

# Assessment of Leadership in Children, Youth and Adults

Thomas Oakland  
University of Florida

Bradd A. Falkenberg  
Child and Adolescent Service Center

Christopher Oakland  
University of Birmingham

## Abstract

Current theories of leadership are reviewed. Eleven leadership measures, six for children and youth and five for adults, are reviewed in the context of these theories and with psychometric standards for test use. In general, the measures are normed inadequately and lack information about reliability and validity. The use of short subscales embedded within other scales, which themselves often are inadequate, should be avoided. Currently, we lack the assessment technology to measure leadership adequately in children and youth. Leadership measures for adults are more promising, especially the Leadership Opinion Questionnaire, the Supervisory Behavior Description Questionnaire, and the Campbell Leadership Index.

The development of talents and abilities constitutes a primary goal of education. Schools are expected to promote cognitive and affective qualities that enhance students' leadership, thus enabling them to assume positions of leadership within the many vocations, professions, and institutions that provide critical services and stability to society. By nurturing leadership abilities in the young, we expect to have the adult leadership necessary for effective governmental, business, commercial, educational, religious, and philanthropic activities.

The purpose of this article is to review existing standardized measures of leadership and to suggest future directions for the assessment of leadership. The review is guided by two important criteria: whether a scale's psychometric qualities meet existing standards of technical adequacy and whether a scale reflects current concepts of leadership.

## Standards of Technical Adequacy

Standards by which test development and use are evaluated are presented in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, &

National Council on Measurement in Education, 1985). Test developers and publishers are expected to present information that fully describes a test's development and normative population, reliability, and validity (especially criterion-related and construct).

## Concepts of Leadership

In the literature on leadership, four concepts or theories of leadership often are discussed: the exertion of power and influence over others, the skillful management of behaviors, the demonstration of personal traits, and the interaction between personal traits and environmental qualities (Stogdill, 1974; Yukl, 1989).

**Leadership as power and influence.** Leadership can be explained in terms of the amount and types of power an individual possesses and how that power is exercised. In this conceptualization, leadership stems from personal qualities (expertise, friendships, charisma), political sources (coalitions, co-optations, institutions), and management position (control over resources, rewards, and information) that influence people and events.

**Leadership as skillful management of behavior.** This orientation emphasizes the importance of three broad qualities that influence leadership abilities: technical, interpersonal, and conceptual. Other narrower competencies

## Putting Research to Use

Leadership can be enhanced by the use of measures that identify persons who evidence leadership potential and that document the effectiveness of programs designed to promote its development. Despite this need, there are few suitable measures of leadership for children and youth. The assessment of leadership must go beyond the use of existing scales and surveys.

The most promising assessments are those guided by situational theories of leadership. These measures, however, have been developed for adults. Educators can advance leadership programs for students by examining the literature on situational leadership and by collaborating with assessment specialists to develop measures within that context. Adult measures like the Leadership Opinion Questionnaire, the Supervisory Behavior Description Questionnaire, and the Campbell Leadership Index reviewed in this article provide models and a place for educators to begin.

Thomas Oakland, Department of Foundations of Education, University of Florida; Bradd A. Falkenberg, Child and Adolescent Service Center, Canton, Ohio; Christopher Oakland, University of Birmingham, England.

Address correspondence to Thomas Oakland, P.O. Box 117047, University of Florida, Gainesville, FL 32611-7047.

Downloaded from [gcq.sagepub.com](http://gcq.sagepub.com) at PENNSYLVANIA STATE UNIV on May 16, 2016

include the ability to work efficiently and proactively, the clarity of an individual's conceptual abilities and orally presented ideas, impact on others, diagnostic abilities, and the ability to properly manage group resources and social power.

**Leadership as personal qualities and traits.** Various traits and skills are thought to influence leadership. These include an individual's adaptability to situations, alertness to social environments, ambition and need to achieve, assertiveness, cooperation, decisiveness, dependability, dominance, energy, persistence, self-confidence, tolerance for stress, and willingness to assume responsibility. Personal trait theories of leadership also include intelligence, conceptual clarity, creativity, diplomacy, fluency in speaking, knowledge about group tasks, administrative abilities, persuasion, social skills, integrity, industry, energy level, assertiveness, emotional stability, conscientiousness, and agreeableness. Hogan, Curphy, and Hogan (1994) offer a succinct review of psychological research on adult leadership as personal qualities and traits.

**Leadership as an interaction between personal qualities and environmental resources and needs.** Theories that emphasize leadership as personal qualities interacting with environmental resources and needs define leadership situationally. These theories emphasize the moderating influences situational variables exert between a leader's behaviors and outcomes. Different situations are assumed to require different patterns of leadership traits and skills. Thus, a detailed review of situational resources and needs is needed prior to the selection of a leader, and leadership is not a set of abstract personal qualities. Instead, it can be identified only in terms of the qualities needed within a well-defined environment. For example, organizations characterized by horizontal or vertical structures differ considerably in their leadership needs. Situational theories of leadership generally are thought to be the most viable and current (Yukl, 1989).

