Measuring trainer fidelity in the transfer of suicide prevention training

Wendi F. Cross, Ph.D., **Anthony R. Pisani, Ph.D.**, **Karen Schmeelk-Cone, Ph.D.**, **Yinglin Xia, Ph.D.**, **Xin Tu, Ph.D.**, **Marcie McMahon, M.A.**,
Center for Employee Wellness at the School of Nursing at the University of Rochester Medical Center (URMC), Rochester, NY, USA

Jimmie Lou Munfakh, B.A., and **Madelyn S. Gould, Ph.D.**

Abstract

Background—Finding effective and efficient models to train large numbers of suicide prevention interventionists, including 'hotline' crisis counselors, is a high priority. Train-the-trainer (TTT) models are widely used but understudied.

Aims—To assess the extent to which trainers following TTT delivered the Applied Suicide Intervention Skills Training (ASIST) program with fidelity, and to examine fidelity across two trainings and seven training segments.

Methods—We recorded and reliably rated trainer fidelity, defined as adherence to program content and competence of program delivery, for 34 newly trained ASIST trainers delivering the program to crisis center staff on two separate occasions. A total of 324 observations were coded. Trainer demographics were also collected.

Results—On average, trainers delivered two-thirds of the program. Previous training was associated with lower levels of trainer adherence to the program. 18% of trainers' observations were rated as solidly competent. Trainers did not improve fidelity from their first to second training. Significantly higher fidelity was found for lectures and lower fidelity was found for interactive training activities including asking about suicide and creating a safe plan.

Conclusions—We found wide variability in trainer fidelity to the ASIST program following TTT and few trainers had high levels of both adherence and competence. More research is needed to examine the cost-effectiveness of TTT models.

Keywords

Suicide prevention; train-the-trainer; fidelity; observations; telephone crisis service

Introduction

Suicide is a significant global public health problem accounting for almost a million deaths each year (World Health Organization, 2012). Telephone crisis lines are widely used across a number of countries (Coveney, Pollack, Armstrong, & Moore, 2012; King, Nurcombe, Bickman, Hides, & Reid, 2003) and is an important component of the United States' (US) national suicide prevention strategy (Covington, Hogan, Abreu, Berman, & Breux, 2011;

Bobevski & Holgate, 1997; Kalafat, Gould, Munfakh, & Kleinman, 2007; Gould, Kalafat, Munfakh, & Kleinman, 2007; Gould et al., 2013). Over 800,000 callers used the US National Suicide Prevention Lifeline ("Lifeline", www.suicidepreventionlifeline.org) crisis services in 2012. Given the numbers of at-risk individuals who access telephone crisis services, it is critical that counselors who respond to callers are well trained in the knowledge, attitudes and skills necessary to assess and intervene with those at- risk for suicide. A key challenge for crisis centers, and for other gatekeeper and community level suicide preventive interventions, is to provide high quality standardized training for a large number of counselors dispersed across the country in a cost effective manner.

One approach to educating a large number of people to learn and then disseminate standardized programs is the train-the-trainer (TTT) model. TTT programs involve master trainers who teach program content, along with the process of how to deliver the training, to others who then conduct their own training to the target audience. Those individuals then use the knowledge and skills obtained through the training to carry out the target behaviors. The TTT model depends upon the assumption that educational interventions can be effectively transmitted across generations of trainers. The main benefit of disseminating an intervention through TTT is the ability to rapidly, and relatively inexpensively, train large numbers of people (Welber, 2002). TTT models are frequently used in health care and health education (Allen, Connelly, Morris, Elmer, & Zwickey, 2011; Assemi, Mutha and Hudman, 2007; Besculides, Trebino, & Nelson, 2011; Byrne, Willis, Deane, Hawkins, & Quinn, 2010), disaster preparedness (Abatemarco, Beckley, Borjan, and Robson, 2007; Becker, 2009; Cross, Cerulli, Richards, He, & Hermann, 2010; Gelkopf, Ryan, Cotton & Berger, 2008); and school prevention and intervention programs (Bowes, Marquis, Young, Holowaty & Isaac, 2009). The underlying assumption of the TTT model is that newly trained instructors deliver competent training that is comparable to the original training with minimal loss of information. Yet this primary assumption has yet to be tested in a large-scale prevention program, raising questions about the ability of community implementers to deliver the standardized training as intended, with fidelity.

Observational methods for measuring fidelity of program implementation are the current 'gold standard' for measuring fidelity because they are objective, unlike self-report measures, which tend to be biased and positively skewed (Carroll et al., 2000; Lillehoj, Griffin & Spoth, 2004; Miller & Mount, 2001; Moore, Beck, Sylvertsen & Domitrovich, 2009). Objective measures of fidelity are also more likely to be linked to outcomes than self-report data (Hansen, Graham, Wolkenstein, & Rohrbach, 1991; Hogue et al., 2008; Knoche, Sheridan, Edwards, & Osborn, 2010; Lillehoj, Griffin & Spoth, 2004; Trepka, Rees, Shapiro, Hardy, & Barkham, 2004). The need for research that examines training behaviors is great and has been underscored in recent publications (Cross et al., 2010; McHugh & Barlow, 2010; Pisani, Cross & Gould, 2012; Segre et al., 2011). Research examining trainer fidelity is particularly critical for intervention research because low fidelity of trainers could lessen the impact of an otherwise effective intervention by distorting research findings and potentially diluting an intervention that could save lives.

