

A known group analysis validity study of the Vanderbilt Assessment of Leadership in Education in US elementary and secondary schools

**Elizabeth Covay Minor · Andrew C. Porter ·
Joseph Murphy · Ellen B. Goldring · Xiu Cravens ·
Stephen N. Elloitt**

Received: 18 September 2013 / Accepted: 30 December 2013 /
Published online: 12 January 2014
© Springer Science+Business Media New York 2014

Abstract The Vanderbilt Assessment of Leadership in Education (VAL-ED) provides educators with a tool for principal evaluation based on principal, teacher, and supervisor reports of principals' learning-centered leadership. In this study, we conduct a known group analysis as part of a larger argument for the validity of the VAL-ED in US elementary and secondary schools. We asked superintendents to select the principals in their district who they believe in performance of their duties are in the top 20 % and the bottom 20 %. We ask how accurately VAL-ED scores can identify membership of the two known groups. Using a discriminant analysis, the VAL-ED places principals in the superintendent groups, on average, 70 % of the time for both elementary and secondary schools. Placement accuracy is greater for the top group than the bottom group.

E. Covay Minor (✉)
College of Education, Michigan State University, 620 Farm Lane, Erickson Hall Room 516A, East
Lansing, MI 48824, USA
e-mail: covayeli@msu.edu

A. C. Porter
University of Pennsylvania, Philadelphia, PA, USA
e-mail: andyp@gse.upenn.edu

J. Murphy · E. B. Goldring · X. Cravens
Vanderbilt University, Nashville, TN, USA

J. Murphy
e-mail: joseph.f.murphy@vanderbilt.edu

E. B. Goldring
e-mail: ellen.goldring@vanderbilt.edu

X. Cravens
e-mail: xiu.cravens@vanderbilt.edu

S. N. Elloitt
Arizona State University, Phoenix, AZ, USA
e-mail: steve_elliott@asu.edu

Keywords Known group study · Learning-centered leadership

Over the last three decades, evidence has been accumulating on the importance of the principal in school improvement (Leithwood and Montgomery 1982; Murphy et al. 1983b; Murphy et al. 1985; Bryk et al. 2010). Research on each new generation of school reform starting with effective schools in the late 1970s (Brookover and Lezotte 1977; Edmonds and Frederiksen 1978) and carrying through to more recent studies (Preston et al. 2012) has linked effective leadership to improvement defined in terms of learning outcomes for students. The cumulative message from increasingly sophisticated studies and well-crafted quantitative reviews is that principals have small but significant indirect effect on student achievement (Hallinger and Heck 1996, 1998; Leithwood et al. 2010; Witziers et al. 2003) and that principals are almost always the maestros of school improvement processes (Bryk et al. 2010; Murphy 2013; Preston et al. 2012).

We have also learned over the same period about the actions of leaders that are connected to highly productive schools, institutions in which all youngsters reach ambitious targets of performance. In particular, instruction focused or learning-centered leadership¹ behaviors have routinely been underscored as critical for principals conducting successful school improvement work (Leithwood et al. 2004; Leithwood et al. 2006; Marks and Printy 2003; May and Supovitz 2011; Preston et al. 2012; Robinson et al. 2008; Supovitz et al. 2009). Numerous portraits, taxonomies, and models of learning-centered leadership have been crafted across these three decades (for more recent formulations, see Murphy et al. 2007; Robinson et al. 2008; Bossert et al. 1982; Hallinger and Murphy 1985; Murphy et al. 1983a for early conceptualizations) with almost all reaching consensus around the importance and centrality of learning-centered leadership behaviors. These behaviors entail such actions as implementing a rigorous curriculum in the school and promoting high quality instruction. One of the key questions facing the field now is how best to develop, incentivize, and support learning-centered behaviors. There is a small set of lever points that can be engaged to influence principal learning-centered leadership (see Murphy and Shipman 1999; Murphy et al. 2000; Young et al. 2009). Some of these leverage points address the “education of leaders” (e.g., preparation, induction, and continuing education). Other leverage points fall into the category of “conditions of work” (e.g., incentives; control over aspects of the job such as hiring). One of the most critical of the leverage points from the second cluster is principal evaluation, a domain that has been amplified in recent times by the larger accountability and effective schools movement and was subsequently nourished by No Child Left Behind, Race To The Top, and federal waivers to states (Murphy 1989; Murphy 1992; Murphy et al. 2012).

As reforms have more recently been directed toward principal evaluation systems, weaknesses have become increasingly visible (Goldring et al. 2009b). Two problems of especial note emerge. First, many systems have not been constructed using empirically or theoretically grounded understanding of effective leadership or findings from the larger body of school improvement. Second, most principal evaluation systems do not adhere to professional standards (AERA, APA, and NCME) for personnel evaluation.

¹ We define principal learning-centered leadership as intentional mediated actions focused on student learning.

Many fall short on the criteria of employing multiple criteria, and most can amass no evidence on the reliability and validity of the procedures employed.

Our team was formed to create an assessment instrument that would overcome these weaknesses.² The tool, the Vanderbilt Assessment for Leadership in Education (VAL-ED) instrument, is a replicable research-based evaluation tool. It is a behavior inventory of learning-centered leadership based on ratings from the principal herself/himself, teachers, and the principal's supervisor. Over the last 4 years, VAL-ED has been used in over 4500 schools throughout the USA. The VAL-ED underwent an extensive construction and development phase, which included a sorting study, cognitive labs, an item bias study, and two pilot test studies. Following the necessary revisions that resulted from these studies, the VAL-ED was considered ready to be tested on a larger scale (Porter et al. 2008). In the spring of 2008, the team conducted a nationally representative field trial, which included 235 schools. Guiding by the Standards for Educational and Psychological Testing (American Educational Research Association et al. 1999), the study examined reliability, validity, bias, scales, norms, performance standards, and score comparability. The study provides evidence of the validity of the VAL-ED's conceptual framework, high reliability, and low standard errors (see, Porter et al. 2010b for more information). Currently, we are engaged in a series of studies that explore the psychometric properties and examine the validity of the instrument the builds off of the initial psychometric development of the instrument (see, Cravens et al. 2013; Porter et al. 2010a, b; Polikoff et al. 2009).

