Development and Validation of the Gatekeeper Behavior Scale

A Tool to Assess Gatekeeper Training for Suicide Prevention

Glenn L. Albright[1], Jesse Davidson[1], Ron Goldman[2], Kristen M. Shockley[3], and Jane Timmons-Mitchell[4]

[1]Department of Psychology, Baruch College, City University of New York, USA
[2]Kognito, New York, NY, USA
[3]Department of Psychology, Baruch College and The Graduate Center, City University of New York, USA
[4]Begun Center for Violence Prevention Research and Education, Jack, Joseph and Morton Mandel School of Applied Social Sciences, Case Western Reserve University, Cleveland, OH, USA

Abstract: *Background:* Community-based gatekeeper trainings are effective tools in increasing gatekeeper skills but few validated measures assess impact. Aims: This study aimed at determining the validity of an 11-item Gatekeeper Behavior Scale (GBS) to assess gatekeeper skills that predict behavior. *Method:* To validate the scale, 8,931 users were administered GBS surveys at pretraining, posttraining, and follow-up periods. The training was one of five from the suite of online At-Risk mental health learning simulations for university faculty/staff or students or high/middle school educators. Results: A confirmatory factor analysis revealed the three-factor model based on the subscales of preparedness, likelihood, and self-efficacy fit the data best. Factor loadings showed all items correlated highly with theoretical constructs ($r \geq .84$, $p < .001$). The GBS had high internal consistency ($\alpha = 0.93$). Criterion-related validity for likelihood to discuss concerns at posttraining was significantly related to approaching students believed to be in psychological distress ($r = .219$, $p < .001$). Likelihood to refer significantly correlated with the number of students referred ($r = .235$, $p < .001$). Convergent validity was established via a correlation between self-efficacy in motivating someone to seek help and general self-efficacy ($r = .519$, $p < .001$). Conclusion: The GBS appears to be a valid tool in measuring the impact of online gatekeeper training simulations and holds promise for assessing other delivery methods.

Keywords: suicide prevention, gatekeeper, assessment, validity, confirmatory factor analysis

The Centers for Disease Control and Prevention (CDC) reported in 2012 that suicide was the 10th leading cause of death for all age groups and among 15- to 24-year-olds accounted for 20% of all deaths annually. The prevalence of suicidal thoughts, attempts, and behavior is greater among adolescents and young adults than among those older than 30 years (Crosby, Han, Ortega, Parks, & Gfoerer, 2011). The cause of suicidal ideation and behavior is complex and multifaceted, but the goal of suicide prevention is simple: to reduce risk factors and promote protective factors (CDC, 2010). Gatekeeper training has been found to be an effective training model for suicide prevention by increasing awareness of suicide and suicidal behaviors and by promoting protective factors within the community as a whole (Chagon, Houle, Marcoux, & Renaud, 2007; Kalafat & Gagliano, 1996; Knox, Litts, Talcott, Feig, & Caine, 2003).

Gatekeeper Training

Gatekeeper programs are designed to increase knowledge of the behavioral and psychological warning signs of suicide and suicidal behaviors, to expand strategies for effectively discussing suicide with at-risk individuals, and to optimize the ability to refer to appropriate support services (Wyman et al., 2008). Research suggests gatekeeper training is effective in providing knowledge about suicide risks and warning signs, increasing suicide prevention skills, and improving self-efficacy (Aseltine & DeMartino, 2004; Cross, Matthieu, Cerel, & Knox 2007; Gould & Kramer, 2001; Issac et al., 2009; Kalafat & Elias, 1996; Mann et al., 2005). However, the link between participation in gatekeeper programs and decreased suicidal behavior has not been conclusively established (Gunnel & Frankel, 1994). This need is reinforced empirically by York and colleagues' (2013) systematic review, which suggests that student curriculum combined with gatekeeper training had positive effects on knowledge and attitudes, but negligible effects on suicide prevention behavior.

DOI: 10.1027/0227-5910/a000382

In order to determine whether gatekeeper training positively affects suicide prevention, researchers need to use a standardized assessment tool that examines the training's impact on the behavior of both the gatekeeper and the person at risk for suicide. In this study we developed such an assessment tool based on Kirkpatrick's (1996) four-level training evaluation model, which identifies levels in evaluating training effectiveness: reaction, learning, behavior, and results. Reaction is the level of satisfaction with the training itself; learning is the impact on attitudes, knowledge, and/or skills; behavior represents the change in actual behavior; and results are the final outcomes such as reduced suicide rates and ideation and better overall mental health (Kirkpatrick & Kirkpatrick, 2006). The learning stage is important to evaluate because behavior will not change if the desired knowledge, skills, and attitudes are not acquired. Thus, an important step is to measure the change in attitudes and intentions so as to identify suicide warning signs, discussing concerns and motivating a person to seek help and knowing where to refer him/her (Albright, Goldman, Shockley, Mcdevitt, & Akabas, 2011; Cross et al., 2011).

