

Predicting the gap: perceptual congruence between American principals and their teachers' ratings of leadership effectiveness

Peter T Goff · Ellen Goldring · Leonard Bickman

Received: 12 August 2013 / Accepted: 23 September 2014 /
Published online: 4 October 2014
© Springer Science+Business Media New York 2014

Abstract The aim of this study is to determine the extent to which principals' self-ratings of leadership effectiveness coincide with their teachers' perceptions of their leadership effectiveness. Furthermore, we explore several characteristics of teachers and principals in an attempt to identify the factors that may predict congruence in perceptions of leadership. This study draws from survey data of 76 principals and over 2100 teachers who completed parallel forms of a 72-item Learning-Centered Leadership survey (VAL-ED[®]) in the USA. Teacher and principal characteristics are incorporated into a multivariate regression analysis. Although there is zero difference in the overall sample, teachers and principals within any given school seldom share the same perspective. Principals' self-efficacy was a strong predictor of principals rating themselves higher than the teachers. Interestingly, the more time a teacher spent with a principal, the less congruence they shared. This research has identified rather large disparities in perceptions of leadership between teachers and their principals. Such a gap suggests that teachers have information and perspectives on school leadership distinct from the principals' information and perspectives. This research provides evidence that structured teacher feedback may provide a useful avenue for principals seeking additional perspectives on their leadership effectiveness.

Keywords Leadership evaluation · Feedback · Instructional leadership · Congruence

P. T. Goff (✉)

Educational Leadership and Policy Analysis, University of Wisconsin-Madison, 270L Education Bldg,
1000 Bascom Mall, Madison, WI 53706, USA
e-mail: pgoff@wisc.edu

E. Goldring

Department of Leadership, Policy and Organizations, Vanderbilt University, 230 Appleton Place,
Nashville, TN 37203, USA
e-mail: ellen.goldring@vanderbilt.edu

L. Bickman

Psychology, Psychiatry and Public Policy, Vanderbilt University, 230 Appleton Place, Nashville, TN
37203-5721, USA
e-mail: leonard.bickman@vanderbilt.edu

Despite a national climate of accountability and data-driven decision making in the USA, few schools support a system of evaluation and feedback for school leaders. For many districts, evaluation of school leaders centers on meeting contractual obligations and has become a legal formality rather than a performance-enhancing process. These assessment systems are not typically designed to improve future performance of leaders, do not provide specific feedback, and rarely inform decisions related to retention and promotion (Ginsberg and Berry 1990; Goldring et al. 2009).

Reeves (2008) surveyed over 500 districts across the USA, finding that principals found their evaluations to be largely positive, accurate, and consistent with job expectations. Despite these merits, few principals found that the evaluative process enhanced their motivation or elicited changes in performance. In keeping with existing research (Leithwood and Montgomery 1986), respondents in Reeves' study thought that existing evaluations were lacking in specificity and therefore failed to direct principals to concrete behaviors to be changed. Based on the survey results, Reeves also noted that most principals reported that they (a) did not receive useful feedback from their evaluations, (b) assessments were inconsequential, and (c) the criteria of their evaluations was unclear.

Until recently, the field of education was not alone in failing to provide substantive assessments of leadership. Since the mid-1990s however, many businesses have been endorsing a policy of multi-source feedback to develop strong personnel evaluation systems (Atwater et al. 1995; Conway and Huffcutt 1997). Multi-source feedback typically entails a self-evaluation of the leader as well as parallel evaluations from subordinates, peers, and/or superiors; when all three sources are engaged, this is known as a 360° feedback. The motivation behind multi-source feedback is that more information regarding leadership efficacy resides within the shared experiences of these individuals than from any one source alone (Atwater et al. 1998). The combined information can be channeled to improve leadership or inform promotion, development, and retention decisions. Research in the private sector supports the use of subordinate feedback to facilitate communication (Boyd and Jensen 1972), provide unique perspectives (Church 1997), and serve as a reliable source of useful information (Smither et al. 2005).

In the field of education, a reaffirmed commitment to accountability (Reauthorization of ESEA 2010), coupled with increased emphases on and funding for pay-for-performance and other educational incentive programs, such as the federal Teacher Incentive Fund, underscores the need for robust performance evaluation and feedback mechanisms for school principals. The availability and effective use of feedback information may provide avenues for improved leadership practices as well as information for accountability purposes. It logically follows that assessments of principals may benefit by incorporating teachers' perspectives. Indeed, there exists a growing trend to supplement principal evaluations by seeking information from parents, school boards, teachers, and principals (Doud and Keller 1998; Goldring et al. 2009). Multi-source feedback is expected to improve not only the quality of principal evaluation but also the quality of leadership within schools.

Research on multi-source feedback suggests that principals' views of their learning-centered leadership alone may be insufficient to expose the domains of leadership behavior most in need of improvement (Church 1997) because self-evaluations are inextricably tied to the leader's own experiences and biases. By contrast, when

principals reflect on their evaluations, multiple perspectives can create contrasts to motivate behavioral change (Bickman 2008; Riemer and Bickman 2011).

Leadership feedback from teachers can play two key roles in the evaluation process. First, teacher feedback may provide an alternative perspective on principals' learning-centered leadership. This "bottom-up" view of leadership may be a more valid evaluative source than is the principal's self-evaluation (Church 1997, p. 986). The larger number of respondents also lends credence to teacher feedback representing a more reliable measure than a principal's self-evaluation (Smither et al. 2005, p. 54). Second, teacher evaluations of principals can become a metric against which principals and superiors can compare and contrast the principal's self-evaluation. Principals who consistently express views of leadership that are not supported by their faculty may re-evaluate their leadership practices. Superintendents and district officials may also want to incorporate a comparison of principal and teacher ratings when considering performance reviews or when they are recommending coaching or professional development opportunities. Research suggests that leaders whose perceptions of their own leadership are dissonant from the views held by their subordinates are less-effective leaders, all else being equal (Atwater et al. 1998; Atwater and Yammarino 1992; Urlick and Bowers 2013). There are plausible exceptions to these findings, such as dissonance in schools where a principal is newly assigned with the goal of changing the professional culture and a school-turnaround effort.

The above research suggests that subordinate ratings may offer a more valid and reliable perspective on leadership than are self-ratings, yet there is no agreed-upon "true" perceptual measure of school leadership. While teacher ratings may not represent an absolute benchmark of leadership, these ratings are reflective of the teachers' perceptions of leadership within the school. Such perceptions shape school culture and affect the professional behavior of teachers. Indeed, Ladd (2011) found that teachers with lower perceptions of working conditions and school leadership were more apt to leave the school. The teacher-principal gap is a meaningful measure of perceptual agreement, representative of the shared environment in which teachers and principals work. The use of teacher-principal congruence as an evaluative, diagnostic, or formative learning tool will be enhanced by investigating how individual traits or attributes and organizational characteristics relate to the congruence measure.

The degree to which a principal's self-evaluation aligns with teachers' evaluations of their leadership is referred to as perceptual congruence, or simply, "the gap" (Benlian 2013; Shope 2013). Although increased attention is being placed on principal evaluation and assessment systems, little is known about the gap between principal self-evaluations and teacher evaluations. There has been limited research about the factors associated with the gap and how the gap impacts behavior of school leaders.

In schools, there has been limited study of the measures of perceptual congruence primarily because multi-rater or multi-source assessments and evaluation feedback is relatively new. Given the scarcity of understanding of measures of perceptual congruence, it is important to better understand the nature of the gap, the size of the gap, and the direction of the gap before we can fully understand the importance of perceptual congruence in developing leadership practice. It is precisely the theoretical and

empirical ambiguity surrounding the interpretation of perceptual agreement that necessitates a systematic inquiry. The aim of this paper is to better understand what may cause differences in perceptual congruence between teachers and principals of the effectiveness of principals' learning-centered leadership.

1 Learning-centered leadership

Learning-centered leadership is characterized by “strong, directive leadership focused on curriculum and instruction from the principal” (Hallinger 2003, p. 329). This model of leadership, also referred to as instructional leadership, is characterized by principals who focus on high standards for student performance, a rigorous curriculum, effective teaching pedagogy, performance accountability, and a strong culture for student learning (Goldring et al. 2009).

This paper uses the Vanderbilt Assessment of Leadership in Education (VAL-ED) to evaluate learning-centered leadership. The VAL-ED (see Porter et al. 2009) is a multi-rater instrument to measure the effectiveness of school leadership behaviors known to influence teacher performance and student learning. The conceptual framework of VAL-ED rests on a set of six core components and six key processes (see Goldring et al. 2009). Core components refer to characteristics of schools that support the learning of students and enhance the ability of teachers to teach. These core components of learning-centered leadership are high standards for student learning, rigorous curriculum, quality instruction, a culture of learning and professional behavior, connections to external communities, and performance accountability. The six key processes outline how leaders create and manage those core components of learning-centered leadership. These key processes are planning, implementing, supporting, advocating, monitoring, and communicating. The conceptual and empirical basis for the VAL-ED lies in the research on schools where students are effective in meeting ambitious learning targets (Murphy et al. 2007).¹

This study uses VAL-ED data collected from elementary and middle schools in one urban district to explore the congruence between principals' and their teachers' views of the principal's learning-centered leadership. The first portion of this study is largely descriptive, painting a picture of the magnitude, direction, and distribution of the gap between teachers and principals within schools and between schools. In the second portion of this study, we identify individual and organizational factors that may explain the wide variation we observe in these perception gaps.