### Research on Leadership

Literature on leadership is voluminous: a recent review cites more than 7,000 books, articles, and presentations (Bass, 1990). Several taxonomies of leadership behavior have been proposed (Borman & Bush, 1993; Davis, Skulse, Helleruik, Gebelcin, & Sheard, 1992; Yukl, Wall, & Lepsinger, 1990). However, most literature on leadership is adult-oriented and is often found in nontrade publications such as *In Search of Excellence*, (Peters & Waterman, 1982), *The Change Masters* (Kanter, 1983), *Leaders: The Strategies for Taking Charge* (Bennis & Nanus, 1985), and *The New Leader* (Morrison, 1992). Thus, a review of existing instruments with attention to measures for children and youth is useful to educators designing and implementing programs to develop leadership in students.

### Procedures

This study was designed to locate and review all existing standardized and commercially available measures specifically assessing leadership in children, youth, and adults. Measures that include a leadership subscale but are not specifically designed to assess leadership were identified and reviewed also. Scales available only in journals or technical reports or those no longer published commercially were excluded. References that both list and review tests were consulted along with test catalogs from various publishers in forming the list of measures. Reviews in the *Mental Measurement Yearbooks* also were considered. All eleven measures included in this review were acquired from test publishers and personally examined.

Some scales measure leadership exclusively while others assess broader qualities (e.g., giftedness or interpersonal behaviors). The latter class of general scales can be subdivided further into those which have an actual leadership subscale or those which do not have a specific leadership subscale but are used to assess leadership. Tests without a specific leadership subscale were examined but not reviewed.

The review organizes leadership tests in three sections: scales specifically designed to assess leadership in children and youth, scales measuring more general qualities which include a leadership subscale for children and youth, and scales measuring adult leadership. Each review includes information on the types of leadership measured, the intended use or purpose of the scale, and information on the scale's psychometric properties. Because some measures report little psychometric data, some descriptions are more brief than others. Tests specifically designed to assess leadership in children and adults are summarized in Table 1.

The inclusion of adult measures was based on two premises. First, some adult measures may be suitable for use in programs for talented students at the secondary levels. Thus, information on them may be relevant to those planning programs at this level. Second, one or more adult measures of leadership may be found to be of superior psychometric quality and thus may serve as models for researchers interested in developing measures of leadership for children. Information on the tests' psychometric qualities is summarized in Table 2. All tests can be self-scored and are appropriate for both individual and group administration.

### Findings

#### Tests of Leadership in Children and Youth

**Leadership Ability Evaluation (LAE)** The LAE (Cassel & Stancik, 1982) measures decision-making patterns or social climate created by someone in a leadership position. This 50-item paper-pencil multiple-choice test is designed to be self-administered by persons beyond grade 8. Based on work by Flanagan (1952), four leadership decision styles are assessed: laissez faire, democratic-cooperative, autocratic-submissive, and autocratic-aggressive. Scores are provided for each of the four decision styles as well as a total score.

**Table 1**  
**Measures of Leadership in Children and Adults**

<i>Title</i>	<i>Author(s)/ Publisher</i>	<i>Age Range</i>	<i>Respondent: Self/Other</i>	<i>Administration/ Scoring Time</i>
Leadership Ability Evaluation	Cassel, Stancik; Western Psychological Services	Grade 9 to Adult	Self	15-20
Leadership Skills Inventory	Karnes, Chauvin; D.O.K. Publishers	Grades 4-12	Self	45
Leadership Appraisal Survey	Hall; Teleometrics International	Adults	Other	Varies
Styles of Leadership Survey	Hall, Williams; Teleometrics International	Adults	Self	Varies
Leadership Opinion Questionnaire	Fleishman; Science Research Associates/London House	Adults	Self	15-25
Supervisory Behavior Description Questionnaire	Fleishman; Science Research Associates/London House	Adults	Other	20-30
Campbell Leadership Index	David Campbell, Allen Kraut; National Computer Systems	Adults	Self & Others	20-30

In addition, responses also can be analyzed according to one of five life activity areas: home and family, work and vocational, play and avocational, school and educational, and community.

The LAE was normed on more than 2,000 typical individuals and 400 leaders. However, information on characteristics of the standardization sample and how they were selected is not provided in the test manual. Split-half coefficients for the four decision-making styles are variable, ranging from .29 to .91, while split-half coefficients for the total scale range from .71 to .91. Test-retest data are not reported. The authors contend that the content validity of the LAE was assured through constructing the scale's items to be consistent with Flanagan's (1952) four-part theory of leadership and through validation methods utilizing four expert judges. However, information on how the items were selected or grouped into subscales as well as empirical evidence to support these four factors is not provided.