Currently, the modes of assessing gatekeeper training vary among researchers (Klimes-Dougan, Klingbeil, & Meller, 2013). Miller, Eckert, and Mazza (2009) found that only 30% of the research studies involving school suicide prevention programs included reliable and valid measures for assessing the training. Assessment criteria have included declarative knowledge (e.g., Cross et al., 2007; Wyman et al., 2008), self-efficacy (e.g., Cross et al., 2011; Matthieu, Cross, Batres, Flora, & Knox, 2008), self-rated skills, attitudes, and knowledge of suicide prevention behavior (e.g., Chagnon et al., 2007; Jacobson, Osteen, Sharpe, & Pastoor, 2012; Stuart, Waalen, & Haelstromm, 2003), and intentions to help (Capp, Deane, & Lambert, 2001). One standardized assessment tool is the Suicide Intervention Response Inventory (SIRI-2; Muthen, Muthen, Neimeyer, & Bonnelle, 1997), which is a promising metric for assessing situational judgments in a counseling setting. Additional skills may be needed for gatekeepers in a variety of other roles, however.

Berman (2009) reports that education alone does not create behavior change in suicide prevention program participants. Gatekeeper training aims to boost action through identifying, motivating, and referring people who may be at risk of suicide. Therefore, motivation theories should be incorporated into the development of an assessment tool that measures intent to act in addition to knowledge gained.

Motivation theories are psychological frameworks aimed at understanding goal-driven behavior and predicting future outcomes. Motivation theories describe several operative means of measuring behavioral intent. *Reasoned action theory* (Fishbein & Ajzen, 1975) posits that behavioral beliefs and subjective norms are the antecedents to behavioral intention, which is the direct precedent to behavior. Behavioral beliefs are theorized to underlie a person's attitude toward performing a behavior, or a person's potential to engage in a behavior. In order to measure participants' beliefs, we measure how prepared they believe they are to engage in suicide prevention behaviors, or their beliefs about potential to engage in this behavior. Secondly, behavioral intentions arise from an understanding of the likelihood that a certain action will produce a specific outcome. An individual's self-reported likelihood about a behavior acts as a measure of their intentions to engage in the behavior or the probability of taking action. *The theory of planned behavior* extends the boundary condition of pure volitional control specified by the theory of reasoned action (Madden, Ellen, & Ajzen, 1992). These authors state that "the more resources and opportunities individuals think they possess, the greater should be their perceived control over the behaviors" (p. 4). Related to this is Bandura's (1977) integrative framework of personal efficacy or perceived behavior control where *self-efficacy* is both a direct and indirect predictor of behavior. When self-efficacy is high and people feel confident in their abilities, this also leads to a sense of control in terms of ability to change their behavior in future circumstances. Self-efficacy is a person's perceived control to engage in behavior, which supports including it in an assessment tool. The combined motivation theories promote three separate concepts of behavioral intent needing assessment: preparedness to act or potentiality, likelihood to act or probability, and self-efficacy or control.

Validity Assessment

In order to complete the development of any assessment tool, it is incumbent on the researchers to address psychometric properties including reliability and validity. All four types of validity are addressed in this study.

Content Validity

Content validity refers to the extent that a measure represents all facets of a theoretical construct. As noted previously, the goal of gatekeeper training is to increase participants' knowledge of the behavioral and psychological warning signs of suicide and suicidal behaviors, to expand their strategies for effectively discussing suicide with at-risk individuals, and to optimize their ability to refer those in distress to appropriate support services (Wyman et al., 2008). We aimed to assess this content validity through the three facets of preparedness, likelihood, and self-efficacy. The preparedness subscale taps into the knowledge

and strategy component, the likelihood and self-efficacy subscales tap into the optimization of ability to refer individuals. These constructs and their accompanying items were developed using subject matter experts as will be described, which we believe provides evidence of content validity as well as construct validity.

Construct Validity

Further evidence for construct validity can be obtained statistically through the use of confirmatory factor analysis (CFA). CFA is a confirmatory technique that is driven by the theoretical framework of the inputted measures (Schreiber, Nora, Stage, Barlow, & King, 2006). It allows the researcher to test how well the data fit specified structural models, which offers insight into the number of distinct factors within a measure and whether items reflect the latent variable they are intended to measure or if they have cross-loadings on other factors.

Additionally, construct validity can be inferred by examining changes in an individual's Gatekeeper Behavior Scale (GBS) score due to an intervention. If the GBS accurately measures the construct of gatekeeper behaviors, it should be sensitive to changes due to training around said behaviors; therefore we expect to see within-subject changes on the GBS when comparing pre- and posttraining scores.

Criterion Validity

Criterion-related validity refers to the relationship between a construct and other constructs that it should theoretically predict. As an indication of criterion-related validity in the present context, we expected that scores on the GBS posttraining assessment would predict actual gatekeeper intervention behaviors as measured 3 months after undergoing training.

Convergent Validity

Convergent validity focuses on the notion that constructs should be correlated with other constructs that are theoretically similar. To test this idea, we focused on the self-efficacy aspect of the GBS, as self-efficacy is a widely used measure in gatekeeper training (Burnette, Ramchand, & Ayer, 2015). As evidence of convergent validity, our measure of specific self-efficacy in the GBS should be similar, but still provide unique information from measures of general self-efficacy.