Specifically, the research questions guiding this study are:

1. What is the size and distribution of the perceptual gap between principals' self-evaluation and teacher evaluations of their principals?
2. What individual and organizational factors contribute to the magnitude, direction, and distribution of the gap?

¹ A detailed analysis of the reliability and validity of the VAL-ED can be found in Porter et al. (2010); A full explication of the conceptual and theoretical foundations of the VAL-ED can be found in Murphy et al. (2007).

2 Explaining perceptual congruence

Several theories support the use of multi-source feedback in leadership development and evaluation. Cognitive dissonance theory suggests that dissimilarity between expected and actual inputs will create the opportunity for reflection and, subsequently, behavioral change (Greenwald and Ronis 1978). The implication is that differences in multi-source ratings may stimulate changes in leadership behaviors, including goal setting, increased communication, and transparency in decision making (Locke et al. 1984; Locke and Latham 1990).

Similarly, self-awareness theory suggests that effective leaders who are self-aware can incorporate observations to generate behavioral change. “Consequently, the self-aware individual is more cognizant of how he or she is perceived by others, which results in more accurate self-assessment” (Atwater and Yammarino 1992, p. 143). From an evaluative standpoint, self-awareness may be regarded as a desirable, even essential, component of a capable leader.

The Contextualized Feedback Intervention Theory (CFIT) synthesizes these dominant theories of professional performance (internal cognitive-affective processes) into a process of guided behavioral change (Bickman et al. 2011; Riemer et al. 2005; Sapyta et al. 2005). CFIT suggests that cognitive dissonance, stimulated by evidence that the principal’s view of their own leadership is somehow incongruous with teachers’ perceptions of their leadership, can be used to motivate behavioral changes that may lead to improved leadership. This “theory weaving” has been used in empirical work by using feedback to change professional practice (Bickman et al. 2011). The fundamental premise underlying these theories is that dissonant cognitions induce a psychologically uncomfortable state of arousal that provides the motivation to reduce dissonance (Festinger 1957). A discrepancy or dissonance between behavior and some standard increases motivation to reduce the dissonance by bringing evaluations from others in line with self-view. The comparison between self-ratings and feedback from others can challenge behavioral patterns and motivate one to rethink a behavior and its impact on others (McCauley and Moxley 1996).

Multi-source feedback may provide an opportunity to stimulate dissonance. Yet, previous research shows that the reactions to multi-source feedback are highly variable, with the outcome conditioned by the particular individual and situation (Walker et al. 2010). Behavior change may depend on self-efficacy or the extent to which individuals believe change is possible (Bandura 1982, 1997; Gist and Mitchell 1992; Locke et al. 1984; Stajkovic and Luthans 1998). If feedback reveals a deficiency but the principal does not believe he or she has the ability to change that specific behavior, feedback may not have the desired impact. A first step in understanding this process behavioral change is to construct a more thorough understanding of the factors associated with teacher-principal discrepancies.

There is empirical research that suggests that some set of leader and subordinate characteristics may predict the degree of perceptual congruence on leadership effectiveness ratings. Since there is limited research on these topics in regard to educational leaders, we rely on the literature on leaders from other sectors, mostly the business sector. Brutus et al. (1999) provide a twofold framework for conceptualizing how various characteristics may manifest differences in leadership perceptions. The first framework component identifies characteristics that may influence how

leaders evaluate their own leadership; the second incorporates characteristics that influence others' ratings of the leader. For example, extreme narcissism in a principal would be expected to inflate self-ratings while depressing subordinate ratings. With principal ratings greater than teacher ratings, this would be expressed as a substantial negative gap. This conceptual framework for understanding influences on the subordinate—self gap from the perspectives of both teacher and principal complements the use of multivariate modeling of the gap advocated by Edwards (1995).

The literature on explaining perceptual congruence in leader evaluation focuses on characteristics of the leader, such as gender, experience, and locus of control; characteristics of the subordinates, such as gender and years of experience working with the leader; and organizational structure such as the number of teachers per school. In the next section, we review principal, subordinate, and organizational factors, exploring how each may influence the teacher-principal gap in their evaluations of principals' learning-centered leadership.

2.1 Principal characteristics

Gender The research on gender and leader-subordinate evaluation is mixed. Gender, may be related to self-evaluation; male leaders may rate themselves higher and therefore exhibit greater discrepancies from their subordinates (although these findings are mixed depending on the task and the evaluation context) (Brutus et al. 1999). However, in instances where data were collected under a promise of confidentiality, as in this study, such differences disappear (Daubman et al. 1992).

Although substantial strides made in gender equity within educational administration over the last several decades, notable differences between male and female leaders may still remain. There is some evidence that school leadership, specifically the behaviors associated with learning-centered leadership, may not manifest equally in male and female principals (Adkison 1981; Glasman 1984; Gross and Trask 1976; Hallinger and Murphy 1985). Leithwood and Jantzi (1990) note that female principals may be more inclined to exhibit leadership behaviors associated with instruction or curriculum, perhaps as the result of gender-related socialization. Hallinger et al. (1996) find weak evidence that teachers may perceive female principals to be more involved with instruction at their schools. Grissom and Keiser (2011) find that female principals encounter more teacher turnover than do their male counterparts.

How these findings bear on learning-centered leadership is not readily evident. Historically, school administration has been male-dominated, with male principals continuing to be over-represented in administration as compared the teaching labor pool. Teaching, specifically elementary instruction, has been a historically female-dominated field, recent changes in the labor force notwithstanding. Learning-centered leadership conceivably melds two gendered domains, making a clear prediction for the role of gender in the self-other evaluation of leadership unclear. If male principals are more likely to rate themselves higher and perceive themselves as more effective learning-centered leaders, while teachers are likely to rate them as less effective on instructionally anchored behaviors, this would imply a larger, more negative gap for male principals.

Leadership experience In education as in other fields, principal salary structures are often structured as to reward administrative experience. Presumably, this relationship exists as a result of the returns to experience in the form of enhanced student outcomes. While efforts to quantify the returns to experience on outcomes have met with limited success (Murphy and Hallinger 1992; Fiedler 1972), Blase and Blase (1999) recommend further research regarding the role of administrative experience on teachers' perceptions of principal leadership.

Prior leadership experience may have an effect on self-ratings as well as on those of their teachers. Research suggests that experience is positively correlated with perceptual evaluations of leadership, and the effect is more pronounced for self-ratings than for subordinate ratings (Brutus et al. 1999). This suggests that influence of prior experiences on current leadership behaviors may loom larger in the mind of the leader than in the eyes of the subordinates.

Duration of administrative experience of principals may follow the general trend noted from other sectors of increasing self-evaluations, that is, self-ratings of higher effectiveness as an instructional leader with increasing experience. To date, there has been no research indicating how principals' self-perception of their instructional leadership changes over the course of their administrative career. Rather, research has tended to focus on the relationship between administrative experience and student achievement. In a working paper from the National Center for Analysis of Longitudinal Data in Education Research Institute, Clark et al. (2009) report that principal experience has a positive impact on student outcomes, such as math learning and student attendance. As with teacher experience, the marginal benefits to experience appear to be largest early on in principals' careers. If this finding is robust across contexts, we may expect that teachers could perceive this difference as well. In such situations, if administrative experience increases both teacher and principal ratings comparably, the impact on the gap will be null.

Principals' teaching experience Principals with substantial classroom experience may focus their leadership on their experiential strengths, thereby emphasizing their commitment to learning-centered leadership (Hallinger and Heck 2010). While it is plausible that teaching experience may be accompanied by higher perceptions of learning-centered leadership by teachers, Hallinger and Murphy (1985) found no evidence to support their hypothesis that teachers' perceptions of leadership are related to the duration of principals' teaching experience. If teachers are unlikely to perceive that principals with more teaching experience make more effective learning-centered leaders while principals perceive their teaching experiences to be informing their leadership, years of teaching experience may be associated with an increasingly negative teacher-principal gap.