Many response options appear too broad. The items may not assess leadership qualities or leadership situations, and several items are similar to the items on Cassel's Test of Social Insight. A previous reviewer criticized the rigid use of a total score, derived from a complicated formula based on the results from one study (Black, 1970). Empirical evidence supporting a clear interpretation of LAE data generally is lacking and much of the test data are difficult to interpret (Gibb, 1970). Hence, the use of the LAE is not recommended.

**Leadership Skills Inventory (LSI)** The LSI (Karnes & Chauvin, 1985) is a 125-item paper-pencil or computer-administered inventory designed to assess leadership abilities of children and adolescents in grades 4 through 12. Children are asked to rate their competencies using a four-point Likert scale. Nine dimensions of leadership are measured: fundamentals of leadership, written communication skills, speech communication skills, values clarification, decision-making skills, group dynamics skills, problem-solving skills, personal development skills, and planning skills. The test results are intended to assist students in learning about and developing their leadership skills.

The LSI was standardized on 452 children from seven geographic regions. Little information is available about the methods used to norm the LSI (Lee, 1989). No information is provided about how the children were chosen or about their age, race, and socioeconomic status.

Most split-half and Kuder-Richardson reliability coefficients are in the .80s for each of the nine leadership dimensions. The test-retest stability of several subscales is lower (Kerr, 1989), with six of the nine subscales reporting test-retest reliability coefficients under .50.

Although the authors state that the nine leadership skills are based on a review of the literature, they do not associate the skills with any identified theory of leadership. In a previous review, Kerr (1989) noted that the use of adult professionals and members of various youth organizations as judges, rather than individuals more knowledgeable about leadership qualities, was questionable. The lack of

**Table 2**  
**Psychometric Properties of Leadership Measures**

<i>Test</i>	<i>Norms</i>	<i>Reliability</i>	<i>Validity</i>
Leadership Ability Evaluation	<i>N</i> = 2,000 typical individuals & 400 leaders; not differentiated by subgroup	split-half for total scale: .71 to .91; split-half for subscales: .29 to .91; no test-retest	content: expert judges; no criterion-related or construct
Leadership Skills Inventory	<i>N</i> = 452 children; not differentiated by subgroup	split-half and Kuder-Richardson coefficients in .80s for each scale; test-retest: 6 of 9 coefficients less than .50	content: questionable; no criterion-related or construct
Leadership Appraisal Survey	<i>N</i> = 3,176 persons rating their leaders; not differentiated by subgroup	median coefficient of stability greater than .70	not available
Styles of Leadership Survey	<i>N</i> = 2,844 leaders; not differentiated by subgroup	median coefficient of stability greater than .70	not available
Leadership Opinion Questionnaire	<i>N</i> = 5,716 individuals in 37 occupations; differentiated by gender & occupational group	internal consistency: .62 to .89; test-retest: .67 to .80	content: excellent; criterion: correlated with other measures; construct: factor analysis
Supervisory Behavior Description Questionnaire	<i>N</i> = 1,048; differentiated by gender & occupational group	split-half: .68 to .98; test-retest: .46 to .87	content: excellent; criterion: correlated with other measures; construct: factor analysis
Campbell Leadership Index	<i>N</i> = 1,709 self-respondents only, 1,411 self-respondents with other respondents, and 6,245 other respondents: selected but not identifiable by subgroups	Median Alpha coefficients of .85 for orientations and .72 for scales. Median interrater reliability is .70. Median test-retest reliability is approximately .89 for self and .87 for others	content: excellent; criterion-related: numerous small studies; construct: factor analysis

concurrent or construct validity data weakens the LSI as a measure of leadership.

#### **Tests for Children and Youth with a Leadership Subscale**

**Eby Gifted Behavior Index** The Eby Gifted Behavior Index (Eby, 1989) consists of a product rating scale and six paper-pencil checklists used to assess gifted behavior in six talent areas: verbal, math-science-problem-solving, musical, visual-spatial, social-leadership, and mechanical-technical-inventiveness. The checklists are intended to be used by teachers familiar with qualities being assessed. Although a specific age range is not reported, the Eby checklists apparently are intended for use with all ages (Sweetland & Keyser, 1991). Scores on the differ-

ent checklists may be used for screening and selection of students for inclusion in gifted programs.

The Eby checklists utilize a five-point Likert scale to assess 10 behavioral attributes identified in a literature review of gifted behavior: perceptiveness, active interaction with the environment, reflectiveness, persistence, independence, goal orientation, originality, productivity, self-evaluation, and communication of findings. Two items associated with each of ten behavioral areas of giftedness appear on each of the six checklists. The Social Leadership Checklist is comprised of 2 items associated with each of these areas for a total of 20 items.

