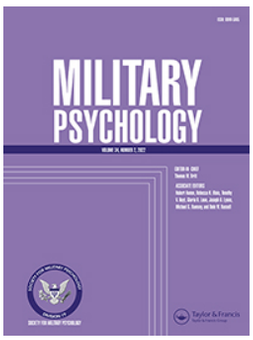

Routledge
Taylor & Francis Group

Military Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmlp20

Validation and refinement of the "Airman Comprehensive Assessment": Evaluating competency proficiencies of enlisted members

Laura G. Barron, Patrick J. Rolwes & Mark R. Rose

To cite this article: Laura G. Barron, Patrick J. Rolwes & Mark R. Rose (2022): Validation and refinement of the "Airman Comprehensive Assessment": Evaluating competency proficiencies of enlisted members, Military Psychology, DOI: 10.1080/08995605.2022.2050639

To link to this article: https://doi.org/10.1080/08995605.2022.2050639

Published online: 25 Apr 2022.

Submit your article to this journal

Article views: 6

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hmlp20

MILITARY PSYCHOLOGY
https://doi.org/10.1080/08995605.2022.2050639

Routledge
Taylor & Francis Group

Check for updates

Validation and refinement of the "Airman Comprehensive Assessment": Evaluating competency proficiencies of enlisted members

Laura G. Barron, Patrick J. Rolwes, and Mark R. Rose

U.S. Air Force, Air Education & Training Command (Randolph AFB) Texas, USA

ABSTRACT

Beyond proficiency on occupationally specific tasks, the U.S. Air Force expects members to develop proficiency on institutionally valued "soft skill" competencies (e.g., Teamwork, Communication, and Initiative) throughout their careers. As such, all E1-E6 members are annually evaluated using Behaviorally Anchored Rating Scales (BARS) designed to measure such competencies. Despite mandated use, these Airman Comprehensive Assessment (ACA) scales previously have not been empirically evaluated. To address this gap, we surveyed Air Force supervisors, using a criterion-related sampling methodology to validate the behavioral anchors for each scale. Supervisors identified two subordinates of the same rank/career field who they viewed as having (a) high potential for future success in an Air Force career or, alternately, (b) lower potential for future career success and rated each subordinate on the individual behaviors that comprise the 12 scales. ACA items were intermixed with scale items previously identified as distinguishing top performers in civilian organizations. Results demonstrate scale reliability and generally validate the ACA competency scales as stronger differentiators of supervisor-rated career potential than competency scales developed for civilian organizations. We provide recommendations for re-calibration of scale anchors based on the relative percentage of high vs. low potential members that demonstrate each behavior, and suggest changes to improve correspondence between measured competency proficiency and supervisor-rated career potential.

ARTICLE HISTORY
Received 1 April 2021
Accepted 21 October 2021

KEYWORDS
Performance appraisal; competency; Behaviorally Anchored Rating Scales (BARS); criterion-related validity

What is the public significance of this article?—This study shows the overall validity of the enlisted performance appraisal form used by the U.S. Air Force, but suggests that certain competencies that are not currently measured (Leadership, Analytical Thinking) may need to be assessed to more effectively differentiate career potential. The study also demonstrates a rigorous method for developing Behaviorally Anchored Rating Scales (rubrics defining competency proficiency levels) that is less resource-intensive than the traditional approach.

Although inherently subjective, performance rating processes merit the same psychometric scrutiny as more objective employment procedures (Bersoff, 1988; Gutman et al., 2011). For instance, the *Principles for the Validation and Use of Personnel Selection Procedures* expressly apply the same scrutiny to appraisals of job performance as to personnel selection methods like aptitude and physical ability tests (see, Gutman, 2004). Despite this, within the U.S. Air Force, and likely much of the U.S. military, the performance appraisal forms required for military members have not been formally evaluated based on rigorous quantitative methods or psychometric standards. In this paper we aim to address this gap, evaluating the reliability and criterion-related validity of the behaviorally anchored ratings scales used for enlisted evaluation, and providing empirically driven recommendations for refining the current enlisted performance appraisal instrument, demonstrating a survey-based criterion-sampling methodology that could be applied by other Services or organizations.

Current U.S. Air Force enlisted evaluation system

Two performance appraisal forms are currently required for supervisors to evaluate enlisted members at the ranks of E-6 (Technical Sergeant) and below, on an annual cycle (see AFI 36–2406, Assistant Secretary of the Air Force for Manpower and Reserve Affairs, 2021). The Enlisted Performance Report (EPR; AF 910) is a formal evaluation of performance that is used administratively as a determinant of selections for promotion, reenlistment, assignments, and other career outcomes. The

CONTACT Laura G. Barron laura.barron@us.af.mil HQ Air Education & Training Command, Bldg 902, Randolph AFB, Texas, USA.

We have no known conflict of interest to disclose. The views expressed are those of the authors and not necessarily those of the United States government, the Department of Defense, or the US Air Force.

This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 USC. 105, no copyright protection is available for such works under US Law.

Airman Comprehensive Assessment (AF 931) is a developmental appraisal instrument used to provide structure and substance to performance feedback discussions as part of a mid-cycle review (approximately 6 months prior to EPR completion). Both forms are based on a common set of competencies intended to be (but never empirically evaluated as) applicable across the ranks of E1-E6 (Airman Basic through Technical Sergeant), regardless of career field. While the EPR relies on narrative comments and a Likert-type rating format (*Met some but not all expectations, Met all expectations, Exceeded some but not all expectations*, or *Exceeded most if not all expectations*) in each competency area, revision to the ACA in 2015 implemented behaviorally anchored ratings scales (BARS) for each individual competency (Carr, 2015, March 11; Keller et al., 2014).

Behaviorally Anchored Rating Scales and typical BARS development process

BARS are intended as a means of reducing the influence of construct-irrelevant variance in job performance rating systems (Kell et al., 2017; Smith & Kendall, 1963). Performance appraisal forms such as those using graphic rating scales, the most common of all rating methods (Guion & Highhouse, 2006), are often vague and ambiguous with performance dimensions and different levels of performance on each dimension poorly defined. Consequently raters are forced to impose their idiosyncratic interpretations of dimensions and levels in making ratings (Kell et al., 2017), potentially resulting in a wide range of errors such as leniency and halo, and reducing inter-rater agreement and differentiation between persons being rated (Borman & Dunnette, 1975). BARS aim to reduce the influence of these rater idiosyncrasies by focusing raters on observable behaviors that exemplify performance at different levels on distinct dimensions (Smith & Kendall, 1963). Because it provides a clear definition and a consistent frame of reference for the competencies and levels of each competency to be rated (Klieger et al., 2018; Smith & Kendall, 1963), the use of BARS is generally considered best practice for competency-based performance assessments.

While earlier evaluations questioned whether rating formats such as BARS actually produce meaningful enhancements in the quality of ratings (Landy & Farr, 1980), recent studies provide evidence that a rigorous BARS development process produces increased relevance, relatively lower measurement bias (e.g., halo, leniency), favorable perceptions of fairness, and potentially decreased subgroup bias (cf., Klieger et al., 2018). Even when used for strictly developmental purposes, BARS have showed some demonstrable advantages. For example, at least over the short term, BARS feedback has shown to produce greater behavioral change among ratees than feedback based on a (carefully constructed) summated Likert-type rating format (Hom et al., 1982). Other benefits of BARS include greater acceptability from a legal standpoint because the approach is behaviorally based, explicit communication of performance expectations, and increased acceptance due to greater participation of workers and supervisors in the development process (Ployhart et al., 2006). Indeed, other branches of the U.S. military make use of competency-based BARS as part of their performance appraisal processes (competencies termed "performance traits" by Navy). For officer evaluation, for example, the Navy, Marine Corps, and Coast Guard all use a BARS format for administrative purposes (NAVPERS 1610/2; NAVMC 10835; CG-5310B).

