How Effective is the Principal? Discrepancy between Teachers' and Principals'

Perceptions of Principal Effectiveness

Claire E. L. Sinnema[a]

Viviane M. J. Robinson[a]

Larry Ludlow[b]

and

Denyse Pope[a]

[a]Faculty of Education, The University of Auckland

[b] Lynch School of Education, Boston College

Corresponding Author:

c.sinnema@auckland.ac.nz

How Effective is the Principal? Discrepancy between New Zealand Teachers' and Principals' Perceptions of Principal Effectiveness

Abstract

Multi-source evaluation of school principals is likely to become increasingly common in education contexts as the evidence accumulates about the relationship between principal effectiveness and student achievement. The purpose of this study was to examine 1) the magnitude and direction of discrepancy between how principals and their teachers perceive the principal's effectiveness and 2) what predicts principals who are at risk because their self-ratings considerably exceed the ratings others give them. We also investigated the appropriateness of various probability cut-levels in analyses to predict over-rating principals. The data sources were ratings by New Zealand principals ($n = 135$) and their teachers ($n = 2,757$) of principal effectiveness—one scale (16 items) of an educational leadership practices survey. On average, both groups rated principals highly and teachers tended to rate their principal higher than the principals rated themselves. There was more variance in teachers' ratings than principals' ratings. The variables of principal age (younger), time in principal role at the school (shorter) and socio-economic status of the school (lower) were all associated with greater magnitudes of discrepancy. Such discrepancies have implications for principals' evaluations, principal development efforts and for school improvement.

Keywords: Leadership; Principal effectiveness; Educational leadership; Self-other agreement; Discrepancy

How Effective is the Principal? Discrepancy between New Zealand Teachers' and

Principals' Perceptions of Principal Effectiveness

There is now a body of evidence showing not only that the quality of school leadership makes a difference to student outcomes, but also identifying how that relationship varies for different types of leadership practice (Grissom & Loeb, 2011; Heck & Hallinger, 2010; V. Robinson, Lloyd, & Rowe, 2008). While much of that research has focussed on distributed, or school-wide leadership, some of it has focussed more specifically on the relationship between principal behaviours and characteristics and student outcomes (Grissom & Loeb, 2011; Heck, 1992; Heck, Marcoulides, & Lang, 1991).

In the context of school improvement, there is substantial evidence about the importance of the principal (Murphy, 2013; V. Robinson & Timperley, 2007). In concluding his comprehensive review of research on school improvement, Joe Murphy writes "The storyline is one in which school leaders provide the dynamism to make all the components of the [school improvement] framework function" (Murphy, 2013, p. 261). A similar conclusion was reached by the authors of a comprehensive empirical study of the organizational conditions required for sustained improvement in student achievement (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). The work of the principal in orchestrating the improvement effort while simultaneously building trust between students, teachers, parents and leaders is central to the success of school improvement efforts (Bryk & Schneider, 2002).

As evidence accumulates about the importance of the principal, policy makers,

professional associations and researchers are raising questions about their effectiveness and

about the quality and utility of systems for their evaluation and development. Recent research

in New Zealand has examined the effectiveness of principals in a range of capabilities that are

central to the task of instructional leadership—addressing concerns (Sinnema, Le Fevre,

Robinson, & Pope, 2013) handling performance issues, responding to disagreement (Le Fevre

& Robinson, 2014) enacting genuine inquiry (Le Fevre, Robinson, & Sinnema, In press) and

dealing with complaints (Robinson & Le Fevre, 2011). Other research has focused on the role

of principal evaluation as a policy instrument to promote such instructional leadership

(Sinnema & Robinson, 2007; Sun, Youngs, Yang, Chu, & Zhao, 2012) and on the accuracy of

principal evaluation instruments for identifying high and low performing principals (Minor et

al., 2014). Others have focused on the psychometric properties of such tools and have found

that many systems employ assessment tools that do not meet accepted standards, and so

provide principal evaluations of questionable accuracy (Ellen Goldring et al., 2009). In

addition, principals in the United States at least, report that their evaluations are of limited

utility in terms of providing challenging feedback that can inform their future development

(Reeves, 2008).

The situation may change if school systems adopt and adapt some aspects of the personnel

evaluation practices developed in the business sector. One such practice is the use of multi-

source evaluations in which the judgments of different role holders about the performance of

the target employee are systematically compared during feedback to the employee (Atwater,

Ostroff, Yammarino, & Fleenor, 1998; Church, 1997). In 360 degree evaluation, for example,

ratings from the leaders' superiors and subordinates are compared with the leader's self-

evaluation (Brett & Atwater, 2001; Smither, London, & Reilly, 2005).

