



IDEA Technical Report No. 20

**Updated Technical Manual  
for the IDEA Feedback System  
for Administrators**

Stephen L. Benton

Dan Li

July 2018

## Table of Contents

Introduction.....	5
Sample Description .....	6
Response Rate and Its Relationship to Ratings .....	6
Characteristics of Administrators, Units, and Institutions.....	8
Administrator Perceptions of the Unit’s Reputation and Expectations at the Time of Appointment ....	9
Summary.....	10
Analysis of 2013 to 2017 Database .....	10
Administrator Self-Ratings of Performance of Administrative Roles.....	10
Administrator Self-Ratings of Personal Characteristics .....	11
Personnel Ratings of the Administrator’s Performance of Administrative Roles .....	12
Personnel Ratings of Leadership Style and Personal Characteristics.....	13
Summary Judgments of the Administrator.....	14
Validity .....	14
Evidence of Content Validity .....	14
Validity Evidence Based on Internal Structure .....	14
Factor Analysis of Self-Ratings of Administrative Roles .....	15
Factor Analysis of Administrator Self-Ratings of Personal Characteristics .....	15
Factor Analysis of Personnel Ratings of the Administrator’s Performance of Administrative Roles .	16
Factor Analysis of Personnel Ratings of the Administrator’s Personal Characteristics.....	17
Correlations Between Personnel Ratings of Administrative Roles and Summary Judgments.....	18
Correlations Between Personnel Ratings of Interpersonal Characteristics, Leadership Style, and	
Summary Judgments .....	19
Summary .....	19
Validity Evidence Based on Relationships to Other Variables .....	19
Correlations Between Administrator and Unit Characteristics, Factor Composite Scale Scores, and	
Summary Judgments .....	20
Comparison of Summary Judgments by Perceptions of the Unit’s Reputation and Expectations at	
Time of Appointment.....	22
Comparisons of Mean Scale Scores by Perceptions of the Unit’s Reputation and Expectations at Time	
of Appointment .....	23
Summary .....	24
Reliability and Standard Errors of Measurement.....	25
Within-Group Interrater Reliability Coefficients and Standard Errors of Measurement on Administrative	
Roles .....	25
Administrator-Level Reliability .....	29
Internal Consistency of Personnel Ratings of Administrator Scales .....	30
Summary.....	30
Criterion-Referenced Interpretations .....	31
References .....	33

*List of Tables and Figures*

Figure 1 Model of Reflective Practice .....	5
Table 1 Descriptive Statistics for Number of Personnel Surveyed and Number of Respondents Within the Unit (N = 1,011).....	6
Table 2 Pearson r Correlations Between Respondent Mean Ratings, Number Surveyed, and Response Rate (N = 1,011).....	7
Table 3 Frequency and Percentage of Administrators From Institutions in Carnegie Classifications .....	8
Table 4 Frequency and Percentage of U.S. Regions Represented by Administrators' Institutions	8
Table 5 Descriptive Statistics for Characteristics of Administrator and Unit .....	9
Table 6 Frequencies and Percentages of the Administrator's Responses to Questions About the Unit .....	9
Table 7 Means and Standard Deviations for Administrator Self-Ratings of Administrative Roles (n = 868).....	11
Table 8 Means and Standard Deviations for Administrator Self-Ratings on Personal Characteristics (n = 855).....	11
Table 9 Means and Standard Deviations for Personnel Ratings of Administrative Roles (N = 1,011) .....	12
Table 10 Means and Standard Deviations for Personnel Ratings of Personal Characteristics (N = 1,011) .....	13
Table 11 Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation: Eigenvalues and Percentages of Variance for Self-Ratings of Administrative Role ....	15
Table 12 Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation: Eigenvalues and Percentages of Variance for Administrator Self-Ratings of Personal Characteristics.....	16
Table 13 Factor Loadings from Principal Components Factor Analysis with Varimax Rotation: Eigenvalues and Percentages of Variance for Personnel Ratings of Administrative Roles .....	16
Table 14 Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation: Eigenvalues and Percentages of Variance for Personnel Ratings of Administrators' Personal Characteristics .....	17
Table 15 Pearson r Correlations Between Personnel Ratings of Administrative Roles and Summary Judgments (N = 1,011).....	18
Table 16 Pearson r Correlations Between Personnel Ratings of the Administrator's Personal Characteristics and Summary Judgments (N = 1,011) .....	19
Table 17 Correlations Between Administrator and Unit Characteristics, Factor Composite Scale Scores, and Summary Judgments .....	21

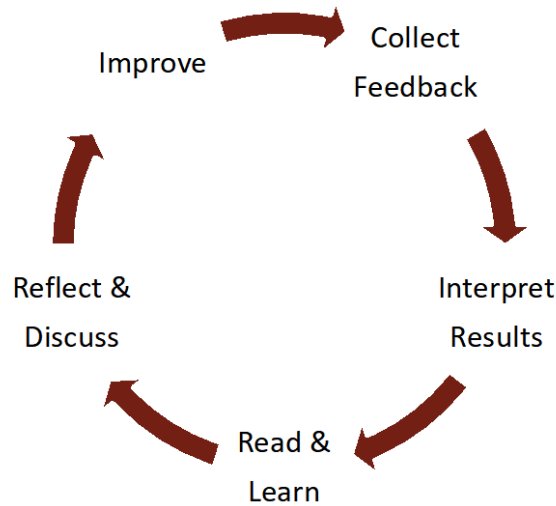
Table 18 Means and Standard Deviations for Summary Judgments by Administrator Perceptions of the Unit at the Time of Appointment.....	22
Table 19 Multivariate and Univariate Analyses of Variance for Administrator Perceptions of the Unit at the Time of Appointment Effects for Summary Judgment Items.....	23
Table 20 Means and Standard Deviations for Scale Scores by Administrator Perceptions of the Unit at the Time of Appointment.....	23
Table 21 Multivariate and Univariate Analyses of Variance for Administrator Perceptions of the Unit at the Time of Appointment Effects for Composite Scores.....	24
Table 22 Estimates of Within-Group Interrater Reliabilities and Standard Errors of Measurement for Personnel Ratings of Administrative Roles, Personal Characteristics, and Summary Judgements (n = 805).....	27
Table 23 Pearson r Correlations Between Personnel Ratings of the Same Administrator on Two Different Occasions (n = 178) .....	29
Table 24 Descriptive Statistics and Cronbach's Alpha for Personnel Ratings of Administrator Scales (N = 1,011) .....	30
Table 25 Percent of Personnel Rating Administrator Positively on Administrative Roles ("More a Strength than a Weakness" or "Definite Strength") and Summary Judgments ("Agree" or "Strongly Agree").....	31
Table 26 Percent of Personnel Rating Administrator Positively on Leadership Characteristics (6 or 7 on a 7-point Semantic Differential Scale) .....	32

## Introduction

This technical report describes the results of analyses performed on data collected from 2013 to 2017, using the IDEA Feedback System for Administrators (FSA). The FSA is used to gather impressions from core constituents about an administrator's performance of relevant administrative roles, as well as her/his leadership style, interpersonal characteristics, and overall job performance. The system is based on a model of reflective practice (see Figure 1), which is consistent with IDEA's longstanding approach to professional development: Improvement is more likely to occur when administrators receive useful feedback that motivates them.

*Figure 1*

*Model of Reflective Practice*



IDEA makes no claim that using the system will result in improvement. Rather, collecting feedback from personnel merely begins the process, as indicated at the top of Figure 1. With the aid of explanations contained in the IDEA [Feedback Report](#), administrators can then interpret results derived from responses to the FSA's *Administrator Information Form* (AIF) and the *Impressions of Administrators* (IA) surveys. However, change is more likely to occur when individuals read and learn about how to improve and then discuss results with a knowledgeable person or supervisor who can help to facilitate self-reflection. The administrator is, at that point, more prepared to improve.