The current focus in validity is on validity as an argument where explicit statements of interpretations of scores are made and evaluated (Kane 1992, 2001). The argument involves four parts including examining evidence of the statements in terms of the interpretation, assessing alternative explanations, examining the consequences of use of the scores, and evaluating whether the interpretations are consistent (American Educational Research Association et al. 1999; Kane 2001). The overall validity argument is focused on the combination of these four parts.

We interpret the principal's scores on the VAL-ED to be an indication of the extent to which the principal is performing actions that promote student learning within their school. Our interpretation of the meaning of the VAL-ED scores is based on an extensive review of the literature (Goldring et al. 2009a), work with practitioners and policymakers (Elliott et al. 2008; Cravens et al. 2013), and the national field trial (Porter et al. 2010b). In an effort to continue to establish the validity and reliability of the VAL-ED as a measure of learning-centered leadership behaviors, the team is conducting a series of five studies that utilize data from schools that have purchased the VAL-ED, and who are considered "real users." These studies include a test-retest study, a convergent-divergent study (Goldring et al. 2013), an examination of how schools use results, and a study examining the value added of principal leadership to student achievement. The current paper is one piece of the larger validity argument, and thus alone does not address the totality of our argument of validity. We will use the combination of these the five studies to triangulate evidence of validity of the VAL-ED to support our argument (Kane 1992). In this paper, we advance that work employing a known group validity design and analysis. The purpose of the known

² We wish to acknowledge the generous support of the Wallace Foundation whose grant to Vanderbilt University made this research-anchored development work possible.

group validity study is to assess the VAL-ED as a measure of principal learning-centered leadership through comparing superintendent assessments of principals to the results of the VAL-ED.

In the balance of this introductory section, we outline the VAL-ED conceptual framework and describe the instrument in more detail. In Section 2, we present our methodology, including data and procedures. Results are then presented. We close with a discussion of our findings.

1 VAL-ED conceptual framework

The conceptual framework of the VAL-ED focuses on leadership behaviors through two dimensions: *core components* and *key processes*. The VAL-ED assesses the effectiveness of learning-centered leadership behaviors which are defined by core components created through key processes. In other words, the core components capture the school characteristics that are important for the instructional environment and student learning, while the key processes capture how the principals develop and nurture those characteristics (Porter et al. 2008). The core components and key processes are based on an extensive review of the literature of the elements that are present in schools with environments conducive for student learning (Goldring et al. 2009a), and fit squarely into the eight major dimensions of behavior of learning-centered leadership as found in past research (Murphy et al. 2007). Additionally, the core components and key processes align with the Interstate School Leaders Licensure Consortium standards (Goldring et al. 2009a). For a full review of the literature supporting the core components and key processes, see Goldring et al. (2009a).

There are six core components: high standards for student performance, rigorous curriculum (content), quality instruction (pedagogy), culture of learning and professional behavior, connections to external communities, and systemic performance accountability.

High standards for student learning indicate a school environment that has clear goals that promote rigorous academic and social learning for their students. Effective schools must also provide a *rigorous curriculum*—challenging academic content—to all of the students within the school. That rigorous curriculum needs to be delivered to students through *quality instruction*, which includes instructional practices that provide students with the opportunities to learn the content. The school culture must promote student learning by being an environment that is a *culture of learning and professional behavior*. Effective schools also involve those outside of the school (i.e., parents, community institutions, etc.) in the goal of promoting student academic and social learning. These *connections to external communities* help to promote community collaboration with a focus on learning goals. Finally, effective schools have a system of *performance accountability*. This system places responsibility on administration, teachers, and students to promote student academic and social learning goals.

The six key processes are planning, implementing, supporting, advocating, communicating, and monitoring. These key processes represent ways that the principals work to ensure that the school develops the characteristics of an

effective school (Goldring et al. 2009a). Effective leaders must *plan*, which means that they are able to articulate the shared goals and mechanisms for developing and maintaining high standards for student achievement. Effective leaders must *implement* or put into practice the needed policies and procedures to realize high standards for student achievement. Leaders must create *supportive* conditions that promote student academic and social learning. These conditions include ensuring that the school has the appropriate financial, political, technological, and human resources to promote student learning. Leaders *advocate* for the best interests of all of their students. Effective leaders must *communicate* with and create systems of exchange among the internal and external school community members. Finally, an effective leader *monitors* what is happening within the school in order to make the best decisions for the school and for student learning. A highly effective principal uses these key processes to create the core components needed to have an effective school. The six core components by six key processes in the conceptual framework define 36 domains of learning-centered leadership behavior.

2 VAL-ED instrument

The VAL-ED instrument is composed of a sample of two leadership behaviors from each of the 36 domains of learning-centered leadership behavior. The principal, the principal's supervisor, and the teachers in the principal's school judge a principal's effectiveness on each of the two behaviors sampled from each of the 36 domains resulting in a 360° assessment. In other words, the VAL-ED is a 360° assessment because all members of the school's professional community participate in the evaluation of the principal. Each respondent is asked to first indicate the source of their evidence for making the effectiveness rating for each behavior. These sources of evidence include "reports from others," "personal observations," "school documents," "school projects or activities," "other sources," or "no evidence." The respondent is then asked to rate the principal's effectiveness for each behavior. The VAL-ED has a 5-point effectiveness scale: 1= ineffective, 2=minimally effective, 3=satisfactorily effective, 4=highly effective, and 5=outstandingly effective. Teachers and supervisors have the option to select "don't know" as an effectiveness rating; however, principals must enter an effectiveness score as they should be able to rate each behavior that they do or do not do. In addition, if principals select "no evidence" as the source of evidence, then they must select "ineffective" as the effectiveness score as in this case no evidence means the action was not completed. Teachers and supervisors can select "ineffective" or "don't know" as the effectiveness score if they do not have a source of evidence. There are two parallel forms of the VAL-ED, each consisting of a different sample of 72 behaviors across the 36 domains. The administration is via the Web or paper and pencil, but virtually all users opt for the Web-based version.

The VAL-ED 360° assessment results are reported in a total score and a subscore for each of the six core components and a subscore for each of the six key processes. Data are aggregated across the principal's response, the supervisor's response, and the average teacher's response so that each respondent type is

weighted equally. In addition, data can be reported for total score and each of the 12 subscales separately by respondent type: the principal, the supervisor(s), and/or the teachers. Results are reported in terms of average effectiveness rating across the relevant behaviors (a scale from 1 to 5) as well as in terms of national norms and nationally derived performance levels (Polikoff et al. 2009; Porter et al. 2010a, b).