By drawing on accepted theories of motivation, we developed a tool that measures behavioral intent to act on gatekeeper behavior, with a focus on preparedness to take action, likelihood of taking action, and efficacy to take action. Taking into account Kirkpatrick's model of evaluation and the prevailing theories of motivation, it is important to measure participants' intentions (preparedness, likelihood, and self-efficacy) to engage in gatekeeper behavior. A satisfactory tool needs to assess how prepared participants are to engage in the desired gatekeeper behaviors, the likelihood that they will act on these behaviors, and their confidence (self-efficacy) to engage in them.

The purpose of the present study is to use theories of motivation to create and validate a comprehensive measure that assesses the effectiveness of gatekeeper programs adhering to Kirkpatrick's evaluation model for training. We refer to this measure as the Gatekeeper Behavior Scale, which is made up of three subscales: preparedness to aid people in psychological distress, likelihood to help those in psychological distress, and self-efficacy in helping those in psychological distress. In assessing validity, we focus on four types of validity: construct and content validity (assessed via CFA), criterion validity (assessed via correlations and regression with behavioral measures), and convergent validity (assessed via correlations with similar but distinct constructs).

Method

Development of the Gatekeeper Behavior Scale

The process of developing and refining the GBS items involved several rounds of revisions from subject matter experts (SMEs) across multiple universities over a 6-month period. These PhD- and MD-level SMEs have backgrounds in psychology, mental health, suicide prevention, gatekeeper training, industrial and organizational psychology, survey design, and behavioral health. The measures were first employed in pilot studies with university students and faculty as well as with secondary school administrators and teachers. The gatekeeper trainings used for this study were drawn from Kognito's (2015) suite of At-Risk mental health learning simulations. Feedback from pilot studies and the SMEs was used to reduce the GBS into three subscales: preparedness, likelihood, and self-efficacy for helping those in psychological distress.

Participants

Between 2010 and 2013, faculty, staff, and students from geographically dispersed areas within the US completed one of five different online gatekeeper trainings based on the following populations: university faculty and staff, high school faculty and staff, middle school faculty and staff, veteran-specific awareness for university faculty and staff, and university students (see Table 1 for demographics). A total of 8,657 participants across the five training

Table 1. Participant demographics[a]

Variable	*N* reporting		%	*M (SD)*
Gender	8,587	Female	70.0%	
		Male	29.6%	
		Transgender	0.4%	
Ethnicity	7,156	Hispanic	14.1%	
Race	7,447	White/Caucasian	82.5%	
		Native American	2.4%	
		Asian	4.4%	
		Black/African American	9.6%	
		Other	1.0%	
Employment role	7,762	Educator	43.9%	
		Staff/admin	32.0%	
		Otherb	16.3%	
Received prior gatekeeper training in suicide prevention	6,271		13.3%	
Received some training to become a mental health practitioner	5,230		15.2%	
Age (years)	5,296			39.48 (10.04)
Organizational tenure (years)	4,519			11.91 (9.67)

Note. [a]Due to different constraints, not all demographics were assessed for all studies. [b]Includes resident assistant or student leader for student population.

simulations were used to assess the validity of the GBS. Participants were recruited mainly by e-mails from school administrators' college/university counseling department heads, nonprofit organizations, and state youth suicide prevention coordinators.

Procedure

Before undergoing the gatekeeper training, participants completed the prestudy measures, detailed here. Participants then completed the training session, which lasted from 45 min to 1 hr. The online training involves participant engagement in role-play conversations with emotionally responsive virtual students who exhibit signs of psychological distress, including suicidal tendencies in conversations. In this process, participants practiced and learned to use motivational interviewing strategies such as open-ended questions and reflective listening in order to effectively broach the topic of psychological distress, discuss their concern with students, and avoid common pitfalls such as attempting to diagnose the problem, being judgmental, or giving unwarranted advice.

Upon completion of the training, participants completed an additional survey. Finally, participants were assessed again 3 months after training (hereafter called the follow-up time point).

To measure construct validity, we used CFA techniques on posttraining scores to assess the distinct factors within the GBS. We also compared changes in individual GBS scores at posttraining and pretraining. To assess criterion validity, we investigated the relationship between GBS scores at posttraining and self-reported gatekeeper intervention behaviors reported at follow-up. To assess convergent validity, we compared the self-efficacy subscale from the posttraining assessment with general self-efficacy items also measured during the posttraining assessment.

Measures

Gatekeeper Behavior Scale

The GBS is composed of 11 items and three subscales. The scale items and response scales are listed in Table 2. To compute the overall composite GBS score, we first averaged all of the items within each subscale of preparedness, likelihood, and self-efficacy. Then, owing to differing response scales, we standardized the mean score for each subscale using the percentage of maximize possible (POMP) method, which expresses the raw scores as a percentage of maximize possible score (Fischer & Milfont, 2010). We then computed an overall GBS score by averaging the three subscales' POMP score. The GBS scale was administered both pre- and posttraining, and in the 3-month follow-up.