Norms and reliability data for the Eby measures are not provided in the manual. Although some evidence for the validity of the Eby scales is provided, all six of the behavioral checklists were grouped together in making this determination. Hence, separate data on the validity of the Social Leadership scale are not available. Six validity studies were conducted using teachers and a set of judges. Outside judges and teachers rated students' art and writing products using the Product Rating Scale. Teachers also rated the same students' behavior using the six behavior checklists. Teacher ratings of students' behaviors reportedly correlate moderately with judges' ratings of student products (average correlation of .46). The average correlation between teacher and judges' ratings of products is .47.

The authors suggest that the Eby checklists and Product Rating Scale are valid measures. However, the scales are not based on an identified theory of giftedness, the statistics supporting the psychometric quality of the individual scales have not been gathered adequately, and the reliability and validity of the Social Leadership Checklist are questionable. Available studies correlate teachers' ratings of students' behavior with products or outcomes rather than with behaviors that reflect leadership. Hence, criterion-related validity of the instrument also is lacking. The use of the Eby Gifted Behavior Index to assess leadership is not recommended.

**Gifted and Talented Screening Form (GTSF)** The GTSF (Johnson, 1979) is a 24-item scale for use with students in grades K through 9. Items from the GTSF are grouped into six content areas, each having four items: academics, intelligence, creativity, leadership, visual and performing arts, and psychomotor-athletics and mechanics. Parents and teachers rate students on a five-point Likert scale based on the frequency that various characteristics of giftedness are observed.

Item selection procedures used information provided by 10 teachers who rated 40 gifted students using the GTSF. The gifted and talented students then were matched by grade with 40 students not in a gifted program. The 24 items which comprise the scale significantly discriminated between the two groups.

Norms are unavailable. Reliability of the GTSF was demonstrated in two ways. Johnson (1979) reports that Kendall correlation coefficients comparing each item with the total score for a sample of 208 children were statistically significant. The split-half coefficient correlating odd and even items was .90. Although estimates of internal consistency for the leadership scale are satisfactory, other reliability indices are lacking.

In terms of content validity, data examining relationships between the GTSF and other measures acquired from more than 18 school districts are reported in the manual. Most relevant to this review, GTSF leadership subscale scores correlate moderately with the Social Interaction and Creativity in Communication System leadership (.27),

communication (.29), and creativity (.21) scales. Lacking norms and suitable evidence of reliability and validity, the use of this measure is questionable.

**Gifted Evaluation Scale (GES)** The 48 item GES (McCarney, 1987), designed to screen for gifted behaviors in ways consistent with the definition of giftedness in Public Law 95-561, includes a 10-item leadership subscale. The scale is designed to be completed by educators who are familiar with the students being rated. Respondents rate the frequency a student exhibits each behavior on a five-point Likert scale.

The GES was normed on 2,276 students in grades K through 12 and age 4.5 to 19 years from 72 school districts in 26 states. Item-total score correlations for the leadership subscale range from .80 to .91 while coefficient alpha, a measure of internal consistency, is .96. The total test-retest reliability for the full scale is .91. Interrater reliability coefficients for various age groups are in the low .90s. Content validity for the GES was established by creating and selecting an item pool according to the suggestions of 37 diagnosticians and educational personnel. Criterion-related validity for the leadership subscale reportedly was established by comparing the GES with the System of Multicultural Pluralistic Assessment (SOMPA). For a sample of gifted students, the correlation between the GES leadership subscale and the SOMPA was .38 ( $p < .05$ ). Despite the leadership subscale having the lowest correlation with the SOMPA, the authors note that the SOMPA contains fewer measures of leadership than the GES. Although the GES items were constructed to be consistent with the federal definition of giftedness, the leadership items were not constructed to be consistent with an identified theory of leadership.

The availability of norms and satisfactory estimates of internal consistency and test-retest reliability are positive features. The authors also have addressed the content validity of the full scale. Construct validity of the leadership subscale is more problematic, and the veracity of the concurrent data is questionable. Despite the authors claims that the GES was compared to the SOMPA, such a comparison does not appear possible. The SOMPA consists of a series of diverse measures (e.g., health history, visual-motor ability, intelligence, adaptive behavior) and does not yield a single score. Its use as a suitable concurrent measure is questionable.

**Scales for Rating the Behavioral Characteristics of Superior Students (SRBCSS)** The 95-item SRBCSS (Renzulli, Smith, White, Callahan, & Hartman, 1976) consists of 10 subscales, 1 of which assesses leadership characteristics. This paper-pencil measure is designed for children and adolescents; however, a specific age range is not specified in the manual. The SRBCSS is intended to solicit teacher judgments in identifying students who might be classified as gifted and talented. The manual reports that



several schools were involved in the field testing of the SRBCSS; however, norms are not reported. The SRBCSS subscales are intended to be interpreted separately and a total score is not provided.