The present study examined transfer of training in the TTT model of the manualized Applied Suicide Intervention Skills Training program (ASIST; Lang, Ramsay, Tanney &

Kinzel, 2008; LivingWorks, 2010) conducted as part of a large randomized control trial of the intervention for crisis hotline counselors during a nation-wide roll out of ASIST across the Lifeline network (Gould et al., 2013). ASIST is an internationally disseminated gatekeeper training program which was modified with input from ASIST developers and training personnel for crisis center training. We examined trainer fidelity while delivering the ASIST program to crisis center staff. The first aim of the study was to assess the extent to which trainers adhered to the training and how competent they were in their delivery of the training. The second aim was to assess whether trainers improve fidelity of program delivery over time. To our knowledge, this is the first rigorously conducted observational investigation of the quality of a TTT program reported in the suicide literature. The methods and findings are generalizable, however, to other programs that use TTT, particularly for those programs that use group-based training formats.

Methods

The training program—The Applied Suicide Intervention Skills Training (ASIST) program (Lang, Ramsay, Tanney & Kinzel, 2008) utilizes a TTT approach (referred to as T4T in ASIST) for new instructors to learn to deliver a 2-day training workshop. New instructor trainees attend a 5-day workshop where they receive instruction on the program content and delivery skills from ASIST master trainers, including a detailed trainer manual, role play and group activities. (For additional information on ASIST training, see LivingWorks, 2010).

Participants—The current study occurred in the context of a RCT to test the impact of the ASIST program, version X.2, on Lifeline's crisis center counselors' interventions and at-risk callers' outcomes (see Gould et al., 2013, for details on methods and recruitment in the parent study.) Seventeen of 18 hotline centers recruited for the dynamic waitlist design randomized trial participated in the current study across three phases. One of the 18 centers was not able to participate due to their loss of funding for their clinical operations. Two staff from each crisis center participated in the TTT program delivered by ASIST master trainers. The trainers were chosen by their centers because they either had experience administering trainings or were in a supervisory role, and were likely to remain at the center to administer future ASIST trainings. Each pair of trainers subsequently conducted and video-taped the 2-day training with their center staff counselors on two separate occasions. Thus, the training was provided to all shifts and the trainers' fidelity was recorded over two separate training events. We have ratings from 17 centers each with two trainers for the first training delivered (Time 1), but due to audio failure during one center's second training we only have 16 centers with 2 trainers for the second training delivered. Therefore, in summary, there were 34 trainers from 17 hotline centers who conducted a total of 66 trainings. To assess for possible differences between the crisis center without video-tape for the second training delivered (Time 2) and the other centers, we compared the center's Adherence and Competence scores at Time 1 to the other centers. We found no significant differences and therefore used all collected data where applicable. A total of 324 recorded observations were coded.

In addition to the trainings from the 17 study centers outlined above, two centers were recruited to pilot the data collection process. Video tapes from these two "development" centers' trainings were used to develop the fidelity measures and for observational coder training purposes. Ratings from the development centers were not included in analyses.

The project's protocol was approved by the Institutional Review Boards of the New York State Psychiatric Institute/Columbia University.

Measures

Fidelity measurement development—We focused on two dimensions of fidelity at the implementer level that are fundamental to the quality of implementation: 1) *adherence* to program content as specified in manuals, and, 2) *competence* in program delivery which describes the quality of the implementation (Cross & West, 2011; Forgatch & DeGarmo, 2011; Schoenwald, Garland, Chapman, Frazier, Sheidow, & Southam-Gerow, 2011; Waltz, Addis, Koerner & Jacobson, 1993). These constructs were used to assess overall fidelity across newly trained instructors delivering the 2-day suicide prevention program to their centers' staff. The 2-day training content is divided into five sections: *Preparing, Connecting, Understanding, Assisting,* and *Networking*. We identified, in collaboration with ASIST program developers, seven segments from the *Understanding* and *Assisting* phases of training for observation and rating. These segments were specifically chosen because they reflected core elements of the program's Suicide Intervention Model (SIM), captured the range of trainer behaviors most relevant to the measurement of adherence to this program, included a range of didactic and active learning activities (role play, simulation), and avoided periods of personal sharing among participants to minimize privacy concerns and sensitivity to recording.

Development of each of the seven Adherence scales involved the following process: isolation of important content and processes delineated in the manual, viewing video tape of master trainers, observation of master trainers during actual trainings, observation of development centers' video tapes, and within-team discussion and refinement. Refinement of each of the seven Adherence scales occurred in collaboration with the developers and the research team. (The seven segments selected for Adherence ratings are listed in Table 1 and detailed in Table 4).

In addition to the seven Adherence scales, we developed one Trainer Competence measure to assess behaviors consistent with effective training. We assessed competence for each segment using the Trainer Competence measure. This measure consists of five items to assess group training facilitation skills: 1) presentation style/delivery; 2) experiential learning skill; 3) group leadership; 4) development and maintenance of a safe and productive learning environment; and, 5) anticipation of and responses to group questions and challenges. Each of these domains was chosen based on the literature on adult learning and effective group-based trainer behaviors (Caffarella, 2002; Wlodkowski, 2008). Each item was rated on a 4-point scale corresponding to "inadequate skills", "some deficiencies", "capable skills" or "proficient" performance. Behavioral descriptions were developed for each rating. For example, we operationalized "inadequate" group leadership as follows: *Facilitator forcefully or negatively attempts to control group, and/or allows the session to*

become chaotic and off task without "course correction," and/or excessively rushed or significantly belabors the topic(s)." "Some deficiencies" were defined as: *Facilitator is somewhat rigid with structure/agenda/group, and/or allows session to veer off topic several times, "course corrects" too late/too often, pacing is acceptable.* We operationalized "capable skills" on the group leadership item as follows: *Facilitator maintains appropriate control and flexibility with group, and keeps group focused moving forward (may make minor adjustments to redirect group), and pacing is acceptable.* "Proficient" group leadership was defined as, *Facilitator subtly and authoritatively manages group, allows deviation from the agenda to enhance learning (where appropriate) but skillfully reworks agenda to conclude the session at an appropriate pace (not too slow or rushed).*

Once the fidelity measures were developed, we coded all segments of the two development centers for Adherence and for Competence and made refinements to the items and coding manual. In an effort to confirm the validity of the measures, we invited the program developers to independently score segments that reflected combinations of low and high levels of Adherence and Competence (i.e., high adherence, low competence). Developer ratings were compared to those of the primary coding team to further confirm the face validity and reliability of the scoring principles and to have a shared understanding of trainer behavioral ratings.