To decrease burden, teachers in a school are randomly grouped into two halves and one half of the teachers rate one behavior for each of the 36 domains and the other half rate the other 36 behaviors. Data for the two random halves of teachers are then merged to form a complete teacher evaluation of the principal. A simulation study revealed no loss in reliability for this approach.

3 Validity

In 2008, the VAL-ED underwent a national field trial. The results provide evidence for the reliability and validity of the VAL-ED as an instrument to assess the learning-centered leadership of principals (Porter et al. 2010b). The VAL-ED is both norm- and criterion-referenced. The national field trial was based on research sites recruited for purposes of studying the VAL-ED. The psychometric properties of the VAL-ED for real users (i.e., districts that have purchased the VAL-ED for their own principal evaluations) remain uninvestigated. This paper takes an important step toward filling that knowledge gap by examining the validity of the VAL-ED with real users as the analytic sample.

We examine the validity by a “known groups” analysis. In this analysis, we compare two sources of information: VAL-ED scores and superintendent categorizations of the principals in their district. We created two “known groups” by having superintendents place their principals into the top performing 20 % of principals within their district and the bottom performing 20 % of the principals in their district separately for elementary schools and secondary schools. No directions were given as to how superintendents were to make their placement of principals into the two known groups. To the extent the VAL-ED accurately predicts superintendent known group categories, the VAL-ED has evidence of concurrent validity (American Educational Research Association et al. 1999). Our study does not, however, identify which of the two sources of information is the more valid, the superintendent designation or the VAL-ED scores. More specifically, we ask:

To what extent is the VAL-ED accurate at discriminating between the two known groups of school principals?

- a. Is this the same for elementary and secondary schools?
- b. Is this the same by role of the rater (teacher, self, or supervisor)?
- c. Is this the same for the top performing principals and the bottom performing principals?

4 Methods

4.1 Data

The data for the analyses in this paper come from two sources: superintendents' reports and VAL-ED scores. To qualify for recruitment in the study, districts had to have purchased and administered the VAL-ED for their own internal use of the instrument. In addition, the district had to have ten or more schools at either the elementary level and/or the secondary level so that the superintendent had a large enough pool of principals for placement into the two groups of top and bottom 20 % of principals. The recruitment process results in six school districts participating in the study. All six school districts are in the South and range in size from about 40 schools to over 300 schools. Two of the districts are in suburban areas, two of the districts are in midsize cities, one is in a rural fringe area, and the final district is in a town.

Each superintendent was contacted by our research team and asked to report their top and bottom principals based on those principals in their school district during the 2010–2011 school year.³ Once the superintendents reported the top and bottom performing 20 % of the principals in their district, the VAL-ED scores for these principals were extracted from the VAL-ED database and labeled as to whether the principal was rated by the superintendent as being in the top 20 % of principals or being in the bottom 20 % of principals. Three of our districts purchased the VAL-ED in the 2009–2010 school year but not 2010–2011, so we used VAL-ED scores from 2009 to 2010. Two principals not in the same school during the 2009–2010 and 2010–2011 school were excluded from our analyses. Our results for mean differences and discriminant analyses are consistent whether we include only those schools with VAL-ED scores from the 2010 to 2011 school year or schools from both years. As the results were consistent following the robustness test, we include all six districts. Our analytic sample includes 190 schools from those 6 school districts. For the elementary schools, there are 57 principals in the bottom 20 % of principals category and 63 principals in the top 20 % of principals category for a total of 120 elementary schools. For secondary schools, there are 35 principals in the top 20 % of principals category and 35 in the bottom 20 % of principals category for a total of 70 schools.

Not all of the schools in our sample had responses from all three rater groups: self-rating from principals, teachers, and supervisors. The 14 schools in District C have no ratings from supervisors. District D, with 15 schools, only has ratings from teachers. While the other districts have some schools that are missing various raters, the schools mostly have ratings from all three groups. When we repeat our analyses with only the schools with the ratings from all three groups our conclusions about the validity of the VAL-ED are unchanged. The results that we report include schools that do not have

³ We recognize that it is possible that the superintendent was aware of the VAL-ED scores prior to categorizations, especially for the three districts with VAL-ED scores from 2009 to 2010. These three districts do have higher accuracy rates than the other three districts. Since the districts with VAL-ED scores from 2009 to 2010 are the three smallest districts, it is possible that the superintendent has a better sense of the principals' effectiveness. However, to control for the possibility that the superintendent used the VAL-ED scores to place principals in the two groups and as a robustness test, we run the analyses without those districts. We reach the same conclusions from our results.

ratings of from all three groups for principals. We allowed our sample sizes to vary so that we use all of the information available.

4.2 Analysis

To explore the extent to which the VAL-ED distinguishes between the top and bottom 20 % of principals, we use multiple methods. First, mean differences are examined for significant differences between the two groups. The total effectiveness score and each of the 12 core components and key processes subscales are examined for significant differences for principal, teacher, and supervisor ratings. In addition, we examine the combination of principal and teacher ratings for significant differences. Normally, the school level rating is used for the aggregate effectiveness ratings; however, the school level rating would include the principal, teachers, and supervisors. If we use ratings from all three groups, there may be redundancy in the school level rating and the supervisors' categorization to the extent that the supervisor was the superintendent who categorized the principals into the top and bottom groups. To avoid redundancy of information we combine only the principal and teacher ratings for our aggregate measure to reflect a school level evaluation.

We use discriminant analysis to estimate a linear equation that most accurately discriminates principals into the two groups based on VAL-ED scores (see Burns and Burns 2008; Norusis 2003). For the discriminant analysis, we rely on the principal and teacher combination of ratings as the school level may include redundancy of the superintendent's ratings. We also report the discriminant analyses for each rater group separately.

5 Results

5.1 Mean differences

Based on our expectations principals who are classified in the bottom 20 % of district principals should have a lower effectiveness rating on the VAL-ED compared with principals rated in the top 20 % of principals. Tables 1 (elementary) and 2 (secondary) provide the mean differences of the principals' total effectiveness score, core components, and key processes. We find that consistently principals categorized as being in the top 20 % of principals within a district have a higher rating on their total effectiveness score, the core components, and the key processes compared with those rated as being in the bottom 20 % of principals.

