take action for better mental health. *American Psychologist*, *67*(3), 231-243. DOI: 10.1037/a0025957

Jorm, A. F., Korten, A. E., Jacomb, P. A., Christensen, H., Rodgers, B., & Pollitt, P. (1997). "Mental health literacy": A survey of the public's ability to recognize mental disorders and their beliefs about the effectiveness of treatment. *Medical Journal of Australia*, *166*(4), 182-186.

Jorm, A. F., Nakane, Y., Christensen, H., Yoshioka, K., Griffiths, K. M., & Wata, Y. (2005). Public beliefs about treatment and outcome of mental disorders: a comparison of Australia and Japan. *BMC Medicine*, *3*(1), 12-26. DOI: 10.1186/1741-7015-3-12.

Jung, H., von Sternberg, K., & Davis, K. (2016). Expanding a measure of mental health literacy: Development and validation of a multicomponent mental health literacy measure. *Psychiatry Research*, *243*, 278-286. DOI: 10.1016/j.psychres.2016.06.034.

Kitzrow, M. A. (2009). The mental health needs of today's college students: Challenges and recommendations. *NASPA Journal*, *46*(4), 646-660. DOI: 10.2202/1949-6605.1310

Lipson, S. K., Speer, N., Brunwasser, S., Hahn, E., & Eisenberg, D. (2014). Gatekeeper training and access to mental health care at universities and colleges. *Journal of Adolescent Health*, *55*(5), 612-619. DOI: 10.1016/j.jadohealth.2014.05.009.

Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the Patient Health questionnaire-9. *Medical Care*, *42(*12), 1194-1201.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1-23.

Mitchell, S. L., Kader, M., Darrow, S. A., Haggerty, M. Z., & Keating, N. L. (2013). Evaluating question, persuade, refer (QPR) suicide prevention training in a college setting. *Journal of College Student Psychotherapy*, *27*(2), 138-148. DOI: 10.1080/87568225.2013.766109

O'Connor, M., & Casey, L. (2015). The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy. *Psychiatry Research*, *229*(1-2), 511-516. DOI: 10.1016/j.psychres.2015.05.064.

Ostini, R., & Nering, M. L. (2005). *Polytomous item response theory models*. Newbury Park, CA: Sage.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, *5*(5), 411-419.

Reis, C., & Cornell, D. (2008). An evaluation of suicide gatekeeper training for school counselors and teachers. *Professional School Counseling*, *11*(6), 386-394. DOI: 10.1177/2156759X0801100605

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1-25.

Sheeran, P., Maki, A., Montanaro, E., Avishai-Yitshak, A., Bryan, A., Klein, W. M., & Rothman, A. J. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. *Health Psychology*, *35*(11), 1178-1218.

Quinnett, P. (2013). *QPR gatekeeper training for suicide prevention: The model, rationale, and theory*. Retrieved from: https://qprinstitute.com/research-theory

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092-1097. DOI: 10.1001/archinte.166.10.1092

Wei, Y., McGrath, P. J., Hayden, J., & Kutcher, S. (2017). Measurement properties of mental health literacy tools measuring help-seeking: A systematic review. *Journal of Mental Health, 26*(6), 543-555.

Wyman, P. A., Brown, C. H., Inman, J., Cross, W., Schmeelk-Cone, K., Guo, J., & Pena, J. B. (2008). Randomized trial of a gatekeeper program for suicide prevention: 1-year impact on secondary school staff. *Journal of Consulting and Clinical Psychology*, *76*(1), 104-115. DOI: 10.1037/0022-006X.76.1.104