Test-retest and inter-rater reliability data were gathered through pre- and post-ratings of a group of fifth and sixth grade gifted students by two sets of teachers. The stability coefficient for the leadership subscale range is .77 while the interrater reliability coefficient is .67. Internal consistency coefficients of the leadership subscale for a group of fourth through sixth grade students range from .30 to .93. Criterion-related validity was studied by determining the instrument's adequacy in discriminating between a gifted and general cohort sample of fifth graders. A one-way analysis of variance was statistically significant ( $p < .01$ ) for the leadership subscale. Teacher ratings on the leadership scale were compared with sociometric peer ratings for three hypothetical leadership situations. Correlations between teacher and peer ratings among fourth through sixth-grade students ranged from .84 (fifth grade) to .23 (sixth grade).

Some support for the content validity of the leadership subscale is available in that the items were written to be consistent with characteristics identified in a literature review on leadership. However, the SRBCSS lacks comprehensive norms, demonstrates variable reliability, and does not report validity data extensively in the manual.

#### **Tests for Assessing Leadership Qualities in Adults**

**Leadership Appraisal Survey (LAS)** The LAS (Hall, 1986), a 12-item paper and pencil self-report inventory, is designed to assess leadership behavior among adults using the Blake-Mouton two-dimensional leadership grid. Intended for use in conjunction with the Styles of Leadership Survey (SLS), leadership styles are measured in terms of concern for people or concern for purpose. Concern for people lies along the vertical axis while concern for purpose lies along the horizontal axis of the grid. Instead of measuring leadership through a self-assessment, the LAS evaluates leadership behavior by asking associates to rate the person's qualities on five leadership styles: directive leadership (low concern for people, high concern for purpose), supportive leadership (high concern for people, low concern for purpose), bureaucratic leadership (low concern for people and purpose), strategic leadership (medium concern for people and purpose), and collaborative leadership (high concern for people and purpose). Scores on each leadership style reflect the strength of each style. Difference scores between the five leadership styles are calculated in order to assess leadership preference. Additional scores are provided in the areas of philosophy, planning and goal setting, implementation, and evaluation.

Norms are provided based on a sample of 3,176 persons rating their leaders. Information on reliability and validity of the LAS is lacking. Although the authors report that the LAS has similar construct and concurrent validities to the

Styles of Management Inventory and the Management Appraisal Survey from which it was adapted, the psychometric evidence supporting these management surveys also is inadequate. Only a few semantic differences exist between the management and leadership scales (Korman, 1978). Little is known about the subjects who constitute the norm group. Korman emphatically warns against the use of these assessments because of their psychometric inadequacies.

**Styles of Leadership Survey (SLS)** The SLS (Hall & Williams, 1986), a 12-item paper and pencil self-report inventory, assesses adult leadership behavior in terms of the previously described Blake-Mouton two-dimensional leadership grid. Personal preferences for each of five styles of leadership are provided. The authors recommend using the SLS in conjunction with the Leadership Appraisal Survey (LAS), which provides a measure of leadership from the perceptions of the supervisor's associates. Scores from the SLS are entered alongside the LAS scores to compare the supervisor's perceptions with associates' perceptions.

Norms are derived from 2,844 leaders in various occupations undifferentiated by subgroup. As with the LAS, information on the reliability and validity of the SLS is limited to a statement about the median coefficient of stability being greater than .70. Although this measure does have some face validity, its psychometric problems are similar to those of the LAS. Thus, the use of both the LAS and SLS seems questionable.

**Leadership Opinion Questionnaire (LOQ)** The LOQ (Fleishman, 1989a), a 40-item adult, paper-pencil test, measures two aspects of leadership: consideration and structure. Consideration refers to the extent individuals have relationships with subordinates characterized by mutual trust, respect, and warmth. Structure refers to the extent individuals structure their own role and their subordinates' roles toward goal attainment. Fashioned after the Leader Behavior Description Questionnaire (which is out of print) and designed for use in conjunction with the Supervisory Behavior Description Questionnaire (SBD), the LOQ was produced in conjunction with the Ohio State University leadership studies. Its primary use is for leaders to quantify their thoughts about how they should be acting in their supervisory role. The LOQ differs from the SBD in its focus on leaders' self-perceptions of how they *should* behave rather than on subordinates' perceptions of their actual behavior (i.e., the SBD). The LOQ has been used in industrial, business, educational, hospital, military, governmental, and research settings for selection, appraisal, counseling, and training of those in supervisory and prospective supervisory roles.

The LOQ manual (Fleishman, 1989a) summarizes various studies documenting the criterion-related validity of the LOQ, analysis of demographic variables, correlations with other measures, means and standard deviations for leaders and supervisors in different occupations, and

updated norm tables. Based on data collected from 5,716 individuals in 37 occupational groups from 17 organizations, norms are differentiated by gender and managerial versus nonmanagerial personnel. Internal consistency coefficients range from .62 to .89 while test-retest reliabilities range between .67 and .80. Construct validity of the LOQ was affirmed through factor analytic studies and item selection procedures. Items were selected for each of the two scales that loaded high on one factor and low on the other (i.e., consideration and structure factors). Subsequent studies indicated the two factors are uncorrelated.