The use of multi-source evaluation in school contexts allows consideration of at least four important matters: a) how principals themselves perceive their effectiveness, b) how teachers perceive the effectiveness of their principal, c) the magnitude and direction of any discrepancy between the two, and d) how to predict which principals are at risk because their self-ratings exceed, by a considerable margin, the ratings others give them. While the purpose of this study is to examine all four issues in the context of a national study of principal effectiveness, our emphasis will be on self-other agreement (SOA) between principals and their teachers and on the identification of those principal and school characteristics which predict discrepancy, particularly for those whom we call the over-raters.

Why Discrepancy Matters

In the following section, we discuss why investigation of self-other discrepancy is important for principal evaluation, principal development and school improvement. We then turn to the research literature on predictors of self-other discrepancy in performance ratings. First, however, we address the conceptual and methodological challenges to the use of difference scores as indicators of perceptual discrepancy, because such scores are central to our own study.

In a series of articles on difference scores, Edwards (1993, 2001) has argued that rather than employ some type of difference score, researchers should enter both ratings as component scores into their analyses. The conceptual argument for doing so is that "the value of the difference is fully determined by the [component values] taken jointly….In short, because a difference score is calculated from its components, it cannot represent a construct that is conceptually or operationally distinct from its components taken jointly" (Brett & Atwater, 2001, p. 279). While the examples Edwards provides illustrate the mathematical logic of his claim, we would argue that to say that a construct is fully determined by its components does

5

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
    Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
    Assessment, Evaluation and Accountability*

not imply that it is conceptually identical. The construct of SOA or discrepancy is richly linked

to the social psychological literature of self-insight (Atwater & Yammarino, 1997) and, at an

organizational level, to the idea of organizational trust (Bryk & Schneider, 2002), and these

linkages are lost unless SOA is directly measured.

The debate about the use of difference scores is particularly relevant to empirical research

on rating congruence and managerial effectiveness (Atwater et al., 1998). The use of a

difference score instead of the component scores obscures information about the respective

contribution of self and other ratings to performance outcomes. Where the relative

contributions of such ratings have been examined, it appears that others' ratings, particularly

those of peers and subordinates, explain far more of the variance in effectiveness than do self-

ratings (Reeves, 2008). These arguments are critical to studies of the consequences of

discrepancy but, as explained above, that is not the focus of this study. Our focus is on the

predictors of discrepancy, not its consequences.

Discrepancy and principal evaluation. One reason for using multiple raters in leadership

evaluation is that leadership is a complex phenomenon, and so the views of any single role

holder are likely to be partial and have limited reliability (Reeves, 2008). While some studies

on educational leadership have found that principals rate themselves lower than the ratings

from key staff in their schools, perhaps due to a halo effect (Day et al., 2011), others have

shown many self-ratings to be affected by leniency bias in the opposite direction (Halverson,

Tonidandel, Barlow, & Dipboye, 2005). This means that leaders are likely to rate their

effectiveness more positively than do their superiors, peers or subordinates. This is especially

true when the purpose of such ratings is believed to be primarily summative rather than

developmental (Smither et al., 2005). A more accurate picture of a principal's performance is

likely to be gained if the inevitably partial perceptions of those in different organizational roles

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
    Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
    Assessment, Evaluation and Accountability*

can be systematically compared. The reliability and validity of the principal's final evaluation

is likely to be substantially increased if the single self-assessment of the principal can be

compared with the ratings provided by a high percentage of the school's teachers.

While multi-source evaluation is likely to produce a more reliable final evaluation of the

principal, there are additional benefits of collecting assessments from others holding different

roles. One of the characteristics of effective leaders is self-awareness, and this is typically

judged by the match between self and other perception. The self-aware individual knows how

they are perceived by others because they have been open to their feedback and have

incorporated it into their own self-perception (Kwan, John, Robins, & Kuang, 2008). This is

not to say that leaders should uncritically accept others' views, because self-ratings should not

be assumed to be less accurate than those of others. The goal is openness to the views of others

and mutual adjustment so that, over time, there is both greater convergence of ratings and

enhancement of the performance skills that are being judged (Fleenor, Smither, Atwater,

Braddy, & Sturm, 2010).