For elementary principals, with the combined principal and teacher ratings as the school level rating, principals placed in the top 20 % of principals in their district had a total effectiveness score of 4.07. In other words, they were rated as highly effective. Conversely, principals in the bottom 20 % of principals have a total effectiveness score of 3.66, a rating between satisfactorily effective and highly effective. Using Cohen's *d* (Cohen 1988), the difference results in an effect size of 0.92, which is a large effect size. The difference is statistically significant at the 0.05 level as are all of the six differences on core components and the six differences on key processes. The effect sizes for the core components and key processes are large as well, with the exception of a medium

Table 1 Mean differences for elementary principals

	Principal and teacher		Principal		Teacher		Supervisor	
	Top N=59	Bottom N=49	Top N=59	Bottom N=49	Top N=63	Bottom N=57	Top N=49	Bottom N=46
Total effectiveness	4.07*** (0.401)	3.66 (0.488)	4.13** (0.501)	3.82 (0.619)	4.01*** (0.442)	3.46 (0.623)	4.23*** (0.575)	3.74 (0.695)
Core components								
High standards	4.16*** (0.397)	3.78 (0.495)	4.23* (0.541)	3.96 (0.599)	4.10*** (0.415)	3.56 (0.628)	4.35*** (0.559)	3.84 (0.656)
Rigorous curriculum	4.04*** (0.410)	3.68 (0.517)	4.09 (0.556)	3.88 (0.666)	3.99*** (0.439)	3.46 (0.628)	4.21** (0.750)	3.70 (0.746)
Quality instruction	4.09*** (0.389)	3.71 (0.462)	4.14* (0.507)	3.88 (0.601)	4.04*** (0.432)	3.51 (0.599)	4.29*** (0.511)	3.77 (0.658)
Culture of learning	4.16*** (0.418)	3.73 (0.537)	4.27** (0.484)	3.93 (0.711)	4.05*** (0.482)	3.49 (0.664)	4.29*** (0.578)	3.76 (0.703)
External community	3.96*** (0.552)	3.52 (0.533)	3.99** (0.626)	3.61 (0.698)	3.95*** (0.471)	3.38 (0.652)	4.11** (0.663)	3.66 (0.755)
Performance accountability	3.99*** (0.432)	3.53 (0.521)	4.03** (0.574)	3.65 (0.706)	3.94*** (0.463)	3.37 (0.622)	4.15** (0.596)	3.71 (0.821)
Key processes								
Planning	4.07*** (0.415)	3.64 (0.487)	4.13** (0.524)	3.77 (0.641)	4.01*** (0.425)	3.46 (0.594)	4.39** (0.554)	3.97 (0.727)
Implementing	4.08*** (0.409)	3.66 (0.493)	4.15** (0.507)	3.83 (0.623)	4.01*** (0.447)	3.45 (0.617)	4.29*** (0.596)	3.71 (0.751)
Supporting	4.13*** (0.393)	3.72 (0.494)	4.23** (0.508)	3.92 (0.579)	4.03*** (0.452)	3.49 (0.659)	4.27*** (0.524)	3.77 (0.752)
Advocating	4.01*** (0.410)	3.67 (0.487)	4.05 (0.544)	3.83 (0.653)	3.96*** (0.432)	3.47 (0.613)	4.07*** (0.670)	3.60 (0.651)
Communicating	4.10*** (0.424)	3.66 (0.523)	4.14** (0.532)	3.81 (0.675)	4.05*** (0.466)	3.48 (0.653)	4.22*** (0.603)	3.70 (0.719)
Monitoring	4.03*** (0.417)	3.61 (0.494)	4.05* (0.548)	3.75 (0.670)	4.01*** (0.450)	3.44 (0.622)	4.16** (0.668)	3.68 (0.757)

Standard deviations are in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2 Mean differences for secondary principals

	Principal and teacher		Principal		Teacher		Supervisor	
	Top N=32	Bottom N=33	Top N=32	Bottom N=33	Top N=35	Bottom N=35	Top N=29	Bottom N=28
Total effectiveness	3.95*** (0.229)	3.57 (0.522)	4.04 (0.419)	3.78 (0.679)	3.86*** (0.326)	3.37 (0.593)	4.21*** (0.512)	3.43 (0.680)
Core components								
High standards	4.03*** (0.250)	3.65 (0.551)	4.09 (0.441)	3.83 (0.686)	3.97*** (0.334)	3.46 (0.595)	4.27*** (0.554)	3.46 (0.720)
Rigorous curriculum	3.94*** (0.258)	3.57 (0.486)	4.02 (0.486)	3.77 (0.654)	3.85*** (0.307)	3.37 (0.567)	4.22*** (0.533)	3.46 (0.680)
Quality instruction	4.03*** (0.238)	3.67 (0.470)	4.17 (0.488)	3.94 (0.593)	3.89*** (0.313)	3.41 (0.583)	4.23*** (0.544)	3.50 (0.673)
Culture of learning	4.09*** (0.245)	3.68 (0.554)	4.27* (0.461)	3.95 (0.761)	3.92*** (0.349)	3.40 (0.623)	4.32*** (0.505)	3.44 (0.722)
External community	3.77** (0.347)	3.39 (0.717)	3.76 (0.522)	3.55 (0.852)	3.77*** (0.358)	3.27 (0.656)	4.04*** (0.548)	3.33 (0.699)
Performance accountability	3.84*** (0.231)	3.45 (0.543)	3.93 (0.456)	3.63 (0.757)	3.76*** (0.355)	3.27 (0.600)	4.20*** (0.562)	3.38 (0.760)
Key processes								
Planning	3.94*** (0.233)	3.52 (0.526)	4.02* (0.476)	3.71 (0.648)	3.86*** (0.314)	3.33 (0.597)	4.30*** (0.583)	3.59 (0.759)
Implementing	3.97*** (0.237)	3.55 (0.541)	4.11* (0.436)	3.79 (0.691)	3.84*** (0.337)	3.32 (0.615)	4.21*** (0.500)	3.36 (0.667)
Supporting	4.01*** (0.247)	3.60 (0.537)	4.16* (0.434)	3.83 (0.723)	3.87*** (0.355)	3.38 (0.621)	4.30*** (0.506)	3.42 (0.619)
Advocating	3.93** (0.240)	3.59 (0.516)	3.99 (0.428)	3.78 (0.696)	3.86*** (0.305)	3.39 (0.564)	4.04*** (0.523)	3.40 (0.682)
Communicating	3.97** (0.258)	3.63 (0.548)	4.03 (0.460)	3.82 (0.737)	3.91*** (0.343)	3.43 (0.593)	4.22*** (0.563)	3.41 (0.776)
Monitoring	3.90** (0.276)	3.54 (0.514)	3.94 (0.456)	3.72 (0.692)	3.86*** (0.335)	3.36 (0.596)	4.19*** (0.566)	3.38 (0.789)

Standard deviations are in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

effect size for advocating ($ES=0.76$). The effect sizes for the core components and the other key processes range from 0.81 to 0.96.