The LOQ requires respondents to indicate how they should behave rather than how they actually act. Thus, the measure does not assess current and actual leadership behaviors (Gibb, 1972). In addition, the scale is subject to social desirability (Kirchner, 1965). Despite some shortcomings, the LOQ appears to be a good measure of leadership. Its superior status is due, in part, to research efforts to determine the dimensions of leader behavior. The LOQ is promising in that it is well constructed and provides considerable psychometric evidence to warrant its use. The instrument is conservative in recommending that local validity studies be carried out if the data are to be utilized for selection (Doppelt, 1965; Gibb, 1972).

**Supervisory Behavior Description Questionnaire (SBD)** The SBD (Fleishman, 1989b), a 40-item paper-pencil questionnaire, is designed to be used in conjunction with the Leadership Opinion Questionnaire (LOQ). Whereas the LOQ focuses on leaders' self-perceptions of how they should behave, the SBD focuses on subordinates' perceptions of their leaders' behavior.

Norms are available for eight groups, including general managerial personnel ( $N=1048$ ) and subgroups such as high school teachers, bank managers, and educational supervisors. Split-half internal consistency reliability coefficients for the SBD range from .68 to .98. Test-retest coefficients for three samples of foremen, after an 11-month interval, ranged from .46 to .87. Inter-rater reliability of ratings of supervisors by more than one subordinate range from .47 to .90. The construct validity of the consideration and structure factors of the SBD was supported through negative correlations between the two scales in six out of seven samples. As with the LOQ, several factor analytic studies are reported which support the two factor structure. The SBD and LOQ provide various norms, estimates of reliability, together with evidence supporting their validity. Their continued use seems warranted.

**Campbell Leadership Index (CLI)** The CLI (Campbell & Kraut, 1991), the most recently developed measure of adult leadership, assesses 22 specific leadership qualities within five broad categories: leadership (including ambitiousness, risk-taking, enthusiasm, resourcefulness, well-connectedness, insightfulness, and persuasiveness), energy (with no subscales), affability (affection, helpfulness, supportiveness,

extroversion, cheerfulness), dependability (candidness, organized, effectiveness, frugality), and reliance (easygoing, adaptability, resiliency, trust). A 100-item adjective checklist is completed both by the target person (i.e., a self-responder rating) as well as by three to five persons who are familiar with the person (observer ratings). According to Campbell, leadership refers to actions which focus resources to create desirable opportunities. The scale is designed for use across a broad range of settings including universities, corporations, military, government agencies, and leadership training courses.

Norms from 3,102 adults comprise the self-responder scale and from 6,245 for the observer rating scale. Thirty standardization samples drawn from various settings include college student leaders, managers with various levels of experience, front line supervisors, and government executives. Test protocols must be returned to the publisher for computer scoring and report generation (which reports both self and observer scores). The process used to obtain the norms and the exact groups on which scores are derived are unclear and confusing. In addition, information on gender, race, and other demographic data is unavailable.

Alpha coefficients range between .56 and .90 with median coefficients of .85 for the 5 broad scales and .72 for the 22 individual scales. Discriminant validity studies of self- and observer-ratings generally converge or agree in expected ways, with correlations ranging between .42 and .66. Off-diagonal correlations were appropriately lower, ranging from .07 to .43.

The CLI is a useful tool for the assessment of leadership qualities in adults. It is based on a well articulated theory, one consistent with previous research on leadership as trait qualities as well as the test's own empirical evidence. The measure generally displays adequate psychometric qualities, including current norms. If used, the measure requires considerable planning and a noticeable financial investment.

## Discussion

In a commentary on gifted education, Shore, Cornell, Robinson, and Ward, (1991) identified leadership programs for talented students as a research priority. The first priority they identified was the need to demonstrate that leadership constitutes an array of qualities that can be assessed suitably. The results from this review suggest that this need remains largely unmet for children, but that it is attainable for adults.

Significant deficiencies exist in the assessment of leadership among children and youth, and few standardized measures of leadership are available. Among the eleven we located that purport to assess leadership, only one, the LSI, was designed specifically to measure leadership in children and youth. Among the remaining ten, four assess various qualities associated with giftedness and typically include one subscale designed to measure leadership. Five are intended to

measure leadership primarily in adults, and one measures leadership skills in individuals grade nine to adult. Given the importance of leadership and its emphasis through federal legislation and state and local school policy, we might expect the number of leadership measures to be greater.