The FSA is intended for this developmental approach. Its unique focus on feedback about the performance of relevant responsibilities and associated personal characteristics and behaviors makes it an extremely useful tool for professional development. As with any effective evaluation system, however, multiple sources of evidence should be collected, analyzed, and interpreted. Making a summative decision about an administrator's performance based solely on the results of the FSA would run counter to its aim.

The FSA assumes that those who observe an administrator over a period of time are in a position to make valid judgments of that administrator's performance and personal characteristics. However, no formal theory underlies the development of the items. The model depicts administrative effectiveness as a function of (a) performance of administrative roles, (b) personal characteristics, and (c) extraneous factors that are beyond the control of the administrator.

In the 2011 revision of the system, slight modifications were made in the wording of administrative roles and overall summary ratings, some personal characteristics were dropped, and new ones were added. In addition, administrators were given the option of completing a self-assessment, which enabled reporting of a gap analysis between self-ratings and average personnel ratings. The purpose of this report is to update relevant statistics associated with using the revised system. For information about the historical background of the system, the most recent revision of the instruments, and additional validity and reliability evidence, see [IDEA Technical Report No. 17](#) (Benton et al., 2011).

### Sample Description

From Spring 2013 to Spring 2017, 1,011 academic administrators from 77 institutions completed the AIF. Data from survey respondents were aggregated for each administrator, and the analyses in this report were performed on the aggregated dataset unless otherwise indicated.

#### *Response Rate and Its Relationship to Ratings*

Table 1 presents descriptive statistics for the number of personnel surveyed and the number responding. Because the variables were discrete counts and positively skewed, medians are reported along with the interquartile range. The number surveyed in each unit (i.e., department, college, institution, etc.) varied greatly from 5 to 2,459. The mean response rate per unit was 75%.

Table 1

*Descriptive Statistics for Number of Personnel Surveyed and Number of Respondents Within the Unit (N = 1,011)*

Variable	<i>Mdn</i>	Interquartile range
Number of personnel surveyed	42	71
Number of respondents within the unit	32	48

Table 2 displays Pearson  $r$  correlation coefficients between the respondent mean ratings on each item on the IA, number of personnel surveyed, and response rate within the unit. The correlations between mean ratings and number surveyed were weak and negative, which indicated ratings of the administrator were slightly lower as unit sample size increased. Numerous explanations could be given for this relationship. Perhaps larger units are more difficult to manage than smaller ones, and personnel are therefore less impressed with the administrator's performance. Alternatively, as unit size increases personnel may have less opportunity to interact with the administrator, which could negatively affect their impressions.

Table 2

*Pearson r Correlations Between Respondent Mean Ratings, Number Surveyed, and Response Rate (N = 1,011)*

Item	Number surveyed	Response rate
1. Communicating a visionary plan for the future	-.12**	.17**
2. Establishing sound priorities	-.15**	.13**
3. Displaying knowledge/expertise required for this position	-.17**	.16**
4. Making wise decisions, judgments, and recommendations	-.18**	.14**
5. Initiating actions that anticipate problems or resolves them	-.16**	.12**
6. Being an effective team member	-.15**	.10**
7. Contributing positively to this institution's image	-.20**	.18**
8. Communicating relevant information to appropriate constituencies	-.15**	.11**
9. Seeking opinions of others before establishing policies or procedures	-.20**	.13**
10. Earning the trust and respect of others	-.16**	.12**
11. Indecisive vs. Decisive	-.08**	.07*
12. Disorganized vs. Organized	-.07*	.05
13. Remote vs. Approachable	-.21**	.15**
14. Unfair vs. Fair	-.19**	.15**
15. Autocratic vs. Democratic	-.17**	.10**
16. Manipulative vs. Straightforward	-.16**	.12**
17. Inconsistent vs. Consistent	-.14**	.08**
18. Ambiguous vs. Clear	-.12**	.07*
19. Self-centered vs. Institution-centered	-.13**	.13**
20. Insensitive vs. Understanding	-.18**	.16**
21. Opinionated vs. Receptive to ideas	-.15**	.10**
22. Untrustworthy vs. Trustworthy	-.18**	.15**
23. Unimaginative vs. Innovative	-.17**	.17**
24. Rigid vs. Flexible	-.16**	.12**
25. Impractical vs. Practical	-.15**	.10**
26. Overall performance	-.16**	.16**
27. Confidence in future leadership	-.15**	.14**

*Note.* \*  $p < .05$ . \*\*  $p < .01$ .

In contrast, correlations between administrator ratings and unit response rate were weak and positive. Administrator ratings were slightly higher as personnel response rate increased. Again, many reasons could exist for the direction of this relationship. Personnel who held favorable impressions of their administrator may have been more likely to respond to the survey. Higher response rates might also suggest that the administrator (or his/her supervisor) carefully identified individuals most likely to respond to the survey. Other administrators, in contrast, may have been less selective and invited individuals who were less likely to respond. However, the

correlations, ranging from .07 to .18, are not high enough to warrant great concern. In addition, the high overall response rate (75%) within the population of personnel invited lessens concern about these relationships.

We also investigated whether response rate was related to the number of personnel surveyed. The correlation was moderate in strength and negative in direction,  $r = -.44$ . As the number of individuals asked to respond increased, the response rate decreased. This speaks to the issue raised in the preceding paragraph about careful selection of the sample. Identifying a sample without concern about the likelihood of individuals responding could contribute to lower response rates. Because response rate affects representativeness, all administrators, especially those in large units, should be encouraged to increase their efforts at improving personnel response rates. In addition, care should be taken in identifying respondents who feel qualified to evaluate the administrator. Those who feel unqualified are probably less likely to respond.

#### *Characteristics of Administrators, Units, and Institutions*

As shown in Table 3, participating administrators came from institutions representing all major Carnegie classifications, although over 80% were from masters- and doctorate-granting institutions. The participants represented 28 states and various regions of the U.S., with 60% from the Midwest and Southeast. Two institutions were outside the U.S. (see Table 4).

Table 3

#### *Frequency and Percentage of Administrators From Institutions in Carnegie Classifications*

Carnegie classification	<i>n</i>	%
Associate	53	5
Baccalaureate	104	10
Masters	259	26
Doctoral	564	56
No response	31	3

*Note.*  $N = 1,011$ .

Table 4

#### *Frequency and Percentage of U.S. Regions Represented by Administrators' Institutions*

Region	<i>n</i>	%
Midwest	10	33
Northeast	5	17
South	2	7
Southeast	8	27
West	3	10
Non-U.S.	2	7

*Note.*  $N = 30$ . Total of percentages is not 100 because of rounding.

Table 5 displays descriptive statistics for characteristics of the administrator and the unit. Of the 1,011 administrators who responded to the AIF, 91% ( $n = 915$ ) answered all open-ended questions shown in the table. One case reported an extreme value (442) for years at the



institution; another reported 142 years in the position at any institution.<sup>1</sup> We, therefore, removed those values prior to conducting the analyses. Responses were still positively skewed, causing the mean to be pulled in the direction of large values and rendering the median the preferred measure of central tendency. Administrators averaged about 12 years of service at the institution and four years in their current position. An average of two individuals, including the current administrator, had served in the position during the past 10 years. The average number of people reporting directly or indirectly to the administrator was 27, with an average of eight reporting directly.

Table 5

*Descriptive Statistics for Characteristics of Administrator and Unit*

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Mdn</i>
Administrator's years at institution	972	13.94	10.44	11.5
Administrator's years in position at institution	969	5.60	4.95	4
Individuals serving in this position in the past 10 years	976	2.51	4.05	2
People in unit reporting directly or indirectly to administrator	947	82.66	204.70	27
People in unit reporting directly to administrator	975	12.29	16.24	8

*Administrator Perceptions of the Unit's Reputation and Expectations at the Time of Appointment*

Administrators responded to several questions about their perceptions of the unit's reputation and expectations at the time of their appointment (see Table 6). Only one in four (25%) perceived a negative campus reputation regarding their unit's importance/influence when they first assumed the position (Question 1). A slightly greater percentage (30%) perceived a negative reputation about the unit's organizational effectiveness (Question 2). From the perspective of most administrators, however, their units had, for the most part, enjoyed positive or at least neutral campus reputations. Two thirds (67%) were expected to make major changes in the unit's activities when they first assumed the position (Question 3).