Despite these advantages, one major criticism is the time-consuming process needed to develop high quality BARS (Bernardin & Smith, 1981; Brannick et al., 2007). Procedures for developing BARS vary, but commonalities in development usually include the following steps:

(1) Apply the critical incident technique (Flanagan, 1954) to collect examples of highly effective and highly ineffective workplace behaviors from subject matter experts. Normally this process is resource intensive, requiring a representative sample of subject matter experts and time for training and generation of critical incidents. In some published examples (e.g., Carretta & Walters, 1991; Klieger et al., 2018), approximately 300–400 critical incidents, provided by 20 or more SMEs, may be needed to define less than 10 competency or performance domains.
(2) Develop a preliminary performance taxonomy by editing the incidents and organizing them into groups based on content similarities, and labeling the groups. In one example of this process (Klieger et al., 2018), six members of the research team each edited, for coherence and relevance, a subset of the 430 incidents that they had been randomly assigned. They then reexamined the edited incidents, grouped them according to themes shared across their behavioral aspects, and wrote succinct definitions of the themes. Next, the incidents and themes were randomly assigned to the same six research members who reviewed the statements and themes and further refined them. Finally, the team met as a whole to discuss problematic statements and reach

consensus. The remaining 398 statements were then transitioned to a new group of SMEs for retranslation.

(3) Use a second group of subject matter experts to organize a randomized set of the critical incidents into the previously defined categories, a process known as retranslation. As an example, Carretta and Walters (1991) first used a team of a personnel research psychologist and two SMEs to reduce the 261 incidents originally generated to 73 based on whether the incidents were irrelevant, unclear, or poorly defined. They then asked 22 SMEs (USAF T-37 instructor pilots) – the third group of SMEs used in their study – to sort 73 index cards with a behavioral example printed on it into one of eight categories.
(4) Compute agreement statistics for the critical incidents based on how they were initially categorized and categorized during the retranslation process. Discard incidents that do not meet a minimum agreement threshold. For example, the group of 22 USAF T-37 instructor pilots in the Carretta and Walters (1991) study reached acceptable agreement (i.e., 70% of the sample resorted the behavior into its original category) on only 37 (50.7%) of 73 behaviors.
(5) Use an additional, independent group of subject matter experts to rate each incident on effectiveness (e.g., 1 = very ineffective to 7 = very effective). In some published examples (e.g., Martin-Raugh et al., 2016) more than 50 additional SMEs have been recruited to complete this step.
(6) Compute the mean effectiveness for each incident, and the degree of agreement on effectiveness by the subject matter experts. Discard incidents that don't meet a predetermined level of subject matter expert agreement. For example, in Martin-Raugh et al. (2016), mean effectiveness ratings for each incident signaled the extent to which an incident described ineffective or effective performance, and the standard deviation of ratings for each incident signaled the amount of agreement among SMEs. Incidents with standards deviation of 1.5 or less, signaling high agreement, were retained.
(7) Use the remaining incidents as the anchors in BARS for each performance category to which they were assigned. Use the incidents' mean effectiveness ratings to determine their placement on the effectiveness continuum. In Martin-Raugh et al. (2016), for example, the lowest band consisted of incidents with mean effectiveness ratings between 1 and 2.5, the medium score band had ratings between 2.5 and 5.5, and the highest score band consisted of incidents with mean effectiveness ratings between 5.5 and 7.

The foregoing discussion, which highlights the rigor of BARS, also illustrates its relative complexity as a multistep process that is normally labor intensive and time consuming.

An alternative method for BARS validation

Absent organizational support for a complete overhaul of the enlisted performance appraisal instruments, and the large time commitment from SMEs that would be required, in the present study we relied on a closed-ended survey of Air Force supervisors that would not require critical incident generation or other (lengthy) narrative responses. To provide a point of comparison, we intermixed behavioral statements from the current Air Force performance appraisal BARS with statements from BARS that had been rigorously developed based on a modified critical incidents methodology (behavioral event interviews) in civilian organizations (Spencer & Spencer, 1993). In doing so, we focused on a subset of competencies that had been previously identified as important for Air Force members, across career fields (Barelka et al., 2019).

Rather than rely on direct SME ratings of the level of effectiveness of each behavior for calibration, we evaluated placement along the effectiveness continuum based on the proportions of supervisor-identified high and low potential members that engaged in each behavior (a form of criterion-related validation). That is, we reasoned that behaviors engaged in by an increasingly high proportion of low-performing (low potential) members should be placed at a lower level along the effectiveness continuum relative to other behaviors on the rating scale. We also reasoned that behaviors that most strongly differentiated high from low-performing (high vs. low potential) members ought to be placed at a higher level along the effectiveness continuum. Because of the "up-or-out" military system, in which enlisted members are rarely if ever in the same job for more than four years, and are expected to take on increasing levels of responsibility in follow-on assignments, we focused on rated "potential for future success across an Air Force career" rather than performance in one's current job.

Method

Overview

Supervisors (N = 2,758) identified two subordinates of the same rank/career field who they viewed as having (a) high potential for future success in an Air Force career or, alternately, (b) lower potential for future career success and rated each (unnamed) subordinate on the individual behaviors that comprise the ACA scales. To help ensure honesty of ratings, supervisors were not required to name or otherwise identify the identity of the (higher or) lower potential members they rated on.

Participants

The full population of 32,568 current Air Force supervisors (civilian, officer, and enlisted; including Active Duty, Guard, and Reserve) was invited to participate in an online survey to support force development. The primary purpose of the survey was to develop a 360 degree developmental feedback tool for Air Force members (of any rank/grade). Although supervisors could rate on members of any rank/grade, to provide recommendations on the AF 910/931 specifically, the analyses reported here are focused on the ranks of grades E2-E6. Of the 6,011 respondents, 2,758 supervisors rated on members at the ranks of E-6 or below.

The 2,758 supervisors who rated on members in the enlisted ranks of E-6 and below included 538 officers, 1,709 enlisted members, and 511 DoD civilians. The most common ranks of respondents were E-5, E-6, and E-7 (341, 588, and 585 respondents, respectively). The most common career fields among respondents were 4N0X1-Aerospace Medical Service (N = 95), 3P0X1-Security Forces (N = 78), and 3D1X2-Cyber Systems Operations (N = 73). By gender, 77.5% of respondents were male and 22.5% female.

Participants were not required to rate on a current supervisee and could rate on any Air Force member they had worked with directly during the career. Participants were directed to identify the rank of the high and low potential ratees at the time they worked with them most extensively – the final sample included ratings on 589 high and low potential E-6 members, 906 high and low potential E-5s, and 1,263 high and low potential members at the rank of E-4 or below.

Respondents reported that, of the identified high potential members, 1,569 had since been competitively promoted to a higher rank; 365 had voluntarily separated from the Air Force with less than 20 years of service, and 23 had involuntarily separated. Of the identified low potential members 416 had since been competitively promoted to a higher rank; 413 had voluntarily separated from the Air Force with less than 20 years of service and 578 had involuntarily separated (including 281 separations for misconduct and 117 for unsatisfactory performance). Current Air Force status was unknown or unreported for 922 high potential and 1,974 low potential members.

Survey content (Behavioral statements)

Spencer and Spencer (1993) *scales*. To evaluate the potential utility of assessing Air Force members on other competencies (not currently evaluated on the AF 931), we included measures of 14 competencies that were previously identified by a large, representative sample of Air Force members as "important for success in an Air Force career" (Barelka et al., 2019). BARS to assess these competencies were selected from a competency dictionary intended for broad applicability across organizations, based on commonalities in 286 competency models developed for a broad range of job types (e.g., technical/professional, human service, sales, and managerial roles; Spencer & Spencer, 1993).

During the Spencer and Spencer (1993) scale development process, 50 verbatim examples from behavioral event interviews of each competency were Q-sorted by several researchers based on the extent to which the behaviors indicated more or less of the target competency. Behavioral statements were scaled in terms of a "*clear progression from lower to higher levels on one or more competencies*" (p. 31). Although Spencer and Spencer scaled certain competencies in terms of Impact, Breadth of Influence, or other unique criteria, all the dictionary behaviors evaluated in this present study had been scaled in terms of *Intensity and Completeness of Action* or (for the more cognitively oriented competencies) *Complexity* (e.g., taking more things, people, data, concepts, or causes into account). Scaling of each competency was intended to be ordinal, ranging from −3 (i.e., based on varying levels of negative or counterproductive behaviors) to +9 (highest level of positive behaviors), depending on the competency; certain scale anchors were defined in terms of multiple statements (e.g., "Systematically breaks down a complex problem into component parts. Uses several techniques to break apart complex problems to reach a solution.") When scale anchors were defined by multiple behavioral statements, these were evaluated separately for the purpose of the present study, for a total of 6 to 20 positive behavioral statements per competency.