We also wish their psychometric qualities were higher. At present, the normative basis of most leadership scales is inadequate. Among the eleven, three are not normed. One test, the LAE, has norms over thirty years old. Among those that have norms, sample sizes for the measures typically are small, averaging 2,539. All provide insufficient information on important demographic factors (e.g., age, gender, ethnicity, geographic region, and income level), thus compromising our ability to know the standards against which testees are compared. Only one instrument, the GES, allows comparison by age by providing separate norm tables for different age groups (boys age 4.5 to 10 and 10 to 18 and girls age 4.5 to 9, 9 to 12, and 12 to 18). The GES also is unique in that it provides demographic data on the norm group such as ethnicity, geographic region, and parental occupation.

Although eight test manuals report information on estimates of internal consistency, and/or stability, the reliability reports are not always comprehensive. The information, when reported, indicates that the internal consistency and stability of most measures are highly variable. In reference to the four subscales that embed their measure of leadership within broader scales, estimates of internal consistency and stability can be expected to be appreciably lower for the subscale than that reported for the broader scale. Among the four, only one reports both internal consistency and stability estimates separately for the leadership subscale.

Issues of validity are somewhat more complex. Little evidence of construct validity exists in most test manuals except for the LOQ, SBD, and CLI. Evidence for concurrent and predictive validity is generally weak. The LOQ, SBD, and GES demonstrated positive correlations with other measures of intelligence and personality. The LAE attempted to demonstrate concurrent and predictive validity by determining whether the test discriminates between leaders and typical persons and whether changes in leadership style occur following training. The CLI reports numerous concurrent studies. As noted in the LOQ manual (Fleishman, 1989a), studies correlating leadership factors with external criteria of effectiveness need to be conducted to better demonstrate the criterion-related validity of these instruments.

Socially desirable responding appears to be a problem with self-report leadership instruments such as the LAE, LSI, SLS, and LOQ. Although most self-report measures assume test-takers will respond truthfully and without bias, the instruments' subjective nature invite socially desirable responding. Kerr (1989) believes that measures like these are not suitable for identification of leadership

characteristics or potential leaders. The CLI suggests using observer ratings in preference to self-ratings to overcome this problem.

Thus, a review of prevailing psychometric qualities among measures of leadership leads to the following conclusions. The measures typically are normed inadequately and lack information about reliability and validity. Those that report information about reliability usually provide variable and sketchy data. Information on validity appears rarely and when reported, is generally insufficient to establish construct validity. The reliance on short subscales embedded within other scales, which themselves are often inadequate, is disturbing and should be avoided. We clearly lack the assessment technology to adequately measure leadership in children and youth.

**Leadership Scales and Leadership Concepts.** Among the 11 tests, none appears to measure leadership as power and influence. None appears to measure leadership as skillful management of behavior. Seven appear to measure leadership as traits. One scale, the CLI, appears to be somewhat promising as a measure that enables one to conceptualize leadership as an interaction between personal and environmental qualities. Three measures lack clarity as to their conceptual base. Given their brevity, the conceptual basis of the four scales that contain leadership subscales are most difficult to evaluate.

The tendency for tests to rely on trait theories of leadership is understandable. Until recently, trait theories of leadership were dominant. In addition, techniques useful in the development of measures that assess traits are readily available and their use widespread. In contrast, techniques used to develop measures that assess interactions between traits and environments are less known and their use more restricted. To be sure, trait concepts of leadership remain important when situational theories are utilized; traits comprise a portion of situational theory. Some measures, such as the LAE which analyzes responses in different settings such as work, home, and community, appear to take into account both trait and situational components of leadership. However, the primary focus of the LAE is to assess leader behavior according to different decision styles or traits.

At present, measures of student leadership based on situational qualities are unavailable. Thus, educators interested in implementing programs based on theories of situational leadership are unable to purchase a turn-key system. For them, four options exist. One is to select on the basis of other qualities for which more suitable measures are available (e.g., specific subject matter or academic aptitudes). In this case, educators are serving academically talented students who may or may not display leadership. While academic programs may be quite worthy, we have not matched the identification of youthful leaders with programs which promote leadership. The decision to abandon programs designed to identify critical leadership qualities



because of inadequate measures clearly is not warranted. Option two is to identify students with leadership potential through scales that are inconsistent with current theory. This option is not recommended as it would jeopardize the program's integrity and minimize its effectiveness by failing to identify latent potential. Option three is to alter the conceptual basis of one's leadership program to be in accord with an existing quality measure of leadership. The selection of this option severely restricts curriculum options and again overemphasizes the prominence of tests.

