

Development and validation of the ASPIRE-VA coaching fidelity checklist (ACFC): a tool to help ensure delivery of high-quality weight management interventions

Laura J. Damschroder, MS, MPH,¹ David E. Goodrich, EdD,¹ Hyungjin Myra Kim, Sc.D,^{1,2} Robert Holleman, MPH,¹ Leah Gillon, MSW,¹ Susan Kirsh, MD, MPH,^{3,4} Caroline R. Richardson, MD,⁵ Lesley D. Lutes, PhD⁶

¹VA Ann Arbor Center for Clinical Management Research, VA Ann Arbor Healthcare System, 2215 Fuller Rd., Mailstop 152, Ann Arbor, MI 48105, USA

²University of Michigan Center for Statistical Consulting & Research, 3550 Rackham CSCAR, 915 E. Washington, Ann Arbor, MI 48109-1070, USA

³Louis Stokes Veterans Affairs Medical Center, 10701 East Blvd., Cleveland, OH 44106, USA

⁴Clinical Consultant Office of Primary Care, VA Central Office, Washington, DC, USA

⁵Department of Family Medicine, University of Michigan School of Medicine, 1018 Fuller St., Ann Arbor, MI 48109-0708, USA

⁶Department of Psychology, East Carolina University, 344-B Rawl, East Fifth St., Greenville, NC 27858-4353, USA

Correspondence to: L Damschroder laura.damschroder@va.gov

Cite this as: *TBM* 2016;6:369–385
doi: 10.1007/s13142-015-0336-x

Abstract

Practical and valid instruments are needed to assess fidelity of coaching for weight loss. The purpose of this study was to develop and validate the ASPIRE Coaching Fidelity Checklist (ACFC). Classical test theory guided ACFC development. Principal component analyses were used to determine item groupings. Psychometric properties, internal consistency, and inter-rater reliability were evaluated for each subscale. Criterion validity was tested by predicting weight loss as a function of coaching fidelity. The final 19-item ACFC consists of two domains (session process and session structure) and five subscales (sets goals and monitor progress, assess and personalize self-regulatory content, manages the session, creates a supportive and empathetic climate, and stays on track). Four of five subscales showed high internal consistency (Cronbach alphas > 0.70) for group-based coaching; only two of five subscales had high internal reliability for phone-based coaching. All five subscales were positively and significantly associated with weight loss for group- but not for phone-based coaching. The ACFC is a reliable and valid instrument that can be used to assess fidelity and guide skill-building for weight management interventionists.

Keywords

Treatment fidelity, Weight management, Process evaluation, Obesity, Intervention

INTRODUCTION

Complex behavioral interventions implemented with higher fidelity are associated with better clinical outcomes [1, 2]. While several obesity interventions have proven to be efficacious in ideal study settings, the effects are often weakened when implemented in clinical settings [3], at least in part because these interventions are delivered with low fidelity. Treatment fidelity is the degree to which an intervention was delivered as designed and intended [4] and measures of fidelity can be used to monitor and enhance reliability and validity of behavioral interventions [5]. Assessing treatment fidelity establishes internal validity of a research study

Implications

Practice: The ACFC is freely available for use by practitioners who design and deliver weight management interventions to patients who may find this tool useful for rating the quality and delivery of coaching in a clinical setting.

Policy: Fidelity checklists like the ACFC need to be used in clinical practices to help ensure transparent and high-quality delivery of weight management interventions.

Research: Validation of the ACFC in diverse settings has the promise of improving the translation of weight management interventions.

and helps to establish whether outcomes are caused by the intervention or by other unmeasured factors [5–9].

Fidelity assessment is essential for moving toward understanding theoretical mechanisms, and more pragmatically, identifying which components of an intervention contribute to outcomes [10, 11]. This knowledge reveals the necessary core components of complex interventions that contribute to expected benefits and thus provides guidance for adaptability of the intervention into diverse settings [12, 13, 8]. For example, one meta-analysis of components associated with effectiveness of parent training interventions found that of 18 intervention components tested across 77 published studies, 5 components were consistently associated with larger effects, 3 components were associated with smaller effects, and the remainder had inconsistent effects [14]. This information can be used to help guide adaptation into clinical settings, for example, by placing priority on implementing the components associated with the largest effects. However, these findings are tenuous because of the lack of indication about the degree to which these components were actually delivered (i.e., lack of fidelity assessment); the authors of this review highlight the

need for more consistent assessment of delivery fidelity [14].

Despite the clear need, exceedingly few studies explicitly report assessment of treatment delivery fidelity [5, 6, 9]. One review found only 30 % of 287 published studies over a 10-year span included a mechanism by which to assess treatment delivery and only 6 % assessed the presence of non-treatment specific effects (e.g., empathy) in delivery [9].

There are few validated instruments with which to measure and report fidelity in the delivery of behavioral interventions [9]. Furthermore, despite hundreds of published studies on behavioral treatments for obesity [15], a literature search found no validated fidelity measures for weight management treatment delivery. Therefore, the purpose of the present study was twofold. The first objective was to develop the Aspiring to Lifelong Health (ASPIRE) Coaching Fidelity Checklist (ACFC) to measure and report fidelity in the delivery of a weight management intervention. ASPIRE is an evidence-based treatment weight management program using a small changes approach that has been shown to be effective in several previous studies across 3, 6, 9, and 12 months [16–18]. The measure for the present study was developed and tested based on the largest and most recent study completed to date with 12-month outcomes. The second objective was to assess the reliability and validity of the ACFC. While there are many facets to assessing fidelity [5, 6, 19–21], the focus of this study was on developing an assessment instrument specifically focused on the delivery of a behavioral intervention for weight management.

SETTING

A randomized controlled trial of the ASPIRE 12-month weight management intervention (ASPIRE-VA) was conducted in two Veterans Affairs (VA) medical centers (VAMC). While more detailed information on the overall study design, rationale, and outcomes are already published, a brief description of the ASPIRE-VA intervention sessions is provided here [18, 22]. Two delivery modes (phone and group) were compared to the usual care MOVE!® Weight Management Program [23] at the two study sites. Unfortunately, resource constraints and logistical challenges made it impossible to assess fidelity of MOVE! treatment delivery. Thus, the focus of this study was on assessing delivery of ASPIRE-VA over the phone and via in-person groups.

The ASPIRE-VA trial recruited 481 patients [18, 22] who were referred to MOVE! to participate in the study. Eligibility criteria were broad [18] and included Veterans with a body mass index (BMI) ≥ 30 kg/m² or a BMI between 25 and 30 kg/m² with at least one obesity-related chronic health condition. Enrollment occurred between January 2010 and October 2011 at two VAMCs (one suburban and one urban) located in the Midwestern US. Participants were randomized to one of three programs: (1) individual treatment by phone (ASPIRE-Phone; $n=162$); (2) in-person group

treatment (ASPIRE-Group; $n=160$); or (3) usual care VA weight treatment (MOVE!; $n=159$). ASPIRE-VA was a two-phased intervention that focused on working with each patient to set goals relative to current behaviors. Following a baseline recording week to help them self-evaluate their behavior, they were encouraged to self-select small, personalized, and attainable dietary and physical activity goals that would induce a daily caloric deficit (e.g., 200–300 kcal/day) sufficient to produce slow, steady weight loss [24, 25]. Participants were predominantly older (55 years on average) males who were racially diverse (41 % African American), with health-related disability (52 %), low socio-economic status (43 % with annual income $< \$20,000$), and high levels of disability (55 %) [18, 22]. Participants in both ASPIRE programs lost a statistically significant amount of weight, but participants in ASPIRE-Group lost nearly twice as much as participants in ASPIRE-Phone at 3 months (-2.2 kg, 95 % CI= $-3.1, -1.2$, in ASPIRE-Group vs -1.4 kg, 95 % CI= $-2.3, -0.5$ in ASPIRE-Phone) and 12 months after baseline (-2.8 kg, 95 % CI= $-3.8, -1.9$, in ASPIRE-Group vs -1.4 kg, 95 % CI= $-2.4, -0.5$ in ASPIRE-Phone)[18].

The intensive 3-month phase of ASPIRE-Phone and ASPIRE-Group consisted of 12 weekly sessions following an orientation session. The follow-up phase of the programs consisted of 13 bi-weekly sessions for 6 months followed by 3 monthly sessions for a total of 28 sessions over 12 months. Non-clinician coaches delivered treatment and participants generally had the same coach for the duration of the program. Planned session duration was 20–30 min for ASPIRE-Phone (which was delivered individually) and 60 min for ASPIRE-Group.