Table 6

*Frequencies and Percentages of the Administrator's Responses to Questions About the Unit*

Question and response options	<i>n</i>	%
1. When you first assumed this position, what was the general campus reputation of your unit in terms of importance/influence <sup>a</sup> :		
Positive	311	32
Neutral	313	32

<sup>1</sup> This variable was, nonetheless, not analyzed because 71% of administrators reported serving 0 years in the position at any institution, which did not make sense.

Negative	241	25
NA/new unit	76	8
Don't know	39	4
2. When you first assumed this position, what was the general campus reputation of your unit in terms of organizational effectiveness <sup>b</sup> :		
Positive	228	23
Neutral	335	34
Negative	291	30
NA/new unit	78	8
Don't know	47	5
3. When you first assumed this position, you were expected to <sup>c</sup> :		
Make major changes in the unit's activities	654	67
Primarily maintain established services and procedures	231	24
Other	94	10

Note. <sup>a</sup>*n* = 980. <sup>b</sup>*n* = 979. <sup>c</sup>*n* = 979. Total of percentages is not 100 because of rounding.

### Summary

Responding administrators came from institutions representing the four major Carnegie classifications, although most were from masters- and doctorate-granting institutions. More than half of the institutions were in the Midwest and Southeast regions of the U.S. On average, this was an experienced group of administrators who had been at their institution more than a decade. Turnover in the current position was relatively low. On average, more than two dozen individuals reported either directly or indirectly to the administrator, with fewer than 10 reporting directly. Only about a quarter of the administrators believed the general reputation of the unit in terms of importance and influence was negative at the time of their appointment. Roughly a third perceived a negative campus reputation about their unit's organizational effectiveness when they assumed the position. Nonetheless, two-thirds believed they were expected to make major changes in the unit's activities.

## Analysis of 2013 to 2017 Database

### Administrator Self-Ratings of Performance of Administrative Roles

Administrators were given the option of rating their performance on each of 10 administrative roles over the previous year. They used the scale 1 = *Definite Weakness*, 2 = *More a Weakness than a Strength*, 3 = *In Between*, 4 = *More a Strength than a Weakness*, and 5 = *Definite Strength*. Table 7 presents means and standard deviations for administrator self-ratings organized into the following five a priori categories that Hoyt and colleagues (Hoyt, Bailey, Pallett, & Gross, 1999) hypothesized characterize effective administrators:

1. *Planner*. Communicates a vision and sets priorities for the unit
2. *Consultant*. Is a team player who makes wise decisions and offers helpful recommendations

3. *Communicator*. Seeks the input of those who will be affected by policies and procedures; communicates relevant information to appropriate constituencies
4. *Expert*. Is knowledgeable and anticipates potential problems
5. *Community builder*. Contributes positively to the institution's image and earns the trust and respect of others

A total of 868 administrators responded to all 10 items. They rated their performance highest on "contributing positively to the institution's image," "displaying knowledge/expertise required for this position," and "being an effective team member." They rated themselves lowest on "communicating a visionary plan for the future."

Table 7

*Means and Standard Deviations for Administrator Self-Ratings of Administrative Roles (n = 868)*

Administrative roles	<i>M</i>	<i>SD</i>
Planner		
Communicating a visionary plan for the future	3.94	0.75
Establishing sound priorities	4.16	0.60
Consultant		
Making wise decisions, judgments, and recommendations	4.28	0.56
Being an effective team member	4.38	0.63
Communicator		
Communicating relevant information to appropriate constituencies	4.09	0.69
Seeking opinions of others before establishing policies or procedures that affect them	4.12	0.72
Expert		
Displaying knowledge/expertise required for this position	4.40	0.60
Initiating actions that anticipate problems or resolving them before they become major concerns	4.04	0.66
Community builder		
Contributing positively to the institution's image	4.55	0.57
Earning the trust and respect of others	4.22	0.64

### **Administrator Self-Ratings of Personal Characteristics**

Administrators were given the option of rating themselves on 15 personal characteristics, using 7-point semantic differential scales with higher scores indicating personal traits that are consistent with exemplary administrators. Antonyms were anchored on opposite ends of a 7-point scale, and administrators selected the number that best described themselves along the continuum. Table 8 presents means and standard deviations for the 855 administrators who provided self-ratings on all characteristics. The highest self-ratings were for being "trustworthy," "institution-centered," and "fair." Administrators rated themselves lowest on being "democratic" and "organized."

Table 8

*Means and Standard Deviations for Administrator Self-Ratings on Personal Characteristics (n = 855)*

Personal characteristic	<i>M</i>	<i>SD</i>
Indecisive vs. Decisive	5.63	0.86
Disorganized vs. Organized	5.50	1.03
Remote vs. Approachable	5.93	1.01
Unfair vs. Fair	6.20	0.77
Autocratic vs. Democratic	5.47	1.05
Manipulative vs. Straightforward	6.09	0.79
Inconsistent vs. Consistent	5.90	0.80
Ambiguous vs. Clear	5.59	0.86
Self-centered vs. Institution-centered	6.22	0.81
Insensitive vs. Understanding	5.91	0.93
Opinionated vs. Receptive	5.67	1.02
Untrustworthy vs. Trustworthy	6.45	0.70
Unimaginative vs. Innovative	5.65	1.00
Rigid vs. Flexible	5.62	0.93
Impractical vs. Practical	6.02	0.84

**Personnel Ratings of the Administrator's Performance of Administrative Roles**

Turning now to personnel ratings of the administrator, the first section of the IA survey asked respondents to rate the administrator's strengths and weaknesses on the same 10 roles on which administrators rated themselves. Table 9 presents means and standard deviations for personnel ratings on each administrative role. The respondents used the same scale of 1 = *Definite Weakness* to 5 = *Definite Strength*, with the additional option of CJ = *Cannot Judge* (designated as system missing for the analyses). They rated administrators highest on "contributing positively to this institution's image" and "displaying knowledge/expertise required for this position," which were also the highest administrator self-ratings. In addition, as did administrators, personnel assigned low ratings to "communicating a visionary plan for the future." This demonstrates concordance between administrator self-ratings and personnel impressions of the administrator. But, faculty staff gave the lowest ratings to "initiating actions that anticipate problems or resolving them before they become major concerns."

Table 9

*Means and Standard Deviations for Personnel Ratings of Administrative Roles (N = 1,011)*

Administrative role	<i>M</i>	<i>SD</i>
Planner		
1. Communicating a visionary plan for the future	3.92	0.56
2. Establishing sound priorities	3.99	0.53
Consultant		

4. Making wise decisions, judgments, and recommendations	4.06	0.55
6. Being an effective “team” member	4.11	0.58
Communicator		
8. Communicating relevant information to appropriate constituencies	4.05	0.54
9. Seeking opinions of others before establishing policies or procedures that affect them	3.91	0.59
Expert		
3. Displaying knowledge/expertise required for this position	4.30	0.51
5. Initiating actions that anticipate problems or resolving them before they become major concerns	3.88	0.57
Community Builder		
7. Contributing positively to this institution’s image	4.39	0.48
10. Earning the trust and respect of others	4.08	0.63

*Note.* Means and SDs were computed by calculating an average response per item after aggregating respondent data for each administrator. Personnel responded to all items on a 1 (*Definite Weakness*) to 5 (*Definite Strength*) scale.

### Personnel Ratings of Leadership Style and Personal Characteristics

Personnel also rated the administrator on the same 15 characteristics administrators rated themselves, using the same 7-point semantic differential scale. Table 10 contains means and standard deviations for ratings of individual items. Personnel gave the highest ratings to the same two characteristics that administrators did: being trustworthy and institution-centered. They gave the lowest for being democratic, consistent with administrator self-ratings. So, again, how administrators perceived themselves corresponded to how others perceived them.