General Self-Efficacy

Overall self-efficacy was measured using the eight items from the New General Self-Efficacy Scale validated by Chen, Gully, Stanley, and Eden (2001). Participants were

Table 2. Gatekeeper Behavior Scale

Subscale	Number	Item	Response Scale
Preparedness	How would you rate your preparedness to:		
	Prep 1	Recognize when a student's behavior is a sign of psychological distress	1 = Very low 2 = Low 3 = Medium 4 = High 5 = Very high
	Prep 2	Recognize when a student's physical appearance is a sign of psychological distress	
	Prep 3	Discuss with a student your concern about the signs of psychological distress they are exhibiting	
	Prep 4	Motivate students exhibiting signs of psychological stress to seek help	
	Prep 5	Recommend mental health support services (such as the counseling center) to a student exhibiting signs of psychological distress	
Likelihood	Like 6	How likely are you to discuss your concerns with a student exhibiting signs of psychological distress?	1 = Very unlikely 2 = Unlikely 3 = Likely 4 =Very likely
	Like 7	How likely are you to recommend mental health/ support services (such as the counseling center) to a student exhibiting signs of psychological distress?	
Self-Efficacy	Please rate how much you agree/disagree with the following statements:		
	Effi 8	I feel confident in my ability to discuss my concern with a student exhibiting signs of psychological distress	1 = Strongly disagree 2 = Disagree 3 = Agree 4 = Strongly agree
	Effi 9	I feel confident in my ability to recommend mental health support services to a student exhibiting signs of psychological distress	
	Effi 10	I feel confident that I know where to refer a student for mental health support	
	Effi 11	I feel confident in my ability to help a suicidal student seek help	

told that different people have different levels of ability for performing different tasks and asked to respond in terms of their abilities (e.g., "In general, I think that I can obtain outcomes that are important to me"). Responses ranged on a 5-point Likert scale from *not at all* to *a very little extent* to *a very great extent*. Internal consistency reliability was high ($\alpha = .94$). The measure was only administered once, during the posttraining survey.

Gatekeeper Intervention Behaviors

Participants were asked to indicate over the time period of the previous two academic months the estimated number of students that they had: (a) been concerned about owing to psychological distress including suicidal ideation, (b) discussed their concerns with, and (c) referred to appropriate services. A behavior composite was created by computing the mean of the three behavior items. Behaviors were assessed twice, in the pretraining survey and the 3-month follow-up.

Results

First, because our data came from multiple samples, we determined the merit of combining multiple smaller samples into one larger sample. We ran an ANOVA for each item to determine whether any significant differences appeared between the five training groups we used. The *F* test indicated that for all 11 items on the GBS mean score there was at least one significant difference ($p < .05$) between groups. Tukey's post-hoc tests help identify the differences between the various training group's mean item scores, but no group consistently scored low or high across all items. To further investigate the issue, since we have a relatively large sample size, we calculated the omega square (ω^2) effect size to determine how much of the variance was explained by the different groups versus the sample size. The ω^2 for all 11 items only ranged between .004 and .050, indicating that the statistical differences we found between modules were based on the large power rather than meaningful variation between the groups. It was then determined sufficient to combine the five different groups into one larger sample for the purposes of measuring the validity of the GBS assessment.

Construct Validity

We tested the construct validity of the GBS using CFA via Mplus version 6 (Muthen & Muthen, 2010). We tested the GBS with multiple factor models (i.e., single factor, two, three, and four factors) to determine the best fit for the data. To evaluate the goodness of fit, we used the chi-squared test (χ^2), comparative fit index (CFI), Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). A large χ^2 result indicates a poor fit of the data, but is recognized as problematic as it is very sensitive to sample size. To combat this drawback, other measures of

Table 3. Summary of confirmatory factor analysis (CFA) results

Model	χ^2	No. of free parameters	*df*	*p*	CFI	TLI	RMSEA
A Three-factor	3,612.02	36	41	< .001	.954	.938	.028
B Two-factor Like/Effi	9,526.93	34	43	< .001	.877	.843	.108
C Two-factor Prep/Like	7,780.49	34	43	< .001	.900	.872	.066
D One-factor	27,644.17	33	44	< .001	.642	.553	.143
E Four-factor (four steps)	25,164.32	36	41	< .001	.674	.563	.130

Note. Items used for CFA are from posttraining assessment. CFI = comparative fit index. TLI = Tucker–Lewis index. RMSEA = root mean square error of approximation. Like = likelihood. Effi = self-efficacy. Prep = preparedness.