ASPIRE-VA was a manualized program whose design aligned with an underlying theoretical framework that described hypothesized mechanisms for behavioral change that are essential for weight loss (Fig. 1). Because this was a pragmatic trial with a goal of designing a low-cost intervention that could be scaled up and implemented in the future, non-clinician lifestyle coaches were hired to deliver the program to evaluate whether less expensive, non-credentialed coaches could deliver ASPIRE-VA effectively. None of the coaches had prior experience or training in delivery of behavioral interventions; three of the coaches had bachelor's degrees in psychology or public health and the fourth coach had a master's degree in public policy. ASPIRE-VA coaches were to provide support to create a reinforcing feedback loop between: self-monitoring of food intake using a stoplight monitoring system for daily food intake and pedometers to monitor daily step counts and other topic-specific behaviors; setting small goals that were personalized relative to baseline behaviors; and maintaining cumulative changes. As shown in Fig. 1, sessions were structured to include three main coaching components: (1) checking in to review self-monitoring and goal attainment since the last session; (2) sharing psychoeducational information in a patient-centered way that built self-

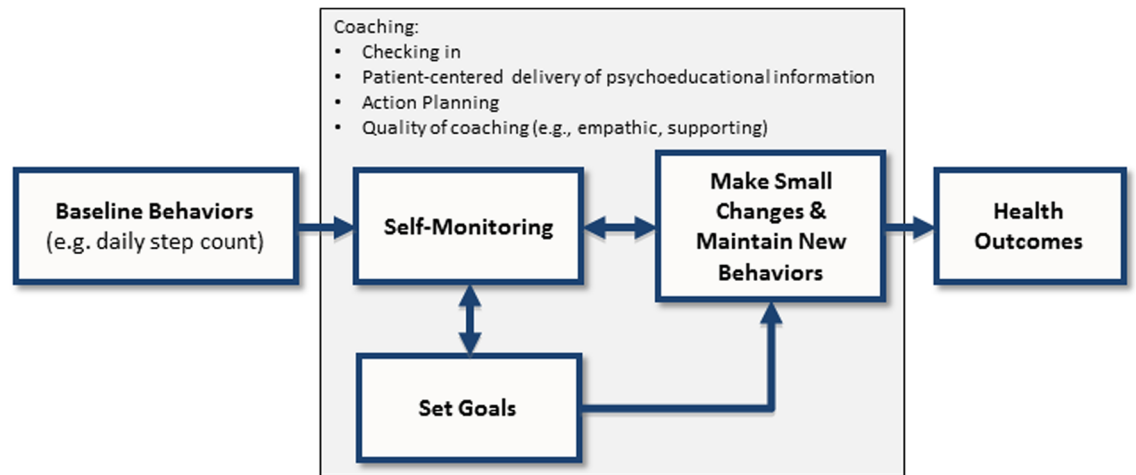


Fig 1 | Theoretical framework for ASPIRE weight management coaching

regulation skills for making and sustaining behavioral changes; and (3) action planning which included setting new goals and using problem-solving techniques to address barriers to achieving goals. A fidelity instrument was needed to assess the extent to which ASPIRE-VA was delivered as designed. Consonant with recognized dual functions of fidelity assessment, fidelity ratings were used to provide feedback [21] to coach supervisors to use in improving coaching skills [6]. Secondly, this study sought to assess the validity and reliability of the ACFC.

METHODS

A valid and reliable measure of fidelity was needed to establish causality of ASPIRE coaching with weight outcomes. Therefore, classical test theory was used to guide development of the ACFC, which assumes that there is a true score that can characterize delivery of ASPIRE-VA, while acknowledging that its measurement will contain error. This theoretical approach guided our analyses in assessing reliability of ACFC fidelity scores. Ratings based on recordings of coaching sessions and outcomes from the two ASPIRE programs were used to develop and assess the ACFC. Methods for developing the ACFC were chosen to create an effective (scientifically validated) and efficient (feasible and useful in routine care) fidelity instrument [26]. This study was approved by the Institutional Review Boards at both of the study sites. Patient consent to be audio-taped during coaching sessions was obtained as part of the informed consent process at the baseline assessment and affirmed by verbal assent for each recorded session.

Purpose of the ASPIRE-VA coaching fidelity checklist (ACFC)

The ACFC was designed to assess the degree to which the interventionists adhered to treatment protocol when conducting sessions with patients in the ASPIRE-VA weight loss program. Fidelity scores were also used to help establish the relationship between

treatment components and outcomes, a first step toward identifying the “active ingredients” of the ASPIRE-VA coaching approach [27]. This will help determine which components can be adapted to other healthcare settings without compromising treatment effectiveness [12].

Identifying fidelity criteria

The three main components of coaching sessions (see first three bulleted items in Fig. 1) were used to guide selection of fidelity criteria. Additional criteria (the fourth bulleted item in Fig. 1) were selected to reflect the expectation that coaches use patient-centered communication techniques throughout each session, including complex reflections to convey empathy, understanding, and support, and open-ended questions to support patient autonomy in a way that encourages and sustains lifestyle changes [28–30].

Assessment items were developed that applied to all patient sessions (not topic-specific) within both phone- or group-based modes of delivery. An initial set of 26 items was developed. Eleven items related to session structure, as described by the three main components above, were developed by the coach supervisors. The goal with these items was to assess whether the coaches covered the actions prescribed by the intervention; for example, “prompts review of goal attainment (i.e., review of goals versus actual).” An additional 15 items were developed to assess how well the coach delivered each aspect of the session. These process items were drawn from a non-validated but published checklist with which a team member had experience in other behavioral intervention trials [31–36]. These items are common to many patient-centered interventions; for example, the health coach “came prepared and organized” and “responded empathically to individual or group member behavior.” Two experienced interventionists (not study coaches) independently rated each of the 26 items using session recordings. After raters scored each session, they met to discuss differences in ratings and discussed disagreements to come

to consensus on a final rating. The raters dropped items if they continued to cause confusion or disagreement about how to apply them, did not uniformly apply to every session with every patient, or overlapped with another item. In addition, preliminary principal component analyses early in the trial led to dropping items with poor reliability. Twenty items were retained and included in subsequent principal component analyses.

For each item, full fidelity was indicated when an interventionist “fully covered/demonstrated” all checklist items. Fidelity assessments were used to identify gaps in delivery that were translated into topics for continuing coach training and skill-building.

Fidelity measurement process

The fidelity checklist was designed for use by expert raters. An observational coding approach was used, based on audio-recorded sessions. This approach is a gold standard for providing objective and highly specific measures [26].

Sample size calculations, based on an assumed proportion of 80 % of sessions being completed, specified 288 observations [37–40], about 5 % of total planned sessions. Recordings were made proportional to the number of group versus phone sessions (a ratio of three phone recordings to one group recording) and proportional to the number of participants expected to be recruited at each of the two study sites. All fidelity assessments were performed by DEG, who is an experienced rater with an educational psychology doctorate degree. He was not involved with treatment supervision and therefore was a true independent rater of treatment fidelity.

Coaches were asked to record phone and group sessions based on a randomized schedule assigned by study investigators. The protocol was adapted to accommodate missed sessions. Coaches were asked to record the next available comparable session.

Scoring treatment fidelity

Each fidelity item in the checklist was rated using a three-point ordinal scale (0=did not cover/demonstrate; 1=partially covered/demonstrated; 2=fully covered/demonstrated). The use of ACFC to establish causality of treatment with weight outcomes demanded valid and reliable measures. Therefore, trained raters assessed fidelity by listening to audio recordings of treatment sessions and ensured there were no missing ratings.

Establishing inter-rater reliability

Inter-rater reliability was established with a second expert rater (LG), who independently assessed 32 % of recorded sessions. The use of two raters highlighted the need for clearly defined criteria by which to score each ACFC item. A manual to help guide [41, 42] assignment of ratings is provided in Appendix 3. This guide helped to ensure high reliability between raters. Comparing Kappa scores across items can be

misleading when there are especially high or low rates of specific values for ratings (e.g., a high majority of ratings equal to 2 (fully covered)) [43]. Thus, for each ACFC item, inter-rater agreement was assessed as raw percent agreement rather than using a kappa score. Percent agreement was based on whether raters agreed that the coach did not cover the item (ratings of 0) versus partially or completely covered the item (ratings of 1 or 2). Pearson correlation coefficients were also computed for each ACFC scale and sub-scale based on the full 0–2 rating scale for each item.

Development of ACFC subscales

The 20 ACFC checklist items were initially organized within two logical domains of fidelity based on adherence to: (1) session structure and (2) session process. Within each domain, principal component analyses (PCA) were conducted for each item using varimax rotation to guide clustering of the items into appropriate underlying constructs. The analyses were done using pooled fidelity data of phone- and group-based sessions for the full 12 months of the program. Summary ACFC domain and subscale scores were computed as sums of item ratings (0–2) within each domain and subscale. Item-rest correlations were evaluated to ensure that each item measured different facets of the subscale construct, but items with item-rest correlations below 0.40 were considered for deletion. Varying numbers of items within each subscale made it difficult to compare fidelity meaningfully across subscales. Therefore, standardized subscale scores were computed so that values ranged from 0 (worst) to 100 (best) by dividing the raw score by the maximum range of values for each subscale.