When the data are disaggregated to respondent type, the differences are again significant and in the predicted direction for total effectiveness, the six core components, and the six key processes for teachers and supervisors. For example, teacher ratings of total effectiveness for elementary school principals in the top 20 % of principals is 4.01 and principals in the bottom 20 % of principals is 3.46—a difference of being highly effective versus satisfactorily effective. The effect size for this difference is 1.02. Consistently, teachers rate principals in the top 20 % of principals to be highly effective while teachers rate principals in the bottom 20 % of principals as being closer to satisfactorily effective. The effect sizes for the core components and key processes are large, ranging from 0.92 to 1.06.

Supervisor ratings of principal's effectiveness follow a similar pattern to that of the teachers' ratings of principals. However, the differences between supervisors' ratings for the principals in the top 20 % of principals and the principals in the bottom 20 % of principals tend to be smaller than the differences for the teacher ratings. Supervisors tend to rate those in the top 20 % of principals as being highly effective and those in the bottom 20 % of principals as being close to highly effective. The effect sizes for the differences for supervisors ratings range from 0.88 for the core component of quality instruction to 0.61 for performance accountability.

When elementary principals rate themselves, there is a slightly different pattern. While we do see that principals who are rated in the top 20 % of principals rate themselves higher than principals in the bottom 20 % of principals, these differences are not always statistically significant. For example, principals in the top 20 % of principals rate themselves at 4.09 in terms of the core component of rigorous curriculum. Principals in the bottom 20 % of principals rate themselves at 3.88 for rigorous curriculum. The difference is *not* statistically significant and has a modest effect size of 0.34. In addition, the difference between the top and bottom 20 % of principals' self-rating on the key process of advocating is not statistically significant and has a modest effect size of 0.37. Overall, elementary principals in the top and bottom 20 % of the district principals tend to rate themselves highly effective or close to highly effective.

The combined principal and teacher ratings for secondary principals for total effectiveness, key processes, and most core components are statistically significantly different between the top and the bottom 20 % of principals. Once again, the principals in the top 20 % of principals are rated significantly higher than those in the bottom 20 % of principals. The total effectiveness score for principals in the top 20 % of principals is 3.95, while the total effectiveness score for principals in the bottom 20 % of principals is 3.57. This difference has an effect size of 0.94, which is considered a large effect size.

For secondary school principals, teachers and supervisors rate the top and bottom 20 % of principals significantly different and in the predicted direction. Principals in the top 20 % of principals are rated closer to being highly effective while those in the bottom 20 % of principals are rated closer to being satisfactorily effective. The differences between the two known groups for teacher and supervisor ratings have large effect sizes ranging from 0.96 to 1.56. While for elementary school principals the differences in teacher ratings tend to be slightly larger than the differences in supervisor ratings, for secondary school principals the difference in effectiveness scores for the

principals in the top 20 % of principals and the principals in the bottom 20 % of principals are larger for the supervisor ratings than for the teacher ratings.

Secondary school principals tend to rate themselves as highly effective or close to highly effective whether or not they are in the top or bottom 20 % of principals. The principals in the top 20 % of principals rate themselves slightly higher than the principals in the bottom 20 % of principals; however, not to a significant extent. The effect size for the difference between top and bottom principal self-rated total effectiveness is 0.55. There are a few exceptions to the nonsignificant findings: the core component of culture of learning ($ES=0.51$) and the key processes of planning ($ES=1.03$), implementing ($ES=1.01$), and supporting ($ES=0.98$) all have significant differences in the expected direction.

5.2 Discriminant analysis

For elementary school principals, the combined principal and teacher VAL-ED scores, on average, accurately predict group placement 69 % of the time (see Table 3). In other words, 69 % of the time, the discriminant analysis places the principal into the same group as the superintendent places the principal. The lowest accuracy rating for elementary principals is 64 % for advocating. The highest accuracy rating is 73 % for quality instruction and performance accountability.

For secondary principals, a similar pattern is observed. The combined principal and teacher VAL-ED scores have an average of 71 % for accuracy of placement. The lowest level of accuracy for secondary principals is for external community at 66 %, and the highest level of accuracy is for planning at 75 %.

Principals as individual raters tend to have the lowest accuracy among the rater groups, which corresponds to the lack of significant differences between the top and bottom scores on total effectiveness, core components, and key processes. Teachers have the highest accuracy ratings, and supervisors' accuracy is between teachers and principals. This pattern is consistent for both elementary and secondary principals.

When considering accuracy of placement separately for the top and bottom 20 % of principals groups, a potentially important pattern emerges. For both elementary and secondary school principals, accuracy of placement is uniformly greater for the top group than the bottom group using the combined principal and teacher ratings (see Table 4).

For elementary school principals, the accuracy rating for the top 20 % of principals is 78 % for total effectiveness. In other words, about 78 % of the time both the VAL-ED discriminant analysis and the superintendent placed the same principals in the top group. The highest accuracy is for the core component of quality instruction and the key processes of implementing and supporting at 80 %. The lowest accuracy rate is for the core components of high standards, external community, and performance accountability and the key processes of advocating, communicating, and monitoring at 73 %.

The elementary principals in the bottom 20 % of principals have a lower total effectiveness accuracy rate at 61 %. The lowest accuracy rate for the bottom 20 % of elementary principals is 53 % for advocating (about chance level), while the highest accuracy rate for the bottom 20 % of principals is 67 % for communicating.