Notably, although the authors asserted their "experience and evidence to indicate the levels, as applied by trained coders, are in the correct order" (p. 24), to our knowledge, no studies have quantitatively evaluated the

appropriateness of the scaling (i.e., extent to which the behaviors represented increasing levels of the competency) or the validity of specific behavioral indicators within each competency. Results in specific organizations from which the generic competency BARS were derived have shown, however, that, overall, employees identified by their management as high performers tend to perform behaviors associated with significantly higher scale values than employees identified as average or below average performers (Spencer & Spencer, 1993).

ACA (AF 931) scales. On the ACA, each competency is defined by a BARS with four scale anchors, intended to represent increasing levels of each competency. As with Spencer and Spencer (1993) scales, the lowest anchor for each competency may be either neutral ("Makes *little or no effort* to . . . ") or negatively worded ("*Fails to* . . . "), while the remaining anchors are defined in terms of positive behaviors. As with the Spencer & Spencer scales, when scale anchors were defined in terms of multiple statements, each statement was evaluated separately, for a total of 3 to 8 positively worded behavioral statements per competency. Although the AF 931 includes BARS for 12 competencies, we excluded from study the 3 competencies that are only rated when members are in a certain status (e.g., based on assignment as a trainer). Three of the ACA scales were intended to assess the competencies included in Spencer and Spencer (Initiative, Teamwork, and Communication), though the competency definitions from the two sources were not directly parallel (see, Table 1).

Rating methodology

To minimize the time required to complete the survey, supervisors received one of five survey versions, with behaviors grouped thematically (see, Table 2); ACA and Spencer & Spencer items were intermixed with additional behaviors from other sources, including the officer version

Table 1. AF 910/931 and Spencer and Spencer (1993) competency definitions.

	Competency	Definition
AF 910/931	Task Knowledge/Proficiency	Quality, quantity, results and impact of the Airman's knowledge and ability to accomplish tasks
	Initiative/Motivation	Degree of willingness to execute duties, motivate team members, and develop innovative new processes.
	Resource Utilization	How effectively the Airman utilizes resources to accomplish the mission (e.g., time management, manpower, and budget).
	Comply with Enforce Standards	Personal adherence and enforcement of fitness standards, dress and personal appearance, customs and courtesies, and professional conduct.
	Communication Skills	How well the Airman receives and relays information, thoughts, and ideas up and down the chain of command (includes listening, reading, speaking, and writing skills); fosters an environment for open dialogue.
	Teamwork (Caring, Respectful, and Dignified Environment)	How well the Airman selflessly considers others, values diversity, and sets the stage for an environment of dignity and respect; to include promoting a healthy organizational climate
	Air Force Core Values	How well the Airman adopts, internalizes, and demonstrates Air Force Core Values of Integrity First, Service Before Self, and Excellence in All We Do.
	Personal and Professional Development	Amount of effort the Airman devoted to improve themselves and their work center/ unit through education and involvement.
	Esprit de Corps and Community Relations	How well the Airman promotes camaraderie, embraces esprit de corps, and acts as an Air Force ambassador.
Spencer and Spencer (1993)	Analytical Thinking	Understanding a situation by breaking it apart into smaller pieces, or tracing the implications of a situation in a step-by-step causal way.
	Conceptual Thinking	Understanding a situation or problem by putting the pieces together, seeing the larger picture.
	Information Seeking	Making an effort to get more information; not accepting situations at "face value."
	Flexibility	Ability to adapt to and work effectively with a variety of situations, individuals, or group.
	Self Control	Ability to keep emotions under control and to restrain negative emotions when tempted, when faced with opposition or hostility from others, or when working under conditions of stress.
	Initiative	A preference for taking action; doing more than is required or expected on the job, doing things that no one has requested which will improve or enhance job results and avoid problems, or finding and creating new opportunities.
	Concern for Order, Quality, and Accuracy	An underlying drive to reduce uncertainty in the surrounding environment.
	Achievement Orientation	A concern for working well or for competing against a standard of excellence.
	Teamwork and Cooperation	A genuine intention to work cooperatively with others, to be part of a team, to work together as opposed to working separately or competitively.
	Developing Others	The intent to teach or to foster the development of one or several other people.
	Team Leadership	Intention to take a role as leader of a team or other group.
	Customer Service Orientation	Focusing efforts on discovering and meeting the customer or client's needs.
	Impact and Influence	An intention to persuade others, convince, influence, or impress others, in order to get them to support the speaker's agenda; or the desire to have a specific impact or effect on others.
	Communication (Listening and Responding)	Ability to hear accurately and understand the unspoken or partly expressed thoughts, feelings, and concerns of others.

Table 2. Item content for evaluation on supervisor survey by survey version.

Source	Items	Competencies/Qualities by Version
Spencer & Spencer scales	178 items	V1: Analytical Thinking; Conceptual Thinking; Information Seeking V2: Flexibility; Self-Control V3: Initiative; Concern for Order, Quality, and Accuracy; Achievement Orientation V4: Teamwork; Develops People; Leadership V5: Impact and Influence; Listening and Responding
AF 931 scales	52 items	V1: Personal and Professional Development V3: Task Knowledge/Proficiency; Initiative/Motivation; Resource Utilization; Comply with Enforce Standards V4: Teamwork V5: Air Force Core Values; Esprit de Corps and Community Relations; Communication

of the Airman Comprehensive Assessment and SNCO promotion criteria. Supervisors rated on a 1–4 Likert scale the extent to which the identified high potential member consistently and effectively demonstrated each individual behavior (1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Somewhat Agree, 4 = Strongly Agree). Supervisors could alternately indicate "Unknown/No opportunity to observe" or indicate if the ratee had had "No opportunity to perform" a given behavior in their job role. Each supervisor rated the identified high potential member before providing ratings on the lower potential member.

Results

Scale reliability

As shown in Table 3, all AF 931 scales demonstrated appropriate internal consistency in ratings of lower potential members (α ranging from .79 for Compliance to .90 for Core Values, Task Knowledge, and Teamwork scales). Although alpha coefficients were somewhat higher for the Spencer and Spencer (1993) scales (ranging from .91 for Communication to .97 for Developing People), this reflected the larger number of items on the Spencer and Spencer scales. Mean item-total correlations ranged from .62 to .77 for the AF 931 scales, and ranged from .60 to .80 for the Spencer and Spencer scales. Internal reliability estimates were substantially lower in ratings of high potential members given substantial range restriction, ranging from .67 to .87 for AF 931 scales and .59 to .96 for Spencer and Spencer scales.

Table 3. Reliability of ACA and Spencer & Spencer scales (E2-E6 ratees).

		Lower Potential		High Potential	
	Items per scale	α	Mean item-total *r*	α	Mean item-total *r*
AF 931 Communication	8	.88	.65	.84	.59
AF 931 Core Values	5	.90	.75	.87	.70
AF 931 Esprit de Corps	3	.83	.69	.84	.70
AF 931 Teamwork	5	.82	.62	.79	.56
AF 931 Resource Utilization	3	.83	.69	.77	.61
AF 931 Task Knowledge	5	.90	.76	.84	.66
AF 931 Compliance	3	.79	.64	.67	.48
AF 931 Personal Development	3	.88	.77	.84	.71
AF 931 Initiative	8	.90	.68	.79	.51
S&S Analytical Thinking	14	.95	.73	.90	.59
S&S Conceptual Thinking	13	.96	.78	.92	.66
S&S Information Seeking	8	.93	.75	.86	.62
S&S Flexibility	7	.94	.80	.87	.65
S&S Self Control	6	.88	.69	.74	.52
S&S Initiative	6	.82	.60	.59	.37
S&S Order, Quality, and Accuracy	10	.94	.76	.92	.70
S&S Achievement Orientation	8	.93	.76	.83	.57
S&S Teamwork	12	.94	.72	.88	.63
S&S Developing Others	18	.97	.80	.96	.74
S&S Leadership	14	.96	.77	.96	.77
S&S Service Orientation	9	.94	.77	.94	.76
S&S Influence	20	.95	.69	.89	.51
S&S Communication	7	.91	.73	.89	.69

Note. Ns = 119–548. Internal reliability could improve minimally (by .01 to .03) with item deletion for AF 931 Personal Development and Spencer & Spencer scales of Self-Control, Initiative, Teamwork, and Communication (Listening and Responding).