Establishing ACFC internal consistency and criterion validity

Internal consistency was established using Cronbach's alpha. Criterion validity was assessed by testing the association between ACFC subscale scores and weight loss outcomes for the 3-month intensive phase of treatment. It was hypothesized that higher fidelity by the interventionists would be associated with more weight loss. Analyses were done separately for the two modes of delivery, group and phone. There were four coaches, each of whom coached participants in the phone- and group-based programs. Average ACFC scores were computed for each coach, for each delivery mode. Therefore, each coach was assigned two sets of ACFC scores: one for phone- and one for group-based coaching. Averages were based on ACFC ratings during the first 3 months (the most intensive coaching phase) of the program. Separate linear mixed-effects models for each ACFC subscale score were fit with participants' 3-month weight loss as the dependent variable, with coaches as random intercepts to account for potential within-coach correlation, and ACFC subscale scores as the primary independent variables. The “unadjusted” estimates were obtained adjusting only for baseline weight. Although this is not the same as percent change, it does evaluate the

change (loss) controlling for baseline weight, and is the typical primary response variable in many research studies focused on weight loss, including the ASPIRE-VA trial [18]. The adjusted estimates were obtained after adjusting for other baseline covariates: age, depression diagnosis, substance use diagnosis, Charlson index, and BMI. Significant parameter estimates for ACFC subscales in the hypothesized direction indicate support for criterion validity for each of the ACFC subscales. All analyses were conducted with STATA 13.1.[44]

RESULTS

The 322 participants in the two ASPIRE-VA programs completed 2,334 total sessions in their first 3 months and 2,183 sessions in the following 9 months for a total of 4,517 sessions over the 12-month program. Overall, 185 (4.1 %) sessions were rated between January 2010 and November 2012. Recordings were made of 134 (5.4 %) ASPIRE-Phone sessions, which involved 76 (57 %) unique participants. Of those participants, 15 (9 %) were recorded three or more times. Recordings were made of 51 (6.2 %) ASPIRE-Group sessions, which involved 154 (84 %) unique participants, and of those participants, 79 (51 %) were recorded three or more times. All ASPIRE-Phone participants and all but one group (in deference to multiple members suffering from post-traumatic stress disorder) assented to be recorded. All recorded sessions were rated using the ACFC.

Reliability

Table 1 shows results of an orthogonal solution from the factor analysis based on a varimax (orthogonal) rotation of 10 items within each of the two logical domains of session structure and session process. The Kaplan–Meyer Olkin measure indicated that our sample size was adequate for the factor analyses within each domain (scores > 0.80). The ACFC checklist is provided in the Appendix 1.

The session structure domain has two subscales. The first subscale comprises five items, which are logically related to goal-setting and monitoring progress toward achieving set goals, and is labeled “Set Goals and Monitor Progress.” The second subscale also comprises five items, which are related to delivering core psycho-educational content, and is labeled “Assess and Personalize Self-regulatory Content.” No items were dropped based on the factor loading or item-total correlation.

The session process domain has three subscales. The first subscale comprises five items, which are logically related to session organization and delivery style, and is labelled “Manage the Session.” The second subscale comprises two items, which are related to empathic and non-judgmental interactions, and is labeled “Creates a Supportive and Empathetic Climate.” The third subscale comprises two items, which are related to keeping the session on

track, and is labeled “Stays on Track.” One item, labeled “Health Coach avoided delving too deeply into psychological issues,” was dropped because of its low item-rest correlation (0.29).

Inter-rater agreement was reasonably high based on percent agreement and Pearson correlations for domain and subscale scores (Table 2). Internal consistency for domains and subscales, reflected by Chronbach alpha, were all 0.75 or higher, indicating high internal consistency of the items within each subscale (Table 3). The only exception was “Stays on Track” within the session process domain, which has an alpha of 0.66. This lower reliability appears to be driven by the exceedingly low reliability for phone delivery (0.19).

Fidelity summary

Standardized scores (Table 3) were lower for ASPIRE-Group delivery compared to ASPIRE-Phone delivery for both session structure ($p < 0.001$, t test for unequal variance) and session process ($p < 0.001$). ASPIRE-Phone sessions had less variation in ratings (reflected by lower standard deviations) for all items, except “Delivers didactic material in a matter of fact and friendly way.” Median scores were at the maximum value for all three subscales in the session process domain and for one of the two subscales in the session structure domain for ASPIRE-Phone. As shown in Table 3, standardized mean scores for ASPIRE-Phone were generally higher than 80 for both the session process and session structure domains, indicating that coaches partially or fully met the objective of each fidelity item. The only exception was “Assess and Personalize Self-regulatory Content.” Coaches received low ratings for the item, “Assesses self-efficacy for change” within this sub-scale, which most likely contributed to the lower score (see Table 2). Session process scores were higher than session structure scores for both delivery modes.

Criterion validity

Table 4 presents adjusted and unadjusted beta values based on linear mixed-effects models for each of the ACFC domain and subscale scores from sessions delivered in the first 3 months. Scores for ASPIRE-Group sessions were significantly associated with weight loss. The beta coefficients presented in Table 4 estimate weight loss in kilograms at 3 months for each one-unit increase in the corresponding subscale score. Approximately one-half kilogram more weight was lost, on average, for each ten-point improvement for each of the two session structure subscales. The single strongest predictor of weight loss was adherence to the “Stays on Track” subscale within the session process domain for ASPIRE-Group sessions; each 10-point increase translated to nearly 3 kg of weight loss at 3 months. ACFC domain and subscale scores were not associated with weight loss for ASPIRE-Phone sessions.

Table 1 | Factor loadings using principal component factors with orthogonal varimax rotation

Checklist dimension summary score					
Individual items	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Session structure					
Set goals and monitor progress (factor 1)					
Determines compliance	0.85	.	NA	NA	NA
Reviews goal attainment	0.90	.	NA	NA	NA
Discusses success and failures	0.66	0.32	NA	NA	NA
Prompts possible goals	0.67	0.38	NA	NA	NA
Agrees on goals	0.60	0.38	NA	NA	NA
Assess and personalize self-regulatory content (factor 2)					
Elicits discussion of content	.	0.85	NA	NA	NA
Assesses knowledge of content	.	0.86	NA	NA	NA
Customizes content	.	0.81	NA	NA	NA
Assesses self-efficacy for change	.	0.52	NA	NA	NA
Identify barriers and problem-solve solutions	0.49	0.56	NA	NA	NA
Variance accounted for	31.8 %	31.2 %			
Eigen value	4.87	1.43			
Session process					
Manage the session (factor 3)					
Organized and prepared	NA	NA	0.80	.	.
Times session delivery well	NA	NA	0.78	.	.
Employs motivational interviewing	NA	NA	0.82	.	.
Seeks clarification	NA	NA	0.84	.	.
Delivers didactic material in a matter of fact and friendly way	NA	NA	0.61	.	.
Creates a supportive and empathetic climate (factor 4)					
Avoids judgmental responses	NA	NA	.	0.82	.
Responds empathetically	NA	NA	0.40	0.71	.
Stays on track (factor 5)					
Stays on topic	NA	NA	0.37	.	0.73
Modulates distractions	NA	NA	.	.	0.81
Variance accounted for			34.2 %	18.7 %	15.3 %
Eigen value			3.99	1.64	1.19

Factor loadings computed separately for adherence to content and coaching quality items

Notes: Period (.) indicates abs(loading)<0.4

NA not applicable

DISCUSSION

The 19-item ACFC was found to be a reliable measure of ASPIRE-VA program coaching treatment fidelity for five sub-scales across two domains. The ACFC assessed structural aspects of coaching (session process and session structure) and session quality (sets goals and monitors progress, assesses and personalizes self-regulatory content, manages the session, creates a supportive and empathetic climate, and stays on track). All sub-scale scores were independently and positively associated with weight loss in ASPIRE-Group, which indicates that these features were “active ingredients” for achieving weight loss [13].

No other published validated tools for group- or phone-based delivery of weight management interventions were found, which is notable in terms of understanding the results from the present study in context. Though a few recent studies have developed fidelity assessment tools based on a self-reported measure of behavioral health provider fidelity [45] or expert

ratings for delivering a complex intervention for chronic pain [46] and a practice change intervention [47] (both of the latter instruments assessed face validity and reliability only), the dearth of validated and published instruments extends beyond weight loss interventions and is well-recognized [48, 49, 9, 50–53]. Clearly, there is a need for validated measures to help determine the effectiveness of weight loss interventions when delivered as they were intended.