Internal locus of control Weiner (1985) theorizes that the causal attributions people make as a consequence of success or failure can be described along a number of dimensions including locus of causality (internal vs. external). The nature of causal attribution along this causal spectrum influences the leaders' dissonance reduction strategy. The causal attribution process also determines how leaders will determine the appropriate action in response to the feedback (Moore 2000). Internal locus of

control promotes self-directed behavior (Bandura 2006), and therefore, principals with greater internal locus of control are likely to be more aware of their abilities and how their abilities influence their immediate environment (Rotter 1954). As internal locus of control increases, self-ratings may more accurately represent a principal's true leadership behaviors. Locus of control is likely to be positively associated with principals' self-evaluations and, to a lesser extent, positively associated with teachers' evaluations of principals. In keeping with self-awareness theory, principals with a strong internal locus of control may then be expected to demonstrate greater congruence with their faculty (Atwater and Yammarino 1992; Bono and Colbert 2005; Brutus et al. 1999).

Self-efficacy Internal locus of control captures one's perception of their opportunity to influence the outcome of events, yet it does not take into account how confident an individual feels about performing certain tasks in a certain environment (Bandura 2006). Self-efficacy can be understood as one's perception of their ability to capitalize upon specific opportunities with tangible, positive results. In short, self-efficacy is a task-dependent phenomenon while locus of control is global. Bandura (1997, p. 3) defines self-efficacy as the "belief in one's capabilities to organize and execute the courses of action required to produce given attainments." Bono and Colbert (2005) consider the integration of self-efficacy "fundamental" to their conceptual framework when examining the longitudinal impact of multi-source feedback on behavioral change in MBA candidates:

"[W]e draw from the voluminous existing literature that addresses the crucial roles played by self-efficacy ... in the initiation and maintenance of behavior change to further understand motivation to change following multi-source feedback." (Bono and Colbert (2005), p. 176)

In an educational context, we may expect higher self-efficacy ratings to correspond with higher self-ratings of effectiveness as learning-centered leaders. We also posit that principals who perceive a wide scope of possible actions are principals who are able to make leadership decisions from a more robust choice set. All else being equal, principals with more choices at their disposal will be able to undertake leadership behaviors that are best suited to the needs of their faculty, parents, and students. If this relationship holds, we would expect principals' self-efficacy to be associated with high self-ratings of learning-centered leadership. We expect principal self-efficacy to be positively associated with both teacher and principal evaluations, with a stronger association for principals.

2.2 Teacher characteristics

Experience Teacher experience may have a notable influence on a principal's learning-centered leadership. Significant returns to teacher experience have been seen to plateau beyond 5 years (Goe 2007); such a phenomenon may indicate that teachers' instructional capacity develops predominantly during these earlier years. When a principal is leading an inexperienced faculty, they may be more inclined to demonstrate learning-centered leadership as a way of guiding and mentoring their newer faculty. Extending

this idea, principals may take a differential leadership approach with new teachers—one that is more focused on developing instructional capacity as compared with their leadership with the more veteran teachers. Whether implemented collectively or individually, this potential variation in leadership behavior could result in higher teacher ratings of learning-centered leadership among novice teachers.

Duration of principal-teacher relationship The final teacher characteristic we examine is the duration of common professional time shared between principals and teachers. It is possible that when teachers have spent more time with the same principal, shared experiences may influence the gap between the teachers' and the principal's ratings of the principal's learning-centered leadership. The impact on the gap may depend on the nature of the relationship, rather than merely shared time alone. On one hand, amiable teacher-principal relationships may have an attenuating effect on the gap; on the other hand, when the teacher and principal do not have a productive working relationship, additional years of experience together could exacerbate differences in ratings.

2.3 Organizational structures

Organizational structure may play a role in moderating the degree of perceptual congruence between leaders and subordinates. The business literature cites internal hierarchy as contributing to greater discrepancy (Brutus et al. 1999). In more hierarchical organizations—those with multiple tiers—individuals in the higher leadership roles may be responsible for more subordinates and/or be more removed from interactions with subordinates. Both scenarios would distance leaders from subordinates and hinder the flow of information that could enable self-awareness, thereby leading to reduced perceptual congruence.

In the case of schools, the degree of hierarchy is relatively similar across schools, as all schools are organized with teachers under principals. However, the research on the organizational structure of schools suggests rather strongly that school size may be a factor regarding levels and types of internal communication and professional culture (Coburn and Russell 2008; Louis et al. 1996). For example, in schools with larger teacher-principal ratios, principal behaviors may not be as visible and transparent to all teachers. Larger schools may also compel principals to spend more time and resources on bureaucratic and organizational tasks, diverting their focus from learning-centered leadership. Therefore, as school size increases, we may observe a reduction in teacher-principal congruence.

Many studies of congruence examine small subordinate-leader clusters, and samples often contain less than ten subordinates per leader. Likely owing to this small cluster phenomenon, the distribution of subordinate responses as a predictor of congruence is absent from the literature. The comparatively large organizational clusters of schools permit an opportunity to pursue this line of inquiry. We hypothesize that greater variation in faculty perceptions of learning-centered leadership may be indicative of a fragmented or polarized school, which we anticipate being associated with larger and more negative principal-teacher gaps.

3 Methodology

3.1 Sample

In the large urban district where our study took place, there were 108 eligible middle and elementary schools. We excluded high schools as well as schools for specific student populations (e.g., alternative schools, special education, and charter schools) due to the organizational complexities or unique contexts of these institutions. Of the 108 principals invited to partake in the study, 76 agreed to participate. All teachers in these participating schools were asked to participate. Members of the project team administered the VAL-ED during the late spring of 2008 at each school site, where the principals and teachers separately completed the survey during a regularly scheduled faculty meeting. 2447 teachers participated in the study. The average response rate for teachers in schools was 87 %, ranging from 63 to 100 %.

In our sample, principals had been leaders at their current schools at an average of 2.4 years and had been principals in the same school district at an average of 5.6 years. Seventy-seven percent of school principals were female and 42 % were white. Fifty percent of teachers had less than 5 years of teaching experience in the school and less than 10 years of experience as a teacher; 42 % of teachers currently hold or have held a leadership position within the school, such as department chair. Similar to national trends in elementary schools, teachers were predominantly female (86 %) and white (73 %).

The average percentage of students in the free and reduced price lunch program was 68 %. Furthermore, 88 and 85 % of students were proficient or advanced on their standardized test in reading and math, respectively.

3.2 Measures

The research presented here is part of an ongoing research project to explore the response of principals to teacher feedback regarding their learning-centered leadership. All regular (not magnet nor charter) elementary and middle schools from one urban, southeastern district were invited to participate in the study during the 2007–2008 academic year.² Teachers and principals in the study responded to a modified version of the VAL-ED and other measures.

The survey is constructed as a six-by-six matrix of core components and key processes. Each point on the matrix contains two items to measure one of the 36 combinations of core components and key processes. For example, at the intersection of “high standards for student learning” and “planning” there are the following two items:

- (a) How effective is the principal (or their designee) at ensuring the school plans rigorous growth targets in learning for all students?
- (b) How effective is the principal (or their designee) at ensuring the school plans targets of faculty performance that emphasize improvement in student learning?

² The data for this paper are part of a larger project studying the effect of feedback and coaching on school leaders where additional data were collected.

The VAL-ED survey consists of 72 items (two items for each intersection of the six-by-six matrix); both principals and teachers take the survey, with the 72 learning-centered leadership items being essentially identical. The teacher version of the survey asks: “How effective is the principal (or their designee) at ensuring the school ...” with responses measured on a 5-point Likert scale, ranging from *ineffective* to *outstandingly effective*. The teacher version includes a sixth response option, “don’t know,” for each item.

The VAL-ED was developed and psychometrically tested in a field test after an extensive item writing and instrument development phase, including a sorting study, two rounds of cognitive interviews, a bias review, and two rounds of small-scale pilots. The field trial sample consisted of more than 270 schools and more than 8000 individual surveys, with 218 complete sets of responses. The sample included urban, suburban, and rural schools; elementary, middle, and high schools; and schools from all regions of the country (Porter et al. 2009). A recent differential item functioning (DIF) by Polikoff et al. (2009) shows that item functionality is robust across a wide spectrum of contextual variation including urbanicity, geographic region, and school-grade level (elementary, middle, and high schools). The Cronbach alpha reliability measure is over 0.95 for both principals and teachers.

Using a classical approach to evaluate the reliability of differences between subscales, Porter et al. (2010) have found that

“[i]n terms of differentiating core components from one another, the reliability of the difference calculations indicates that Culture of Learning and Professional Behavior, Connections to External Communities, and Performance Accountability are reliably distinguished from each other and from the other core components across respondent groups. For these three core components, all but two of the comparisons with other core components have reliability greater than 0.50 across the six form-by-respondent analyses. For teachers in particular, the reliabilities are high for these three core components, with all above 0.68 except for the comparison between Culture of Learning and Quality Instruction on both forms. Connections to External Communities is the best differentiated from the other core components across respondents and forms. The other three core components, Rigorous Curriculum, High Standards for Student Learning, and Quality Instruction, are well differentiated from one another for teachers but not as well differentiated for principals and supervisors. Overall, however, the results suggest that the core components can be reliably distinguished from one another, especially for teachers. Comparatively, these are strong reliabilities of differences.” (Porter et al. 2010)

We found a similar validity structure to the data in this paper based on factor analysis.