Criterion-related validity

As shown in Table 4, all AF 931 and Spencer and Spencer scales effectively differentiated high and low potential members, but the magnitude of the difference varied greatly by scale. Of the nine AF 931 scales, Task Knowledge, Initiative, and Compliance most strongly differentiated high from lower potential members (Cohen's *d*s = 4.02 to 4.28), while Esprit de Corps was a far weaker differentiator (*d* = 2.52). As shown in Table 5, this pattern of findings was consistent across ranks (E-4 and below, E-5, and E-6). (Results by behavioral item are available from the authors upon request).

Overall, the Spencer and Spencer scales demonstrated lower validity than the AF 931 scales. Mean differences between high and lower potential members exceeded four standard deviations for three AF 931 scales, but none of the Spencer and Spencer scales; mean differences exceeded 3 standard deviations for 8 of 9 AF 931 scales, but only 8 of 14 Spencer and Spencer scales. Notably, for all three competencies that appear in both the AF 931 and in Spencer and Spencer (1993), the AF 931 scales demonstrated greater validity: Initiative (*d* = 4.14 vs. 3.59), Communication (*d* = 3.37 vs. 2.73), and Teamwork (*d* = 3.35 vs. 2.93).

Table 4. Criterion-related validity of Airman Comprehensive Assessment (AF 931) and Spencer and Spencer (1993) scales (E2-E6 ratees).

	Lower Potential		High Potential		
	Mean	*SD*	Mean	*SD*	Cohen's *d*
AF 931 Task Knowledge	1.62	0.63	3.79	0.35	4.28
AF 931 Initiative	1.66	0.58	3.63	0.34	4.14
AF931 Compliance	1.58	0.61	3.69	0.43	4.02
AF 931 Personal Development	1.55	0.62	3.62	0.49	3.70
AF 931 Resource Utilization	1.63	0.67	3.63	0.45	3.52
AF 931 Communication	1.91	0.62	3.63	0.38	3.37
AF 931 Teamwork	1.83	0.59	3.62	0.44	3.35
AF 931 Core Values	1.83	0.69	3.71	0.42	3.29
AF 931 Esprit de Corps	1.91	0.77	3.56	0.59	2.52
S&S Achievement Orientation	1.65	0.60	3.53	0.41	3.64
S&S Initiative	1.61	0.56	3.42	0.44	3.59
S&S Leadership	1.78	0.63	3.64	0.46	3.40
S&S Analytical Thinking	1.75	0.61	3.50	0.39	3.40
S&S Information Seeking	1.76	0.63	3.54	0.41	3.39
S&S Service Orientation	1.85	0.67	3.72	0.43	3.32
S&S Conceptual Thinking	1.78	0.65	3.51	0.40	3.24
S&S Order, Quality, and Accuracy	1.73	0.62	3.51	0.48	3.20
S&S Developing Others	1.83	0.64	3.51	0.47	3.00
S&S Flexibility	1.86	0.72	3.57	0.41	2.94
S&S Teamwork	2.04	0.65	3.61	0.40	2.93
S&S Communication	1.98	0.69	3.56	0.45	2.73
S&S Self Control	2.02	0.70	3.28	0.52	2.03
S&S Influence	2.12	0.64	2.85	0.55	1.23

Note. 307–587 raters (survey respondents) per competency scale.

Table 5. Criterion-related validity of Airman Comprehensive Assessment (AF 931) and Spencer and Spencer (1993) scales, by ratee rank.

	E-4 and below	E-5	E-6
Competency	Cohen's *d*	Cohen's *d*	Cohen's *d*
AF 931 Task Knowledge	4.29 (#1)	4.94 (#1)	3.53 (#3)
AF931 Compliance	4.10 (#2)	4.17 (#3)	3.65 (#2)
AF 931 Initiative	4.07 (#3)	4.43 (#2)	3.93 (#1)
AF 931 Personal Development	3.81 (#4)	3.80 (#4)	3.40 (#4)
AF 931 Resource Utilization	3.77 (#5)	3.57 (#5)	3.03 (#7)
AF 931 Teamwork	3.66 (#6)	3.55 (#6)	2.97 (#8)
AF 931 Core Values	3.59 (#7)	3.14 (#8)	3.13 (#5)
AF 931 Communication	3.53 (#8)	3.36 (#7)	3.12 (#6)
AF 931 Esprit de Corps	2.66 (#9)	2.18 (#9)	2.35 (#9)
S&S Achievement Orientation	3.60 (#1)	3.83 (#2)	3.47 (#2)
S&S Initiative	3.57 (#2)	3.91 (#1)	3.29 (#6)
S&S Analytical Thinking	3.46 (#3)	3.40 (#6)	3.36 (#4)
S&S Information Seeking	3.42 (#4)	3.35 (#7)	3.42 (#3)
S&S Conceptual Thinking	3.26 (#5)	3.18 (#9)	3.29 (#5)
S&S Order, Quality, and Accuracy	3.22 (#6)	3.47 (#5)	2.90 (#9)
S&S Service Orientation	3.16 (#7)	3.64 (#3)	3.15 (#7)
S&S Leadership	3.14 (#8)	3.53 (#4)	3.75 (#1)
S&S Communication	3.09 (#9)	2.53 (#12)	2.39 (#12)
S&S Flexibility	2.93 (#10)	3.07 (#11)	2.73 (#11)
S&S Developing Others	2.91 (#11)	3.18 (#10)	2.93 (#8)
S&S Teamwork	2.85 (#12)	3.21 (#8)	2.77 (#10)
S&S Self Control	2.07 (#13)	2.02 (#13)	1.98 (#13)
S&S Influence	1.11 (#14)	1.34 (#14)	1.32 (#14)

Note. 136–285 raters for E-4 and below; 99–187 raters for E-5; 62–124 raters for E-6.

Of the 14 Spencer and Spencer (1993) scales, Achievement Orientation, Initiative, Leadership, and Analytical Thinking most strongly differentiated high from lower potential members (*d*s = 3.40 to 3.64), while Influence and Self Control were the weakest differentiators (*d* = 1.23 and 2.03, respectively). Although findings for the Spencer and Spencer (1993) scales were generally consistent across ranks, the competency of Leadership was a notable exception. As shown in Table 5, Leadership was the strongest differentiator of career potential for E-6s, but was a somewhat lesser differentiator for lower ranks.

Scale calibration

To determine if scale anchors are appropriately calibrated, we considered two criteria. First, positive behaviors associated with higher scale anchors ought to be performed by relatively fewer low potential members than behaviors associated with lower scale anchors. This is similar to the assumption one would apply when evaluating items for inclusion on a unidimensional aptitude or knowledge test: lower aptitude/less knowledgeable individuals should answer more of the easy (less complex) items correctly than they do the difficult (more complex) items. Second, positive behaviors associated with higher scale anchors (items intended to be more complex or more difficult) ought to more strongly differentiate high from low potential members than do the behaviors associated with lower scale anchors.

Overall, both conditions indicating proper calibration were fully met (i.e., for all scale anchors) for three of the nine AF 931 scales: Resource Utilization, Personal Development, and Teamwork. For example, the AF 931 Resource Utilization BARS defines three increasing levels of demonstrated proficiency: (1) "Makes good use

of available time and other resources within Airman's control," (2) "Seeks better ways to more effectively utilize time and resources," and (representing the highest level of Resource Utilization) (3) "Sought after as an utilization expert in saving time, equipment, manpower, and budget with impact outside of work center or unit." As shown in the Appendix, lower potential members were more likely to "Make good use of available time and other resources … " (16.9% of lower potential members) than they were to "Seek better ways to more effectively utilize time and resources" (13.2% of lower potential members). In turn, lower potential members were more likely to "Seek better ways to more effectively utilize time and resources" than they were to demonstrate the highest scale anchor "Sought after as an utilization expert … " (only 9.4% of lower potential members). Additionally, there was an increasing level of differentiation between high and low potential members at higher scale anchors.