The fourth option, to take the best existing measures and supplement them by developing additional assessment procedures, is the most defensible professionally. Not unexpectedly, considerable work is needed to develop and implement identification procedures for situationally based leadership programs. Locally developed measures of leadership are likely to have psychometric qualities inferior to those reviewed in this article. Programs seeking this direction for assessing leadership need to invest in developing, implementing, and evaluating identification measures and procedures. Some of the adult leadership measures reviewed here, including the LOQ, SBD, and CLI, could serve as models and a launching point for constructing leadership measures for children and youth.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bass, B. M. (1990). Bass and Stogdill's handbook of leadership (3rd ed.). New York: Free Press.
- Bennis, W. G., & Nanus, B. (1985). *Leaders: The strategies for taking charge*. New York: Harper & Row.
- Black, J. D. (1970). Review of the Leadership Ability Evaluation. In O. K. Buros (Ed.), *Personality and test reviews* (pp. 1083-1085). Highland Park, NJ: Gryphon Press.
- Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, 6, 1-21.
- Campbell, D., & Kraut, A. (1991). *Campbell Leadership Index*. Minneapolis, MN: National Computer Systems, Inc.
- Cassel, R. N., & Stancik, E. J. (1982). *The Leadership Ability Evaluation-Revised: Manual*. Los Angeles: Western Psychological Services.
- Doppelt, J. E. (1965). Review of the Leadership Opinion Questionnaire. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (pp. 1371-1372). Lincoln, NE: The University of Nebraska Press.
- Eby, J. W. (1989). *Eby Gifted Behavior Index: Administration manual*. East Aurora, NY: D.O.K. Publishers.
- Flanagan, J. A. (1952). *USAF research contract on leadership structure*. Lackland Air Force Base: The Air Force Personnel and Training Research Center.
- Fleishman, E. A. (1989a). *Leadership Opinion Questionnaire: Examiner's manual*. Park Ridge, IL: Science Research Associates.
- Fleishman, E. A. (1989b). *Supervisory Behavior Description Questionnaire: Examiner's manual*. Park Ridge, IL: Science Research Associates.
- Gibb, C. A. (1970). Review of the Leadership Ability Evaluation. In O. K. Buros (Ed.), *Personality and test reviews* (p. 1085). Highland Park, NJ: Gryphon Press.
- Gibb, C. A. (1972). Review of the Leadership Opinion Questionnaire. In O. K. Buros (Ed.), *The seventh mental measurements yearbook* (pp. 1529-1531). Lincoln, NE: The University of Nebraska Press.
- Hall, J. (1986). *Leadership Appraisal Survey*. The Woodlands, TX: Teleometrics International.
- Hall, J., & Williams, M. S. (1986). *Styles of Leadership Survey*. The Woodlands, TX: Teleometrics International.
- Hogan, R., Curphy, G., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American Psychologist*, 49(6), 493-504.
- Johnson, D. L. (1979). *Gifted and Talented Screening Form: Instruction manual*. Chicago: Stoelting Co.
- Kanter, R. M. (1983). *The change masters*. New York: Simon & Schuster.
- Karnes, F. A., & Chauvin, J. C. (1985). *Leadership Skills Inventory: Administration manual*. East Aurora, NY: D.O.K. Publishers.
- Kerr, B. A. (1989). Review of the Leadership Skills Inventory. In J. C. Conoley, J. J. Kramer, & L. L. Murphy (Eds.), *The tenth mental measurements yearbook* (pp. 438-439). Lincoln, NE: The University of Nebraska Press.
- Kirchner, W. K. (1965). Review of the Leadership Opinion Questionnaire. In O. K. Buros (Ed.), *The sixth mental measurements yearbook* (pp. 1372-1373). Lincoln, NE: The University of Nebraska Press.
- Korman, A. K. (1978). Review of the Styles of Leadership and Management. In O. K. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 1763-1764). Lincoln, NE: The University of Nebraska Press.
- Lee, S. W. (1989). Review of the Leadership Skills Inventory. In J. C. Conoley, J. J. Kramer, & L. L. Murphy (Eds.), *The tenth mental measurements yearbook* (pp. 439-441). Lincoln, NE: The University of Nebraska Press.
- McCarney, S. B. (1987). *The Gifted Evaluation Scale: Technical manual*. Columbia, MO: Hawthorne Educational Services.
- Peters, T. J., & Waterman, R. H. (1982). *In search of excellence*. New York: Harper & Row.
- Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., & Hartman, R. K. (1976). *Scales for Rating the Behavioral Characteristics of Superior Students*. Mansfield Center, CT: Creative Learning Press.
- Shore, B., Cornell, D., Robinson, A., and Ward, V. (1991). *Recommended practices in gifted education*. New York: Teachers College Press.
- Stogdill, R. M. (1974). *Handbook of leadership: A survey of theory and research*. New York: Free Press.
- Sweetland, R. C., & Keyser, D. J. (Eds.). (1991). *Tests* (3rd ed.). Austin, TX: Pro-Ed.
- Yukl, G. A. (1989). *Leadership in organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Yukl, G. A., Wall, S., & Lepsinger, R. (1990). Preliminary report on the validation of the management practices survey. Clark & M. B. Clark (Eds.), *Measures of leadership* (pp. 223-238). West Orange, NJ: Leadership Library of America.