3.3 Dependent variables

Typically difference scores are calculated as a simple difference, an absolute difference, a squared difference, or as an index comparing response profiles (Cronbach and Gleser 1953). Each of these approaches poses methodological complexities and introduces

limitations to the analysis. For example, Atwater and Yammarino (1992) use a measure that transforms the gap from a continuous variable to a categorical variable (over-raters, agreement, and under-raters). This approach solves some of the mathematical problems associated with difference measures, yet now the measure ignores magnitude of the gap and relies on an arbitrary judgment to establish the categorical cut-points. As is typical of difference measures, we expect lower reliability in the difference measure than in either of the two constituent measures yielding statistically more conservative point estimates (Pedhazur 1982). A second consideration is that the variance in the difference measure will more closely resemble the variance of the measure with the greatest variance (Edwards 1995). In our case, we can view the simple difference in each school as a rescaled version of the original teacher survey response; here, the within-school variance of the difference measure is identical to the within-school variance of the teacher response.

A third complication common to all difference measures is the ambiguity of a non-significant coefficient. Because each coefficient is impacted by a teacher component and a principal component, offsetting effects will yield null results (Edwards 1995). For example, school size may have a significant and positive relationship with teacher responses and school size may also have a significant and similarly positive relationship on principal responses. The difference between the two estimates would sum to zero; we may likely conclude that school size has no impact on the teacher-principal gap. While such a conclusion is not, directly an error, it omits information that may be useful to the line of inquiry.

A similar situation arises when considering the origin of the effect. Continuing the example above, a significant coefficient on school size may indicate that school size may significantly predict the gap from the teacher's perspective, from the principal's perspective, or both. Using difference measures alone would not allow the research to distinguish between these various contributions (Edwards 1995).

In order to minimize existing limitations of the various difference measures, we apply the recommended procedure of modeling subordinate and self-evaluations jointly in a multivariate model and determining the effect of each coefficient as the difference between the two measures.

The dependent variables in this study are the means of the 72-item VAL-ED perceptions of leadership effectiveness, for teachers (T) and principals (P). Ultimately, we are interested in degree of perceptual congruence (the gap) between teachers' ratings of their principal and the principals' self-ratings. We define this gap as a simple difference:

$$\text{Gap} = \text{VALED}_{\text{Teacher}} - \text{VALED}_{\text{Principal}}$$

For 357 of the 2447 records, there was a small degree of missing data. Principals' reported years of teaching experience was missing for seven principals, associated with 238 teachers; all other variables were missing at rates less than 3 %. These missing data were imputed using a MCMC multiple imputation procedure with five imputations for records with missing data (Rubin 1987). Adequate chain convergence was determined graphically (Brooks and Gelman 1998) and confirmed using the Gelman-Rubin diagnostic (Gelman and Rubin 1992). Autocorrelation was determined graphically and the thinning interval was adjusted accordingly. Findings using imputed and unimputed

values are statistically and substantively consistent. The Cronbach alpha for the teachers' rating of principals is 0.97; for principals' self-rating, 0.98.

3.4 Independent variables

Table 1 (below) shows that there are over three times as many female principals as male. The average male principal in this study has almost 12 years of administrative experience and 11 years of teaching experience; the average female principal has 8 years of administrative experience and just over 15 years of teaching experience.

Table 2 presents information on the scales used to measure internal locus of control and self-efficacy, which we operationalize as a perceived mastery scale. Both scales were constructed from six items, each with 5-point Likert response options (1=not at all; 5=a great deal). Internal locus of control was ascertained by asking principals, "During the last few months, how much has your leadership effectiveness been due to..." Where items included options such as "my effort," "my leadership style," and "my attitudes."

The perceived mastery scale is based on the work of Tschannen-Moran and Gareis (2004), who created and validated an 18-item scale to measure the self-efficacy of school leaders. The scale contains items such as "In your current role as principal, to what extent can you facilitate student learning in your school? ... generate enthusiasm for a shared vision for the school?" Our study uses six of the items in the Tschannen-Moran and Gareis self-efficacy scale; however, for reasons pertaining to the use of this survey in the larger scope of research, our prompt is quasi-retrospective: "In the past few months, in your current role as principal, to what extent can you ..." Because of the limitations of this scale to fully capture self-efficacy, which may also depend on task context and goal or task difficulty, we label this construct "perceived mastery."

As may be expected from an inflated self-rating bias, both locus of control and perceived mastery scales are skewed toward the higher, more positive ratings. Despite this truncated use of the response scale, both constructs present an acceptable degree of internal reliability (0.78 or higher).

Table 1 Principal gender and experience

| | <i>N</i> |
|---------------------------|----------|
| Gender | |
| Male | 16 |
| Female | 60 |
| Administrative experience | |
| 0–4 years | 13 |
| 5–9 years | 37 |
| 10 ⁺ years | 26 |
| Teaching experience | |
| 0–4 years | 7 |
| 5–14 years | 29 |
| 15–24 years | 36 |
| 25 ⁺ years | 4 |

Table 2 Principal locus of control and perceived mastery

| | Items | Alpha | Unimputed mean | Imputed mean | Standard deviation | Minimum | Maximum |
|---|-------|-------|----------------|--------------|--------------------|---------|---------|
| Internal locus of control | 6 | 0.84 | 4.069 | 4.070 | 0.55 | 2.3 | 5 |
| During the past few months, how much has the level of your leadership effectiveness been due to ... (a) my effort, (b) my leadership style, (c) my knowledge, (d) my natural talent, (e) skills that I have, and (f) my attitudes? | | | | | | | |
| Perceived Mastery | 6 | 0.78 | 4.066 | 4.066 | 0.47 | 3 | 5 |
| In the past few months, in your current role as principal, to what extent have you been able to ... (a) facilitate student learning, (b) generate enthusiasm toward a shared vision for the school, (c) manage change in your school, (d) create a positive learning environment in your school, (e) raise student achievement on standardized tests, and (f) motivate teachers? | | | | | | | |

With respect to teacher characteristics, 27 % of teachers had less than 3 years of experience and 25 % had 16 years or more of experience. By contrast, when we examine the years of experience working with the current principal, the majority of all teachers in this sample (89 %) have less than 3 years. It should be noted that this includes prior experience with the principal at other schools.

School size was determined by student enrollment using district-provided administrative data records. The average school enrollment was 475 with a standard deviation of 176. Enrollment is expressed as enrollment per 100 students to maintain a comparable scale with other coefficients.

3.4.1 Analytic approach

As mentioned previously, there are notable challenges to an analysis of difference measures. We first present descriptive results of perceptual congruence within and between schools. We use kernel density and histogram plots to convey the magnitude and variation of the gap. We characterize the gap as the simple difference between teacher and principal ratings as well as the absolute value of the difference and present descriptive results for both measures.

The analysis also employs multivariate multiple regression models to examine the regression coefficients for teachers and principals separately (Edwards 1995). First, both teacher and principal equations (Eqs. 1 and 2, below) are estimated independently at the teacher level. Next, the difference between regression coefficients are tested using Seemingly Unrelated Estimation (SUE), a technique that adjusts regression coefficients for shared variance between Eqs. 1 and 2 while also making precision adjustments for multiple observations and non-independence of observations within schools (Fiebig 2001; Zellner 1962). We employ this strategy to examine the impact of principal, teacher, and school factors on teacher and principal ratings individually and simultaneously estimate the impact of each factor on the gap. This approach allows our analytical strategy to be consistent with theory by integrating the flexibility for the impact (coefficient) of a given factor to be different for principal and teachers, while still providing standard errors to test for the significance of the difference between

coefficients. This strategy has been used in the past to investigate difference measures within clustered groups, such as parent–child reporting of support (Mandemakers and Dykstra 2008).

Teachers

$$\text{VAL-ED}(t)_{ij} = b_0 + b_1(\text{Teacher Vars})_{ij} + b_2(\text{Principal Vars})_j + b_3(\text{School Vars})_j + e_{ij} \quad (1)$$

Principals

$$\text{VAL-ED}(p)_{ij} = \alpha_0 + \alpha_1(\text{Teacher Vars})_{ij} + \alpha_2(\text{Principal Vars})_j + \alpha_3(\text{School Vars})_j + u_{ij} \quad (2)$$

The dependent variables for Eqs. 1 and 2, “VAL-ED(*t*)” and “VAL-ED(*p*)” represent teachers’ rating of principals instructional leadership (72-item VAL-ED mean) for teacher *i* at school *j* and the principals’ self-rating of their own instructional leadership (72-item VAL-ED mean), respectively. “Teacher Vars” represents a vector of six dummy variables representing various levels of teacher experience (6–10, 11–15, 16–20, and 20 years or more) and the amount of time a teacher has spent with the principal (6–10 and 11 years or more). “Principal Vars” represents a vector of five principal characteristics: internal locus of control, self-efficacy, gender, years of teaching experience, and years of administrative experience. “School Vars” represents the school-level variables, which in this case is limited to school size and the school-level standard deviation of teacher ratings. As mentioned previously, Eqs. 1 and 2 are estimated separately at the teacher level. Application of SUE procedure allows us to estimate the difference between the teacher and principal coefficients in Eqs. 1 and 2, while concurrently generating cluster adjusted standard errors for the difference measure (teacher-level models are corrected for clustering).