For other AF 931 scales, these calibration criteria would recommend some modifications – either deletion of certain behaviors (where a given scale anchor is defined by multiple behavioral statements) or re-aligning certain behaviors to define higher or lower level of demonstrated competency proficiency. For example, on the AF 931 "Actively participates; organizes and occasionally leads team building and/or community events to foster esprit de corps" was intended to represent a higher level of Esprit de Corps than "Fosters esprit de corps through involvement in base and/or community events." However, results showed that a greater percentage of lower potential members "Actively participate … " (32.1% of low potential members) than "Foster esprit de corps … " (28.5% of low potential members). Additionally, "Foster[ing] esprit de corps … " was a greater differentiator of high vs. lower potential members than "Actively participat[ing] … " As such, results would suggest the need for swapping the behaviors associated with these two scale anchors. See Appendix.

Notably, even the published Spencer and Spencer (1993) scales, which were rigorously developed using best practices, including the use of behavioral event interviews (modified critical incidents technique), and Q-sorting of behaviors by multiple researchers, showed some potential for improved calibration based on this criteria. See Appendix.

Discussion

Although the AF 931 BARS were developed without the types of rigorous, empirically driven process that is normally recommended as best practice (c.f., Klieger et al., 2018; Smith & Kendall, 1963), the current study generally supports the reliability and validity of these scales. Enlisted members who Air Force supervisors identified as having a "high potential for success in an Air Force career" were rated as more consistently and effectively demonstrating behaviors associated with the nine competencies included on the AF 931 than members of the same rank and career field with "lower potential." Given the assessment's focus on broader competencies, rather than specific job tasks or technical skills (e.g., proficiency with specific equipment) we would anticipate the overall validity of these scales to be relatively stable over time.

Certain AF 931 competencies were far stronger differentiators of career potential than others, however. The AF 931-rated competencies of Initiative, Task Knowledge, and Compliance most strongly differentiated career potential; Esprit de Corps showed the least differentiation. This may be problematic to the extent that all of these competencies may be presumed to be of equal importance to ratees by virtue of their inclusion on the same appraisal instrument (without weights as to relative importance). To the extent that supervisors may be limited in the time available to conduct performance appraisal feedback sessions, focusing discussion on those competencies that are the strongest differentiators of success may be most fruitful. Considerations of relative competency importance may be especially important in performance appraisal systems like those used by the Navy (i.e., NAVPERS 1610/2) that compute average scores across competencies on performance evaluation reports (for promotion consideration) based on equal weights.

One key question addressed in this study was the extent to which published, rigorously developed competency scales intended for broad use across civilian organizations (Spencer & Spencer, 1993) would demonstrate validity in the Air Force context. Our results generally demonstrated that these more generic competency scales were effective differentiators of supervisor-rated potential for Air Force career success. However, overall results suggest that the competency behaviors included on the AF 931 are stronger differentiators of career potential than these generic scales. Despite intermixed item content and no reference to the AF 931 in communication with study participants, we found that the behaviors included on the AF 931 Initiative, Communication, and Teamwork scales all better differentiated ratings of Air Force career potential than the Initiative, Communication, and Teamwork scales developed by Spencer and Spencer for use across organizations.

Perhaps more importantly, this study also suggests that some of the strongest differentiators of Air Force career success may not be captured by the current AF 910 and AF 931. Notably, many of the generic Spencer & Spencer scales demonstrated greater validity than at least some AF 931 competency measures. Among these, Leadership and Analytical Thinking stand out as competencies that are quite conceptually distinct from those assessed on the AF 931, but which were stronger differentiators of Air Force career potential than nearly half (4 of 9) of the AF 931 competencies. Future revisions to the AF 931 ought to incorporate Leadership and Analytical Thinking competencies in some form, potentially prioritizing these over current AF 931 competencies that are lesser differentiators of career potential (e.g., Esprit de Corps).

Lastly, despite overall appropriate reliability and validity of the AF 931 BARS, we identified several opportunities to improve calibration of the specific behavioral scale anchors, using a logical, but, to our knowledge, novel methodology to help ensure that behavioral scale anchors truly represented a "*clear progression from lower to higher levels*" on each competency based on the relative percentage of high vs. low potential members that effectively demonstrate each behavior. Notably even the rigorously developed, and widely cited, BARS published in Spencer and Spencer (1993) showed some potential need for recalibration of behaviors to represent relatively higher or lower levels of demonstrated competency. While six of nine AF 931 scales would require recalibration of at least one behavior, all 14 of the Spencer and Spencer scales included in this study would require some recalibration based on this methodology.

Enduring utility of BARS for performance assessment

It has been over forty years since Landy and Farr (1980) called for a research moratorium on performance appraisal rating formats. Their paper (which by now has been cited over 2,000 times) famously reached some rather discouraging conclusions about the utility of BARS:

> "We know that the rater should have a clear understanding of the rating task, that the anchors on the scale should be rigorously developed, and that those anchors should be more than simple descriptive labels such as poor, average, outstanding, and so forth. Nevertheless, even when all of these suggestions are taken into account, evidence suggests that their effect may be minimal. Data reviewed . . . indicate that about 4% – 8% of the variance in ratings can be explained on the basis of format. When one considers that the designs that yielded these estimates were seldom sensitive enough to identify true levels of performance, even this effect may be an overestimate." (p. 43)

Although not all researchers have heeded Landy and Farr's (1980) call for a research moratorium, to our knowledge, there has not been empirical evidence gathered since then that would reach a fundamentally different conclusion (see, Debnath et al., 2015 for a more recent review). Rather, much BARS research since then has shifted from a focus on strictly psychometric issues (e.g., method variance and halo) to consider broader advantages of BARS relative to other performance rating formats (Debnath et al., 2015). Such advantages include improved face validity, more positive employee reactions (MacDonald & Sulsky, 2009), and greater employee acceptance of BARS performance feedback (Krein, 1990). Others have noted the greater legal defensibility of performance appraisal methods like BARS that focus on behaviors rather traits and avoid global assessment (Malos, 1998), and provide "informational justice" in explaining personnel decisions (Debnath et al., 2015).

Perhaps more importantly though is the potential role of BARS in employee development. In that light we note that there is another common term often used to refer to Behaviorally Anchored Rating Scales in educational and training settings: a rubric (though certainly not all rubrics are behaviorally anchored, many are). For example, rather than evaluate a student's presentation by simply assigning a letter grade or its numeric equivalent, instructors are often encouraged to use a rubric that behaviorally defines poor, average, and outstanding performance on each rated area (e.g., eye contact, use of visual aids, body language, etc.). While such rubrics undoubtedly make grading more efficient for instructors, and help provide justification in case students dispute their grade, when provided to students they also have educational or developmental value in clearly communicating performance standards (Andrade, 2000). Beyond assessing individual tendencies, formal performance appraisal forms like the AF 931 provide an important opportunity to communicate what the organization values and expects in its members (see, Carr, 2015, March 11). Ultimately, by making criteria explicit, BARS can be thought of inherently providing formative feedback (see, Taras, 2005). That is, specific, carefully defined behavioral anchors can educate employees on what constitutes effective performance and by extension how they can improve. Careful attention to how each

competency is behaviorally defined allows for consistent, accurate feedback on how a member's behaviors need to change to improve their performance and likelihood of career success.

Addressing the primary disadvantage of BARS

Echoing Landy and Farr (1980), even the biggest proponents of BARS (e.g., Debnath et al., 2015) note the high cost (SME time and effort) associated with rigorous BARS development. Indeed, use of the traditional critical incidents methodology to develop BARS requires training SMEs to write a large number of detailed behavioral examples, grouping and editing 100s of qualitative behavioral examples, and recruiting two additional independent SME samples to complete subsequent retranslation and effectiveness rating tasks. As an alternative, the quantitative survey-based criterion-sampling methodology used here demonstrates a potential means to achieve the benefits of BARS (c.f., Klieger et al., 2018) with a substantially less time-consuming and labor intensive process.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, *57*(5), 13–18. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.452.5684&rep=rep1&type=pdf

Assistant Secretary of the Air Force for Manpower and Reserve Affairs. (2021, January 13). Officer and enlisted evaluations systems (Air Force Instruction 36-2406). https://static.e-publishing.af.mil/production/1/af_a1/publication/afi36-2406/afi36-2406.pdf

Barelka, A., Barron, L., Coggins, M., Hernandez, S., & Kulpa, P. (2019). *Development and validation of Air Force Competency Model* (Technical Report No. 1). HQ AETC A3/6.