Trainer Demographics—A survey assessed trainer demographics prior to the 5-day TTT that included gender, race/ethnicity, highest education level, years of (non-ASIST) suicide prevention training experience, and years of experience in social services.

Procedures

Coding manual, training and procedures—We developed a rater coding manual which delineated Adherence scales for each of the seven segments as well as the Trainer Competence measure. The coding rules were initially developed based on materials and observations of the two development centers' training tapes, and were further clarified through coding team consensus meetings and discussion. Exemplars from the development centers were added to the coding manual to further clarify coding rules.

The project coordinator (MM) and first author (WC) trained the coding team. All coders: 1) reviewed the scales, the program manual, the master trainer DVDs, the coding manual, exemplars with the project coordinator and, 2) scored several segments of development center videotapes. Before rating independently, coders had to meet an acceptable level of inter-rater reliability (ICC = .60 on segments) with the team's consensus ratings. Segments for rating were randomly assigned and coders met on an ongoing basis to determine consensus.

Coder drift—Several procedures were put in place to guard against coder drift and to maintain a high level of inter-rater reliability over the course of the study. For example, just over 20% of all segments were randomly assigned for recoding to assess for coder drift over the two-year period of observational coding. Thus, segments that were coded during the first months of the study were surreptitiously inserted into coder assignments during the mid and later coding periods. Re-coded scores were then compared to initial consensus scores. We

found almost no 'drift' over time, with ICCs of .89 for Adherence and .95 for Competence across all scales. In addition, because these measures were developed for the present study, and we wanted to be confident in our results, we double coded nearly half (49.6%) of the total observations. Overall, average fidelity measure ICCs ranged from .80-.98 for Adherence scales and from .71-.92 for Competence scales across segments rated (see Table 1). If inter-rater reliability did not reach or fell below an ICC of 0.6, segments for that scale were team coded and consensus was reached via team discussion. Consensus scoring continued until inter-rater reliability achieved an acceptable level; consensus scores were used in the final data set.

Statistical Analyses—Exploratory factor analysis of the Trainer Competence scale was first conducted using the five items on the scale, followed by a confirmatory factor analysis (CFA), to confirm the assumed underlying construct. The number of factors was chosen based on an eigenvalue-one criterion (the principal component with an eigenvalue greater than 1.00 was retained), the scree test (Cattell, 1966), the proportion of variance accounted for, and interpretability criteria. In addition to the chi-square statistic, CFI, TLI, and RMSEA (Brown, 2006; Byrne, 2010) were also used to verify the construct. Since the clustered data structure was designed with two trainers nested within each center, and each trainer conducted two trainings, linear mixed-effect models (LMM) were used to determine predictors of the observed Adherence and Competence outcome variable. These models examined random and fixed effects of center and trainer, and fixed effects of time, gender, years of (non-ASIST) suicide prevention training experience, and segment. Statistical analyses were performed using SPSS version 19 (IBM, Armonk, NY, USA), SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Mplus 7 (1998-2012 Muthen & Muthen, Los Angeles, CA). All statistical tests were 2-sided; p values of less than .05 were considered to be statistically significant.

Results

Scale psychometrics

Exploratory factor analysis of the five items on the Competence scale resulted in a single factor with an eigenvalue of 1.19, with all other values below 0, indicating that the scale reflects one unified concept. This factor accounted for 54.76% of the common variance. Cronbach's alphas for the competence scale within segment ranged from .61 to .77. Within both EFA and CFA, our hypothesis of one-factor for the competence scale was verified. For CFA, the test of one-factor structure yielded a χ^2 (5) = 2.355, p = 0.798, CFI and TLI values of 1.000 and 1.022, and an RMSEA value of 0.000 (90 % CI= 0.000, 0.051), all of which suggests that our hypothesized one-factor model was sufficiently parsimonious and well-fitting. In contrast to the Competence scale, the items for Adherence measures are causal indicators – that is, they are not expected to be correlated with each other (Smith & McCarthy, 1995). For example, a trainer's adherence to one item during the *Asking about thoughts of Suicide* segment (i.e., asks why a caregiver needs to ask about suicide) is not necessarily predictive of adherent delivery of another item (i.e., discusses the benefits of asking directly). Internal consistency analysis is therefore not appropriate for the Adherence measures and was not conducted.

We found that adherence and competence were unique, but related concepts, as total Adherence and Competence correlated at r = .49, p<.001 indicating a 25% overlap between the two measures.

Trainer demographics

Analyses of the trainers' pre-training demographic variables found trainer Competence and Adherence were not predicted by highest education level, years of experience in social services, or training hours (non-ASIST). Adherence ratings for the first training conducted (i.e., Time 1) did, however, differ by gender (t (32) = 2.129, p <.05), with females showing greater Adherence (mean (SD) = 70.53 (13.130)) than males (mean (SD) = 58.51 (19.460)). Adherence ratings at Time 1 also differed by years of training experience (r=-0.43, p=.011); those with fewer years of training experience showing greater Adherence (range of 0 – 30 hrs; mean=10.21, median=10, SD=8.18.). Because of the these significant findings, in the LMM models for Adherence below, we ran models as described, but also included gender and years of training experience to determine if these variables would affect the relationships with adherence.