The teacher-principal gap is the algebraic difference between Eqs. 1 and 2. When we subtract the principal equation from the teacher equation, the resulting dependent variable becomes our gap measure (VAL-ED(*t*)_{*ij*} – VAL-ED(*p*)_{*ij*}). On the right-hand side, the difference between each of the regression coefficients represents the effect of the respective regression coefficient on the gap. For example, the effect of school size on the gap is determined by subtracting the effect of school size in the principal equation from the effect of school size in the teacher equation. The teacher-principal gap is summarized in Eq. 3, below.

The Teacher-Principal Gap

$$\begin{aligned} \text{Gap}_{ij} &= (\text{VAL-ED}(t)_{ij} - \text{VAL-ED}(p)_{ij}) \\ &= (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)(\text{Teacher Vars})_{ij} + (\beta_2 - \alpha_2)(\text{Principal Vars})_j \\ &\quad + (\beta_3 - \alpha_3)(\text{School Vars})_j + (e_{ij} - u_{ij}) \end{aligned} \quad (3)$$

The null hypothesis of a standard gap model (Eq. 3) would posit that the coefficients from the teacher model would be equal in magnitude and opposite in sign to the corresponding coefficients in the principal model. The null hypothesis of the difference between teacher and principal coefficients is tested in the SUEs procedure.

4 Results

We first present the descriptive results regarding the nature of the perceptual gap. Across all schools, the teacher-principal gap is centered on zero, with 95 % of the values falling between negative 1.50 and positive 1.50. This 3-point range is equivalent to the difference between the minimally effective (2) and outstandingly effective (5) ratings. Figures 1 and 2 present a histogram and kernel density distribution of both the gap and the absolute value of the gap, respectively. The simple gap shown in Fig. 1 portrays the symmetry of over- and under-raters. The variance the gap also portrays a balanced symmetry with 52 % of the variance occurring between schools and 48 % occurring within schools. This even balance of within to between variance suggests that the factors that contribute to the gap are likely a combination of factors that vary within schools (such as teacher experience) and those that vary between schools (such as principal gender).

Figure 2 illustrates the absolute value of the gap, shown here to better illustrate the extent of overall congruence. A simple difference, especially with aggregated data, can mask the extent of agreement. An extreme example, to motivate this point, would be a school with a polarized faculty where half the teachers rate the principal a 2 and half the teachers rate the principal a 5. If the principal rates himself as a slightly more than satisfactory leader (e.g., 3.5), the teacher-principal gap as a simple difference would be zero, creating the illusion of congruence within the school. The absolute value of the gap allows us to inspect this overall congruence, however the cost of the absolute measure is the loss of perspective with respect to how teachers scored relative to principals (e.g., above or below).

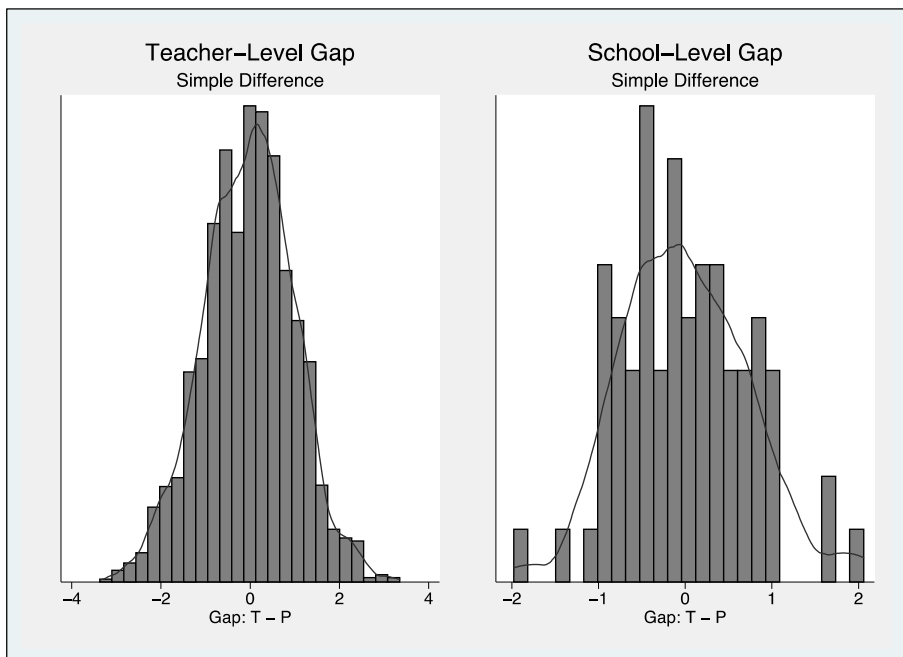


Fig. 1 Perceptual congruence—simple difference at the teacher level and school level

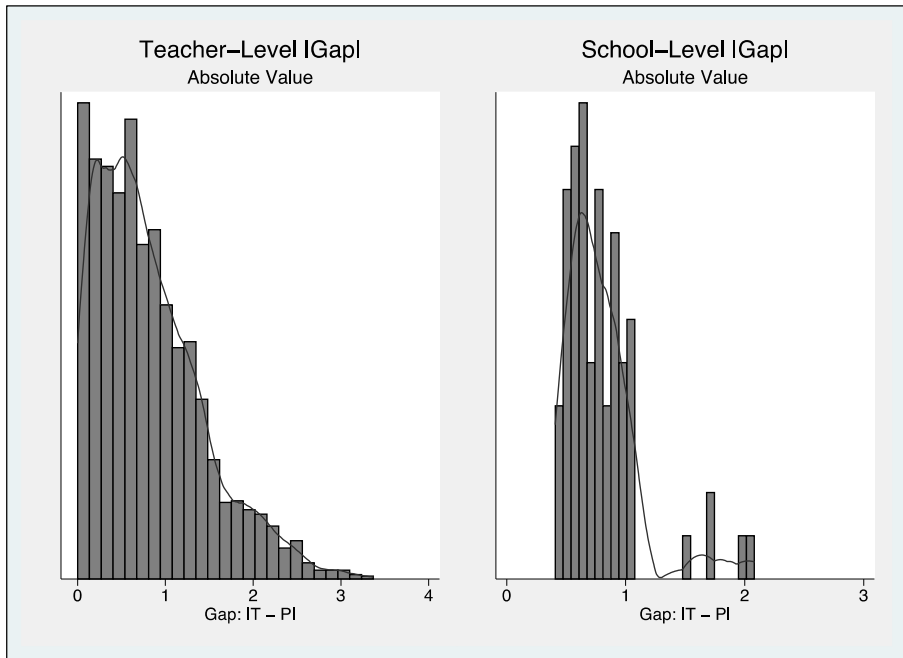


Fig. 2 Perceptual congruence—absolute difference at the teacher level and school level

From Fig. 2, we see that the absolute value of the gap has a lesser degree of variation between schools than within, relative to the simple difference: 28 versus 72 %, respectively. The absence of absolute gaps smaller than 0.40 at the school-level suggests that there is notable variation within schools between principal and teacher perspectives. In the case of the simple difference, schools with gaps between -0.4 and 0.4 may be cases where there is less congruence than is apparent at first glance.

In sum, we note that the distribution of the gap is normally distributed and is centered close to zero, with as many teachers rating above their principals' self-rating as there are teachers rating below the principal. Comparisons between the simple and absolute gap lead us to observe that, when aggregated to the school level, a small teacher-principal gap may be an artifact of offsetting scores. When we ignore the direction of the gap and focus only on the magnitude, we see that the schools in our sample have a mean gap of 0.80 and range from 0.41 to 2.05. The minimum value of 0.41 in Fig. 2 indicates that schools with near-zero gaps (as a simple difference) are schools where the faculty tends to be equally distributed above and below the principal's self-rating, yet substantial variation around that value remains. The presence of such variation implies that the average teacher-principal gap alone may not reveal the whole story regarding teacher-principal congruence. One way to explore this further is to examine the distribution of gap scores within schools.