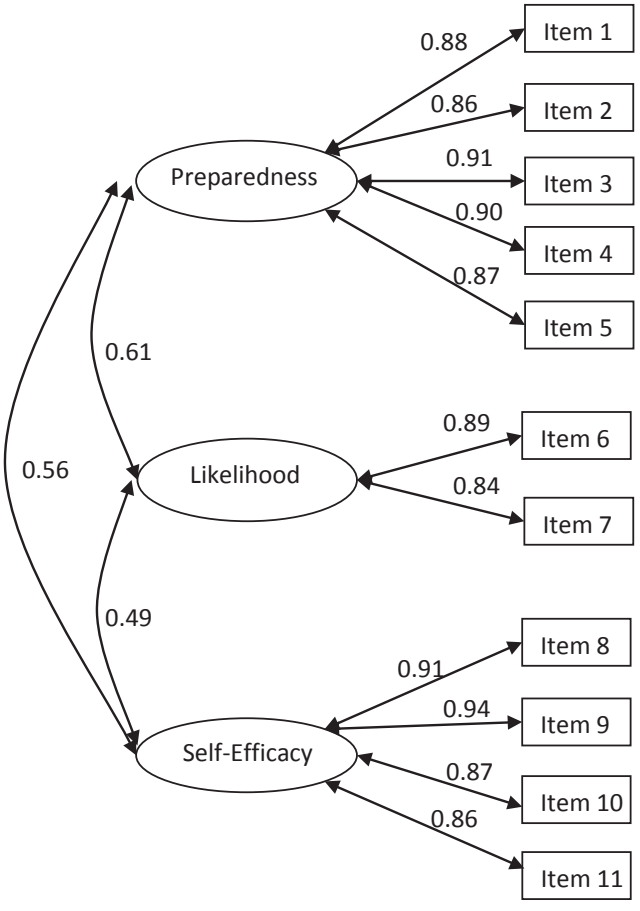

Figure 1. Three-factor CFA model. The standardized model coefficients are all $p < .001$.

fit have been developed (Albright & Park, 2009). CFI, TLI, and RMSEA are all relatively insensitive to sample size. The CFI ranges from 0 to 1 with a value of .90 or greater indicating an acceptable fit. An RMSEA value of about .06 or less is considered to indicate a close fit of the model. There are no hard-set rules for determining goodness of fit, but Albright and Park (2009) state that, in general, "when RMSEA values are close to .06 or below and CFI and TLI are close to .95 or greater, for example, the model may have a reasonably good fit" (p. 7). Therefore, we report the χ^2, RMSEA, CFI, and TLI to better determine the best model.

The results of the CFA analyses are presented in Table 3. Model A, the a priori model, included preparedness, likelihood, and self-efficacy as three separate factors. The single-factor model included all GBS items on a single factor (Model D). We tested two two-factor models: one with preparedness items on the first factor and likelihood/self-efficacy items on the second factor (Model B) and one with preparedness and likelihood measures on the first factor and self-efficacy on the second factor (Model C). The one-factor model (Model D) of all items together produced a poorer fit than both the two-factor and the three-factor models. Lastly, we investigated a four-factor model (Model E) based on the four specific gatekeeper behaviors that included separate factors of recognizing, discussing, motivating, and referring people in psychological distress. The three-factor model (A) produced the best fit on all three fit indices (CFI, TLI, and RMSEA). The standardized factor loadings for Model A are presented in Figure 1. Furthermore, the GBS was found to have high internal consistency ($\alpha = 0.93$). In summary, taking into account the high internal reliability and the CFA models, we determined that the three-factor model of the GBS best represents the intended theoretical framework delineated by the three subscales of preparedness, likelihood, and self-efficacy.

In addition to the CFA to assess construct validity, we investigated how gatekeeper training affects participants' GBS score. A repeated sample *t* test was conducted to compare participants' GBS scores before and after undergoing gatekeeper training. Participants' composite GBS score was significantly higher after training ($M = 78.9$, $SD = 14.9$) than before training, $M = 63.7$, $SD = 18.2$; $t(5176) = -71.9$, $p < .01$ and Cohen's $d = -1.02$. The large effect size indicates that the significant difference in GBS scores between the pre- and posttraining periods are not due to sample size. These gains on the GBS due to training indicate the additional construct validity of the assessment.

Criterion Validity

To explore the criterion-related validity, we used zero-order correlations and linear regression to determine how

Table 4. Correlations among variables

	GBS: Comp[2]	Prep: Comp[3]	Like: Comp[3]	Effi: Comp[3]	Beh: Comp[4]	Beh. Conc[5]	Beh. Appr[5]	Beh. Refer[5]	Prep1[6]	Prep2[6]	Prep3[6]	Prep4[6]	Prep5[6]	Like6[6]	Like7[6]	Effi8[6]	Effi9[6]	Effi10[6]	Effi11[6]
GBS: Comp[2]	—																		
Prep: Comp[3]	.852	—																	
Like: Comp[3]	.846	.548	—																
Effi: Comp[3]	.891	.697	.608	—															
Beh: Comp[4]	.270	.224	.217	.254	—														
Beh. Concern[5]	.260	.214	.215	.241	.985	—													
Beh. Approach[5]	.278	.232	.221	.264	.986	.964	—												
Beh. Refer[5]	.257	.214	.203	.244	.977	.938	.943	—											
Prep1[6]	.761	.914	.481	.615	.198	.189	.205	.189	—										
Prep2[6]	.735	.893	.456	.593	.156	.142	.166	.151	.846	—									
Prep3[6]	.791	.922	.507	.654	.206	.195	.211	.200	.803	.766	—								
Prep4[6]	.790	.919	.511	.653	.210	.200	.217	.200	.775	.734	.833	—							
Prep5[6]	.793	.896	.538	.652	.216	.214	.220	.203	.736	.712	.787	.832	—						
Like6[6]	.802	.535	.936	.577	.214	.212	.221	.195	.483	.456	.505	.495	.491	—					
Like7[6]	.779	.489	.935	.559	.194	.192	.194	.186	.415	.395	.442	.459	.513	.749	—				
Effi8[6]	.817	.664	.559	.894	.224	.211	.236	.212	.603	.570	.653	.617	.574	.564	.479	—			
Effi9[6]	.829	.641	.582	.920	.237	.228	.246	.224	.550	.537	.594	.598	.631	.535	.552	.796	—		
Effi10[6]	.760	.575	.515	.874	.216	.205	.221	.210	.504	.484	.523	.531	.569	.468	.493	.671	.762	—	
Effi11[6]	.783	.618	.519	.888	.191	.178	.197	.189	.548	.536	.576	.591	.557	.497	.473	.733	.735	.691	—