For secondary school principals, the total effectiveness accuracy rate for the principals in the top 20 % of principals is 75 %, about the same as for elementary school

Table 3 Percentage of accurate placement based on the combined principal and teacher VAL-ED score

	Elementary				Secondary			
	Principal and teacher N=108	Principal N=108	Teacher N=120	Supervisor N=95	Principal and teacher N=65	Principal N=65	Teacher N=70	Supervisor N=57
Total effectiveness	70	60	71	66	71	62	69	68
Core components								
High standards	68	58	69	69	73	62	70	70
Rigorous curriculum	68	59	70	65	68	55	69	67
Quality instruction	73	63	72	66	72	57	71	68
Culture of learning	69	60	71	64	71	58	70	72
External community	69	65	69	66	66	54	73	63
Performance accountability	69	60	70	66	71	58	71	67
Key processes								
Planning	69	65	68	63	75	58	71	67
Implementing	70	59	73	62	72	63	69	68
Supporting	70	62	67	63	69	63	70	72
Advocating	64	51	70	69	74	60	69	65
Communicating	70	61	68	63	68	55	70	67
Monitoring	66	59	72	66	68	58	71	67

Table 4 Percent accurate placements by top 20 % or the bottom 20 %

	Elementary						Secondary									
	Principal and teacher N=108		Teacher N=120		Supervisor N=95		Principal and teacher N=65		Principal N=65		Teacher N=70		Supervisor N=57			
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom		
Total effectiveness	78	61	63	57	75	67	69	63	75	67	72	52	74	63	76	61
Core components																
High standards	73	61	54	63	76	61	69	70	81	63	69	55	77	63	76	64
Rigorous curriculum	78	55	64	53	76	63	71	59	72	64	63	48	74	63	72	61
Quality instruction	80	65	59	67	75	68	65	67	78	67	59	55	74	69	76	61
Culture of learning	78	59	59	61	75	67	63	65	78	64	66	52	77	63	76	68
External community	73	63	66	63	73	65	78	54	72	61	56	52	80	66	69	57
Performance accountability	73	65	58	63	76	63	73	59	78	64	59	58	74	69	72	61
Key processes																
Planning	75	61	69	59	75	61	65	61	84	67	63	55	77	66	69	64
Implementing	80	59	64	53	78	67	61	63	81	64	72	55	74	63	69	68
Supporting	80	59	64	59	73	60	61	65	75	64	72	55	77	63	76	68
Advocating	73	53	47	55	75	65	80	59	78	70	69	52	74	63	69	61
Communicating	73	67	64	57	73	63	65	61	72	64	63	48	74	66	79	54
Monitoring	73	57	61	57	76	67	67	65	69	67	69	48	74	69	72	61

principals. The highest accuracy rate for those principals is 84 % for the key process of planning. The lowest accuracy rate for the secondary principals in the top group is 69 % for the key process of monitoring.

For total effectiveness, the accuracy rate for secondary principals in the bottom 20 % of principals is higher than the accuracy rate for elementary principals in the bottom 20 % of principals at 67 versus 61 %. The highest accuracy rate for the bottom 20 % of secondary principals is for total effectiveness, quality instruction, planning, and monitoring at 67 %. The lowest accuracy rating for the bottom 20 % of principals is 61 % for external community.

For elementary principals, teachers as raters have a higher accuracy in rating principals in the top 20 % of principals than in the bottom 20 % of principals. This pattern is consistent for teacher and supervisor ratings of secondary principals. However, there are instances where the principals and supervisor ratings of elementary school principals are more accurate for the bottom 20 % of principals than for the top 20 % of principals.

In supplemental analyses (results available upon request), we divided the principals' VAL-ED scores into two performance level categories (proficient or not proficient) as another way to examine the relationship between the VAL-ED and superintendents' categorizations. As part of the development of the VAL-ED, a 22 member panel of principals, teachers, supervisors, leadership researchers, and state level policymakers determined cut score values from the VAL-ED for four categories of principal effectiveness: distinguished, proficient, basic, and below basic (Elliott et al. 2008; Cravens et al. 2013). The VAL-ED total effectiveness score cut scores values are as follows: distinguished=4.00, proficient=3.60, basic=3.29. Principals scoring below 3.29 are considered to be below basic (Elliott et al. 2008). We collapse these four performance levels into proficient and not proficient by combining distinguished and proficient into one category and basic and below basic into another category. When we compare the principals' in our study proficient and not proficient categories to the top 20 % and bottom 20 % of principals categories, we find that 50 % of elementary principals that are placed in the bottom 20 % of principals and 45 % of the secondary principals that are placed in the bottom 20 % of principals have VAL-ED scores that are considered to be proficient. These percentages are much larger than the elementary and secondary principals in the top 20 % of principles who have VAL-ED scores that are not proficient, 14 % and 3 % respectively. The principals in the top 20 % of principals are more likely to have VAL-ED scores that are proficient but the principals in the bottom 20 % of principals are almost equally like to be proficient or not proficient. With such a large percent of bottom 20 % of principals being proficient, it may have been more difficult to discriminant these proficient principals from proficient principals in the top 20 % of principals resulting in a lower accuracy rate for the bottom 20 % of principals.

6 Discussion

The known group validity study of the VAL-ED provides additional evidence of the validity of the VAL-ED as an instrument to measure principal effectiveness for learning-centered leadership. We find that VAL-ED performance distinguishes between

principals placed in the top 20 % performance in their district by the superintendent and the principals placed in the bottom group. The rate of accuracy is about 70 %, overall. Placement into the superintendent's identified known group was more accurate for the top group than the bottom group.

The direction of the mean differences on the VAL-ED were consistently in the predicted direction whether looking at VAL-ED results for total effectiveness or disaggregated by respondent type and whether looking at total score or subscale scores of core components or key processes. However, principal ratings of themselves are less able to significantly differentiate between the two known groups. It may be that principals are less able to give an unbiased assessment of their effectiveness than are their teachers and their supervisors. The ratings that principals give themselves are fairly high, hovering around 4.00 on the 5-point scale, suggesting that principals believe that they are effective. This is consistent with noneducation literature that finds that leaders tend to rate themselves in an inflated manner (Atwater and Yammarino 1992; Kruger and Dunning 1999).