In the ASPIRE-VA trial, ratings were used to identify gaps in coaching skills that supervisors remediated through targeted skill-building approaches. Lower and more variable ratings for ASPIRE-Group compared to ASPIRE-Phone highlighted the challenges that our non-clinician interventionists had when coaching groups of patients, many of whom had complex medical and socio-economic circumstances. Discussions with the interventionists led to innovations, such as the use of a sign-in checklist for each participant to indicate progress toward their dietary, walking, and other goals. Supervisors guided the interventionists

Table 2 | ACFC inter-rater agreement and ratings based on ratings of recordings of phone and group sessions

	Inter-rater agreement		Phone <i>n</i> =134		Group <i>n</i> =51	
	(%) ^a	Pearson correlation ^b	Mean	SD	Mean	SD
Session structure (0–20)		0.82	16.27	3.68	12.45	5.75
Set goals and monitor progress (0–10)		0.80	8.90	1.70	6.47	3.37
Determines compliance	91.5		1.93	0.28	1.20	0.85
Reviews goal attainment	88.1		1.82	0.52	1.08	0.93
Discusses success and failures	84.5		1.75	0.58	1.31	0.73
Prompts possible goals	96.6		1.80	0.42	1.45	0.76
Agrees on goals	93.2		1.60	0.60	1.43	0.76
Assess and personalize self-regulatory Content (0–10)		0.79	7.37	2.52	5.98	2.84
Elicits discussion of content	93.2		1.78	0.53	1.55	0.54
Assesses knowledge of content	91.5		1.63	0.61	1.37	0.69
Customizes content	84.7		1.71	0.59	1.41	0.73
Assesses self-efficacy for change	76.3		0.86	0.89	0.61	0.87
Identify barriers and problem-solve solutions	84.7		1.38	0.76	1.04	0.85
Session process (0–18)		0.62	16.64	1.88	13.84	3.62
Manage the session (0–10)		0.65	8.95	1.62	7.37	2.53
Organized and prepared	94.9		1.89	0.32	1.47	0.58
Times session delivery well	91.5		1.60	0.56	1.24	0.71
Employs motivational interviewing	91.5		1.72	0.47	1.33	0.65
Seeks clarification	91.5		1.78	0.51	1.47	0.61
Delivers didactic material in a matter of fact and friendly way	88.1		1.93	0.28	1.86	0.40
Create a supportive and empathetic climate (0–4)		0.69	3.89	0.40	3.76	0.68
Avoids judgmental responses	100		1.93	0.29	1.92	0.34
Responds empathetically	100		1.96	0.19	1.84	0.42
Stays on track (0–4)		0.53	3.81	0.51	2.71	1.33
Stays on topic	89.8		1.87	0.33	1.31	0.71
Modulates distractions	86.4		1.93	0.35	1.39	0.72

^a Percent agreement between two independently rated sets of items on 0 (not covered) versus 1–2 (partially or completely covered)

^b Correlation of ratings between two independently rated sets of items based on the 0–2 completion scale (not covered, partially covered, and completely covered)

about how to use this information to segue into salient topics in an efficient but personalized way.

There are a few possible explanations for why ASPIRE-Phone scores did not predict weight outcomes. First, ACFC scores were significantly higher compared to ASPIRE-Group and had much lower variation, indicating a ceiling effect. In fact, median scores were at the maximum value for all three subscales in the session process domain and for one of the two subscales in the session structure domain for ASPIRE-Phone. Second, ASPIRE-Phone participants experienced significantly lower weight loss compared to ASPIRE-Group [18]. Therefore, it would be difficult to detect effects of coaching for this group.

The ACFC was designed to assess fidelity of coaching specifically within the ASPIRE-VA program. However, many of the characteristics and flow of the sessions are common to other behavioral interventions targeted to weight loss, e.g., creating a supportive and empathetic climate. Future research is needed to determine the extent to which the ACFC may be useful for other similar interventions. Another area of development is to evaluate the use of the ACFC as a self-

assessment tool or for use by clinical supervisors who may not have the level of expertise of our raters. One VAMC is currently piloting the ASPIRE-Group program. A health psychologist supervisor is using the ACFC to rate recordings of group sessions. The supervisor has found the ACFC valuable for assessing and then providing feedback and helping build skills collaboratively with the interventionists. This is one example of how the ACFC could be used in a clinical setting in a feasible and useful way. A user-friendly version of the ACFC, with items rearranged sequentially to follow the flow of each session is provided in Appendix 2. Issues to be attentive to include receptivity of interventionists to being rated in varying contexts, ethical considerations for coaches and patients (e.g., being recorded, potential distraction from treatment), and exploring alternative modes of use (e.g., a checklist for self- or peer-ratings or direct observations).

While there are strengths of this study, there are also several notable limitations that deserve attention. Though ACFC ratings were associated with weight loss, factors not measured by the ACFC may have

Table 3 | ACFC domain composition, reliability, and standardized means by delivery mode

	Number of items	Score range	Cronbach alpha	Standardized scores	
				Mean	Standard deviation
Group					
Session structure	10	0–20	0.93	64.0	30.4
Set goals and monitor progress	5	0–10	0.92	64.4	35.0
Deliver psycho-educational content	5	0–10	0.82	63.6	28.6
Session process	9	0–18	0.82	76.2	19.1
Manage the session	5	0–10	0.91	74.4	25.2
Creates a supportive and empathic climate	2	0–4	0.85	91.0	22.7
Stays on track	2	0–4	0.77	66.0	29.7
Phone					
Session structure	10	0–20	0.80	83.1	17.7
Set goals and monitor progress	5	0–10	0.67	89.4	16.4
Assess and personalize self-regulatory content	5	0–10	0.75	76.8	23.9
Session process	9	0–18	0.69	91.6	10.5
Manages the session	5	0–10	0.75	88.9	15.5
Creates a supportive and empathic climate	2	0–4	0.68	95.8	12.2
Stays on track	2	0–4	0.19	94.2	15.2
Total					
Session structure	10	0–20	0.88	77.8	23.4
Set goals and monitor progress	5	0–10	0.85	82.4	25.5
Assess and personalize self-regulatory content	5	0–10	0.79	73.1	25.8
Session process	9	0–18	0.80	87.3	15.0
Manages the session	5	0–10	0.85	84.9	19.7
Creates a supportive and empathic climate	2	0–4	0.78	94.4	15.8
Stays on track	2	0–4	0.66	86.4	23.7

contributed to weight outcomes. In addition, the ACFC was not used to differentiate the ASPIRE-VA programs from the usual care MOVE! program because it was not feasible to conduct fidelity ratings for

Table 4 | Adjusted and Unadjusted beta coefficients for association of ACFC domains with weight loss at 3 months (based on N=90 ratings of recordings)

	Unadjusted ^a	Pvalue	Adjusted ^b	Pvalue
Session structure				
Phone				
Adherence to content	-0.017	0.657	-0.021	0.539
Set goals and monitor progress	-0.040	0.433	-0.044	0.321
Assess and personalize self-regulatory content	-0.007	0.825	-0.010	0.717
Coaching quality	-0.021	0.804	-0.029	0.706
Manages the session	-0.022	0.653	-0.026	0.549
Creates a supportive and empathic climate	0.028	0.846	0.017	0.892
Stays on track	0.076	0.435	0.082	0.354
Session process				
Group				
Adherence to content	-0.047	0.013	-0.048	0.004
Set goals and monitor progress	-0.046	0.015	-0.048	0.005
Assess and personalize self-regulatory content	-0.048	0.012	-0.049	0.003
Coaching quality	-0.071	0.013	-0.073	0.004
Manages the session	-0.059	0.012	-0.060	0.004
Creates a supportive and empathic climate	-0.049	0.024	-0.050	0.008
Stays on track	-0.291	0.024	-0.294	0.009

Beta coefficients are based on standardized scores that range from 0 to 100 and are based on the interaction of fidelity by visit

^a Adjusted for baseline weight only

^b Adjusted for baseline weight, age, depression diagnosis, substance use diagnosis, Charlson index, and BMI

MOVE!. In addition, a relatively small sample of sessions were recorded and rated, and coaches were responsible for recording a sample of their own sessions. Recent recommendations have specified recording every session and randomly selecting recordings to rate. Moreover, some items had lower than optimal inter-rater agreement (e.g., the stays on track subscale). Reliability, and the possibility of dropping these items altogether, needs further evaluation. More exploration is needed as to whether the ACFC performed better for group- than for phone-based coaching due to measurement error, lack of variation, or ceiling effect for the measure in the present study, or the fact that weight change for ASPIRE-Phone was relatively small (−1.4 kg at 3 months) [18]. It is also important to note that the ACFC assesses fidelity of treatment delivery. Other dimensions of program fidelity [5] will be evaluated in a future article.

CONCLUSIONS

The ACFC is the first validated checklist to assess fidelity of a patient-centered coaching approach for weight management. It is a reliable measure of coaching fidelity for group-based coaching and for coaching individuals over the phone. Higher ACFC ratings were associated with significant and clinically mean-

ingful weight loss for group-based delivery of ASPIRE-VA coaching. The ACFC assessment highlights “active ingredients” for successful weight loss. It can be used to reliably assess quality of delivery of self-management support for weight loss and is an effective tool for supervisors to build interventionist skills over time. It is most helpful when done as close to “real-time” as possible so that timely feedback can be given to interventionists and supervisors to improve skills [9] and enhance patient outcomes.