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored ratings scales (BARS). *Journal of Applied Psychology*, *66*(4), 458–463. https://doi.org/10.1037/0021-9010.66.4.458

Bersoff, D. N. (1988). Should subjective employment devices be scrutinized? It's elementary, my dear Ms. Watson. *American Psychologist*, *43*(12), 1016–1018. https://doi.org/10.1037/0003-066X.43.12.1016

Borman, W. C., & Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. *Journal of Applied Psychology*, *60*(5), 561–565. https://doi.org/10.1037/0021-9010.60.5.561

Brannick, M. T., Levine, E. L., & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management*. Sage.

Carr, J. (2015, March 11). *How to: The Airman Comprehensive Assessment*. Air Force News Service. https://www.af.mil/News/Article-Display/Article/580200/how-to-the-airman-comprehensive-assessment/

Carretta, T. R., & Walters, L. C. (1991). *The development of behaviorally anchored ratings scales (BARS) for evaluating USAF pilot training performance* (pp. AD–A239 969). Air Force Systems Command. https://apps.dtic.mil/dtic/tr/fulltext/u2/a239969.pdf

Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes behaviorally anchored rating scales so perennial as an appraisal method? *International Journal of Business and Social Science*, *6*(2), 16–25. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1071.5824&rep=rep1&type=pdf

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327–358. https://doi.org/10.1037/h0061470

Guion, R. M., & Highhouse, S. (2006). *Essentials of personnel assessment and selection*. Erlbaum.

Gutman, A. (2004). Ground rules for adverse impact. *The Industrial/Organizational Psychologist*, *41*(3), 109–119. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.489.7073&rep=rep1&type=pdf

Gutman, A., Koppes, L. L., & Vodanovich, S. J. (2011). *EEO law and personnel practices* (3rd ed.). Taylor & Francis.

Hom, P. W., DeNisi, A. S., Kinicki, A. J., & Bannister, B. D. (1982). Effectiveness of performance feedback from behaviorally anchored rating scales. *Journal of Applied Psychology*, *67*(5), 568–576. https://doi.org/10.1037/0021-9010.67.5.568

Kell, H. J., Martin-Raugh, M. P., Carney, L. M., Inglese, P. A., Chen, L., & Feng, G. (2017). *Exploring methods for developing behaviorally anchored rating scales for evaluating structured interview performance*. ETS Research Report Series (RR-17-28). ETS. https://onlinelibrary.wiley.com/doi/full/10.1002/ets2.12152

Keller, K. M., Robson, S., O'Neill, K., Emslie, P., Burgette, L. F., Harrington, L. M., & Curran, D. (2014). *Promoting airmen with the potential to lead: A study of the air force master sergeant promotion system*. RAND Corporation. https://apps.dtic.mil/sti/pdfs/ADA611614.pdf

Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). *Development of the behaviorally anchored rating scales for the Skills Demonstration and Progression Guide*. ETS Research Report Series (RR-18-24). ETS. https://doi.org/10.1002/ets2.12210

Krein, T. J. (1990). Performance reviews that rate an A. *Personnel*, *67*(5), 38–40.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, *87*(1), 72–107. https://doi.org/10.1037/0033-2909.87.1.72

MacDonald, H., & Sulsky, L. M. (2009). Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science*, *41*(4), 227–240. https://doi.org/10.1037/a0015165

Malos, S. B. (1998). Current legal issues in performance appraisal. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 49–94). Jossey-Bass.

Martin-Raugh, M., Tannenbaum, R. J., Tocci, C. M., & Reese, C. (2016). Behaviorally anchored rating scales: An application for evaluating teaching practice. *Teaching and Teacher Education*, *59*, 414–419. https://doi.org/10.1016/j.tate.2016.07.026

Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory* (3rd ed ed.). Erlbaum.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, *47* (2), 149–155. https://doi.org/10.1037/h0047060

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. Wiley.

Taras, M. (2005). Assessment – Summative and formative – Some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Appendix. Examples of appropriately calibrated ACA BARS

Resource Utilization	Current AF-931 BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
Sought after as an utilization expert in saving time, equipment, manpower, and budget with impact outside of work center or unit.	3 *(Most Proficient)*	94.5%	9.4%	85.1%
Seeks better ways to more effectively utilize time and other resources.	2	98.2%	13.2%	85.0%
Makes good use of available time and other resources within Airman's control.	1 *(Least Proficient)*	98.6%	16.9%	81.6%

Personal Development	Current BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
Relentlessly pursues personal and professional development of themselves and others; efforts result in significant positive impact to unit and/or Air Force.	3 *(Most Proficient)*	97.1%	6.6%	90.6%
As a driven Airman, exceeds both professional and personal development goals with positive impact on individual performance or mission impact.	2	98.3%	8.3%	90.0%
Establishes goals and progresses to meet those goals for personal and/or professional development.	1 *(Least Proficient)*	97.1%	16.4%	80.7%

Appendix. Examples of ACA scales for potential recalibration (compliance, esprit de corps)

Compliance	Current BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
As the model Airman, raises the standard in all areas for others to emulate; coaches others.	3 *(Most Proficient)*	97.7%	5.2%	92.6%
Consistently meets all standards.	1 *(Least Proficient)*	97.0%	15.5%	81.5%
Meets all/surpassed some standards of fitness, conduct, appearance, and behavior; influences others by example.	2	95.2%	18.7%	76.5%

Esprit de Corps	Current BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
Epitomizes an Air Force ambassador; consistently and selflessly leads efforts that inspire esprit de corps with significant impact to the mission and/or community.	3 *(Most Proficient)*	94.0%	16.0%	78.0%
Fosters esprit de corps through involvement in base and/or community events.	1 *(Least Proficient)*	92.1%	28.5%	63.6%
Actively participates; organizes and occasionally leads team building and/or community events to foster esprit de corps.	2	93.9%	32.1%	61.8%

Appendix. Examples of Spencer & Spencer (1993) scales for potential recalibration

Analytical Thinking	Current BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
Organizes, sequences, and analyzes extremely complex interdependent systems.	6 *(Most Proficient)*	92.9%	9.8%	83.2%
Uses several analytical techniques to break apart complex problems.	4	94.3%	10.2%	84.2%
Uses several analytical techniques to identify several solutions and weighs the value of each.	5	93.9%	10.2%	83.8%
Systematically breaks down multidimensional problems or processes into component parts.	5	95.9%	10.6%	85.3%
Generally anticipates obstacles and thinks ahead about next steps.	3	97.0%	13.0%	84.0%
Makes long chains of causal connections.	4	87.8%	14.5%	73.3%
Systematically breaks down a complex problem or process into component parts.	4	97.4%	16.0%	81.4%
Recognizes several likely causes of events or several consequences of actions.	3	97.4%	17.8%	79.7%
Breaks down a complex task into manageable parts in a systematic way.	3	97.7%	18.6%	79.1%
Analyzes relationships among several parts of a problem or situation.	3	96.5%	21.0%	75.5%
Breaks problems into simple lists of tasks or activities.	1 *(Least Proficient)*	98.1%	27.2%	70.8%
Analyzes relationships among a few parts of a problem or situation.	1 *(Least Proficient)*	90.8%	30.1%	60.7%
Sets priorities for tasks in order of importance.	2	98.7%	31.5%	67.1%
Makes simple causal links (A causes B) or pro-and-con decisions.	2	95.5%	36.5%	59.1%

Leadership	Current BARS Level	% High Potential Members Engage in Behavior (If Known/Observable)	% Low Potential Members Engage in Behavior (If Known/Observable)	Difference Δ
Uses complex strategies to promote team morale and productivity (e.g., team assignments, cross-training).	4	88.30%	5.50%	82.80%
Communicates a compelling vision that generates excitement, enthusiasm, and commitment to the group mission.	7 *(Most Proficient)*	94.20%	8.20%	86.00%
Inspires and motivates through interactions.	7 *(Most Proficient)*	96.00%	9.90%	86.10%
Ensures that others buy into leader's mission, goals, agenda, climate, tone, and policies.	6	93.20%	11.00%	82.20%
Sets a good example (e.g., models desired behavior).	0 *(Least Proficient)*	98.50%	11.70%	86.80%
Protects the organization and its reputation vis-a-vis the larger organization or the community at large.	5	96.20%	16.40%	79.80%
Manages meetings, states agenda and objectives, controls time, and makes assignments.	1	96.10%	16.40%	79.70%
Lets people who are affected by decisions know what is happening.	2	97.00%	18.60%	78.40%
Obtains needed personnel, resources, and information for the organization.	5	96.00%	19.60%	76.40%
Makes sure group members have all the necessary information.	2	97.80%	19.90%	77.80%
Explains reasons for decisions.	2	94.70%	21.20%	73.50%
Uses formal authority and power in a fair and equitable manner.	3	97.10%	22.20%	74.90%
Makes sure the practical needs of the organization are met.	5	98.40%	32.20%	66.20%
Makes a personal effort to treat all group members fairly.	3	97.80%	33.60%	64.30%
Ensures that organizational tasks are completed.	6	97.10%	33.80%	63.40%