Center- and Trainer-level Effects

To examine if crisis center had an effect on trainer Adherence and Competence we ran mixed models with center as a fixed effect and trainer as a random effect (Table 2). For Adherence, there was a significant center effect (F(16,14.2) = 4.25, p = 0.005) and a trend for an effect of years of training experience (F (1,14) = 3.71, p=.075). Nevertheless, trainer effect accounted for 56.6 % of the variability in Adherence above and beyond variance accounted for by center and years of training experience. Therefore, while the culture at centers may have some influence on how adherent trainers were, a significant amount of the variability in adherence came from the trainers themselves and their experience with other types of training. Experience had a negative estimate (-.0034), however, indicating that *more* experience with *previous training* was associated with *less* adherence to the specific program under study (ASIST). For Competence, since center effect was not significant (F(16,16.5) = 1.13, p = 0.406), the 77.76 % of variability in Competence was mainly accounted for by trainer. This indicates that quality of program delivery was not due to the center, previous training or other external variables but mainly due to inherent trainer qualities prior to the TTT.

Overall Trainer-level Effects

To further examine the main effect of trainer, models were first run with trainer as a random effect (and gender and years of training experience as fixed effects when predicting Adherence). Trainer accounted for 81.6% of the variability in Adherence (p <.0001), with significant main effects of gender (F(1,30.1) = 4.31, p=.047) and years of training experience (F(1,30) = 4.40, p=.045). Specifically, females showed greater Adherence than males and, once again, previous suicide prevention training was associated with *less* adherence to the ASIST model. For Trainer Competence, trainer accounted for 78% of the variability (p < .001). When trainer is treated as a fixed effect, significant differences

amongst trainers were found for both Adherence and Competence; $F(33,32) = 12.23$, $p < .0001$ for Adherence, and $F(33,32) = 4.35$, $p < .0001$ for Competence.

As we could not specifically examine which trainers were significantly higher than others due to the low sample size, we categorized trainers by mean levels of Competence and mean levels of Adherence into three categories: low, medium, and high, respectively (Table 3). For Adherence, trainers were grouped in terms of percent of program content delivered as follows: below 60%, 60-75%, and above 75% based on the distribution of scores into comparably-sized groups and interpretation of levels of adherence indicative of unacceptable, minimally acceptable, and sufficient levels of adherence. There is precedence in the literature to consider a program delivered if at least two thirds of the content is presented (Sholomskas, et al., 2005). For Competence, trainers were grouped into three categories based on the range of scores (5-20) and distributions: below 13, 13-15, and above 15. Interpretation of scores as representative of unique levels of performance corresponding to scores in each domain as: predominantly deficient (3 or more scores in the lower ranges), more capable than deficient (3 or more scores in the capable range) and predominantly capable (4 or more scores in capable or skilled range) levels of competence, respectively. Categories were confirmed with statistical analyses using ANOVA which showed that they were significantly different from one another and thus meaningfully distinct groups (i.e., high, medium, low competence). With center as a covariate, these groups were found to have significantly different levels of Adherence ($F (2, 44) = 8.20$, $p = .001$) and Competence ($F (2, 47) = 25.69$, $p < .0001$). Center was close to significant for Adherence ($F (16, 44) = 1.85$, $p =0.055$), but not for Competence ($F (16, 47) = 0.85$, $p =0.628$). Table 3 illustrates the overall distribution of trainer fidelity in terms of adherence to the manual and competence in delivery combined. As the table shows, only twelve percent (12%) of average trainer ratings were both high in adherence and solidly competent. Less than 10% of average trainer ratings fell in the high adherence and low competence cell, and none fell in the low adherence and high competence cell.

Fidelity scores from training Time 1 to Time 2

Each trainer conducted two trainings in the context of the study which provided the opportunity to examine if trainer fidelity improved from Time 1 to Time 2. The average number of days between the two trainings was 45.4 days (sd= 47.92). Using a mixed model, with trainer as a random effect and training time as a fixed effect, there was a significant trainer effect (est= 0.01666, p<0.001), but training time was not significant ($F(1,31.5) = 1.70$, p=0.20) for Adherence. Trainer accounted for 81.8% of the variability in Adherence from Time 1 to Time 2, and both gender ($F(1,30.2) = 4.42$, p=.044) and years of training experience ($F(1,30) = 4.49$, p=.043) had significant main effects where females had higher Adherence, and those with more experience had lower Adherence. For Competence, there was also a significant trainer effect (est=2.61, p<0.001) with training time not significant ($F (1,29.8) = 3.08$, p=0.090). Trainer accounted for 78.7% of the variability in Competence from Time 1 to Time 2 (see Table 2). Thus, fidelity did not improve across two administrations.

Fidelity across program segments

We examined if there are systematic differences in fidelity by segment as measured by Adherence and Competence and found significant differences amongst segments on Adherence with F(6,358) value of 5.99, p < .0001. The highest level of Adherence (84.18%) was for *Process of an Intervention*, a training segment that is lecture-based. The lowest levels of Adherence (62.44%, and 60.17%, respectively) were in *Contracting a Safeplan* and *Asking about Thoughts of Suicide*, two segments with significant active learning and practice elements. When trainer was added as a random effect, the segment differences in Adherence remained significant (p < .001). In addition, there was a significant experience effect (F (1,27.5) = 4.91, p = .035); more previous experience was associated with lower levels of adherence. The trainer random effect, however, dropped to 23.5% of variance explained, indicating that some of the trainer effect on Adherence was explained by these underlying differences in segment-level Adherence.

We also found significant differences amongst segments on Trainer Competence with an F(6,292) value of 4.28, p < .001. The highest Competence score (14.15) was in the *Process of an Intervention* segment, while the lowest scores (12.98 and 12.33, respectively) were in *Contracting a Safeplan* and *Asking about Thoughts of Suicide*. These findings are consistent with the Adherence results. When trainer was added as a random effect, the segment difference in Competence was still significant (p < .001). Table 4 provides descriptions for each of the segments, the types of learning activities in each, the expected and actual timeframe for each and corresponding Adherence and Competence scores. We also indicate the relative rank of each segment for both measures from highest (1) to lowest (7) ranking.