Acknowledgments: This work was funded by Veteran Affairs Health Services Research & Development (IBB 09-034) and Quality Enhancement Research Initiative (QLP 92-024) programs. The ASPIRE-VA, for which the Coaching Fidelity Checklist was developed, is registered on clinicaltrials.gov with Trial Registration Number: NCT00967668NCT00967668IRB. The views expressed in this article are those of the authors and do not represent the views of Veterans Affairs. We thank the Veterans who agreed to participate in this study the ASPIRE-VA coaches who were committed to serving them.

Conflict of interest: Laura J. Damschroder, David E. Goodrich, Hyungjin Myra Kim, Robert Holleman, Leah Gillon, Susan Kirsh, Caroline R. Richardson, and Lesley D. Lutes declare that they have no conflict of interest.

Adherence to ethical standards: All procedures followed were in accordance with study protocols approved by the Institutional Review Boards at the two study sites for the original study. Informed consent and assent to be recorded was obtained from all participants included in the study.

Appendix 1: ASPIRE-VA coaching fidelity checklist (ACFC) ordered by subscale

Session Date ___ / ___ / ___ Interventionist _____ Rater _____ Session Number ___ Duration _____ Mode _____

Session structure

Scoring:

- 0: Did not cover=this topic or focus point did not happen at all
- 1: Partially covered=this happened to some extent with the individual, or in a group setting, occurred for some but not all group members, all of the time (e.g., interventionist facilitated discussion, but only among certain members of the group)
- 2: Fully covered=the topic was met fully for individual or all participants in a group

Set goals and monitor progress

- __a–Health Coach determines level of self-monitoring compliance (i.e., average daily step counts, Stoplight levels)
- __b–Health Coach prompts review of goal attainment (i.e., review of goals versus actual)
- __c–Health Coach elicits discussion about successes/failures since last session and initiates problem-solving approach when necessary to address barriers
- __d–Health Coach prompts participant(s) to identify *specific* small change goals for the next

week (a “*menu of ideas*” for steps, Stoplight colors, and/or current week topic)

__e–Health Coach and participant(s) come to agreement on small change goals

Assess and personalize self-regulatory content

__a–Health Coach elicits discussion of a psycho-educational topic that develops a self-management skill or changes cognitions

__b–Health Coach assesses participant(s)’ knowledge of topic and degree of relevance to their weight management practices

__c–Health Coach customizes session content to individual/group situation to increase knowledge or skills

__d–Health Coach assesses participant(s)’ self-confidence/readiness to follow-through on plan

__e–Health Coach helps participant(s) identify barriers to success and problem-solve possible solutions to these contingencies

Session process

Scoring:

- 0: Did not demonstrate=this process objective or component was not demonstrated at all

- 1: Inconsistently demonstrated=this happened to some extent, but not or all group members, all of the time
- 2: Demonstrated consistently through entire session= objective was demonstrated consistently and appropriately throughout the entire session
- 9: NA

Manage the session

- __a–Health coach came prepared and organized
- __b–Time was allocated appropriately in order to cover the appropriate content focus points for that session
- __c–Health coach delivered didactic material in a matter of fact and friendly way
- __d–Health coach facilitated discussion and interaction using open-ended questions, affirmations, reflections, summaries
- __e–Health coach elicits clarification of participant(s)’ engagement by seeking feedback about didactic content

Stay on track

- __a–Health coach addressed process (tangential) issues but did not allow them to disrupt content agenda
- __b–Health coach modulated distractions (e.g., side bar conversations, interruptions by family members)

Create a supportive and empathetic climate

- __a–Health coach avoided judgmental feedback on participant(s) contributions
- __b–Health coach responded empathically and accurately to individual or group member behavior (verbal, nonverbal)

Appendix 2. ASPIRE-VA coaching fidelity checklist (ACFC) items ordered for clinical use

Session Date ___/___/___ Interventionist _____ Rater _____ Session Number ___ Duration _____ Mode _____

Session core elements

Scoring:

- 0: Did not cover=this topic or focus point did not happen at all
- 1: Partially covered=this happened to some extent, but not or all group members, all of the time (e.g., interventionist facilitated discussion, but only among certain members of the)
- 2: Fully covered=the goal was met fully for all participants

Checking in on self-monitoring and goal attainment

- __a–Interventionist determines level of self-monitoring compliance (i.e., average daily step counts, Stoplight levels)
- __b–Interventionist prompts review of goal attainment (i.e., review of goals versus actual)
- __c–Interventionist elicits discussion about successes/failures since last session and initiates problem-solving approach when necessary to address barriers

Assess and personalize self-regulatory content

- __a–Interventionist elicits discussion of a psycho-educational topic that develops a self-management skill or changes cognitions
- __b–Interventionist assesses participant(s)’ knowledge of topic and degree of relevance to their weight management practices

- __c–Interventionist customizes session content to individual/group situation to increase knowledge or skills

Action planning and session wrap-up

- __a–Interventionist prompts participant(s) to identify *specific* small change goals for the next week (a “menu of ideas” for steps, Stoplight colors, and/or current week topic)
- __b–Interventionist assesses participant(s)’ self-confidence/readiness to follow-through on plan
- __c–Interventionist helps participant(s) identify barriers to success and problem-solve solutions to these contingencies
- __d–Interventionist and participant(s) come to agreement on small change goals

Notes for coach:

Scoring:

- 0: Did not demonstrate=this process objective or component was not demonstrated at all
- 1: Inconsistently demonstrated=this happened to some extent, but not or all group members, all of the time
- 2: Demonstrated consistently through entire session= objective was demonstrated consistently and appropriately throughout the entire session
- 9: NA

Manage the session

- __a—Interventionist came prepared and organized
- __b—Time was allocated appropriately in order to cover the appropriate content focus points for that session
- __c—Interventionist delivered didactic material in a matter of fact and friendly way
- __d—Interventionist facilitated discussion and interaction using open-ended questions, affirmations, reflections, summaries
- __e—Interventionist elicits clarification of participant(s)' engagement by seeking feedback about didactic content

Stay on track

- __a—Interventionist addressed process (tangential) issues but did not allow them to disrupt content agenda
- __b—Interventionist modulated distractions (e.g., side bar conversations, interruptions by family members)

Create a supportive and empathetic climate

- __a—Interventionist avoided judgmental feedback on participant(s) contributions
- __b—Interventionist responded empathically and accurately to individual or group member behavior (verbal, nonverbal)

Notes for coach:

Appendix 3. ASPIRE-VA coaching fidelity checklist (ACFC) guidelines for assigning ratings—items ordered for clinical use

Heading information

- a. Session date: enter date of session being rated
- b. Interventionist: enter initials of interventionist
- c. Rater: enter initials of the individual assigning the ratings for the session
- d. Session number: enter the session/topic number or week (e.g., week 1)
- e. Duration: enter the number of minutes from start to finish
- f. Mode: enter group or phone

There should be no *not applicable* (N/A) for any core element. To minimize subjectivity, this checklist is designed to determine whether a core element occurred. Rather than using a dichotomous rating of covered or not covered, *partially covered* is added to refer to situations where the interventionist begins to cover a point but is diverted from proceeding by a tangential issue or other reason.

Checking in on self-monitoring and goal attainment

- __a—Interventionist determines level of self-monitoring compliance (i.e., average daily step counts, Stoplight levels)

0: Did not cover—The interventionist did not ask the patient to provide average weekly attainment of step counts and Stoplight goals based on their ASPIRE tracking log.

- 1: Partially covered—Interventionist began to assess self-monitoring compliance but failed to (1) obtain at least step count and stoplight goals; (2) did not collect data from all group members; (3) or is diverted off topic and did not come back to compliance check
- 2: Fully covered—basic logging compliance was obtained from participant or all group participants

__b—Interventionist prompts review of goal attainment (i.e., review of goals versus actual)

0: Did not cover—The interventionist did not prompt/or review participant(s)' actual behavior versus planned goal

- 1: Partially covered—Interventionist did not fully review actual vs. planned goal attainment by (1) only reviewing progress for one of two key behaviors (step counts, red goal, green goal); (2) did not collect data from all group members; (3) or was diverted off topic and does not come back to compliance check
- 2: Fully covered—Interventionist obtained/reviewed participant(s)' self-reported goal attainment relative to goals for phone participant or all group participants

__c—Interventionist elicits discussion about successes/failures since last session and initiates problem-solving approach when necessary to address barriers

0: Did not cover—The interventionist did not elicit participant(s)' perspective on specific barriers or facilitating factors to goal attainment to identify strategies to overcome problems to successful goal attainment in the future; nor did the participant volunteer these barriers out of natural group/phone discussion.