Discriminate analysis results are consistent with mean effectiveness comparisons and add an estimate of accuracy of VAL-ED placements into the known groups. For the total effectiveness score across teachers and principals, the VAL-ED was 70 % accurate in placing principals into the correct known group for elementary principals and a nearly identical 71 % for secondary principals. The accuracy of placement ranges from 64 to 75 % across elementary and secondary principals. Since there are two options for categorization (top 20 % and bottom 20 % of principals), there is a 50/50 chance that scores on the VAL-ED will place the principal in the correct known group. For total effectiveness score, each core component, and each key process, the accuracy between the VAL-ED and the superintendent is greater than 50 % and statistically significant with relatively small variance across results. Considering that superintendents were not told to place their principals into two groups based on their learning-centered leadership, it is likely that superintendents drew upon additional information about the quality of the principals when placing the principals into the top and bottom 20 % of principals categories. As the VAL-ED measures learning-centered leadership, we would not expect 100 % accuracy between superintendents and the VAL-ED. The overall accuracy rate of 70 % provides evidence that the VAL-ED and the superintendents' judgments of principal quality, at least partially, measure the same construct.

To determine whether some of the core components or some of the key processes are more useful for discriminating between the two known groups, we ran the discriminate analyses for the all of the core components together and all of the key processes together. The standardized canonical discriminant function coefficients give us a measure of which core component or key processes is most important for discriminating the two known groups (IBM Corporation 2011). At the elementary school level, the data suggest that for principals, rigorous curriculum was more important in discriminating the two known groups than the other core components. For key processes at the elementary school level, there was some evidence that for the teacher sample advocating was more important. At the secondary level for core components, culture of learning as rated by the supervisor as slightly more important than the other five core components. For the teacher raters, the key process of planning is slightly more important. Generally, however, each of the six core components and each of the six key processes were equally important in discriminating the two known groups for both elementary and secondary schools.

The VAL-ED is better able to place principals in the top 20 % of principals group than in the bottom 20 % of principals group. This difference in accuracy is slightly more pronounced for elementary school principals. Furthermore, the difference in accuracy is robust across total score and all 12 subscales. We do show that there is more variation in the bottom 20 % of principals group and that about half of the bottom 20 % of principals group has VAL-ED scores that are considered proficient.

Why the VAL-ED is more accurate in identifying principals that fall into the superintendent's top 20 % performance category than the superintendent's bottom 20 % of principals effectiveness category is difficult to determine. We do know that effectiveness rating scales tend toward a positive bias overall, but this would suggest that if the difference in accuracy is due to performance on the VAL-ED instrument, the VAL-ED would be more accurate for the bottom group than the top group. Returning to our earlier observation that our analyses do not determine which is the more valid, VAL-ED scores or superintendent assessments of top and bottom 20 % of principals on effectiveness, allows for the possibilities that the superintendents are better able to determine who are the best performers, but they have difficulty or are unwilling to identify who are the bottom performers. We are in the process of conducting a study in which we will determine the VAL-ED's ability to distinguish between principals who are more and less effective in terms of value added to student achievement for their schools. Perhaps these data will help to clarify the answer to the question. If the VAL-ED is able to predict differences in value added to student achievement for the school and to do so equally across the range of between-school variance in value added to student achievement, then perhaps the differences in accuracy noted here for the top versus the bottom group is more a function of the superintendent placement into groups (i.e. the quality of our criterion) than it is the quality of the VAL-ED.

The results reported here suffer from the possibility that the superintendent and the supervisor who rated the principal were one and the same, creating a potential redundancy in analyses using the VAL-ED total effectiveness score aggregated across all three respondent groups. The VAL-ED is designed for this total effectiveness across all three respondent groups data to be the best aggregate-level evaluation of a principal and so, arguably, the best data to discriminate more and less effective principals. As we feared redundancy could result in a spuriously large known group difference for the supervisor data, we focused our attention on the principal and teacher data and left the supervisor data out of the aggregate results. Nevertheless, when looking at the data from Table 1, the difference between the known groups on the 5-point effectiveness scale is 0.49 for supervisor data, 0.55 for teacher data, and 0.31 for principal data. From Table 2, we see that for secondary principals, the difference was 0.78 for supervisor data, 0.49 for teacher data, and 0.26 for principal data. Thus, only the data for secondary school principals suggests there might be a superintendent-supervisor redundancy bias, at least as seen in VAL-ED scores. Furthermore, the size of the difference between the two known groups was similar for elementary and secondary principals for the teacher data and for the principal data.

We also analyze accuracy of placement into two known groups separately by each of the six districts to see if there were any results at the district level which would suggest redundancy in the superintendent data and the VAL-ED data. For the two largest districts, there was no suggestion of redundancy. The same was true for one

substantially smaller district. For the other three small districts, the accuracy of the VAL-ED placement into the known groups was sufficiently high for the elementary school sample and for the secondary school sample, that we eliminated these data and redid the analyses. The results remained unchanged.

Overall, the VAL-ED reliably measures differences in effectiveness as seen by superintendents for both elementary schools and secondary schools and across the total and 12 subscales of the VAL-ED. These findings provide additional evidence in support of the VAL-ED being a valid tool for assessing learning-centered leadership behaviors of principals; however, we will add to this evidence as we complete our other studies. The known group validity of the VAL-ED for this study based on superintendent placing their principals in the top and bottom 20 % performance groups was both strong and consistent across levels of schooling and subscales of the VAL-ED.

Past research on principal evaluation has found a need for a systematic approach to assessing principal effectiveness. The VAL-ED is a principal evaluation tool that is embedded in the research literature surrounding effective schools and student learning and the only principal assessment instrument having well-documented known group validity.