Discussion

Telephone crisis services are a critical part of suicide prevention in the US and globally. Crisis counselors, and other community-based gatekeepers, must be trained to assess and intervene with at-risk individuals. Training large numbers of people in a standardized approach, however, poses immense logistic and resource challenges. Train-the-trainer models (TTT) are considered to be highly efficient. The actual effectiveness of this approach, however, is undermined if the subsequent training is not faithful to the program content or is presented in a way that knowledge and skills are not acquired or implemented by the intended audience. It is critical, therefore, to assess the fidelity of the training provided by newly trained instructors in order to know if the intervention was delivered as intended.

We found wide variability across trainers in terms of adherence to the program content and competence in the delivery of the program. On average, the majority of trainers delivered about two-thirds of the examined program segments, an acceptable level of program delivery (Sholomskas et al., 2005) although in some cases none of the essential content was provided by trainers to staff. We also found that previous crisis or suicide prevention training was associated with lower levels of trainer adherence to the ASIST model. This finding may be explained in terms of a "primacy effect" of previously learned material (Tulving, 2008) that essentially interfered with new learning. In addition, trainers with affiliations with another program may experience motivated resistance to the 'new'

intervention training – a response that has been noted among clinicians required to learn new psychotherapy treatments (Wiltsey-Stirman, Miller, Toder, Calloway, Beck, Evans & Crits-Christoph, 2012). These trainers may have found their previous training to be more appropriate to their setting or of higher quality. They may have been critical of the ASIST program given their experiences with another model. This situation is likely to arise whenever a suicide prevention program is implemented. Program developers and trainers may need to consider previous training as a potential obstacle to new trainers' adherence and consider strategies to ameliorate the impact during TTT. The finding that females were more adherent on average is consistent with findings from a previous TTT study (Cross et al., 2010), although the underlying reasons for a gender difference is not clear. Additional research to examine the factors associated with the gender difference is recommended.

In contrast to the generally high level of adherence to program content, only 18% of the trainers' observations were rated as solidly competent in facilitation and process skills (e.g., engaging participants in simulations), which are central to effective adult learning and training programs. These competencies are not easily inculcated in the relatively brief training provided, especially for a complex intervention such as the one under investigation. It is likely that personal comfort and skill facilitating group activities characterized some trainers prior to TTT who then developed specific competencies as a result of training. Trainer competencies may be particularly challenging to attain in TTT programs. In essence, it may be unreasonable to expect training to be "one size fits all." In fact, our findings are consistent with implementation science models (e.g., Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005) that highlight interventionist selection as a critical component of implementation. Future trainings could select potential trainers based on existing competencies such as group facilitation skills and comfort conducting active learning activities. This would allow the TTT program to focus specifically on learning to deliver program content with fidelity. Future TTT studies could test the impact of trainer selection on training and fidelity outcomes.

Trainers did not improve fidelity from their first to their second training, suggesting that mere practice is not sufficient to improve TTT fidelity. Our findings call into question the common practice of trainer certification based on number of trainings delivered. The training literature, moreover, indicates that improvement in performance is not likely without expert guidance and specific feedback (Beidas & Kendall, 2010; Burns, Peters & Noell, 2008) and may actually deteriorate over time (Cross et al., 2011). Several studies demonstrate the importance of feedback through supervision, consultation, and/or coaching based on observations, as critical for interventionist fidelity in psychotherapy (Hepner et al., 2011). A challenge to trainers across a variety of suicide prevention programs, from community-based gatekeeper to clinical skill training, is to develop strategies to incorporate cost-effective feedback on core competencies to enhance and maintain fidelity over time.

Trainer fidelity varied with types of training. Higher scores on both measures of fidelity were achieved on segments involving lecture material (i.e., *Process of Intervention*) while segments with a high degree of active learning or 'hands on' experience were ranked lowest in terms of observed ratings. In fact, the segments focused on *"Asking about Suicide"* and *"Contracting a Safeplan"* were delivered with the least faithfulness to content adherence

and competence overall. These highly interactive training activities involved role play practice asking about suicide directly and creating a Safeplan. They clearly require additional training support to achieve fidelity and the goals of training. One approach to address the problem of low levels of fidelity for some segments of the program would be to modify the TTT to focus specifically on empirically demonstrated challenges associated with segments. In addition, it seems clear that individualized feedback to trainers about fidelity, either during training or program implementation, would be effective for improving transfer of training. Fidelity measures, such as those used in research, could be modified for clinical consultation purposes either through live observation or video-tape reviews as part of a certification process. In fact, a certification process based on performance rather than program completion may be important for some TTT suicide prevention programs.

The present study has several limitations. While our overall sample of observations was high (N=324), 34 trainers were observed making our effective sample relatively small. Moreover, we selected seven segments of the training (in collaboration with the ASIST developers and trainers) to reflect variability of training activities and core content, and did not code other components of the training program. It is possible that those segments would have yielded different findings. Lack of variability in various demographic and other characteristics of our sample may have undermined our ability to find relationships between these variables and trainer's Adherence and Competence scores. Strong relationships were consistently found with previous training, however, lending support for these findings. Despite these limitations, the use of observational methods and rigorously developed fidelity measures, including examination of psychometric properties, is a major strength of the study.

The relationship between trainer fidelity and counselor outcomes is not addressed in the current study. How faithful trainers must be to a program in order to be a true test of the intervention and to assess the relationship between trainer fidelity and trained counselor outcomes on the crisis line has yet to be established. We found that two segments -- *'Asking about suicide' 'Creating a Safeplan'* --- had the lowest fidelity ratings in the TTT. It is intriguing to note that Gould et al. (2013) found after the ASIST training counselors were not more likely to assess suicide risk (e.g., asking about and exploring plans, prior attempts) or to contract for safety outcomes, two outcomes that are directly linked to the observed and rated training segments. Future research will explore if adherence to specific segments of ASIST's content or competence in the delivery of ASIST is related to counselor behaviors on crisis calls.