- 1: Partially covered—Interventionist elicited participant(s)' perspective on specific barriers or facilitating factors to goal attainment to identify strategies to overcome problems to successful goal attainment in the future but did not obtain specific response from participant(s) because interventionist: (1) did not collect data from all group members; (2) or was diverted off topic and did not obtain sufficient understanding of barriers/feedback to offer reinforcement to facilitators and/or problem solving response to barriers. Score as "1: if participant(s) voluntarily discussed issue but interventionist did not elicit further elaboration or problem-solving for participant(s) and passively permitted unguided discussion of barriers.

- 2: Fully covered=Interventionist elicited participant(s)' perspective on specific barriers or facilitating factors to goal attainment to reinforce successes and to identify strategies to overcome problems to successful goal attainment in the future for phone participant or all group participants

Assess and personalize self-regulatory content

a—Interventionist elicits discussion of a psycho-educational topic that develops a self-management skill or changes cognitions

- 0: Did not cover=The interventionist did not engage participant(s) with key questions that stimulated a discussion of a self-management session topic that was designed to foster development of specific behavioral skills or awareness of psychosocial topic for self-regulation of weight (i.e., no specific ASPIRE session module was discussed). An undesirable example would be to ask an open-ended question about a topic but not follow-up with substantive prompts and queries that customize the topic to participant or group interests or needs (see minimum duration).
- 1: Partially covered=The interventionist engaged participant(s) with key questions to initiate a discussion about a self-management session topic that is designed to foster development of a specific behavioral skills or awareness of psychosocial topic for self-regulation of weight (i.e., a specific ASPIRE session module was discussed). However, after discussion is initiated, substantive engagement is curtailed due to discussion of tangential issues, interventionist changed topic before participant(s) adequately expressed how topic pertains to their life situation, and/or participant(s) replies with short answers or denied relevance.
- 2: Fully covered=The interventionist fully engaged participant(s) in a discussion of a self-management session topic that is designed to foster development of a specific behavioral skills or awareness of psychosocial topic for self-regulation of weight (i.e., a specific ASPIRE session module was discussed). Interventionist skillfully uses key open-ended questions to initiate participant dialogue using techniques like elicit–provide–elicit, chunk–check–chunk, summaries, and reflective listing that encouraged participant(s) to explore and draw own conclusions as to the personal relevance of content. “Substantive discussion” should last at least 5 min in phone session and at least 30 min (phase I groups) or 15–30 min (phases II and III groups).

b—Interventionist assesses participant(s)' knowledge of topic and degree of relevance to their weight management practices

- 0: Did not cover=The interventionist did not attempt to elicit specific thoughts, assumptions, meanings, or behaviors related to the current ASPIRE content in order to customize the material in a way that was relevant to participant(s) or to help overcome resistance to exploring the topic.

- 1: Partially covered=The interventionist used appropriate techniques (e.g. open-ended questions, reflections, metaphors) to elicit specific thoughts, assumptions, meanings, or behaviors related to the current ASPIRE content but had difficulty establishing a focus or focused on thoughts/behaviors that were irrelevant to the participant(s)' life situation.
- 2: Fully covered=The interventionist skillfully focuses on key thoughts, assumptions, behaviors, etc. that made the topic relevant to participant(s)' weight loss efforts and which offered an opportunity for change to help support progress towards healthy weight loss. Interventionist also asked for participant(s) understanding of session content, asking permission to share information or advice to correct misperceptions or inaccuracies.

c—Interventionist customizes session content to individual/group situation to increase knowledge or skills

- 0: Did not cover=The interventionist made no effort to customize ASPIRE content in participant workbook to information elicited from participant(s) about topic relevance (e.g., thoughts, beliefs, available resources/constraints, etc.) and workbook exercises are not modified to encourage participant(s) to consider topic as another strategy to help achieve weight loss goals.
- 1: Partially covered=The interventionist appropriately customizes discussion of ASPIRE content in participant workbook to information elicited from participant(s) about topic relevance but does not explore or encourage participant(s) to take actionable steps to integrate this information into their weight loss efforts.
- 2: Fully covered=The interventionist appropriately customizes discussion of ASPIRE content in participant workbook to information elicited from participant(s) about topic relevance and supports participant(s)' efforts to take actionable steps to integrate this new content information into their weight loss efforts.

Action planning and session wrap-up

a—Interventionist prompts participant(s) to identify *specific* small change goals for the next week (a “menu of ideas” for steps, Stoplight colors, and/or current week topic)

- 0: Did not cover=The interventionist did not help participant(s) identify a list of potential ideas for daily or weekly behavior changes (i.e., menu of ideas) that could be implemented before next ASPIRE session.
- 1: Partially covered=The interventionist either appropriately helped some participant(s) but not all participants summarize a list of patient-cited ideas for change and choose goals (group only), or interventionist chose goal for patient rather than patient choosing own goal.
- 2: Fully covered=The interventionist skillfully focuses helped participant(s) summarize a list of possible goals for behavior change (from current or prior sessions) and helped participant(s) identify specific, a behavior to be implemented before the next ASPIRE session. Interventionist may also query whether there was a potential goal behavior not on the menu of ideas list to be added.

b–Interventionist assesses participant(s)’ self-confidence/readiness (motivation) to follow-through on plan

- 0: Did not cover=The interventionist took no action to elicit participant motivation (self-confidence, readiness, interest using 1-10 rulers or verbal inquiry) to implement goals selected for next ASPIRE session.
- 1: Partially covered=The interventionist elicited participant motivation to implement goals selected for next ASPIRE session but did not use 1–10 rulers to help identify barriers, beliefs, or self-confidence to carry out action steps to goal.
- 2: Fully covered=The interventionist elicited participant motivation and ability to implement goals selected for next ASPIRE session by using 1-10 rulers to help identify barriers, benefits, and solutions depending on participant response.

c–Interventionist helps participant(s) identify barriers to success *and* problem-solves possible solutions to these contingencies

- 0: Did not cover=The interventionist took no action to use problem-solving theory model to help participant(s) identify strategies to overcome potential barriers cited to implementing their goals before the next ASPIRE session. Patient action steps were vague and did not fit into a coherent strategy to help participant(s) achieve specific ASPIRE goals for weight loss.
- 1: Partially covered=The interventionist successfully elicited potential barriers/challenges to attaining short-term goals and encouraged participant(s) to identify strategies to overcome potential barriers cited by participant(s) to implementing their goals

before the next ASPIRE session but no agreement on specific steps participants will take for these contingencies.

- 2: Fully covered=The interventionist successfully elicited potential barriers/challenges to attaining short-term goals and encouraged participant(s) to identify strategies to overcome potential barriers cited by participant(s) to implementing their goals before the next ASPIRE session and encouraged exploration of hypothetical scenarios of what steps needed to be taken to be successful at making a change and a elicited a set of specific actions that participant(s) will take to overcome goals to achieve their goal (what, when, where). When applicable, interventionist asks permission to share advice on what others have done or to provide expert information.

d–Interventionist and participant(s) come to agreement on small change goals

- 0: Did not cover=There was vague or poor agreement in action steps (measureable goal) behaviors to be implemented by participant in interim period before next session. Recorded goals may be interventionist initiated with tacit acceptance by participant(s).
- 1: Partially covered=Interventionist and participant identified ASPIRE goals to be implemented before next session and tracked with monitoring log. However, interventionist did not summarize these goals, how they will be achieved, and tracked and verified for understanding and accuracy with participant(s) (e.g., “have I gotten that right?”)
- 2: Fully covered=Interventionist and participant collaboratively identified ASPIRE goals to be implemented before next session and tracked with monitoring log. Interventionist summarized the action planning steps to be implemented and elicited verification from participant that what he/she has described is accurate. Also, checks “Is there anything else?”

Session conduct that promotes high-quality delivery

Manage the session

a–Interventionist came prepared and organized

- 0: Did not demonstrate=Interventionist presents in a disorganized, flustered, or anxious state. Session lacks focus, direction, and interventionist is unable to speak authoritatively about planned session content or the personal details of individual participant(s).
- 1: Inconsistently Demonstrated=Interventionist was able to deliver sections of session well

(i.e., check-in, content, or action planning) but appears flustered and confused delivering information customized to individual session or participant during one section of session (e.g., lack of familiarity or ability to customize session content with outline or lack of participant progress in ASPIRE (participant knowledge)).