Acknowledgments The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants # R305A0803070, R305B100013-01, and R305E100008 of the U.S. Department of Education. The authors declare a potential conflict of interest (e.g., a financial relationship with the commercial organizations or products discussed in this article) as follows: The Vanderbilt Assessment of Leadership in Education (VAL-ED) instrument is authored by Drs. Porter, Murphy, Goldring, and Elliott and copyrighted by Vanderbilt University, all of whom receive a royalty from its sales by Discovery Education Assessment. The VAL-ED authors and their research partners have made every effort to be objective and data based in statements about the instrument and value the independent peer review process of their research. With any publication, readers should judge the facts and related materials for themselves.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, *45*, 141–164.
- Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64.
- Brookover, W. B., & Lezotte, L. W. (1977). *Changes in school characteristics coincident with changes in student achievement*. East Lansing: College of Urban Development, Michigan State University.
- Burns, R., & Burns, R. (2008). *Business research methods and statistics using SPSS*. London: Sage Publications Ltd.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cravens, X., Goldring, E., Porter, A., Polikoff, M., Murphy, J., & Elliott, S. (2013). Standard setting for principal leadership assessment: a deliberative process. *Educational Administration Quarterly*, *49*(1), 124–160.
- Edmonds, R., & Frederiksen, J. R. (1978). *Search for effective schools: The identification and analysis of city schools that are instructionally effective for poor children*. Cambridge, MA: Center for Urban Studies, Harvard University.

- Elliott, S. N., Murphy, J., Goldring, E., & Porter, A. (2008). *VAL-ED users' guide*. Nashville: Discovery Education Assessments.
- Goldring, E., Cravens, X., Murphy, J., Elliott, S., Porter, A., & Carson, B. (2009b). The evaluation of principals: what and how do states and urban districts assess? *Elementary School Journal*, *110*(1), 19–32.
- Goldring, E., Cravens, X., Porter, A. C., Murphy, J., & Elliott, S. N. (2013). Instructional leadership and emotional intelligence: the convergent and divergent validity of the Vanderbilt Assessment of Leadership in Education Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Goldring, E., Porter, A. C., Murphy, J., & Elliott, S. (2009a). Assessing learning-centered leadership: connections to research, professional standards, and current practice. *Leadership and Policy in Schools*, *8*(1), 1–36.
- Hallinger, P., & Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, *32*(1), 5–44 (JM).
- Hallinger, P., & Heck, R. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, *9*(2), 157–191.
- Hallinger, P., & Murphy, J. (1985, November). Assessing the instructional management behavior of principals. *Elementary School Journal*, *86*(2), 217–247.
- IBM Corporation. (2011). Standardized canonical discriminant function coefficients. In IBM SPSS Statistics (Version 20) [Computer Software].
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.
- Leithwood, K., & Montgomery, D. J. (1982). The role of the elementary school principal in program improvement. *Review of Educational Research*, *52*(3), 309–339.
- Leithwood, K., Jantzi, D., & McElheron-Hopkins, C. (2006). The development and testing of a school improvement model. *School Effectiveness and School Improvement*, *17*(4), 441–464.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning*. Minneapolis, MN: Center for Applied Research and Educational Improvement, University of Minnesota.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Educational Administration Quarterly*, *46*(5), 671–706.
- Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370–397.
- May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*(2), 332–352.
- Murphy, J. (1989, Fall). Educational Reform in the 1980s: explaining some surprising success. *Educational Evaluation and Policy Analysis*, *11*(3), 209–223.
- Murphy, J. (1992). School effectiveness and school restructuring: contributions to educational improvement. *School Effectiveness and School Improvement*, *3*(2), 90–109.
- Murphy, J. (2013). The architecture of school improvement. *Journal of Educational Administration*, *51*(3), 252–263. [Invited Article].
- Murphy, J., & Shipman, N. J. (1999, September). The Interstate School Leaders Licensure Consortium: a standards-based approach to strengthening educational leadership. *Journal of Personnel Evaluation in Education*, *13*(3), 205–224.
- Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. (2007, April). Leadership for learning: a research-based model and taxonomy of behaviors. *School Leadership & Management*, *27*(2), 179–201.
- Murphy, J., Goldring, E., & Porter A. C. (2012). Building productive principal evaluation systems. Nashville: Vanderbilt University, VALED.
- Murphy, J., Hallinger, P., & Mesa, R. P. (1985, Summer). School effectiveness: checking progress and assumptions and developing a role for state and federal government. *Teachers College Record*, *86*(4), 615–641.
- Murphy, J., Hallinger, P., Weil, M., & Mitman, A. (1983a, Fall). Instructional leadership. A conceptual framework. *Planning and Changing*, *14*(3), 137–149.
- Murphy, J., Hallinger, P., Weil, M., & Mitman, A. (1983b, Fall). Problems with research on educational leadership: issues to be addressed. *Educational Evaluation and Policy Analysis*, *5*(3), 297–305.
- Murphy, J., Yff, J., & Shipman, N. J. (2000, January–March). Implementation of the interstate school leaders licensure consortium standards. *The International Journal of Leadership in Education*, *3*(1), 17–39.

- Norusis, M. (2003). *SPSS 12.0 Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall.
- Polikoff, M. S., May, H., Porter, A. C., Elliott, S. N., Goldring, E., & Murphy, J. (2009, November). An examination of differential item functioning on the Vanderbilt Assessment of Leadership in Education. *Journal of School Leadership, 19*(6), 661–679.
- Porter, A. C., Murphy, J., Goldring, E., Elliott, S. N., Polikoff, M. S., & May, H. (2008). Vanderbilt assessment of leadership in education: technical manual. Retrieved from http://www.valed.com/documents/VAL-ED_TechManual_030509.pdf.
- Porter, A. C., Polikoff, M. S., Goldring, E., Murphy, J., Elliott, S. N., & May, H. (2010a, April). Developing a psychometrically sound assessment of school leadership: the VAL-ED as a case study. *Educational Administration Quarterly, 46*(2), 135–173.
- Porter, A. C., Polikoff, M. S., Goldring, E. B., Murphy, J., Elliott, S. N. & May, H. (2010b). Investigating the validity and reliability of the Vanderbilt Assessment of Leadership in Education. *The Elementary School Journal, 111*(2), 282–313.
- Preston, C., Goldring, E., Guthrie, J., & Ramsey, R. (2012). Conceptualizing essential components of effective high schools. Paper presented at the National Conference on Achieving Success at Scale: research in scaling up effective schools. Nashville, TN.
- Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635–674.
- Supovitz, J., Sirinides, P., & May, H. (2009). How principals and peers influence teaching and learning. *Educational Administration Quarterly, 46*(1), 31–56.
- Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly, 39*(3), 398–425 (JM).
- Young, M. D., Crow, G., Murphy, J., Ogawa, R. (Eds.). (2009). *Handbook of research on the education of school leaders*. New York: Routledge.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.