Table 10

*Means and Standard Deviations for Personnel Ratings of Personal Characteristics (N = 1,011)*

Personal characteristic	<i>M</i>	<i>SD</i>
Indecisive vs. Decisive	5.72	0.62
Disorganized vs. Organized	5.77	0.68
Remote vs. Approachable	5.83	0.76
Unfair vs. Fair	5.83	0.71
Autocratic vs. Democratic	5.32	0.86
Manipulative vs. Straightforward	5.75	0.74
Inconsistent vs. Consistent	5.75	0.71
Ambiguous vs. Clear	5.63	0.73
Self-centered vs. Institution-centered	5.90	0.73
Insensitive vs. Understanding	5.79	0.75
Opinionated vs. Receptive	5.60	0.80
Untrustworthy vs. Trustworthy	5.92	0.76
Unimaginative vs. Innovative	5.58	0.71

Rigid vs. Flexible	5.48	0.79
Impractical vs. Practical	5.83	0.67

### Summary Judgments of the Administrator

Personnel responded to two items designed to assess their summary judgments of the administrator's effectiveness. The items stated, "Overall, this administrator has provided excellent leadership" (overall performance) and "I have confidence in the administrator's ability to provide future leadership in this position" (future leadership). Response options were: 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *In Between*, 4 = *Agree*, 5 = *Strongly Agree*, and CJ = *Cannot Judge*. Overall, personnel agreed that the administrator was providing excellent leadership ( $M = 4.16$ ,  $SD = 0.59$ ) and that they had confidence in the administrator's future leadership ( $M = 4.10$ ,  $SD = 0.58$ ). The correlation between mean ratings on the two summary judgments was .98.

### Validity

Validity refers to evidence that supports the proper use and interpretation of scores obtained from an assessment. The validity of any measure depends on how it is used in decision making. The IDEA FSA instruments are primarily intended for formative and developmental purposes. The system is intended to measure (a) personnel perceptions of the administrator's effectiveness and (b) the congruence between the administrator's self-ratings and personnel ratings of his/her performance (i.e., gap analysis). The *Feedback Report* provides administrators with suggestions for interpreting personnel feedback along with insights for improving administrative performance. If institutions choose to use the report for summative decisions, the IDEA Center strongly recommends that additional indicators of effectiveness the institution deems appropriate also be considered.

#### Evidence of Content Validity

Content validity primarily refers to the wording and format of items in a survey. Evidence of content validity can also come from the judgments of experts (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US), 2014). See [IDEA Technical Report No. 17](#) for evidence of content validity.

#### Validity Evidence Based on Internal Structure

Evidence for the validity of an instrument can also come from its internal structure or observed relationships among items. One aspect of item interrelationships concerns whether the underlying structure is comprised of a single dimension or is multidimensional. In the previous version of the IA, personnel ratings of the administrator's performance of administrative roles were unidimensional, whereas ratings of the administrator's personal characteristics were multidimensional (Benton et al., 2011). For this report, we began by investigating the internal structure of administrator self-ratings, and then conducted analyses to confirm previous findings related to personnel ratings of the administrator.

### *Factor Analysis of Self-Ratings of Administrative Roles*

We employed principal components analysis (PCA) to determine the underlying structure of administrator self-ratings on the 10 administrative roles. The scree plot, prior to rotation, revealed the first factor had an eigenvalue of 4.30, which explained 43.02 percent of the variance. This was followed by a large drop off to a second factor that explained 10.24 percent (eigenvalue = 1.02). A leveling off was then observed, and no other factors had eigenvalues greater than or equal to 1.0. Following varimax rotation, two principal factors emerged. The first factor, which consisted of seven items, concerned Relations with Others and included roles such as “Being an effective team member,” “Earning the trust and respect of others,” and “Seeking opinions of others before establishing policies or procedures that affect them.” The second factor pertained to Planning and was comprised of two items: “Communicating a visionary plan for the future” and “Establishing sound priorities.” One role, “Displaying knowledge/expertise required for this position,” was moderately correlated with both factors. Table 11 presents component matrix coefficients along with eigenvalues and percent of variance explained following varimax rotation.

Table 11

*Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation: Eigenvalues and Percentages of Variance for Self-Ratings of Administrative Roles*

Item	Factor loading	
	Relations with others	Planning
Being a team member	<b>.75</b>	.14
Earning trust/respect	<b>.74</b>	.20
Seeking others' opinions	<b>.74</b>	.03
Making wise decisions	<b>.63</b>	.28
Communicating relevant information	<b>.59</b>	.30
Anticipating problems	<b>.56</b>	.34
Contributing positively to image	<b>.52</b>	.43
Displaying knowledge/expertise	.49	.48
Communicating visionary plan	.02	<b>.88</b>
Establishing sound priorities	.32	<b>.68</b>
Eigenvalue	3.32	2.00
% of variance	33.22	20.04

*Note.* Boldface indicates highest factor loadings.

### *Factor Analysis of Administrator Self-Ratings of Personal Characteristics*

PCA was also employed to determine the underlying structure of administrator self-ratings of personal characteristics. The scree plot, prior to rotation, revealed the first factor had an eigenvalue of 6.87, which explained 45.80 percent of the variance. This was followed by a large drop off to a second factor that explained 9.74 percent (eigenvalue = 1.46). A leveling off was then observed, and no other factors had eigenvalues greater than or equal to 1.0. Following

varimax rotation, two principal factors emerged. The first factor contained many of the qualities found in the Interpersonal Characteristics dimension found previously in personnel ratings of the administrator (Benton et al., 2011): insensitive vs. understanding, opinionated vs. receptive to ideas, and autocratic vs. democratic. Likewise, the second factor in many ways mirrored that of Leadership Style, which Benton et al. (2011) previously reported in personnel ratings: indecisive vs. decisive, ambiguous vs. clear, and inconsistent vs. consistent. Two characteristics—fair vs. unfair and self-centered vs. institution-centered—were associated with both factors. Table 12 presents rotated component matrix coefficients, eigenvalues, and percent of variance explained for each factor following varimax rotation.

Table 12

*Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation: Eigenvalues and Percentages of Variance for Administrator Self-Ratings of Personal Characteristics*

Item	Factor loading	
	Interpersonal Characteristics	Leadership Style
Insensitive vs. Understanding	<b>.80</b>	.24
Opinionated vs. Receptive	<b>.77</b>	.21
Autocratic vs. Democratic	<b>.75</b>	.14
Remote vs. Approachable	<b>.75</b>	.17
Rigid vs. Flexible	<b>.67</b>	.26
Unfair vs. Fair	.56	.54
Self-centered vs. Institution-centered	.52	.47
Indecisive vs. Decisive	.00	<b>.75</b>
Ambiguous vs. Clear	.30	<b>.73</b>
Inconsistent vs. Consistent	.38	<b>.72</b>
Disorganized vs. Organized	.06	<b>.67</b>
Manipulative vs. Straightforward	.36	<b>.65</b>
Untrustworthy vs. Trustworthy	.46	<b>.60</b>
Impractical vs. Practical	.40	<b>.59</b>
Unimaginative vs. Innovative	.27	<b>.47</b>
Eigenvalue	4.20	4.13
% of variance	27.99	27.54

*Note.* Boldface indicates highest factor loadings.

*Factor Analysis of Personnel Ratings of the Administrator's Performance of Administrative Roles*

To determine whether the unidimensionality of personnel ratings of administrative roles could be confirmed from Benton et al.'s (2011) analysis, we again performed PCA. The results showed that a single dimension, Performance of Administrative Roles, best explains the internal structure of the ratings (see Table 13).