- 2: Demonstrated consistently=Interventionist is able to confidently lead the session, providing a focused structure to session that provided participant(s) with clear sense of the session agenda, customized data (taking from general to personally specific), and provided session specific information/or advice when needed that fit within the personal situation/goal progress of participant(s).

b–Time was allocated appropriately in order to cover the appropriate content focus points for that session

- 0: Did not demonstrate=Interventionist made no attempt to structure session time. Session seemed aimless/without focus.
- 1: Inconsistently Demonstrated=Session had some focus (topic), but interventionist had problems with structure or pacing (e.g., too little structure, too slowly paced, or too rapidly paced and ending early without adequate topic discussion).
- 2: Demonstrated consistently=Interventionist used time efficiently by tactfully limiting tangential or unproductive discussion and by pacing the session as rapidly as needed for the participants(s) to cover check-in, session content, and action planning. Adequate time was given to ensure topics were adequately covered to participant(s) satisfaction and understanding.

c–Interventionist delivered didactic material in a matter of fact and friendly way

- 0: Did not demonstrate=Interventionist relied primarily on persuasion, lecturing, or debating participant. Interventionist did more talking than participant and expert role of telling participant(s) (imposing viewpoint) put participant(s) on the defensive.
- 1: Inconsistently Demonstrated=Interventionist for the most part, helped participant(s) see new perspectives through guided discovery (engaged discussion with peers, considering alternatives, weight pros/cons, etc., asking permission to share advice or info) rather than taking adversarial or educational orientation. Used opened-ended questions appropriately but on occasions, relied too heavily on persuasion or advice giving than self-discovery.

- 2: Demonstrated consistently=Interventionist was skillful in using guided discovery to engage participant(s) to explore challenges with weight loss, consider new information/skills, and to draw their own conclusions. Participants were treated as experts about their own behavioral change interests and were treated with respect, sincerity, and positive regard/optimism. Interventionist achieved a good balance between skillful use of open-ended questioning and reflective listening, and use of information provision strategies to stimulate engagement in intervention program.

d–Interventionist facilitated discussion and interaction using open-ended questions, affirmations, reflections, and summaries

- 0: Did not demonstrate=Interventionist demonstrated poor use of patient centered communication techniques (MI). Participant exhibited a tendency to talk more than patient, relied on closed-ended questions, and had difficulty demonstrating active, empathetic listening skills to encourage active discussion (affirmations, reflections, and summaries).
- 1: Inconsistently Demonstrated=Interventionist demonstrated a basic but inconsistent ability to use patient centered communication skills to foster active discussion of session components. Interventionist regularly used open-ended questions to initiate discussion but had difficulty using affirmations, complex reflections, and summaries to skillfully guide discussion, elicit deeper participant reflection regarding a response, or to clarifying understanding.
- 2: Demonstrated consistently=Interventionist displayed skillful use of MI techniques to engage participant(s) in collaborative discussion regarding their lifestyle changes. Interventionist relied on reflections to obtain deeper reflection from open-ended questions. Reflections were frequently paired with affirmations to convey empathy and understanding and summaries were used periodically during session to review key points and verified both interventionist and participant(s) understanding of content.

e–Interventionist elicits clarification of participant(s)' engagement by seeking feedback about didactic content

- 0: Did not demonstrate=Interventionist did not ask for feedback to determine participant(s)' understanding of, or response to a session or goal setting objective.
- 1: Inconsistently Demonstrated=Interventionist elicited some feedback from participant(s), regarding satisfaction/buy-in to

session topics. Interventionist inconsistently responded to participant(s) indications of discounting, resisting, or minimizing value of content.

- 2: Demonstrated consistently=Interventionist was adept at eliciting and responding to verbal and non-verbal feedback through session (e.g., elicited reactions to session, regularly checked for participant reactions (reflections, summaries)) to insure learning atmosphere is positive and participant(s) are receptive to information being presented.

Stay on track

__a–Interventionist addressed process (tangential) issues but did not allow them to disrupt content agenda

- 0: Did not demonstrate=Interventionist made no attempt to redirect discussion of topics unrelated to ASPIRE or psycho-behavioral self-management topics for weight maintenance/loss. Session rambled and some participants became visibly irritated by digressions from main topic or monopolization of discussion by one or two participants.
- 1: Inconsistently Demonstrated=Interventionist covered the primary focus points related to check-in, session content, and action planning but at times, Interventionist had problems with participants getting off-topic by telling unrelated stories or discussing issues unrelated to ASPIRE or current discussion point raised by Interventionist that had little focus or benefit to other participants.
- 2: Demonstrated consistently=Interventionist respectfully acknowledged participants' contributions to discussion topics but skillfully redirected discussion of tangential issues by asking permission to revisit the topic later in the session (or after session) or by cutting the participant off respectfully. Session remained focused on key topics generated by session content or problems encountered with self-monitoring (check-in) or action planning.

__b–Interventionist modulated distractions (e.g., side bar conversations, interruptions by family members)

- 0: Did not demonstrate=Interventionist made no attempt to use group management skills to cut off side conversations that interrupted a participant's response to a discussion topic or to remind disruptive participants' that their behavior (e.g. interrupting a speaker) violates the group's rules. On phone sessions, interventionist made no attempt to ask participant to manage disrupting

elements in background (pets, TV or radio noise, or conversations with people in background).

- 1: Inconsistently Demonstrated=Interventionist was moderately successful at using group management skills to cut off side conversations or disruptive participant's actions that violate the group's rules. On phone sessions, interventionist was mostly successful at managing disrupting background noises or factors.
- 2: Demonstrated consistently=Interventionist respectfully acknowledged participants' contributions to discussion topics but skillfully asked permission to modulate distractions or informs disruptive group members of inappropriate behaviors. Session is focused and respectful of participant(s) rights.

Create a supportive and empathetic climate

__a–Interventionist avoided judgmental feedback on participant(s) contributions

- 0: Did not demonstrate=Interventionist demonstrated inability to keep personal opinions or advice out of session and was not able to demonstrate acceptance of patient struggles, warmth, and genuine concern for participant(s). Interventionist seemed hostile, demeaning, unnecessarily critical, or some way destructive to the participant.
- 1: Inconsistently Demonstrated=Interventionist was not destructive to participant(s)' welfare but at times, exhibited statements that undermined patient trust that the interventionist would be helpful, optimistic, warm, concerned, and empathetic. At times, interventionist appeared inpatient, insincere, critical, or had difficulty conveying confidence in participant(s)'s ability to make changes.
- 2: Demonstrated consistently=Interventionist displayed high levels of warmth, concern, confidence, genuineness, and professionalism in helping participant(s) work through painful issues without conveying judgment or superiority.

__b–Interventionist responded empathically and accurately to individual or group member behavior (verbal, nonverbal)

- 0: Did not demonstrate=Interventionist repeatedly failed to understand what the patient explicitly said and consistently missed the literal and implied emotions underlying a point. Had difficulty reflecting or rephrasing what patient said and often missed more subtle communication (e.g., tone and sarcasm of participant(s) comments).
- 1: Inconsistently Demonstrated=Interventionist was usually able to reflect or rephrase (s)

participant's perspective as noted by what but often missed more subtle communications that need to be reflected by reflections of emotional content or what was not said. (But often unaware of shifts in level of participant impatience conveyed by sarcasm, digressions, tone of voice).