Note. Data are from participants with both pre- and posttraining assessment data (N = 5,174). GBS = Gatekeeper Behavior Scale. Like = likelihood. Effi = self-efficacy. Beh = behavior. Prep = preparedness. POMP = percentage of maximum possible. [2] = Composite is average of GBS's three dimension POMP scores from post-training. [3] = Composite is average of all dimension items from posttraining. [4] = Composite computed by averaging dimension items from follow-up training (N = 1,008). [5] = From follow-up assessment (N = 1,008). [6] = From posttraining assessment (N = 5,174).
All correlations are significant at $p < .01$.

well the posttraining assessment's preparedness, likelihood, and self-efficacy dimensions predict participants' engagement in gatekeeper intervention behaviors at the 3-month follow-up.

The GBS significantly correlates with the three gatekeeper intention behaviors of being concerned ($r = .26$, $p < .01$), approaching ($r = .28$, $p < .01$), and referring ($r = .26$, $p < .01$). We also investigated the relationship between the GBS at the item level and the three behavioral items. Table 4 shows the correlations between the GBS composite score, three subscale scores, behavioral composite, three behavioral items, and 11 items from the GBS. All the individual items significantly correlate with each of the gatekeeper intention behaviors and the behavioral composite, providing evidence of criterion validity that those with higher GBS scores are more likely to engage in gatekeeper intervention behaviors.

Multiple regression analysis was used to test whether the three dimensions' mean scores from the GBS at posttraining significantly predicted participants' self-reported gatekeeper intervention behaviors composite at follow-up. The results of the regression indicated that all three predictors explained 7.4% of the variance, $R^2 = .074$, $F(3, 1002) = 26.57$, $p < .001$. Preparedness significantly predicted gatekeeper intervention behavior ($\beta = .30$, $p < .05$) as did Likelihood ($\beta = .43$, $p < .05$) and Self-Efficacy ($\beta = .80$, $p < .01$). As anticipated, higher participant ratings on the GBS dimensions from the posttraining assessment predicted more self-reported gatekeeper intervention behavior as measured at the 3-month follow-up.

Convergent Validity

For convergent validity, we found a robust correlation between our measure of self-efficacy in motivating someone to seek help and general self-efficacy ($r = .43$, $p < .01$). This correlation is high enough to demonstrate convergent validity but not so high as to suggest complete overlap in the two constructs. The self-efficacy component of the GBS is therefore unique from general self-efficacy.

Discussion

A validation study of the 11-item GBS was conducted in order to create a unique tool to measure the effectiveness of gatekeeper training by assessing participants' skills, attitudes, and intentions, which predicts engagement to help those in psychological distress and those at risk of suicide. To our knowledge, such a standard and validated assessment does not exist even though some researchers have used tools validated on different populations or constructed tools based on previous research.

The CFA findings show that the GBS measures three distinct components of suicide prevention behavior that include preparedness to recognize and motivate someone who may be in psychological distress, the likelihood that participants will engage in the suicide prevention behavior, and the self-efficacy or confidence that they can guide a suicidal person to seek help. Through exploring different factor models, the data indicate the best fit is the three-factor model that was theorized based on motivational theories of behavior measuring potentiality, probability, and control. Items in each GBS subscale correlate with one another more highly than they do with other items, indicating three separate constructs are being measured. The high Cronbach's a for all 11 items suggests the GBS measures an overall concept of suicide prevention intentions. The significant criterion-related validity suggests that the GBS measures the probability of engaging in the wanted behavior. Also, preparedness, likelihood, and self-efficacy provide significant predictive ability for gatekeeper behavior, accounting for 7% of the variance in the composite behavior score. While the regression analysis only explains about 7% of the variability in the behavior outcome, it is considered within an acceptable range for predicting such complex behavior providing validity for the three measures to predict gatekeeper behavior. The significant convergent validity suggests that the GBS measures self-efficacy, but more specifically confidence in employing gatekeeper behavior rather than general self-efficacy.