Table 13



*Factor Loadings from Principal Components Factor Analysis with Varimax Rotation:  
Eigenvalues and Percentages of Variance for Personnel Ratings of Administrative Roles*

Item	Performance of Administrative Roles
Making wise decisions	.97
Earning trust/respect	.97
Establishing priorities	.96
Being team member	.96
Anticipating problems	.95
Communicating relevant information	.95
Contributing positive image	.94
Displaying knowledge/expertise	.93
Seeking others' opinions	.92
Communicating vision	.87
Eigenvalue	8.89
% of variance	88.86

*Factor Analysis of Personnel Ratings of the Administrator's Personal Characteristics*

Benton et al.'s (2011) factor analysis of personnel ratings on 13 of the current 15 characteristics revealed that separate dimensions existed for Interpersonal Characteristics and Leadership Style. Both scales had high internal consistency. In the 2012 revision of the FSA, the characteristic "passive vs. active" was dropped, and three new characteristics were added: "unimaginative vs. innovative," "impractical vs. practical," and "rigid vs. flexible." PCA conducted on the current 15 characteristics revealed, prior to rotation, an initial factor with an eigenvalue of 12.15 that explained 80.98 percent of the variance. Following a large drop off, the second factor explained 8.91 percent with an eigenvalue of 1.34. Table 14 presents component matrix coefficients, following varimax rotation, along with eigenvalues and percent of variance explained. Consistent with previous findings (Benton et al., 2011), the first factor concerned Interpersonal Characteristics, such as "autocratic vs. democratic," "insensitive vs. understanding," and "opinionated vs. receptive to ideas." The second factor, as had been found in the earlier analysis, dealt with Leadership Style as it pertained to such qualities as being "indecisive vs. decisive," "disorganized vs. organized," and "ambiguous vs. clear."

Table 14

*Factor Loadings from Principal Components Factor Analysis Following Varimax Rotation:  
Eigenvalues and Percentages of Variance for Personnel Ratings of Administrators' Personal Characteristics*

Item	Interpersonal Characteristics	Leadership Style
Autocratic vs. Democratic	<b>.93</b>	.27
Insensitive vs. Understanding	<b>.92</b>	.32
Opinionated vs. Receptive	<b>.91</b>	.32
Rigid vs. Flexible	<b>.90</b>	.32
Remote vs. Approachable	<b>.89</b>	.31
Unfair vs. Fair	<b>.86</b>	.46

Untrustworthy vs. Trustworthy	<b>.82</b>	.50
Self-centered vs. Institution-centered	<b>.82</b>	.45
Manipulative vs. Straightforward	<b>.80</b>	.51
Impractical vs. Practical	<b>.72</b>	.64
Indecisive vs. Decisive	.14	<b>.95</b>
Disorganized vs. Organized	.33	<b>.84</b>
Ambiguous vs. Clear	.59	<b>.77</b>
Inconsistent vs. Consistent	.65	.71
Unimaginative vs. Innovative	.52	.67
Eigenvalue	8.54	4.94
% of variance	56.96	32.93

*Note.* Boldface indicates highest factor loadings.

#### *Correlations Between Personnel Ratings of Administrative Roles and Summary Judgments*

Another aspect of internal structure can be found in the relationships between personnel ratings of administrative roles and summary judgements of the administrator. An assumption of the FSA is that ratings of administrative roles are positively correlated with overall summary judgments. The coefficients presented in Table 15 confirm that the relationships are positive and strong ( $r$  coefficients ranged from .86 to .95). Average ratings on all 10 administrative roles and the total scale score showed high positive correlations with overall impressions of the administrator's performance and confidence in the administrator's future leadership.

Table 15

#### *Pearson $r$ Correlations Between Personnel Ratings of Administrative Roles and Summary Judgments ( $N = 1,011$ )*

Administrative role	Summary judgements	
	Overall performance	Future leadership
Communicating visionary plan for future	.87	.86
Establishing sound priorities	.95	.94
Displaying knowledge/expertise	.92	.92
Making wise decisions	.95	.95
Initiating actions that anticipate problems	.94	.93
Being an effective team member	.94	.93
Contributing positively to institution's image	.93	.93
Communicating relevant information	.92	.91
Seeking others' opinions	.87	.87
Earning trust and respect	.95	.95
Total scale	.98	.98

*Note.* All  $r$  coefficients are significant at  $p < .001$ .

*Correlations Between Personnel Ratings of Interpersonal Characteristics, Leadership Style, and Summary Judgments*

Another assumption of the FSA is that personnel ratings of Interpersonal Characteristics and Leadership Style should be positively correlated with overall summary judgments. Table 16 confirms there are strong positive correlations.

Table 16

*Pearson r Correlations Between Personnel Ratings of the Administrator's Personal Characteristics and Summary Judgments (N = 1,011)*

Personal characteristic	Summary Judgments	
	Overall performance	Future leadership
Interpersonal Characteristics	.90	.90
Remote vs. Approachable	.81	.80
Unfair vs. Fair	.88	.88
Autocratic vs. Democratic	.80	.80
Manipulative vs. Straightforward	.88	.88
Self-centered vs. Institution-centered	.86	.85
Insensitive vs. Understanding	.84	.84
Opinionated vs. Receptive to ideas	.83	.83
Untrustworthy vs. Trustworthy	.89	.90
Rigid vs. Flexible	.83	.83
Impractical vs. Practical	.91	.91
Leadership Style	.93	.92
Indecisive vs. Decisive	.75	.74
Disorganized vs. Organized	.78	.77
Inconsistent vs. Consistent	.91	.90
Ambiguous vs. Clear	.92	.91
Unimaginative vs. Innovative	.85	.85

*Note.* All coefficients are significant at  $p < .001$ .

*Summary*

PCA revealed that administrator self-ratings of administrative roles are multidimensional, with one factor pertaining to Relations with Others and the other to Planning. In contrast, personnel ratings of administrative roles were most likely measuring one underlying dimension or factor, Performance of Administrative Roles. Both administrator self-ratings and personnel ratings of the administrator's personal characteristics loaded on one of either two dimensions: Interpersonal Characteristics and Leadership Style. Finally, personnel ratings on all IA items and subscales were highly correlated with two summary judgments, which provides additional evidence of validity based on internal structure.

**Validity Evidence Based on Relationships to Other Variables**

Relationships of items and scales to variables external to the instrument can provide another important source of validity evidence (American Educational Research Association et

al., 2014). In this section, we present evidence of when group differences are absent—which supports consistency in interpretations across demographic variables—and evidence of when expected group differences are present.

*Correlations Between Administrator and Unit Characteristics, Factor Composite Scale Scores, and Summary Judgments*

Pearson  $r$  correlations (see Table 17) indicated negligible relationships between summed aggregated personnel ratings on Performance of Administrative Roles, Interpersonal Characteristics, and Leadership Style and the following characteristics: administrator's years at the institution, years in the position at the institution, years in the position at any institution, number of individuals who have occupied the position in the past 10 years, number of people who reported directly or indirectly to the administrator, number who reported directly to the administrator, number surveyed, and response rate. The absence of any meaningful relationships provides validity evidence that the instrument is not strongly biased toward administrator years of service, position turnover, size of unit, number surveyed, and response rate.

Table 17

*Correlations Between Administrator and Unit Characteristics, Factor Composite Scale Scores, and Summary Judgments*

Administrator and unit characteristic	Factor composite scale score			Summary judgement	
	Administrative Roles	Interpersonal Characteristics	Leadership Style	Overall performance	Future leadership
Administrator's years at institution	.14	.13	.12	.14	.11
Administrator's years in position at institution	.02	-.02	-.01	.03	.01
Administrator's years in position at any institution	-.12	-.13	-.10	-.12	-.10
Number who served in this position past 10 years	-.04	-.03	-.03	-.04	-.04
Number of people reporting directly or indirectly to admin	-.09	-.11	-.06	-.08	-.09
Number of people reporting directly to admin	-.01	.00	.01	.01	.00
Number of personnel surveyed	-.17	-.18	-.13	-.15	-.16
Response rate	.14	.13	.10	.14	.16

*Note.* Correlations higher than .09 or lower than -.09 are significant at  $p < .001$ .