The between school variation (standard deviation of 0.75) is comparable to the within school variation (mean standard deviation of 0.69). The school-level gap and the school-level standard deviation of the gap are negatively correlated (-0.57 , $p < 0.000$). This negative correlation may be interpreted to mean that in schools where the teacher ratings are greater than the principal self-rating (i.e., a positive gap), there is greater

consensus regarding learning-centered leadership (i.e., a small standard deviation) than when the principal self-rating is greater than the faculty mean (i.e., a negative gap). Hypothesizing further, this may indicate that principals tend to receive high average ratings of leadership (relative to their self-rating) when there is a shared opinion of high leadership quality throughout the school. However, when principals receive lower teacher ratings (relative to their self-rating), there may be a polarized faculty opinion of school leadership (i.e., a larger variance in the gap). High faculty consensus alone however does not infer highly effective leadership. Figure 3 has been constructed to help convey how teacher ratings and principal ratings vary with the extent of congruence.

Figure 3 presents a scatterplot of teacher ratings versus principal ratings of leadership. The vertical line represents average principal ratings; the horizontal line represents average teacher ratings. The diagonal line represents the region where principal ratings would be equal to teacher ratings—a zero gap line. As would be expected from the kernel density plots in Fig. 1, we see here that roughly half the schools are above the diagonal line, displaying a net positive gap (e.g., teacher ratings > principal ratings), and the other half with a negative gap below the line. As discussed previously, we expect the strongest leaders to be in schools where low gaps are found alongside high teacher ratings. This would indicate ratings of high quality leadership from a source understood to be informed and reliable coupled with leaders who demonstrate moderate to high self-awareness relative to their faculty. These select schools, with small or somewhat positive gaps are depicted as white dots on Fig. 3.

In Fig. 3, we also see that both principals and teachers ratings cluster around the satisfactory level of effectiveness (3- on the 5-point scale). We can observe that, as principals rate themselves higher, the gap becomes increasingly negative. In the lowest quartile schools (by mean principal rating) the average gap is 0.78; in the highest quartile schools, the average gap is -0.60 . It appears that teachers may be more

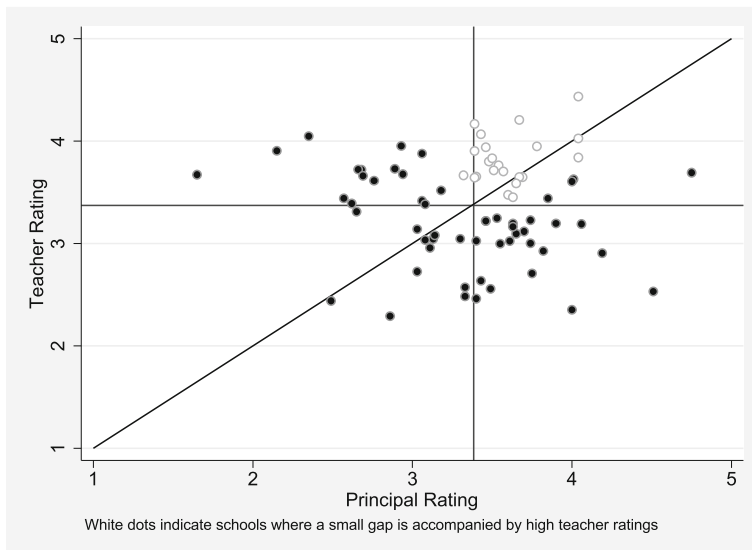


Fig. 3 Principal and teacher ratings of leadership

moderate in their ratings than are principals. Principals who give themselves low ratings tend to have teacher means that are slightly higher; principals who rate themselves highly tend to have teacher means that are slightly lower. To an extent, this is an artifact of the instrument: on a 5-point scale, principals who rate themselves a 5 can only have non-positive gaps. Additionally, the teacher values are averages across the whole school, which may cause a tempering of aggregate scores relative to any one extreme individual rating.

5 Explaining variation in congruence

Next, we use teacher level and school-level variables to explain the widely varying relationship between teacher and principal perceptions of the effectiveness of learning-centered leadership in the context of the multivariate regression models. Table 3 shows the results for the multivariate regression with teacher ratings of principal leadership as the dependent variable in the first column and principal self-ratings in the second column. The third column shows the gap, calculated as the teacher coefficient less the principal coefficient. These three columns correspond to Eqs. 1, 2, and 3 described previously, with the standard errors representing the adjusted standard errors as a result of the SUE procedure. The coefficients on Table 3 represent the extent to which individuals deviate from average responses, conditional on the other covariates. First, we describe how teacher and school characteristics influence teacher ratings, then explain their relationship to principal self-ratings, and lastly we turn to the influence of teacher and school-level variables on the gap, that is differences between principal and teacher ratings of principal effectiveness.

Predictors of teachers' ratings Looking first at the determinants of teacher ratings of leadership in column one, we note the negative and significant coefficients on the first two dummy variables for teaching experience. This indicates that experienced teachers (those with 6–10 or 11–15 years of experience) tend to be more critical of their principal's school leadership effectiveness than are teachers with 1–5 years of teaching experience. Continuing down this column, we see that a teacher's time spent with the principal does not appear to have a significant influence on the mean teacher rating. Looking to how principal characteristics may relate to teacher ratings, principals' locus of control, gender, teaching experience, and administrative experience have no significant relationship to teachers' ratings of their principals' leadership effectiveness. A principals' perceived mastery however does show a positive relationship to teacher ratings. Teachers who work with principals who perceive that they are more able to capitalize on opportunities related to specific, school-related aspects of leadership tend to rate their principal higher on measures of learning-centered leadership.

School size does not appear to have any relationship to teacher ratings of their principals' leadership effectiveness; however, when we examine within school variation in teacher ratings, we see an interesting phenomenon at play. Teacher ratings of the principal tend to decrease as within-school variation increases. This implies that higher levels of within-school consensus accompany higher principal ratings while lower scoring schools are more likely to be divided over perceptions of leadership. From a conceptual standpoint, this finding makes sense as we expect that good leaders are able

Table 3 Predicting ratings of learning-centered leadership—multivariate regression results

| | Teachers ($p> t $) | Principals ($p> t $) | Gap ($p>Chi^2$) |
|-----------------------------------|----------------------|-----------------------------|-----------------------------|
| Teachers' teaching experience | | | |
| 6–10 years | -0.076* (0.040) | -0.030 ⁺ (0.093) | -0.046 (0.448) |
| 11–15 years | -0.150** (0.001) | -0.078 ⁺ (0.073) | -0.072* (0.019) |
| 16–20 years | -0.115 (0.163) | 0.085* (0.022) | -0.200 (0.148) |
| 20+years | -0.029 (0.486) | 0.050 ⁺ (0.095) | -0.079 (0.935) |
| Teachers' time with the principal | | | |
| 6–10 years | -0.002 (0.977) | -0.108 (0.231) | -0.106 (0.962) |
| 11+years | 0.133 (0.306) | 0.021 (0.818) | 0.112 (0.475) |
| Locus of control | -0.006 (0.955) | 0.080 (0.524) | -0.086 (0.470) |
| Perceived mastery | 0.292* (0.023) | 0.540*** (0.000) | -0.248 (0.255) |
| Female principal | 0.143 (0.311) | 0.133 (0.257) | 0.010 (0.627) |
| Teaching experience | -0.003 (0.956) | -0.004 (0.548) | 0.001 (0.363) |
| Admin experience | -0.011 (0.276) | 0.012 (0.461) | -0.023 ⁺ (0.089) |
| School enrollment | 0.038 (0.172) | -0.035 (0.362) | -0.073 (0.495) |
| Teacher variation (SD) | -2.457*** (0.000) | 0.567 (0.195) | -3.024*** (0.000) |
| Constant | 3.745** (0.000) | -0.500 (0.493) | |
| N teachers | 2447 | 2447 | |
| N principals | 76 | 76 | |

⁺ $p<0.10$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$

to build consensus, cultivate a common vision, and align faculty towards a unified school mission.

Predictors of principals' self-ratings Results for principal self-ratings show a pattern similar to those for teacher ratings. Here, we see evidence that principals working with teachers having 6–10 and 11–15 years of experience are associated with somewhat lower self-ratings as compared with principals working with more junior faculty. Interestingly, principals working with some of the most experienced teachers, those with 16–20 and 20 years or more of experience tend to rate themselves as more effective than principals working with more junior faculty. All of these coefficients are less than 0.10, suggesting that the statistical significance may outweigh the substantive significance of these results. As with teachers, there appears to be no relationship between shared time and principal self-ratings. Examining principal factors, we see that, as with teacher ratings, principals' locus of control, gender, teaching experience, and administrative experience have no apparent relationship to principal self-ratings of their effectiveness. Principals' perceived mastery appears to be strongly related to principal self-ratings, where a 1-point increase on the perceived mastery-scale corresponds to over a half of a point increase on the principal's self-rating of learning-centered leadership effectiveness.

Again paralleling the teacher findings, school size does not appear to have any relationship with principal self-ratings. Diverging from the teacher results Table 3

shows that principal self-ratings are not sensitive to the within-school variation in teacher ratings.