Appendix. Recommended modifications to ACA scales

Communication	
3 (*Highest Level)*	Sought out by leaders for various communication forums; Remarkable communicator, mentor, and teacher.
2	Encourages and considers others' input; Clearly conveys complex information in a concise manner; Improves communication skills in themselves and others.
1	Conveys most information in an understandable manner; Makes some effort to improve communication skills.
0 *(Lowest Level)*	Inarticulate; does not assimilate or convey.
Core Values	
3 (*Highest Level)*	As an influential leader, inspires others to embody the Air Force Core Values.
2	Embodies the Air Force Core Values of Integrity, Service Before Self, and Excellence; **Demonstrates personal conduct that exudes Air Force Core Values for others to emulate.**
1	Consistently demonstrates the Air Force Core Values, both on and off duty; **Encourages others to uphold Air Force Core Values.**
0 *(Lowest Level)*	Fails to adhere to the Air Force Core Values.
Espirit de Corps	
3 (*Highest Level)*	Epitomizes an Air Force ambassador; consistently and selflessly leads efforts that inspire esprit de corps with significant impact to the mission and/or community.
2	**Fosters esprit de corps through involvement in base and/or community events.**
1	**Actively participates; organizes and occasionally leads team building and/or community events to foster esprit de corps**.
0 *(Lowest Level)*	Makes little or no effort to promote esprit de corps or act as an Air Force ambassador.
Task Knowledge	
3 (*Highest Level)*	Knowledge and skills impact far beyond those of peers; efforts directly elevate unit's impact on mission success; **Exceeds performance expectations associated with current grade.**
2	Routinely delivers high-quality work; .
1	Demonstrates acceptable ability and consistently produces good quality, quantity, results, and impact; **Meets established suspenses.**
0 *(Lowest Level)*	Demonstrates insufficient ability; requires reaccomplishment of tasks and more guidance or experience.
Compliance	
3 (*Highest Level)*	As the model Airman, raises the standard in all areas for others to emulate; coaches others.
2	**Consistently meets all standards.**
1	**Meets all/surpassed some standards of fitness, conduct, appearance, and behavior; influences others by example.**
0 *(Lowest Level)*	Fails to meet some or all standards.
Initiative	
3 (*Highest Level)*	Drives innovative environments; **Routinely seeks out new ways to execute mission; Self-starter on task completion.**
2	Demonstrates an inspired work ethic; **Aggressively seeks to improve others' motivation; Mindful of others' needs; develops new processes.**
1	Displays good effort in performance of assigned tasks; Mindful of others' needs, develops new processes; **Proactively assists team members.**
0 *(Lowest Level)*	Displays little or no effort in accomplishing duties, lacks motivation, and does not display initiative.

Notes. Bolded behavioral statements recommended for re-alignment based on relatively higher or lower level of competency proficiency than represented on current AF 931. Recalibration recommendations are based on the proportion of lower-potential members performing and increasing differentiation between high and low potential members (must meet both criteria to recommend recalibration).

Appendix. Recommended modifications to Spencer & Spencer scales

Analytical Thinking	
6 *(Highest Level)*	Organizes, sequences, and analyzes extremely complex interdependent systems; **Uses several analytical techniques to break apart complex problems.**
5	Uses several analytical techniques to identify several solutions and weighs the value of each; Systematically breaks down multidimensional problems or processes into component parts.
4	Makes long chains of causal connections; Systematically breaks down a complex problem or process into component parts; **Generally anticipates obstacles and thinks ahead about next steps.**
3	Recognizes several likely causes of events or several consequences of actions; Breaks down a complex task into manageable parts in a systematic way; Analyzes relationships among several parts of a problem or situation.
2	Sets priorities for tasks in order of importance;; **Breaks problems into simple lists of tasks or activities.**
1	Analyzes relationships among a few parts of a problem or situation; **Makes simple causal links (A causes B) or pro-and-con decisions.**
0 *(Lowest Level)*	Does each thing as it comes up; responds to immediate needs or requests.

Conceptual Thinking	
7 *(Highest Level)*	Creates new models or theories that explain a complex situation or problem and reconciles discrepant data.
6	Formulates a useful explanation for complex problems, situations, or opportunities; Generates and tests multiple concepts, hypotheses, or explanations for a given situation; **Identifies problems and situations not obvious to others and not learned from education or experience.**
5	
4	Pulls together ideas, issues, and observations into single concept or a clear presentation. **Identifies useful relationships among complex data from unrelated areas.**
3	Applies and modifies complex learned concepts or methods appropriately; Applies complex concepts (e.g., "root-cause analysis," "portfolio analysis," "natural selection").
2	Sees crucial differences between current situation and things that have happened before; Observes discrepancies, trends, and interrelationships in data; **Uses "rules of thumb," common sense, and past experiences to identify problems or situations.**
1	**Applies knowledge of past discrepancies, trends, and relationships to look at different situations.**
0 *(Lowest Level)*	Thinks very concretely; **Sees essential similarities between current and past situations.**

Information Seeking	
7 *(Highest Level)*	**Personally establishes ongoing systems or habits for various kinds of information gathering.**
6	**Gets out personally to see problem situations, questions those closest to the problem when others might ignore these people.**
5	**Involves others who would not normally be involved and gets them to seek out information.**
4	Calls on others, who are not personally involved, to get their perspective, background information, or experience; **Does formal research through review of written materials.**
3	Asks a series of probing questions to get at the root of a situation or problem, below the surface presentation.
2	
1	Asks direct questions of people directly involved in the situation; Consults available resources.
0 *(Lowest Level)*	Does not seek out additional information about a situation, other than what has been given.

Flexibility	
6 *(Highest Level)*	Makes large or long-term adaptations in own or partnering organization in response to the needs of the situation.
5	Makes smaller or short-term adaptations in own or partnering organization in response to the needs of the situation; **Changes own behavior or approach to suit the situation.**
4	Adapts own strategies, goals, or projects to situations.
3	
2	Pinch-hits by doing coworkers' tasks as necessary during an emergency; Depending on the individual situation, adapts actions to accomplish organization's larger objectives.
1	Recognizes the validity of others' viewpoints.
0	Always follows procedures.
−1 *(Lowest Level)*	Despite obvious problems, retains same point of view; does not recognize others' views as valid.

Self-Control	
6 *(Highest Level)*	In very stressful situations, calms others and controls own emotions; **Uses stress management techniques to control responses, prevent burnout, and deal with ongoing stresses effectively.**
5	Controls strong emotions or other stress and takes action to respond constructively to the source of the problems.
4	
3	Controls strong emotions, such as anger or extreme frustration, and continues discussions or other processes fairly calmly; **Resists the temptation to engage in inappropriate involvements or impulsive behavior.**
2	Controls strong emotions, such as anger, extreme frustration, or stress, but does not take constructive action to address the source of the problems.
1	
0	Avoids people or situations that provoke negative emotions.
−1 *(Lowest Level)*	"Burns out" or breaks down under stress; Expresses feelings, such as frustration and/or other negative emotions, inappropriately; Loses control; own emotions interfere with work effectiveness; Gets inappropriately involved personally with subordinates, peers, or others.