*Comparison of Summary Judgments by Perceptions of the Unit's Reputation and Expectations at Time of Appointment*

We conducted multivariate analyses of variance (MANOVA) to investigate whether mean scores on the summary judgment items varied by the administrator's perceptions of the unit's reputation and effectiveness, as well as expectations of the chair at the time of appointment. Personnel summary judgments on overall performance and future leadership served as dependent variables. Administrators choosing "NA/new unit," "Don't know," or "Other" were not included in the analyses.

We considered effect sizes less than .01 to be of little practical significance. Partial eta squared ( $\eta^2$ ) served as the measure of effect size, and Tukey's B test ( $\alpha = .05$ ) was used to make pairwise comparisons between means following significant effects. Table 18 presents means and standard deviations.

Table 18

*Means and Standard Deviations for Summary Judgments by Administrator Perceptions of the Unit at the Time of Appointment*

Administrator perception of the unit	<i>n</i>	Overall performance		Future leadership	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Importance					
Positive	311	4.19	0.52	4.25	0.53
Neutral	313	4.15	0.56	4.20	0.55
Negative	241	3.97	0.64	4.03	0.65
Effectiveness					
Positive	228	4.18	0.53	4.24	0.53
Neutral	335	4.16	0.56	4.21	0.56
Negative	291	3.99	0.62	4.05	0.63
Expectation					
Major changes	654	4.08	0.59	4.13	0.59
Status quo	231	4.22	0.52	4.28	0.52

The results of the MANOVA and subsequent univariate analyses are presented in Table 19. The following meaningful effects emerged:

1. Administrators received higher ratings on summary judgments when—at the time of appointment—they perceived the general campus reputation of the unit, in terms of *importance*, to be positive or neutral rather than negative. Specifically, personnel gave higher ratings on both overall performance,  $\eta^2 = .025$ , and future leadership,  $\eta^2 = .025$ , to administrators who perceived the reputation as either positive or neutral,  $p < .001$ .
2. Administrators received higher ratings on summary judgments when—at the time of appointment—they perceived the general campus reputation of the unit, in terms of *effectiveness*, to be positive or neutral rather than negative. Specifically, personnel gave higher ratings on both overall performance,  $\eta^2 = .02$ , and future leadership,  $\eta^2 = .019$ , to administrators who perceived the reputation as either positive or neutral,  $p < .001$ .

3. Administrators received higher ratings on summary judgments, when—at the time of appointment—they perceived that they were expected to primarily “maintain established services and procedures” rather than “make major changes.” Specifically, personnel gave higher ratings on both overall performance,  $\eta^2 = .013$ , and future leadership,  $\eta^2 = .012$ , when the perception was to primarily maintain the status quo,  $p < .001$ .

Table 19

*Multivariate and Univariate Analyses of Variance for Administrator Perceptions of the Unit at the Time of Appointment for Summary Judgment Items*

Source	Multivariate		Univariate			
	F	$\eta^2$	Overall performance		Future leadership	
			F	$\eta^2$	F	$\eta^2$
Perceived importance	5.49***	.013	11.04***	.025	10.88***	.025
Perceived effectiveness	4.61***	.011	8.73***	.020	8.44***	.019
Perceived expectation	5.73**	.013	11.48***	.013	11.08***	.012

Note. \*\*\* $p < .001$ , \*\* $p < .01$ . Multivariate  $F$  ratios were generated from Pillai’s statistic.

*Comparisons of Mean Scale Scores by Perceptions of the Unit’s Reputation and Expectations at Time of Appointment*

MANOVAs were conducted to examine whether mean scale scores varied as a function of the administrator’s perceptions of the unit’s reputation and effectiveness, as well as expectations at the time of appointment. Personnel mean scale scores on Administrative Roles, Interpersonal Characteristics, and Leadership Style served as dependent variables. Administrators choosing “NA/new unit,” “Don’t know,” or “Other” were not included in the analyses. We again considered effect sizes less than .01 to be of little practical significance. Table 20 presents means and standard deviations.

Table 20

*Means and Standard Deviations for Scale Scores by Administrator Perceptions of the Unit at the Time of Appointment*

Administrator perception	n	Administrative Role		Interpersonal Characteristics		Leadership Style	
		M	SD	M	SD	M	SD
<b>Importance</b>							
Positive	311	41.53	4.69	58.75	6.00	28.89	2.79
Neutral	313	41.01	5.01	57.76	7.08	28.64	3.07
Negative	241	39.46	5.74	55.29	7.89	27.86	3.44
<b>Effectiveness</b>							
Positive	228	41.50	4.89	58.85	6.37	28.90	2.93
Neutral	335	41.15	4.91	58.00	6.78	28.68	3.02
Negative	291	39.67	5.60	55.48	7.86	28.02	3.29
<b>Expectation</b>							
Major changes	654	40.40	5.28	56.56	7.42	28.42	3.08

Status quo	231	41.92	4.63	59.47	5.89	28.99	2.96
------------	-----	-------	------	-------	------	-------	------

The results of the MANOVA and subsequent univariate analyses are presented in Table 21. The following meaningful effects were found:

1. Administrators received higher ratings on all three scale scores when—at the time of appointment—they perceived the general campus reputation of the unit, in terms of *importance*, to be positive or neutral rather than negative. Specifically, personnel gave higher ratings on Administrative Roles,  $\eta^2 = .026$ , Interpersonal Characteristics,  $\eta^2 = .039$ , and Leadership Style,  $\eta^2 = .018$ , to administrators who perceived the reputation as either positive or neutral,  $p < .001$ .
2. Administrators received higher ratings on all three scale scores when—at the time of appointment—they perceived the general campus reputation of the unit, in terms of *effectiveness*, to be positive or neutral. Specifically, personnel gave higher ratings on Administrative Roles,  $\eta^2 = .023$ , Interpersonal Characteristics,  $\eta^2 = .038$ , and Leadership Style,  $\eta^2 = .014$ , to administrators who perceived the reputation as either positive or neutral,  $p < .01$ .
3. Administrators received higher ratings on two of three scale scores, when—at the time of appointment—they perceived that they were expected to primarily maintain “the status quo.” Specifically, personnel gave higher ratings on both Administrative Roles,  $\eta^2 = .017$  and Interpersonal Characteristics,  $\eta^2 = .032$ , when administrators perceived they were expected to primarily maintain the status quo than make major changes,  $p < .001$ .

Table 21

*Multivariate and Univariate Analyses of Variance for Administrator Perceptions of the Unit at the Time of Appointment for Composite Scores*

Source	Multivariate		Univariate					
			Administrative Roles		Interpersonal Characteristics		Leadership Style	
	<i>F</i>	$\eta^2$	<i>F</i>	$\eta^2$	<i>F</i>	$\eta^2$	<i>F</i>	$\eta^2$
Perceived importance	6.04***	.021	11.68***	.026	17.27***	.039	7.99***	.018
Perceived effectiveness	6.65***	.023	9.83***	.023	16.76***	.038	6.03**	.014
Perceived expectation	14.01***	.046	14.93***	.017	29.16***	.032	6.04*	.007

*Note.* \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ . Multivariate *F* ratios were generated from Pillai’s statistic.

### *Summary*

Personnel gave the highest ratings to administrators who, at the time of appointment, perceived the unit’s reputation in terms importance and effectiveness as either positive or neutral and who believed they were expected to maintain the status quo rather than make major changes. This suggests that perceptions of the unit and expectations placed on the administrator upon appointment are important extraneous factors to consider when interpreting findings from the *Feedback Report*. Other characteristics of the administrator and unit—e.g., years of service, position turnover, and size of unit—were only negligibly related to ratings. These very weak



correlations offer evidence of consistency in interpretations of scores across levels of administrators' experience and unit characteristics.

### Reliability and Standard Errors of Measurement

Reliability refers to the consistency of measures. Users of the FSA, for example, need to know how consistently personnel within the unit rate their performance. Are certain items or scales more reliable than others? Does the number of raters in the unit affect reliability? Standard error of measurement, the flipside of reliability, informs users that all scores contain some amount of error. Those who interpret scores from the *IDEA Feedback Report* should, therefore, take into account the imperfect nature of any measure.