Predictors of the gap When we turn our attention to the difference between the teacher and principal coefficients, shown in the third column of Table 3, we observe one marginally significant coefficient among more experienced (11–15 years) teachers. It appears that more experienced teachers may have a significantly negative teacher-principal gap as compared with their colleagues in their first few years of teaching. The next set of factors address the associations between principal characteristics and the teacher-principal gap. Two findings are of particular note for their lack of significance: internal locus of control and gender. The significance of these two measures tends to be fairly robust in the multi-source rating literature, especially with regard to self-evaluations, in which males and individuals and internal locus of control tend to be associated higher self-ratings (Atwater and Yammarino 1992; Bono and Colbert 2005; Brutus et al. 1999). Here, we see no evidence of such trends; not in teacher ratings, principal self-ratings, nor in the extent of congruence between them.

Although the duration of a principal's administrative experience was not a significant predictor in either the teacher or principal rating models, principals' administrative experience does appear to be a modest predictor of teacher-principal congruence. To put the coefficient into context, principals' years of administrative experience ranges from 1 to 25 years with a standard deviation of 5.4 years. This translates to an eta-squared effect size of 0.062 or a partial eta-squared of 0.095. This effect is modest to small, yet notable for the unexpected direction of this coefficient. Prior research suggests that leaders may tend to inflate their self-ratings as they accrue experience, yet in our analysis we see an unusual finding in that principals appear to have no significant bias in their self-ratings attributable to their administrative experience.

The final category of factors that may influence the teacher-principal gap pertain to school-level variables. As noted above, we see no effect of schools size on either teacher or principal ratings. This may be somewhat surprising given the variation in school size across the sample; we have small schools with as few as 200 students to schools just shy of 900. Nonetheless, we observe no relationship between the size of the organization and the extent of teacher-principal congruence.

The teacher-principal gap shows a strong relationship to the distribution of scores where greater within-school variation in teachers' perceptions of leadership corresponds to a more negative teacher-principal gap. Schools with teachers expressing a wide variety of perceptions regarding learning-centered leadership are also schools that tend to have principal self-ratings exceeding those of teacher ratings.

6 Discussion

With our first research question we set out to describe the gap between teacher and principal perceptions of learning-centered leadership. We find that the gap is normally distributed with a mean of zero and standard deviation of 0.75 units. The variation within schools is similar to the variation between schools, and little of that variation can be readily explained through descriptive statistics alone.

Based on the literature, the effect of principal gender on the direction or magnitude of the perception gap in the context of educational leadership was uncertain. The regression results presented above suggest that principal gender does not play a substantial role in either teachers' perceptions of leadership effectiveness as measured by the VAL-ED, or in principals' self-ratings. It is worth mentioning that male principals only constitute 21 % of the sample (16 of 76 principals) and we may lack the statistical power to adequately detect differences between groups. However, if this sample is representative, it is unlikely that gender plays a prominent role, as the coefficient is small, implying a negligible effect size for the simple difference of the gap.

Principals' teaching experience was found to have no bearing on the extent of perceptual congruence while self-efficacy was associated with principals rating themselves higher than did their teachers (a negative gap). This finding may reflect recent work examining the determinants of teacher quality where several studies have found that teachers' experience is only loosely related to quality measures, especially beyond the initial 3–5 years. Although we may have thought that higher quality teachers may rely on their instructional expertise and demonstrate superior instructional leadership as principals, such a hypothesis cannot be borne out in this context if the correlation between teaching experience and teaching quality is thin.

The effect of perceived mastery on the gap of -0.248 stems from a greater effect on principal self-ratings (0.540) than for teacher ratings (0.292). The perceived mastery hypothesis was that principals with greater perceived mastery would represent leaders who may harbor inflated views of their competencies, thereby leading to self-ratings higher than those of their faculty. We would expect principals who perceive themselves to be more efficacious to also rate themselves higher as leaders (i.e., the self-rating bias is in the same direction in the gap and in self-ratings); however, the positive coefficient on teacher ratings merits further consideration.

The principal perceived mastery scale is comprised of items that primarily pertain to teacher motivation, cultivating a sense of shared vision, and motivating student learning. These task-specific items may manifest in principals' behaviors more concretely and thus may be more tightly tied to teacher perceptions than are some of the more global items on the VAL-ED, such as the effective planning of schedules to promote quality instruction. The second interesting consideration is that the correlation between principals' perceived mastery and teachers' ratings of learning-centered leadership is higher than the correlation between teachers' ratings and principals' self-ratings. This suggests that the ways in which principals conceptualize leadership and perceive leadership behaviors may be, in some ways, distinct from how teachers conceptualize and perceive their principal's leadership behaviors. This possibility raises interesting considerations for how we define teacher-principal congruence and how we construct measures to tap this concept.

When we consider shared time together, we see no significant results from teachers, principals, or on the congruence between the two. This may be indicative of heterogeneity of leadership as there is no reason to believe that spending more time together would foster exclusively negative or positive perspectives of leadership on the part of the teachers. If shared time between principal and teacher is reflective of an underlying organizational stability (e.g., low principal and teacher turnover), we may expect a positive relationship between shared time and perspectives of leadership. The absence

of such a relationship raises interesting questions as to how the teacher-principal relationship develops over time.

We have detected a negative gap associated with the proportion of teachers in a school with 11 to 15 years of experience. A closer examination of the principal and teacher coefficients shows that principals in these schools are generally rating themselves somewhat lower while teachers in this experience bracket also tend to rate principals somewhat lower. These coefficients are relative to the excluded group, the teachers with five or fewer years of experience. Schools with low proportions of new teachers are also likely to be schools with higher achieving students and perhaps more congenial working conditions (Loeb et al. 2005). If these are higher status schools, it follows that principals may see their appointments at these schools as some signal regarding their leadership ability (which may or may not be true) and rate themselves higher than principals at low-status schools (those with typically high proportions of inexperienced teachers).

Measures of variance are seldom seen in the literature on multi-source feedback. This is primarily owing to the relatively small number of subordinates in the majority of these studies, typically less than ten and often closer to five. This study has the advantage of having between 17 and 55 teachers per school (mean=35) and can therefore produce reasonable estimates of within group variation in leadership ratings. Here, we noted rather strong evidence that increased variation within schools was negatively associated with the extent of teacher-principal congruence. If congruence is operationalized as teacher-principal agreement, variation within school can be operationalized as teacher-teacher agreement and likewise thought of as a potential component of effective leadership. This analysis provides preliminary evidence for further consideration of within school agreement as an additional measure of leadership effectiveness.

Regarding school size, our prediction was that increasing student enrollment would increase the gap, likely in the negative direction. This hypothesis was based on the literature suggesting that principals in small schools would be better able to convey their leadership goals and communicate effectively with their faculty. Increased communication would attenuate the gap as school enrollment declined. The analysis results presented in Table 3 suggests that school size has no appreciable impact on perceptual congruence. There are two plausible reasons why school size may still be correlated with a smaller teacher-principal gap. First, the survey asks that teachers evaluate the principal's leadership in various domains, although this can be through the direct action of the principal or through the principal's designee. Such questioning then captures individual leadership behaviors as well as those distributed throughout administrative staff and leadership teams. In many school districts, as schools increase in size they are provided funds to hire additional administrative support staff. This increase in administrative personnel, coupled with the survey's distributive framing of leadership may effectively mask or ameliorate any impact of school size on the teacher-principal gap.

The second reason school size may be a factor in determining perceptual congruence without significant results here is tied to this particular sample. This study includes only elementary and middle schools, the largest of which has fewer than 60 teachers. Were we to expand this sample to include high schools where the number of teachers is typically larger and represents a more complex organizational structure involving subject-level departments, we may well find that school size does correlate with the

teacher-principal gap. The conclusion we draw from this finding is that the gap itself appears unchanged by the moderate variation in size found in these elementary and middle schools. In future research, controlling for the number of assistant principals may reveal if additional administrative personnel mitigates increases in schools size.

Future research This line of inquiry was predicated on the assumption that knowledge of teacher-principal congruence is meaningful in and of itself as it may reveal how leadership manifests within the school. Literature from business and management has presented this congruence as beneficial as an extension of self-awareness theory. There are, however, notable limitations to this approach. In an ideal setting, measurement of teacher-principal congruence would ask principals how they feel their teachers would rate them in addition to their own assessment of their leadership. Looking one step beyond, we could ask principals for the proportion of their faculty that would rate them at various levels along the 5-point scale. This approach would allow for the possibility where principals are making hard and potentially unpopular decisions that may lead to divisions in the faculty or low teacher ratings. Such a method would facilitate a more direct test of self-awareness theory as it pertains to self-other congruence. This would be the preferred approach to measure school leadership in challenging contexts such as during school reconstitution, new schools, or turnaround schools. In each of these scenarios, leadership is a crucial element to success and the use of self-other comparisons with attention to the extent of congruence may shed valuable light on a critical situation.