Initiative	
6 *(Highest Level)*	**Exceeds job description; starts and carries through new projects.**
5	**Exceeds job description (e.g., takes on extra tasks).**
4	**Works extra hours, nights, and weekends, as needed, to complete work when not required to do so.**
3	**Gets others involved in unusual extra efforts (e.g., enlists the help of family, coworkers, community members).**
2	**Completes assignments without constant supervision.**
1	**Acts without formal authority, takes personal risks, and bends the rules to get the job done to meet the needs of the job.**
0	
−1 *(Lowest Level)*	Shirks or tries to get out of required work; **Requires constant supervision.**

Precision	
7 *(Highest Level)*	Deduces new needs (not having to do with order) from perceived disorder; Puts new, detailed, complex systems in place to increase order and improve quality of data.
6	
5	Monitors data, discovers weaknesses or missing data, and seeks out information to keep order; has a general concern for increasing order in existing systems; **Works for clarity; seeks roles, expectations, tasks, and data that are crystal clear, preferably in writing.**
4	Keeps clear, detailed records of own or others' activities; Monitors quality of others' work; checks to ensure that procedures are followed.
3	Double-checks accuracy of information and own work.
2	**Monitors progress of a project against milestones or deadlines; Develops and uses systems to organize and keep track of information.**
1	Maintains an orderly workspace (e.g., desk, files, and tools are in good order).
0	Lacks concern with order, but it does not cause problems.
−1 *(Lowest Level)*	Lacks concern with order, despite problems caused by disorder.

Achievement Orientation	
7 *(Highest Level)*	Commits significant resources and/or time in the face of uncertainty to improve performance, try something new, reach a challenging goal, while also taking action to minimize the risks involved.
6	Makes decisions, sets priorities, or chooses goals on the basis of inputs and outputs: makes explicit considerations of potential return on investment or cost-benefit analysis; **Encourages and supports subordinates in taking entrepreneurial risks.**
5	Sets and acts to reach challenging goals for self or others (e.g., to improve productivity by 15% in 6 months).
4	Makes specific changes in the system or in own work methods to improve performance without setting any specific goal.
3	Uses his or her own specific methods of measuring outcomes against a standard of excellence not imposed by management (e.g., time or money spent, outperforming others).
2	
1	Works toward implicit standards of excellence; tries to do the job well or right; **Works to meet a standard set by management (e.g., manages to a budget, quotas, quality requirements).**
0	Works hard, but gives no evidence of a standard of excellence for work outputs.
−1 *(Lowest Level)*	Shows no special concern with work; does only what is required.

Teamwork	
7 *(Highest Level)*	**Publicly credits others who have performed well; Encourages and empowers others; makes them feel strong or important.**
6	**Keeps people informed and up to date about the group process; shares all relevant or useful information.**
5	**Speaks of team members in positive terms; Shows respect for others' intelligence by appealing to reason; Expresses positive expectations of others.**
4	Solicits ideas and opinions to help form specific decisions or plans; Invites all members of a group to contribute to a process; **Brings conflict within the team into the open and encourages or facilitates a beneficial resolution of conflicts; Participates willingly, supports team decisions, is a "good team player" who does fair share of the work; Protects and promotes group reputation with outsiders.**
3	**Publicly credits others who have performed well; Encourages and empowers others; makes them feel strong or important.**
2	**Acts to promote a friendly climate, good morale, and cooperation.**
1	**Shows respect for others' intelligence by appealing to reason; Speaks of team members in positive terms.**
0	
−1 *(Lowest Level)*	Disruptive; causes trouble within team; **Neutral, passive, does not participate, or is not a member of any team.**

Develops People	
9 *(Highest Level)*	Gives others rewards for good performance (e.g., arranges promotions or developmental experience).
8	
7	Arranges successful experiences for others to build up their skills and confidence; Identifies a training or developmental need and designs or establishes new programs or materials to meet it; Designs significantly new approaches to teaching traditional materials.
6	Arranges appropriate and helpful assignments, formal training, or other experiences for the purpose of fostering the other person's learning and development.
5	Breaks difficult tasks into smaller components or uses other training strategies.
4	Gives specific positive or mixed feedback for developmental purposes; **After assessing others' competence, delegates full authority and responsibility with the latitude to do a task in their own way, including the opportunity to make and learn from mistakes in a noncritical setting.**
3	Gives directions or demonstrations with reasons or rationale included as a training strategy; Gives practical support or assistance to make others' job easier (e.g., volunteers additional resources, tools, information, or expert advice); Asks questions, gives tests, or uses other methods to verify that others have understood explanation or directions; **Gives negative feedback in behavioral rather than personal terms and expresses positive expectations for future performance; Gives detailed instructions and/or on-the-job demonstrations.**
2	**Makes people work out answers to problems themselves so they really know how, rather than simply giving them the answer; Gives individualized suggestions for improvement.**
1	Makes positive comments regarding others' abilities or potential even in "difficult" cases; **Reassures others after a setback; Tells how to do the task; makes specific helpful suggestions.**
0 *(Lowest Level)*	Makes no explicit efforts to develop others; Focuses on doing his or her own job well, setting a good example.

Leadership	
7 *(Highest Level)*	Inspires and motivates through interactions; Communicates a compelling vision that generates excitement, enthusiasm, and commitment to the group mission.
6	Ensures that others buy into leader's mission, goals, agenda, climate, tone, and policies; **Uses complex strategies to promote team morale and productivity (e.g., team assignments, cross-training); Sets a good example (e.g., models desired behavior).**
5	Protects the organization and its reputation vis-a-vis the larger organization or the community at large.
4	**Manages meetings, states agenda and objectives, controls time, and makes assignments; Lets people who are affected by decisions know what is happening.**
3	Uses formal authority and power in a fair and equitable manner; **Obtains needed personnel, resources, and information for the organization; Makes sure group members have all the necessary information.**
2	Explains reasons for decisions; **Makes sure the practical needs of the organization are met; Makes a personal effort to treat all group members fairly.**
1	**Ensures that organizational tasks are completed.**
0	
−1 *(Lowest Level)*	Refuses or fails to lead (e.g., won't provide direction or mission statements when others need them).

Service Orientation	
5 *(Highest Level)*	**Monitors others' satisfaction; Corrects problems promptly and undefensively.**
4	**Maintains clear communication with others regarding mutual expectations.**
3	**Follows through on others' inquiries, requests, and complaints; keeps others up-to-date about progress.**
2	Distributes helpful information to others; **Makes concrete attempts to add value and make things better for others in some way; Expresses positive expectations about others; Makes self fully available when others are going through a critical period.**
1 *(Lowest Level)*	**Provides friendly, cheerful service.**

Influence	
8 *(Highest Level)*	Structures situations or jobs or changes organizational structure to encourage desired behavior; Uses complex influence strategies tailored to individual situations (e.g., chains of indirect influence – "get A to show B so that B will tell C such-and-such"); **Uses "group process skills" to lead or direct a group; Includes careful preparation of data in presentations.**
7	Uses experts or other third parties to influence; Makes three or more different arguments or points in presentations or discussions**; Uses complex political maneuvering to reach a goal or have an effect.**
6	Anticipates and prepares for other's reactions during discussions and presentations.
5	Models behavior desired in others in order to have a specific impact.
4	Adapts a presentation or discussion to appeal to the interest and level of others; During presentations and discussions, anticipates the effect of an action or other detail on people's image of the speaker.
3	Makes two or more different arguments or points in presentations or discussions**; Takes well thoughtout, unusual, or dramatic actions in order to have a specific impact; Makes complex, staged arguments to persuade others.**
2	Uses direct persuasion in a discussion or presentation (e.g., appeals to reason, data, larger purpose; uses concrete examples, visual aids, demonstrations).
1	
0	Shows no attempt to influence or persuade others; **Assembles political coalitions to influence others, Builds "behind-the-scenes" support for ideas.**
−1 *(Lowest Level)*	Focuses on personal power and position regardless of organizational damage; **Makes no apparent adaptation to the level and interests of the audience; Deliberately gives or withholds information to have specific effects; Expresses intention to have a specific effect or impact, but takes no specific action.**

Communication	
4 *(Highest Level)*	**Actively seeks to understand others (e.g., in order to influence, develop, help, or lead others).**
3	Uses understanding based on listening and observation to predict and prepare for others' reactions.
2	**Uses understanding to explain other's past behavior; Picks up on clues to feelings or meanings, or listens when approached by others; Reflects people's concerns; easy to talk to.**
1	**Makes self available to listen (e.g., has an "open door"); Goes out of the way to initiate conversations.**
0	Makes no attempt to listen to the concerns of others.
−1 *(Lowest Level)*	Offends others; makes them "close up."

Notes. Bolded behavioral statements recommended for re-alignment based on relatively higher or lower level of competency proficiency than represented on current Spencer & Spencer Scales. Recalibration recommendations are based on the proportion of lower-potential members performing and increasing differentiation between high and low potential members (must meet both criteria to recommend recalibration).