- 2: Demonstrated consistently=Interventionist displayed ability to understand participant(s)' perspective thoroughly and was skillful at expressing this understanding through appropriate verbal reflections and non-verbal responses to participant(s). Participants talked less and used more reflections, encouraging statements (uh-huh, yes, I see,) and voice to convey genuine interest and sympathetic understanding to participant discussion.
1. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol*. 2008; 41(3-4): 327-350. doi:10.1007/s10464-008-9165-0.
 2. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clin Psychol Rev*. 1998; 18(1): 23-45.
 3. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implement Sci IS*. 2013; 8(1): 117. doi:10.1186/1748-5908-8-117.
 4. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. 2011; 38(2): 65-76. doi:10.1007/s10488-010-0319-7.
 5. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol*. 2004; 23(5): 443-451.
 6. Borrelli B. The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *J Public Health Dent*. 2011; 71(s1): S52-S63. doi:10.1111/j.1752-7325.2011.00233.x.
 7. Faulkner MS. Intervention fidelity: ensuring application to practice for youth and families. *J Spec Pediatr Nurs JSPN*. 2012; 17(1): 33-40. doi:10.1111/j.1744-6155.2011.00305.x.
 8. Bopp M, Saunders RP, Lattimore D. The tug-of-war: fidelity versus adaptation throughout the health promotion program life cycle. *J Prim Prev*. 2013; 34(3): 193-207. doi:10.1007/s10935-013-0299-y.
 9. Borrelli B, Sepinwall D, Ernst D, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *J Consult Clin Psychol*. 2005; 73(5): 852-860. doi:10.1037/0022-006x.73.5.852.
 10. Hardeman W, Michie S, Fanshawe T, et al. Fidelity of delivery of a physical activity intervention: predictors and consequences. *Psychol Health*. 2007; 23(1): 11-24.
 11. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol*. 2008; 27(3): 379-387. doi:10.1037/0278-6133.27.3.379.
 12. Damschroder LJ, Hagedorn HJ. A guiding framework and approach for implementation research in substance use disorders treatment. *Psychol Addict Behav J Soc Psychol Addict Beh*. 2011; 25(2). doi:10.1037/a0022284.
 13. Cohen DJ, Crabtree BF, Etz RS, et al. Fidelity versus flexibility: translating evidence-based research into practice. *Am J Prev Med*. 2008; 35(5 Suppl): S381-S389. doi:10.1016/j.amepre.2008.08.005.
 14. Wyatt Kaminski J, Valle L, Filene J, et al. A meta-analytic review of components associated with parent training program effectiveness. *J Abnorm Child Psychol*. 2008; 36(4): 567-589. doi:10.1007/s10802-007-9201-9.
 15. Leblanc ES, O'Connor E, Whitlock EP, et al. Effectiveness of primary care-relevant treatments for obesity in adults: a systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2011; 155(7): 434-447. doi:10.7326/0003-4819-155-7-201110040-00006.
 16. Lutes LD, Winett RA, Barger SD, et al. Small changes in nutrition and physical activity promote weight loss and maintenance: 3-month evidence from the ASPIRE Randomized Trial. *nn. Behav Med*. 2008; 35(3): 351-357.
 17. Lutes LD, Daiss SR, Barger SD, et al. Small changes approach promotes initial and continued weight loss with a phone-based follow-up: nine-month outcomes from ASPIRES II. *Am J Health Promot*. 2012; 26(4): 235-238. doi:10.4278/ajhp.090706-QUAN-216.
 18. Damschroder LJ, Lutes LD, Kirsh S, et al. Small-changes obesity treatment among veterans: 12-month outcomes. *Am J Prev Med*. 2014; 47(5): 541-553. doi:10.1016/j.amepre.2014.06.016.
 19. Gearing RE, El-Bassel N, Ghesquiere A, et al. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clin Psychol Rev*. 2011; 31(1): 79-88. doi:10.1016/j.cpr.2010.09.007.
 20. Carroll C, Patterson M, Wood S, et al. A conceptual framework for implementation fidelity. *Implement Sci IS*. 2007; 2(1): 40.
 21. Nelson MC, Cordray DS, Hulleman CS, et al. A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *J Behav Health Serv Res*. 2012; 39(4): 374-396. doi:10.1007/s11414-012-9295-x.
 22. Lutes LD, Dinatale E, Goodrich DE, et al. A randomized trial of a small changes approach for weight loss in veterans: design, rationale, and baseline characteristics of the ASPIRE-VA trial. *Contemp Clin Trials*. 2013; 34(1): 161-172. doi:10.1016/j.cct.2012.09.007.
 23. Kinsinger LS, Jones KR, Kahwati L, et al. Design and dissemination of the MOVE! Weight-management program for veterans. *Prev Chronic Dis*. 2009; 6(3): A98.
 24. Hill JO. Can a small-changes approach help address the obesity epidemic? A report of the Joint Task Force of the American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *Am J Clin Nutr*. 2009; 89(2): 477-484.
 25. Hill JO, Peters JC, Wyatt HR. Using the energy gap to address obesity: a commentary. *J Am Diet Assoc*. 2009; 109(11): 1848-1853. doi:10.1016/j.jada.2009.08.007.
 26. Schoenwald SK, Garland AF, Chapman JE, et al. Toward the effective and efficient management of implementation fidelity. *Adm Policy Ment Health*. 2011; 38(1): 32-43. doi:10.1007/s10488-010-0321-0.
 27. Colquhoun H, Leeman J, Michie S, et al. Towards a common terminology: a simplified framework of interventions to promote and integrate evidence into health practices, systems, and policies. *Implement Sci*. 2014; 9(1): 51.
 28. Resnicow K, McMaster F. Motivational Interviewing: moving from why to how with autonomy support. *Int Behav Nutr Phys Act*. 2012; 9: 19. doi:10.1186/1479-5868-9-19.
 29. Moyers TB, Miller WR, Hendrickson SM. How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. *J Consult Clin Psychol*. 2005; 73(4): 590-598. doi:10.1037/0022-006x.73.4.590.
 30. Resnicow K, McMaster F, Rollnick S. Action reflections: a client-centered technique to bridge the WHY-HOW transition in Motivational Interviewing. *Behav Cogn Psychother*. 2012; 40(4): 474-480. doi:10.1017/S1352465812000124.
 31. Bauer MS, McBride L. Structured group psychotherapy for bipolar disorder: the life goals program. 2nd ed. New York: Springer Publishing Co.; 2003.
 32. Bauer MS, McBride L, Williford WO, et al. Collaborative care for bipolar disorder: part II. Impact on clinical outcome, function, and costs. *Psychiatr Serv*. 2006; 57(7): 937-945. doi:10.1176/appi.ps.57.7.937.
 33. Kilbourne AM, Goodrich DE, Lai Z, et al. Life goals collaborative care for patients with bipolar disorder and cardiovascular disease risk. *Psychiatr Serv*. 2012; 63(12): 1234-1238. doi:10.1176/appi.ps.201100528.
 34. Kilbourne AM, Goodrich DE, Lai Z, et al. Randomized controlled trial to assess reduction of cardiovascular disease risk in patients with bipolar disorder: the self-management addressing heart risk trial (SMAHRT). *J Clin Psychiatry*. 2013; 74(7): e655-e662. doi:10.4088/JCP.12m08082.
 35. Kilbourne AM, Post EP, Nosske A, et al. Improving medical and psychiatric outcomes among individuals with bipolar disorder: a randomized controlled trial. *Psychiatr Serv*. 2008; 59(7): 760-768. doi:10.1176/appi.ps.59.7.760.
 36. Waxmonsky J, Kilbourne AM, Goodrich DE, et al. Enhanced fidelity to treatment for bipolar disorder: results from a randomized controlled implementation trial. *Psychiatr Serv*. 2014; 65(1): 81-90. doi:10.1176/appi.ps.201300039.
 37. Desu MM, Raghavarao D. Sample size methodology. New York: Academic Press; 1990.
 38. Machin D, Campbell M, Fayers P, et al. Sample size tables for clinical studies. 2nd ed. Malden: Blackwell Science; 1997.
 39. Hahn GJ, Meeker WQ. Statistical Intervals. New York, NY: Wiley.
 40. Power Analysis & Sample Size (PASS) Software. Kaysville, UT: NCSS; 2005.
 41. Piette JD, Richardson C, Himle J, et al. A randomized trial of telephonic counseling plus walking for depressed diabetes patients. *Med Care*. 2011; 49(7): 641-648. doi:10.1097/MLR.0b013e318215d0c9.

42. Young J, Beck A. Cognitive therapy scale rating manual. Philadelphia: University of Pennsylvania Psychotherapy Research Unit; 1980.
43. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005; 37(5): 360-363.
44. StataCorp. *Intercooled stata 13.1*. 11th ed. College Station: Stata-Corp LP; 2013.
45. Beehler GP, Funderburk JS, Possemato K, et al. Psychometric assessment of the primary care behavioral health provider adherence questionnaire (PPAQ). *Transl Behav Med*. 2013; 3(4): 379-391. doi:10.1007/s13142-013-0216-1.
46. Mars T, Ellard D, Carnes D, Homer K, Underwood M, Taylor SJ. Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity. *BMJ Open*. 2013; 3(11).
47. Yamada J, Stevens B, Sidani S, et al. Content validity of a process evaluation checklist to measure intervention implementation fidelity of the EPIC intervention. *Worldviews Evid-Based Nurs*. 2010; 7(3): 158-164. doi:10.1111/j.1741-6787.2010.00182.x.
48. Breitenstein SM, Fogg L, Garvey C, et al. Measuring implementation fidelity in a community-based parenting intervention. *Nurs Res*. 2010; 59(3): 158-165. doi:10.1097/NNR.0b013e3181dbb2e2.
49. Breitenstein SM, Gross D, Garvey CA, et al. Implementation fidelity in community-based interventions. *Res Nurs Health*. 2010; 33(2): 164-173. doi:10.1002/nur.20373.
50. Goode AD, Reeves MM, Eakin EG. Telephone-delivered interventions for physical activity and dietary behavior change: an updated systematic review. *Am J Prev Med*. 2012; 42(1): 81-88. doi:10.1016/j.amepre.2011.08.025.
51. Mowbray C, Holter M, Teague G, et al. Fidelity criteria: development, measurement, and validation. *Am J Eval*. 2003; 24: 315-340.
52. Schinckus L, Van den Broucke S, Housiaux M, et al. Assessment of implementation fidelity in diabetes self-management education programs: a systematic review. *Patient Educ Couns*. 2014; 96(1): 13-21. doi:10.1016/j.pec.2014.04.002.
53. Swanson E, Wanzek J, Haring C, Ciullo S, McCulley L. Intervention fidelity in special and general education research journals. *J Spec Ed*. 2011; 0022466911419516. doi:10.1177/0022466911419516.