7 Conclusions

This paper set out to explore the extent of congruence between principals' perceptions of their own learning-centered leadership and the perceptions of their faculty. We also proposed several explanatory variables that may account for some of the variance in these self-other gaps. We found substantial variation in the gap, with as much variation within schools as between. The simple difference of the gap is normally distributed and centered almost precisely on zero. A portion of this variation can be explained through the addition of teacher-level and school-level covariates. Notably, principal's perceived mastery, the extent of principals' administrative experience, and within school variation in teacher ratings were significant in predicting the gap.

The primary implication of this research is that there is often a large, measurable gap in the perceptions of learning-centered leadership between teachers and principals. This difference implies that teachers may be seeing and interpreting elements of learning-centered leadership differently than are principals. Proponents of multi-source feedback would argue that the gap is a signal indicating that teachers may be revealing information that the principal does not have. Incorporating such information into systematic feedback may be a strong component in a larger effort to reform and improve instructional leadership and can support mechanisms for open dialog, communication and trust building.

Future research may continue in this direction, exploring aspects such as longitudinal trends in examining the gap or whether the gap induces behavioral change when tied to feedback, as predicted by cognitive dissonance theory.

Acknowledgments This study was made possible through the generous support of the Improving Principal Leadership through Feedback and Coaching project from the U.S. Department of Education's Institute of Education Sciences (IES). Funding for Peter Goff was provided by an IES grant through Vanderbilt's ExpERT program for doctoral training (R305B040110).

References

- Adkison, J. A. (1981). Women in school administration: a review of the research. *Review of Educational Research*, 51(3), 311.
- Atwater, L., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: does it really matter? *Personal Psychology*, 51, 577–598.
- Atwater, L., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self- and follower ratings of leadership. *Personnel Psychology*, 48(1), 35–59.
- Atwater, L., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personal Psychology*, 45, 141–163.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, 5, 307–337.
- Benlian, A. (2013). Are we aligned... enough? The effects of perceptual congruence between service teams and their leaders on team performance. *Journal of Service Research*. doi:10.1177/1094670513516673.
- Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(10), 1114.
- Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatric Services*, 62(12), 1423–1429.
- Blase, J., & Blase, J. (1999). Principals' instructional leadership and teacher development: teachers' perspectives. *Educational Administration Quarterly*, 35(3), 349.
- Bono, J. E., & Colbert, A. E. (2005). Understanding responses to multisource feedback: the role of core self-evaluations. *Personnel Psychology*, 58(1), 171–203.
- Boyd, B. B., & Jensen, J. M. (1972). Perceptions of the first-line supervisor's authority: a study in superior-subordinate communication. *The Academy of Management Journal*, 15(3), 331–342.
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434–455.
- Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *The Journal of Management Development*, 18(5), 417–435.
- Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: a meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4), 331–360.
- Church, A. H. (1997). Do you see what I see? an exploration of congruence in ratings from multiple perspectives. *Journal of Applied Psychology*, 27(11), 983–1020.
- Clark, D., Martorell, P., & Rockoff, J. (2009). *School Principals and School Performance*. CALDER. Available from <http://www.urban.org/url.cfm?ID=1001427>.
- Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, 50(6), 456–473.
- Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, 30(3), 203–235.
- Daubman, K. A., Heatherington, L., & Ahn, A. (1992). Gender and the self-presentation of academic achievement. *Sex Roles*, 27(3–4), 187–204.
- Doud, J. L., & Keller, E. P. (1998). *The K-8 principal in 1998: a ten-year study*. Alexandria, VA: National Association of Elementary School Principals.

- Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. *Organizational Behavior and Human Decision Processes*, 64(3), 307–324.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Fiebig, D. G. (2001). Seemingly unrelated regression. *A companion to theoretical econometrics*, 101–121.
- Fiedler, F. E. (1972). The effects of leadership training and experience: a contingency model interpretation. *Administrative Science Quarterly*, 17, 453–470.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- Ginsberg, R., & Berry, B. (1990). The folklore of principal evaluation. *Journal of Personnel Evaluation in Education*, 3, 205–230.
- Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: a theoretical analysis of its determinants and malleability. *Academy of Management Review*, 17, 183–211.
- Glasman, N. S. (1984). Student achievement and the school principal. *Educational Evaluation and Policy Analysis*, 6(3), 283–296.
- Goe, L. (2007). *The link between teacher quality and student outcomes: a research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., & Carson, B. C. (2009). The evaluation of principals: what and how do states and urban districts assess leadership? *The Elementary School Journal*, 110(1), 19–39.
- Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of cognitive dissonance: case study of the evolution of a theory. *Psychological Review*, 85(1), 53.
- Grissom, J. A., & Keiser, L. R. (2011). A supervisor like me: race, representation, and the satisfaction and turnover decisions of public sector employees. *Journal of Policy Analysis and Management*, 30(3), 557–580.
- Gross, N. C., & Trask, A. E. (1976). *The sex factor and the management of schools*. New York: John Wiley & Sons.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329–352.
- Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership, and student reading achievement. *The Elementary School Journal*, 96(5), 527–549.
- Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: understanding the impact on school capacity and student learning. *School Leadership and Management*, 30(2), 95–110.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, 86(2), 217–247.
- Ladd, H. F. (2011). Teachers' perceptions of their working conditions: how predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, 33(2), 235–261.
- Leithwood, K., & Jantzi, D. (1990). Transformational leadership: how principals can help reform school cultures. *School Effectiveness and School Improvement*, 1(4), 249–280.
- Leithwood, K. A., & Montgomery, D. J. (1986). *Improving principal effectiveness: the principal profile*. Toronto, Ontario: Ontario Institute for Studies in Education.
- Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, 69(2), 241.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, 80(3), 44–70.
- Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal*, 33(4), 757–798.
- Mandemakers, J. J., & Dykstra, P. A. (2008). Discrepancies in parent's and adult child's reports of support and contact. *Journal of Marriage and Family*, 70(2), 495–506.
- McCauley, C. D., & Moxley, R. S. (1996). Developmental 360: how feedback can make managers more effective. *Career Development International*, 1(3), 15–19.
- Moore, J. E. (2000). Why is this happening? A causal attribution approach to work exhaustion consequences. *Academy of Management Review*, 25, 335–349.
- Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007). Leadership for learning: a research-based model and taxonomy of behaviors. *School Leadership and Management*, 27(2), 179–201.
- Murphy, J., & Hallinger, P. (1992). The principalship in an era of transformation. *Journal of Educational Administration*, 30(3), 77–78.
- Polikoff, M. S., May, H., Porter, A. C., Elliott, S. N., & Goldring, E. (2009). An examination of differential item functioning on the Vanderbilt assessment of educational leadership. *Journal of School Leadership*, 19(6), 662–679.

- Porter, A. C., Murphy, J., Goldring, E., Elliott, S. N., Polikoff, M. S., & May, H. (2009). *VAL-ED technical manual*. Nashville, TN: Vanderbilt University.
- Porter, A. C., Polikoff, M. S., Goldring, E., Murphy, J., Elliot, S. N., & May, H. (2010). Investigating the validity and reliability of the Vanderbilt assessment of leadership in education. *The Elementary School Journal*, *111*(2), 282–313.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction*. Holt Rinehart and Winston.
- Reauthorization of ESEA, <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>. Accessed 20 March 2010
- Reeves, D. B. (2008). *Assessing educational leaders: evaluating performance for improved individual and organizational results* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Riemer, M., & Bickman, L. (2011). Using program theory to link social psychology and program evaluation. *Social Psychology and Evaluation*, 104.
- Riemer, M., Rosof-Williams, J., & Bickman, L. (2005). Theories related to changing clinician practice. *Child and Adolescent Psychiatric Clinics of North America*, *14*(2), 241–254.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys* (vol. 519). Wiley Online Library.
- Rotter, J. B. (1954). *Social learning and clinical psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: theory, research, and practice. *Journal of Clinical Psychology*, *61*(2), 145–153.
- Shope, S. C. (2013). *Developing an instrument to measure perceptual congruence among K-12 public school principals*. Doctoral dissertation, Ohio University.
- Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personality Psychology*, *58*, 33–66.
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: a meta-analysis. *Psychological Bulletin*, *124*, 240–261.
- Tschannen-Moran, M., & Gareis, C. R. (2004). Principals' sense of efficacy: assessing a promising construct. *Journal of Educational Administration*, *42*(5), 573–585.
- Urlick, A., & Bowers, A. J. (2013). What are the different types of principals across the United States? A latent class analysis of principal perception of leadership. *Educational Administration Quarterly*, 0013161X13489019.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573. doi:10.1037/0033-295X.92.4.548.
- Walker, A. G., Smither, J. W., Atwater, L. E., Dominick, P. G., Brett, J. F., & Reilly, R. R. (2010). Personality and multisource feedback improvement: a longitudinal investigation. *Journal of Behavioral and Applied Management*, *11*(2), 175–204.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.