#### Within-Group Interrater Reliability Coefficients and Standard Errors of Measurement on Administrative Roles

Evidence of reliability can come from the consistency in personnel ratings within the same unit. Consistency at the unit level is important because it enables evaluators to make interpretations that generalize across most personnel. The reliability of personnel ratings within the unit can be estimated by applying the *within-group interrater reliability coefficient* (James, Demaree, & Wolf, 1993). Coefficients range from 0 to 1.00; the higher the value, the more consistent personnel are in their impressions of the administrator. We began by selecting administrators for whom the number of raters equaled 16 or more. For administrators with more than 16 raters, 16 were randomly selected. We then computed the within-group interrater reliability coefficient on each administrative role for the personnel in that grouping. Following this, we applied the following single-item interrater agreement formula to compute the reliability coefficients for the items found in Table 25:

$$r_{WG(l)} = 1 - (S_{x_j}^2 / \sigma_{EU}^2)$$

where  $r_{WG(l)}$  is the within-group interrater reliability for a group of  $K$  judges on a single item  $X_j$ , and  $S_{x_j}^2$  is the observed variance of  $X_j$ .  $\sigma_{EU}^2$  is the variance of  $X_j$  that would be expected if all judgments solely resulted from random measurement error. Thus  $\sigma_{EU}^2 = (A^2 - 1)/12$  where  $A$  corresponds to *the number of alternatives in the response scale for  $X_j$* , which is presumed to vary from 1 to  $A$ . For example, for the item "Communicating a visionary plan for the future" the standard deviation for all administrators with 16 randomly selected raters was 0.58. Since this item was measured with a 5-point scale, its  $\sigma_{EU}^2$  is calculated as  $(5^2 - 1)/12 = 2$ . The computation for within-group interrater reliability was as follows:

$$r_{WG(l)} = 1 - (0.58^2/2) \approx .83$$

The *Spearman - Brown prophecy formula* was then used to estimate the reliability of the mean scores on each item for administrators with 8, 32, and 64 raters:

$$r_{sp} = \frac{n r_x}{1 + (n - 1) r_x}$$

where  $r_x$  is the reliability of a rating based on  $x$  raters and  $n$  is the number of raters we estimate the reliability for divided by  $x$ . For example, if computing the reliability of ratings for eight raters based on the reliability of 16 raters,  $n = 8/16 = 0.5$ . For the item we computed above,

$$r_{sp} = \frac{0.5 \times .83}{1 + (0.5 - 1) \times .83} \approx .71$$

Likewise, the reliability of the item with 32 raters is estimated as follows,

$$r_{sp} = \frac{(32 / 16) \times .83}{1 + (32 / 16 - 1) \times .83} \approx .91$$

*Standard errors of measurement* (SEM) were then calculated:

$$SEM = s\sqrt{1 - r_{xx}}$$

where  $s$  refers to the standard deviation of the item and  $r_{xx}$  denotes the single-item reliability. For the item above, when 16 raters were randomly selected,

$$SEM = 0.58\sqrt{1 - .83} \approx 0.24$$

As shown in Table 22, the reliability coefficients ( $r$ ) are approximately .80 or higher when the number of raters is at least 16. The standard error of measurement (SEM) provides a meaningful way of interpreting the amount of error for each item. An administrator's "true score," or average, on any item could theoretically be obtained if that item were administered to the same group of personnel an infinite number of times. The SEM indicates that roughly 68% of the time an administrator's true score average rating would fall within the range of  $\pm 1$  standard error around the obtained average rating.<sup>2</sup> The standard errors of measurements in Table 22 indicate that, for most items, the average SEM was approximately .30 or lower when the number of raters was at least 16.

---

<sup>2</sup> The 90% confidence interval would be 1.645 standard deviations around the obtained average rating.

Table 22

*Estimates of Within-Group Interrater Reliabilities and Standard Errors of Measurement for Personnel Ratings of Administrative Roles, Personal Characteristics, and Summary Judgements (n = 805)*

Item	<i>M</i>	<i>SD</i>	Number of Raters							
			8 Personnel		16 Personnel		32 Personnel		64 Personnel	
			<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>	<i>r</i>	<i>SEM</i>
<b>Administrative roles</b>										
Communicating visionary plan for future	3.91	0.58	.71	0.31	.83	0.24	.91	0.17	.95	0.13
Establishing sound priorities	3.97	0.55	.74	0.28	.85	0.21	.92	0.16	.96	0.11
Displaying knowledge/expertise	4.28	0.54	.75	0.27	.86	0.20	.92	0.15	.96	0.11
Making wise decisions	4.03	0.58	.71	0.31	.83	0.24	.91	0.18	.95	0.13
Initiating actions that anticipate problems	3.85	0.60	.70	0.33	.82	0.25	.90	0.19	.95	0.14
Being an effective team member	4.08	0.61	.68	0.35	.81	0.27	.90	0.20	.94	0.14
Contributing positively to institution's image	4.36	0.52	.76	0.25	.87	0.19	.93	0.14	.96	0.10
Communicating relevant information	4.02	0.56	.73	0.29	.84	0.22	.91	0.16	.96	0.12
Seeking others' opinions	3.86	0.63	.67	0.36	.80	0.28	.89	0.21	.94	0.15
Earning trust and respect	4.04	0.66	.64	0.40	.78	0.31	.88	0.23	.93	0.17
<b>Personal characteristics</b>										
Indecisive vs. Decisive	5.72	0.65	.81	0.28	.90	0.21	.94	0.15	.97	0.11
Disorganized vs. Organized	5.78	0.72	.77	0.34	.87	0.26	.93	0.19	.96	0.14
Remote vs. Approachable	5.78	0.83	.70	0.45	.83	0.35	.91	0.26	.95	0.19
Unfair vs. Fair	5.78	0.79	.73	0.41	.84	0.31	.91	0.23	.96	0.17
Autocratic vs. Democratic	5.27	0.93	.64	0.56	.78	0.44	.88	0.33	.93	0.24
Manipulative vs. Straightforward	5.71	0.80	.72	0.42	.84	0.32	.91	0.24	.95	0.17
Inconsistent vs. Consistent	5.73	0.76	.75	0.38	.86	0.29	.92	0.21	.96	0.15
Ambiguous vs. Clear	5.60	0.78	.74	0.40	.85	0.30	.92	0.22	.96	0.16
Self-centered vs. Institution-centered	5.87	0.80	.73	0.42	.84	0.32	.91	0.23	.95	0.17
Insensitive vs. Understanding	5.75	0.80	.72	0.43	.84	0.32	.91	0.24	.95	0.17
Opinionated vs. Receptive	5.56	0.89	.67	0.51	.80	0.40	.89	0.29	.94	0.21

Untrustworthy vs. Trustworthy	5.86	0.83	.71	0.45	.83	0.35	.91	0.26	.95	0.19
Unimaginative vs. Innovative	5.53	0.77	.74	0.39	.85	0.29	.92	0.22	.96	0.16
Rigid vs. Flexible	5.43	0.85	.69	0.47	.82	0.36	.90	0.27	.95	0.20
Impractical vs. Practical	5.80	0.73	.77	0.35	.87	0.26	.93	0.19	.96	0.14
Summary judgements										
Overall performance	4.07	0.60	.69	0.34	.82	0.26	.90	0.19	.95	0.14
Future leadership	4.13	0.61	.68	0.34	.81	0.27	.90	0.20	.95	0.14

*Note.* SEM = Standard error of measurement.

### Administrator-Level Reliability

Unit-level reliability, as indicated by within-group interrater reliability and SEM, is a necessary condition for administrator-level reliability, which is consistency in ratings of the same administrator across more than one assessment. Some measure of consistency in measures of the same thing is important for having confidence in inferences made about how good of a job the administrator is doing. We determined administrator-level reliability by first identifying administrators who had been rated at least twice ( $n = 178$ ). For those that had been rated more than twice, we selected the most recent two occasions. We then computed the Pearson  $r$  coefficient between paired average personnel ratings on each item and the three factor composite scale scores (i.e., Administrative Roles, Interpersonal Characteristics, and Leadership Style). As shown in Table 23, personnel ratings of the same administrator on both occasions were strongly correlated, with  $r$ s ranging from .81 to .90.