To determine the benefits of gatekeeper training, we need to more accurately assess suicide prevention behavioral intentions before training to determine if intentions are learned after training, and, more importantly, to identify actual changes in participants' behavior as theorized by Kirkpatrick and Kirkpatrick (2006). By evaluating whether the training increases intentions to act, we can more accurately link the training to actual behavioral changes in the results phase of gatekeeper programs such as changes in suicide prevention behavior, more open communication, referrals, and identification of those at risk. The GBS appears to be a valid tool for assessing a person's gatekeeper behavior to question, discuss, motivate, and refer those in psychological distress to mental health support services.

Limitations and Future Research

Further studies need to be conducted to validate the tool on other forms of gatekeeper training such as face-to-face training. Also, to further validate the GBS we need to compare it with other performance-based criteria such as referrals reported by mental health support services or

reductions in suicidal behavior. Another limitation is that participants self-selected to take the training and therefore may have been more motivated and not representative of the general population within an educational setting. Also, demographic data showed the participant population was skewed toward being female and middle aged and was not evenly divided between educational settings, which also could have compromised the generalizability.

Furthermore, although not included in this particular study, we recognize the need to compare the GBS with other gatekeeper trainings in further studies.

To examine the potential of the GBS as a predictive tool, the authors are conducting a regression analysis to predict changes in gatekeeper behavior. Additionally, we are investigating any differences in population subsets including measures across all ethnicities, gender, jobs, or ages. A more in-depth dissection of the GBS data is needed to determine whether there is cultural bias in the wording of the questions.

Conclusion

The validation of the GBS tool may allow researchers to generalize their findings to the larger population. Through validation, researchers can use the GBS to explore how gatekeeper trainings affect different populations or how different practices affect participants' behavioral intentions. By focusing on intent to engage in specific behavior, we can more appropriately assess the usefulness of gatekeeper programs since the main purpose of suicide prevention training is to get people to take action and intercede on behalf of those at risk.

References

Albright, J. J., & Park, H. M. (2009). Confirmatory factor analysis using Amos, LISREL, Mplus, SAS/STAT CALIS (UITS Working paper). Indiana University, Center for Statistical and Mathematical Computing. Retrieved from http://www.indiana.edu/~statmath/stat/all/cfa/index.html

Albright, G., Goldman, R., Shockley, K., Mcdevitt, F., & Akabas, S. (2011). Using an avatar-based simulation to train families to motivate veterans with post-traumatic stress to seek help at the VA. *Games for Health, 1*(1), 21–28. doi:10.1089/g4h.2011.0003

Aseltine, R. H., & DeMartino, R. (2004). An outcome evaluation of the SOS suicide prevention program. *American Journal of Public Health*, *94*, 446–451.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Berman, A. L. (2009). School-based suicide prevention: Research advances and practice implications. *School Psychology Review, 38*(2), 233

Burnette, C., Ramchand, R., & Ayer, L. (2015). *Gatekeeper training for suicide prevention: A theoretical model and review of the empirical literature.* Santa Monica, CA: Rand National Defense Research Institute.

Centers for Disease Control and Prevention. (2010). Web-based Injury Statistics Query and Reporting System (WISQARS). Atlanta, GA: National Center for Injury Prevention and Control. Retrieved from http://www.cdc.gov/injury/wisqars/index.html

Capp, K., Deane, F. P., & Lambert, G. (2001). Suicide prevention in Aboriginal communities: Application of community gatekeeper training. *Australian and New Zealand Journal of Public Health*, *25*(4), 315–321.

Chagon, F., Houle, J., Marcoux, I., & Renaud, J. (2007). Control-group study of an intervention training program for youth suicide prevention. *Suicide and Life-Threatening Behavior*, *37*, 135–144.

Chen, G., Gully, S., M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, *4*, 62–83.

Crosby, A. E., Han, B., Ortega, L. A. G., Parks, S. E., & Gfoerer, J. (2011). Suicidal thoughts and behaviors among adults aged ≥18 years – United States, 2008–2009. *CDC Morbidity and Mortality Weekly Report* (no. SS-13). Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6013a1.htm

Cross, W., Matthieu, M. M., Cerel, J., & Knox, K. L. (2007). Proximate outcomes of gatekeeper training for suicide prevention in the workplace. *Suicide and Life-Threatening Behavior*, *37*(6), 659–670. doi:10.1521/suli.2007.37.6.659

Cross, W., Seaburn, D., Gibbs, D., Schmeelk-Cone, K., White, A. M., & Caine, E. D. (2011). Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. *The Journal of Primary Prevention, 32*(3), 195–211. doi:10.1007/s10935-011-0250-z

Fischer, R., & Milfont, T. L. (2010). Standardization in psychological research. *International Journal of Psychological Research*, *3*(1), 88–96.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gould, M. S., & Kramer, R. A. (2001). Youth suicide prevention. *Suicide and Life-Threatening Behavior*, *31*(Suppl), 6–31. doi:10.1521/suli.31.1.5.6.2421

Gunnell, D., & Frankel, S. (1994). Prevention of suicide: Aspirations and evidence. *BMJ (Clinical Research Ed.), 308*(6938), 1227–1233.

Isaac, M., Elias, B., Katz, L. Y., Belik, S., Deane, F. P., Enns, M. W., Swampy Creek Suicide Prevention Team. (2009). Gatekeeper training as a preventative intervention for suicide: A systematic review. *Canadian Journal of Psychiatry*. *Revue Canadienne De Psychiatrie*, *54*(4), 260.