In conclusion, there was wide variability in trainer fidelity to the ASIST program following TTT. Few observed trainings showed trainers with high levels of both adherence and competence when delivering the training to staff and they did not show improvements over time. Adherence to the manualized content was more successful than trainer competence as the majority of trainers delivered an average of over two-thirds of the program content. Several suggested enhancements to the TTT model are offered which may contribute to improvements in the transfer of training and fidelity of the ASIST and other similar programs. Despite rapid dissemination, TTT programs need further study to ascertain the cost-effectiveness of the model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Abatemarco A, Beckley J, Borjan M, Robson M. Assessing and improving bioterrorism preparedness among first responders: A pilot study. Journal of Environmental Health. 2007; 69(6):16–22. [PubMed: 17265726]

Allen ES, Connelly EN, Morris CD, Elmer PJ, Zwickey H. A train-the-trainer model for integrating evidence-based medicine into a complementary and alternative medicine training program. Journal of Science and Healing. 2011; 7(2):88–93.

Assemi M, Mutha S, Hudmon KS. Evaluation of a train-the-trainer program for cultural competence. American Journal of Pharmaceutical Education. 2007; 71:110. [PubMed: 19503694]

Beidas RS, Kendall PC. Training clinicians in evidence-based practice: A critical review of studies from a systems-contextual perspective. Clinical Psychology: Science & Practice. 2010; 17:1–30. [PubMed: 20877441]

Becker SM. Psychosocial care for women survivors of the tsunami disaster in india. American Journal of Public Health. 2009; 99(4):654–658. [PubMed: 19150896]

Besculides M, Trebino L, Nelson H. Successful strategies for educating hard-to-reach populations: Lessons learned from the Massachusetts train-the-trainer project using the Helping you Take Care of Yourself curriculum. Health Education Journal. Mar.2011 10.1177/0017896911398982

Bobevski I, Holgate AM. Characteristics of effective telephone counselling skills. British Journal of Guidance & Counselling. 1997; 25(2):239–249.

Bowes D, Marquis M, Young W, Holowaty P, Isaac W. Process evaluation of a school-based intervention to increase physical activity and reduce bullying. Health Promotion Practice. 2009; 10(3):394–401. [PubMed: 18349146]

Brown, TA. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.

Burns MK, Peters R, Noell GH. Using performance feedback to enhance implementation fidelity of the problem-solving team process. Journal of School Psychology. 2008; 46(5):537–550. [PubMed: 19083371]

Byrne, BM. Structural Equation Modeling With Mplus. New York: Routledge: Taylor & Francis Group; 2010.

Byrne MK, Willis A, Deane FP, Hawkins B, Quinn R. Training inpatient mental health staff how to enhance patient engagement with medications: Medication alliance training and dissemination outcomes in a large US mental health hospital. Journal of Evaluation in Clinical Practice. 2010; 16(1):114–120. [PubMed: 20367823]

Caffarella, RS. Planning programs for adult learners: A practical guide for educators, trainers and staff developers. 2nd. San Francisco: Jossey-Bass; 2002.

Cattell RB. The scree test for the number of factors. Multivariate behavior Research. 1966; 1:245–276.

Conner KR, Gunzler D, Tang W, Tu XM, Maisto SA. Test of a clinical model of drinking and suicidal risk. Alcoholism: Clinical and Experimental Research. 2011; 35:60–65.

Carroll KM, Nich C, Sifry RL, Nuro KF, Frankforter TL, Ball SA, Fenton L, Rounsaville BJ. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. Drug and Alcohol Dependence. 2000; 57:225–238. http://dx.doi.org/10.1016/S0376-8716(99)00049-6. [PubMed: 10661673]

Coveney CM, Pollock K, Armstrong S, Moore J. Callers' experiences of contacting a national suicide prevention helpline: Report of an online survey. Crisis, The Journal of Crisis Intervention and Suicide Prevention. 201210.1027/0227-5910/a000151

Covington, D.; Hogan, M.; Abreu, J.; Berman, A.; Breux, P. Suicide Care in Systems Framework. National Action Alliance: Clinical Care & Intervention Task Force; 2011. Retrieved January 9, 2013, from http://actionallianceforsuicideprevention.org/sites/actionallianceforsuicideprevention.org/files/taskforces/ClinicalCareInterventionReport.pdf

Cross W, Cerulli C, Richards H, He H, Hermann J. Predicting dissemination of a disaster mental health "train-the-trainer" program. Disaster Medicine and Public Health Preparedness. 2010; 4(4): 339–343. [PubMed: 21149237]

Cross W, West J. Examining implementer fidelity: Conceptualizing and measuring adherence and competence. Journal of Children's Services. 2011; 6(1):18–33.

Cross W, Seaburn D, Gibbs D, Schmeelk-Cone K, White AM, Caine ED. Does practice make perfect? A randomized control trail of behavioral practice on suicide prevention gatekeeper skills. Journal of Primary Prevention. 2011; 32:195–211.10.1007/s10935-011-0250-z [PubMed: 21814869]

Dobson D, Cook TJ. Avoiding type III error in program evaluation: Results from a field experiment. Evaluation and Program Planning. 1980; 3:269–276.

Fixsen, DL.; Naoom, SF.; Blase, KA.; Freidman, RM.; Wallace, F. Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network; 2005.

Forgatch MS, DeGarmo DS. Sustaining fidelity following the nationwide PMTO implementation in Norway. Prevention Science. 2011; 12(3):235–246. [PubMed: 21671090]

Gelkopf M, Ryan P, Cotton SJ, Berger R. The impact of "training the trainers" course for helping tsunami-survivor children on Sri Lankan disaster volunteer workers. International Journal of Stress Management. 2008; 15(2):117–135.