Table 23

*Pearson  $r$  Correlations Between Personnel Ratings of the Same Administrator on Two Different Occasions ( $n = 178$ )*

Item	$r$
Administrative Roles	.87
Communicating a visionary plan for the future	.84
Establishing sound priorities	.85
Displaying knowledge/expertise required for this position	.86
Making wise decisions, judgments, and recommendations	.85
Initiating actions that anticipate problems or resolves them	.81
Being an effective team member	.88
Contributing positively to this institution's image	.86
Communicating relevant information to appropriate constituencies	.83
Seeking opinions of others before establishing policies or procedures	.84
Earning the trust and respect of others	.88
Interpersonal Characteristics	.90
Remote vs. Approachable	.87
Unfair vs. Fair	.87
Autocratic vs. Democratic	.87
Manipulative vs. Straightforward	.87
Self-centered vs. Institution-centered	.90
Insensitive vs. Understanding	.88
Opinionated vs. Receptive to ideas	.86
Untrustworthy vs. Trustworthy	.89
Rigid vs. Flexible	.88
Impractical vs. Practical	.87
Leadership Style	.88
Indecisive vs. Decisive	.85
Disorganized vs. Organized	.87

Inconsistent vs. Consistent	.85
Ambiguous vs. Clear	.85
Unimaginative vs. Innovative	.84
Summary judgements	
Overall performance	.86
Future leadership	.85

*Note.* All coefficients are significant at  $p < .001$ .

### Internal Consistency of Personnel Ratings of Administrator Scales

The preceding sections have provided evidence of the reliability of individual items. But, the FSA also contains three subscales, which are combinations of items that measure a common construct. Internal consistency, which estimates how consistently items within a subscale measure the same construct, is computed with *Cronbach's alpha* ( $\alpha$ ) (Cronbach, 1951) or the *intraclass coefficient*. Coefficients range from 0 to 1.00, and high values are desirable. One would expect, for example, that items loading on the same factor should be highly intercorrelated. Although the *Feedback Report* does not provide subscale scores, we provide intraclass coefficients here as evidence of the internal consistency of the various dimensions in personal ratings of administrators.

We computed means, standard deviations, and Cronbach's alphas on the 10-item Administrative Roles scale, and the Interpersonal Characteristics and Leadership Style scales that comprised the administrator's personal characteristics. As displayed in Table 24, all three scales exhibited high internal consistency (Cronbach's  $\alpha$ s ranged from .95 to .99).

Table 24

*Descriptive Statistics and Cronbach's Alpha for Personnel Ratings of Administrator Scales (N = 1,011)*

Scale	<i>M</i>	<i>SD</i>	Cronbach's alpha	<i>SEM</i>	Number of items
Administrative Roles	40.69	5.22	.99	0.52	10
Interpersonal Characteristics	57.27	7.20	.99	0.72	10
Leadership Style	28.46	3.13	.95	0.70	5

### Summary

Three sources of evidence were provided regarding the reliability of personnel ratings of the administrator. First, the within-group interrater reliability coefficient showed that responses to individual IA items from personnel in the IDEA database have high reliability at the unit level, especially when the number of raters is greater than or equal to 16. Similarly, standard errors of measurement supported the dependability of individual items, especially as the number of raters increases. Second, test-retest correlation coefficients revealed that personnel ratings had good stability at the level of the administrator. Third, Cronbach's  $\alpha$  coefficients were high for the Administrative Roles, Interpersonal Characteristics, and Leadership Style scales.

Multiple sources of evidence have, thus, been presented to support the reliability of personnel responses to items in the IA. Ultimately, though, the user must determine whether local scores are sufficiently trustworthy to warrant use and interpretation. The number of raters

and the care administrators and personnel take in completing the ratings are important considerations.

### Criterion-Referenced Interpretations

When identifying strengths regarding administrative roles, the IDEA Feedback System for Administrators Report applies the criterion of 70%. That is, administrative roles rated by at least 70% of the personnel as either “More a Strength than a Weakness” or “Definite Strength” (4 or 5 on a 5-point scale) are considered a positive rating. Table 25 lists the percentage of personnel giving positive ratings for each administrative responsibility and the two overall summary measures, using individual ratings as the unit of analysis. The mean and median percent of positive ratings across all roles are 72.2% and 71.9% respectively.

Table 25

*Percent of Personnel Rating Administrator Positively on Administrative Roles (“More a Strength than a Weakness” or “Definite Strength”) and Summary Judgments (“Agree” or “Strongly Agree”)*

Item	%
1. Communicating a visionary plan for the future	69.2
2. Establishing sound priorities	70.2
3. Displaying knowledge/expertise required for this position	79.9
4. Making wise decisions, judgments, and recommendations	71.5
5. Initiating actions that anticipate problems or resolving them before they become major concerns	65.7
6. Being an effective team member	73.3
7. Contributing positively to the institution’s image	81.5
8. Communicating relevant information to appropriate constituencies	72.2
9. Seeking opinions of others before establishing policies or procedures that affect them	65.0
10. Earning the trust and respect of others	71.4
26. Overall, this administrator has provided excellent leadership	72.8
27. I have confidence in the administrator’s ability to provide future leadership in this position	74.1

When identifying leadership characteristics as strengths, IDEA applies the criterion of 60%. That is, leadership characteristics rated positively by at least 60% of the personnel (6 or 7 on a 7-point semantic differential scale) are considered a strength. Table 26 below lists the percentage of individuals giving positive ratings for each leadership characteristic. The mean and median percent of positive ratings across all roles are 65.0% and 66.7% respectively.

Table 26

*Percent of Personnel Rating Administrator Positively on Leadership Characteristics (6 or 7 on a 7-point Semantic Differential Scale)*

Item	%
11. Indecisive (1) vs. Decisive (7)	67.0
12. Disorganized (1) vs. Organized (7)	68.8
13. Remote (1) vs. Approachable (7)	66.4
14. Unfair (1) vs. Fair (7)	67.0
15. Autocratic (1) vs. Democratic (7)	53.3
16. Manipulative (1) vs. Straightforward (7)	66.7
17. Inconsistent (1) vs. Consistent (7)	66.9
18. Ambiguous (1) vs. Clear (7)	63.1
19. Self-centered (1) vs. Institution-centered (7)	72.3
20. Insensitive (1) vs. Understanding (7)	66.5
21. Opinionated (1) vs. Receptive to ideas (7)	62.4
22. Untrustworthy (1) vs. Trustworthy (7)	70.2
23. Unimaginative (1) vs. Innovative (7)	58.8
24. Rigid (1) vs. Flexible (7)	56.9
25. Impractical (1) vs. Practical (7)	68.2



## References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for Educational and psychological testing*. Washington, DC: American Educational Research Association.
- Benton, S. L., Gross, A. B., Pallett, W. H., Song, J., Webster, R., & Guo, M. (2011). *IDEA Technical Report No. 17: Technical Manual for the Revised IDEA Feedback for Administrators System*. Manhattan, KS: The IDEA Center.
- Blevins, C., & Mullen, L. (2015). Jane, John ... Leslie? A historical method for algorithmic gender prediction. *Digital Humanities Quarterly*, 9(3). Retrieved from <http://www.digitalhumanities.org/dhq/vol/9/3/000223/000223.html>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <http://doi.org/10.1007/BF02310555>
- Hoyt, D. P., Bailey, M., Pallett, W. H., & Gross, A. B. (1999). The IDEA system for evaluating and improving the administrative performance of department chairpersons: Revising the DECAD form. Manhattan, KS: The IDEA Center.
- James, L. R., Demaree, R. G., & Wolf, G. (1993).  $r_{wg}$ : An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78(2), 306–309. <http://doi.org/10.1037//0021-9010.78.2.306>
- Mullen, L. (2015). gender: Predict gender from names using historical data.