Jacobson, J. M., Osteen, P. J., Sharpe, T. L., & Pastoor, J. B. (2012). Randomized trial of suicide gatekeeper training for social work students. *Research on Social Work Practice, 22*(3), 270–281. doi:10.1177/1049731511436015

Kalafat, J., & Elias, M. J. (1996). The use of simulations in an educational context: Broad and narrow foci. *Suicide and Life-Threatening Behavior*, *25*, 123–133.

Kalafat, J., & Gagliano, C. (1996). The use of simulations to assess the impact of an adolescent suicide response curriculum. *Suicide and Life-Threatening Behavior*, *26*(4), 359–364.

Kirkpatrick, D. L. (1996). Great ideas revisited. Techniques for evaluating training programs. Revisiting Kirkpatrick's four-level model. *Training and Development*, *50*(1), 54–59.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.) San Francisco, CA: Berrett-Koehler Publishers.

Klimes-Dougan, B., Klingbeil, D. A., & Meller, S. J. (2013). The impact of universal suicide-prevention programs on the help-seek-

ing attitudes and behaviors of youths. *Crisis*, *34*(2), 82–97. doi:10.1027/0227-5910/a000178

Knox, K. L., Litts, D. A., Talcott, G. W., Feig, J. C., & Caine, E. D. (2003). Risk of suicide and related adverse outcomes after exposure to a suicide prevention program in the US Air Force: Cohort study. *British Journal of Medicine*, *327*, 1376–1378.

Kognito. (2015). *Products*. Retrieved from http://www.kognito.com/products/

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin, 18*(1), 3–9. doi:10.1177/0146167292181001

Mann, J. J., Marusic, A., Mehlum, L., Patton, G., Phillips, M., Rutz, W., & Malone, K. (2005). Suicide prevention strategies: A systematic review. *JAMA*, *294*(16), 2064–2074. doi:10.1001/jama.294.16.2064

Matthieu, M. M., Cross, W., Batres, A. R., Flora, C. M., & Knox, K. L. (2008). Evaluation of gatekeeper training for suicide prevention in veterans. *Archives of Suicide Research, 12*(2), 148–154. doi:10.1080/13811110701857491

Miller, D. N., Eckert, T. L., & Mazza, J. J. (2009). Suicide prevention programs in the schools: A review and public health perspective. *School Psychology Review, 38*(2), 168.

Muthen, L. K., & Muthen, B. O. (2010). Mplus (Version 6) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Neimeyer, R. A., & Bonnelle, K. (1997). The suicide intervention response inventory: A revision and validation. *Death Studies, 21*(1), 59–81. doi:10.1080/074811897202137

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338.

Stuart, C., Waalen, J., & Haelstromm, E. (2003). Many helping hearts: An evaluation of peer gatekeeper training in suicide risk assessment. *Death Studies*, *27*(4), 321–333.

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

York, J., Pearson, J., Lamis, D. A., Friedman, L., Berman, A. L., Joiner, T. E., Gutierrez, P. M. (2013). A systematic review process to evaluate suicide prevention programs: A sample case of community-based programs. *Journal of Community Psychology*, *41*(1), 35–51. doi:10.1002/jcop.21509

Received May 8, 2014
Revision received October 30, 2015
Accepted November 2, 2015
Published online June 1, 2016

About the authors

Glenn Albright is an associate professor and former chair of the Department of Psychology at Baruch College of the City University of New York, NY. His research examines the impact of virtual human role-play simulations in bringing about sustained changes of behavior in the areas of health and behavioral health.

Jesse Davidson is HR Manager of Advanced Analytics at Pfizer, New York, USA. Previously he was a research assistant at Columbia University and survey consultant for Kognito. He holds an MS in Industrial/Organizational Psychology from Baruch College and a BA in Psychology and Dramatic Arts from the University of North Carolina.

Ron Goldman is an innovator, tech executive, data scientist, and social entrepreneur. Ron co-founded and has served as CEO of Kognito, New York, USA, since 2003 where he has pioneered the use of gaming and simulation technology for learning experiences in social, emotional, and physical health. Ron is a member of the advisory board of the *Games for Health Journal.*

Kristen Shockley is Associate Professor of Psychology at Baruch College of the City University of New York, NY. She received her PhD in industrial/organizational psychology from the University of South Florida, and her main areas of research are aimed at understanding how individuals manage work and family and career development.

Dr. Timmons-Mitchell is a child clinical psychologist and Associate Clinical Professor of Psychology in the Department of Psychiatry, Case Western Reserve University School of Medicine in Cleveland, OH, and an adjunct faculty member at the Mandel School. She is Evaluator of the Ohio Suicide Prevention Foundation Garrett Lee Smith Grant from SAMHSA.

Glenn Albright
Baruch College, City University of New York
Department of Psychology
One Bernard Baruch Way
New York, NY 10010
Tel. +1 (646) 312-3780
Fax +1 (646) 312-3781
E-mail glenn.albright@baruch.cuny.edu