Gould MS, Kalafat J, Munfakh JL, Kleinman M. An evaluation of crisis hotline outcomes, Part II: Suicidal Callers. Suicide and Life Threatening Behavior. 2007; 37(3):338–352. [PubMed: 17579545]

Gould MS, Cross W, Pisani AR, Munfakh JL, Kleinman M. Impact of Applied Suicide Intervention Skills Training (ASIST) on national suicide prevention Lifeline counselor interventions and suicidal caller outcomes. Suicide and Life Threatening Behaviors. 2013 Article first published online: 25 JUL 2013. 10.1111/sltb.12049

Hansen WB, Graham JW, Wolkenstein BH, Rohrbach LA. Program integrity as a moderator of prevention program effectiveness: results for fifth grade students in the adolescent alcohol prevention trial. Journal of Studies on Alcohol. 1991; 52(6):568–579. [PubMed: 1758184]

Hepner KA, Hunter SB, Paddock SM, Zhou AJ, Watkins KE. Training addiction counselors to implement CBT for depression. Admin Policy Mental Health. 2011 Online May 31, 2011.

Hogue A, Henderson CE, Dauber S, Barajas PC, Fried A, Liddle HA. Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. Journal of Consulting & Clinical Psychology. 2008; 76:544–555.10.1037/0022-006X.76.4.544 [PubMed: 18665684]

Jaccard, JC. LISREL approaches to interaction effects in multiple regression. Thousand Oaks, CA: Sage; 1996.

Kalafat J, Gould MS, Munfakh JLH, Kleinman M. An evaluation of crisis hotline outcomes, Part I: Non-Suicidal Crisis Callers. Suicide and Life Threatening Behavior. 2007; 37(3):322–337. [PubMed: 17579544]

King R, Nurcombe R, Bickman L, Hides L, Reid W. Telephone counseling for adolescent suicide prevention: Changes in suicidality and mental state from beginning to end of a counseling session. Suicide and Life Threatening Behavior. 2003; 33(4):400–411. [PubMed: 14695055]

Kline, RB. Principles and practice of structural equation modeling. 2nd. New York: Guilford; 2004.

Knoche LL, Sheridan SM, Edwards CP, Osborn AQ. Implementation of a relationship-based school readiness intervention: A multidimensional approach to fidelity measurement for early childhood. Early Childhood Research Quarterly. 2010; 25(3):299–313. [PubMed: 20824112]

Lang, WA.; Ramsay, KF.; Tanney, BC.; Kinzel, T. Applied Suicide Interventions Skills Training: Trainer Manual. Calgary, Alberta, CA: Living Works Education; 2008.

Lillehoj CJ, Griffin KW, Spoth R. Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. Health Education & Behavior. 2004; 31:242–257.10.1177/1090198103260514 [PubMed: 15090124]

LivingWorks. 2010. Retrieved January 9, 2013, from http://www.livingworks.net/

McHugh RK, Barlow DH. Dissemination and implementation of evidence-based psychological interventions: A review of current efforts. American Psychologist. 2010; 65(2):73–84. [PubMed: 20141263]

Moore, JE.; Beck, TC.; Sylvertsen, A.; Domitrovich, C. Paper presented to the 17th Annual Meeting of the Society for Prevention Research. Washington, DC, USA: 2009. Making sense of implementation: multiple dimensions, multiple sources, multiple methods.

Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: Issues and research recommendations. Clinical Psychology: Science and Practice. 2005; 12(4):365–383.

Pisani AR, Cross WF, Watts A, Conner K. Evaluation of the Commitment to Living (CTL) curriculum: A 3-hour training for mental health professionals to address suicide risk. Crisis: The Journal of Crisis Intervention and Suicide Prevention. 2012; 33(1):30–38.

SAS. SAS Version 9.2. Cary, NC: SAS Institute Inc; 2008.

Schoenwald SK, Garland AF, Chapman JE, Frazier SL, Sheidow AJ, Southam-Gerow MA. Toward effective and efficient measurement of implementation fidelity. Administration and Policy in Mental Health and Mental Health Services Research. 2011; 38(1):32–43.10.1007/s10488-010-0321-0 [PubMed: 20957425]

Sholomskas DE, Syracuse-Siewert G, Rounsaville BJ, Ball SA, Nuro KF, Carroll KM. We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. Journal of Consulting and Clinical Psychology. 2005; 73:106–115. [PubMed: 15709837]

Smith GT, McCarthy DM. Methodological considerations in the refinement of clinical assessment instruments. Psychological Assessment. 1995; 7(3):300–308.

SPSS. SPSS Version 16.0 [Computer software]. Chicago, IL: SPSS Inc; 2007.

Trepka C, Rees A, Shapiro DA, Hardy GE, Barkham M. Therapist competence and outcome of cognitive therapy for depression. Cognitive Therapy and Research. 2004; 28(2):143–157.

Tulving, E. On the law of primary. In: Gluck, MA.; Anderson, JR.; Kosslyn, SM., editors. Memory and mind: A festschrift for Gordon H Bower. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, US; 2008. p. 31-48.

Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. Journal of Consulting & Clinical Psychology. 1993; 61(4):620–630. [PubMed: 8370857]

Welber M. Save by growing your own trainers. Workforce. 2002; 81:44–48.

Wiltsey-Stirman S, Miller CJ, Toder K, Calloway A, Beck AT, Evans AC, Crits-Christoph P. Perspectives on cognitive therapy training within community mental health settings: Implications for clinician satisfaction and skill development. Depression Research and Treatment. 2012:11. Article ID 391084. 10.1155/2012/391084

Wlodkowski, RJ. Enhancing adult motivation to learn: A comprehensive guide for teaching all adults. San Francisco, CA: Wiley & Sons; 2008.

World Health Organization. Mental Health: Suicide Prevention (SUPRE). World Health Organization; 2012. Retrieved January 3, 2013

Table 1

Inter-rater reliability of Adherence and Competence scales for coded segments

Training program segments	Inter-rater Reliability[*] Adherence Measures	Inter-rater Reliability[*] Competence Measures
Exploring Invitations	.85	.92
Asking about Suicide	.98	.87
Listening to Reasons	.93	.85
Contracting a Safeplan	.94	.89
Process of Intervention	.80	.83
Ambivalence Simulation	.80	.71
Bridge Simulation	.95	.81

[*]
Average Measures' Intra-Class Correlations

Table 2

Competence and Adherence Based on Linear Mixed Effects Model

	Competence	Adherence
Trainers nested within Centers		
Fixed part	F (DF= num, den)	F (DF= num, den)
Center	1.13 (16, 16.5)	4.25 (16, 14.2)[*]
Random part	Coeff. (s.e)	Coeff. (s.e)
σ^2_u (Trainer)	2.53 (1.02)[*]	0.005 (0.003)[*]
σ^2_e (Residual)	0.72 (0.18)[*]	0.004 (0.0009)[*]
Time Nested within Trainer		
Fixed part	Coeff. (s.e)	Coeff. (s.e)
Intercept	11.84 (0.63)	0.81 (0.05)
Training time	-0.88 (0.61)	-0.02 (0.02)[*]
Random part		
σ^2_u (Trainer)	2.50 (0.73)[*]	0.017 (0.005)[*]
σ^2_e (Residual)	0.65 (0.17)	0.004 (0.0009)

[*] p value <.05;

Adherence models include gender and years training experience.

Table 3

Distribution of trainer Adherence and Competence scores by category

Average Trainer Competence	Average Trainer Adherence			
	Low (<60%)	Medium (60-75%)	High (above 75%)	Totals
More "deficient" than "capable" scores (<13)	17.6%	14.7%	8.8%	41.1%
More "capable" than "deficient" scores (13-14.8)	11.8%	14.7%	14.7%	41.2%
Solidly "capable" scores (15)	0%	5.9%	11.8%	17.7%
Totals	29.4%	35.3%	35.3%	100%

<13= more ratings of "deficient" or "inadequate" than "capable"

13-14.8= more ratings of "capable" than "deficient"

15+= all ratings at least "capable"

[*] Categories devised based on team discussion and confirmed with statistical analyses (all categories are significantly different from each other)

Table 4

Segment description and descriptive statistics

Program Segment: Content and learning activities	# of Adherence Items	Manual prescribed Duration/Avg. actual duration (SD)	% Adherence Mean (SD), Rank & Range, Median (Md)	Avg. Competence Mean (SD), Rank & Range, Median (Md)
Process of Intervention: The "process" elements of an intervention (i.e. that process is fluid and involves movement back and forth between phases of SIM). Remaining "in sync" with the person at risk/what to do if one gets out of sync. All didactic	9	40 min./26.3 (10.90)	84.18 (13.33) Rank: 1 Range: 56-100 Md: 88.89	14.15 (1.81) Rank:1 Range:7-18 Md: 14.0
Exploring Invitations: The first phase of the Suicide Intervention Model (SIM) is "Connecting" and the caregiver's task is to "explore invitations" that a person at risk may be expressing. The caregiver needs to explore the meaning of an event to a person at risk (i.e. may signify loss). Caregivers need to focus on the individual when conducting an intervention and to proceed as if anyone could be at-risk. Full segment includes a role play (not coded). Primarily didactic; includes an active learning component ("newsprint" activity) not coded.	6	16 min./27.0 (6.56)	75.42 (24.45) Rank: 2 Range: 17-100 Md: 83.33	13.48 (2.11) Rank: 4 Range: 9-20 Md: 14.0
Listening to Reasons: The next phase of SIM is "Understanding", the concern is whether or not someone will understand their "reasons" for and against dying and the caregiver task is to "listen to/ listen for" these reasons. Caregivers need to spend sufficient time listening to a person at risk's reasons for dying and that doing so can help uncover potential reasons for living. Includes didactic and some active learning; Participants practice listening to/for the reasons for dying and living.	8	20 min./25.2 (8.09)	75.31 (20.57) Rank: 3 Range: 25-100 Md: 75.0	13.7 (2.03) Rank:3 Range:8-19 Md: 14.0
Ambivalence Simulation: Participants practice reflecting messages that may be conveyed by a person at risk with a focus on identifying messages of death, life, and ambivalence in various statements. Almost entirely simulation.	9	15 min./18.4 (4.51)	72.22 (24.86) Rank: 4 Range: 0-100 Md: 66.67	13.76 (2.14) Rank: 2 Range: 9-19 Md: 14.0
Bridge Simulation: Participants engage in an in-depth, realistic simulation to practice moving through the SIM, addressing each of the person-at-risk's concerns and caregiver tasks. Systematically debrief each participant and the group about the experience. Almost entirely simulation.	14	40 min./35.6 (8.32)	67.29 (24.11) Rank: 5 Range: 8-100 Md: 75.0	13.39 (1.97) Rank: 5 Range: 8-17 Md: 14.0
Contracting a Safeplan: The last phase of SIM is "Assisting" and that the next person at risk concern is a "Safeplan" and the caregiver task is to "contract" a plan for safety. The importance of establishing a Safeplan that specifically addresses each element of risk identified in the previous section of the intervention. Participants practice addressing each element of a Safeplan and gaining a person at risk's consent to each element of a plan. Includes didactic and active learning activities; participants practice addressing each element of a Safeplan.	15	40 min./34.8 (10.64)	62.44 (19.94) Rank: 6 Range: 27-100 Md: 62.14	12.98 (2.21) Rank: 6 Range:8-18 Md: 13.0
Asking about Suicide: Asking a person at risk about suicide is vitally important and that it must be done directly, clearly, and non-judgmentally. Includes didactic and active learning; participants practice asking directly about suicide.	9	10 min./9.0 (3.97)	60.17 (26.9) Rank: 7 Range: 0-100 Md: 63.33	12.33 (2.37) Rank: 7 Range: 7-18 Md: 12.5