The relationship between self-other discrepancy and employee effectiveness has been

empirically contested in the last two decades (Atwater et al., 1998). In some studies, SOA has

been unrelated to leader effectiveness, while in others, the relationship between SOA and

effectiveness has been much stronger. According to Atwater and colleagues, the conflicting

results may be due to an unclear conceptualization of the relative contribution of self-ratings,

others' ratings and difference variables to the explanation of variance in effectiveness. In

addition, the picture is further confused by measurement processes that may be mismatched to

the implicit or explicit conceptual framework employed (Atwater et al., 1998).

The relationship between discrepancy and individual employment outcomes varies

according to classification of the focal leader as an under, over, or in-agreement rater (Atwater

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
        Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
        Assessment, Evaluation and Accountability*

et al., 1998). While overrating may be associated with such positive outcomes as confidence

and risk taking, it is also associated with discounting of criticism that, over time can reduce

effectiveness. In a business context, at least, consistently inflated self-evaluations are

associated with low performance and career failure (Bryk et al., 2010). While the predominant

conclusion is that persistent overrating is undesirable, the findings are mixed (Fleenor et al.,

2010; Ostroff, Atwater, & Feinberg, 2004). The same is true for under-raters, with some

studies showing no relationship with employment outcomes (Ellen Goldring et al., 2009), but

others showing negative relationships with performance improvement (Atwater & Waldman,

1998; Atwater & Yammarino, 1992).

Discrepancy and principal development. In multi-source evaluations, principals not

only learn about their teachers' perceptions but, in addition, have the opportunity to compare

them with their own self-evaluation. Large or unexpected discrepancies between the two create

cognitive dissonance, which may prompt either a defensive rebuttal of the feedback or

productive inquiry into its origins (Smither et al., 2005).

The social psychology of dissonance reduction and research on response to feedback

suggests the conditions required if discrepant feedback is to lead to improved principal

performance (Smither et al., 2005). First, recipients must accept others' ratings as an accurate

indication of their perceptions, rather than be preoccupied with any methodological

shortcomings of the data under discussion. For example, principals may question the validity of

teacher ratings if they believe the teachers have not had sufficient chance to observe what the

principal does. Second, a discrepancy is motivating, and therefore more likely to lead to

improved performance, when it signals that the recipient falls short of a goal or standard that

they value. If, for example, a recipient strongly espouses being trusted by their staff and

discovers he is an over-rater on this indicator, he is more likely to act on the basis of this

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

feedback if being trusted is important to him. Such feedback requires the principal to be provided with a measure of discrepancy—a difference score that indicates the direction and magnitude of the difference between their own and others' views. Third, in addition to accepting the perceptual validity of the feedback, and valuing the standard it embodies, the recipients must believe that they have or can gain the capability to reduce the discrepancy. In other words, the principal must have a high sense of self-efficacy with respect to improving the climate of trust (Smither et al., 2005). If, however, the principal attributes the lack of trust to teacher characteristics that he cannot change, then the discrepancy will not be motivating even if the principal strongly espouses creating a more trusting climate.

When multi-source information is available, principals' development goals can be based not only on principals' own perceptions of their practice, but also on the perceptions of those whom they lead. If the goal setting process includes examination of the recipient's current capability and offers relevant learning opportunities, recipients are more likely to develop a stronger sense of efficacy and become internally committed to their development goals (Latham & Locke, 2006, 2007; Locke & Latham, 1990). Recent descriptive research on principals' evaluation and goal setting practices suggests that there is little focus on principal capability and it is largely assumed that they have the knowledge and skills required to achieve their personal and school development goals (Sinnema & Robinson, 2012). Such assumptions are more likely to be questioned if the evaluation process includes discussion of discrepancies between principals' self-ratings and those provided by their teachers. In addition to the extensive literature on goal setting, research on response to feedback provides many insights into how discussion of discrepancy can lead to learning and convergence rather than defensiveness and persistent discrepancy.

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

Discrepancy and school improvement. Discrepant teacher-principal perceptions of the principal's effectiveness have important implications for school improvement. School improvement involves two key leadership tasks – improving the technical core of teaching, learning and teacher learning while simultaneously building the relational trust that provides the social resource for the hard work of improvement (Bryk et al., 2010). The level of organizational trust reflects the extent to which all those in a role set (leaders, teachers, parents and students in the case of a school) perceive each other as fulfilling the expectations and obligations of their role. Given the interdependence of those roles in achieving joint work, each member is vulnerable to failures in the other's performance (Goddard, Tschannen-Moran, & Hoy, 2001). In the case of principals, those who overrate their level of effectiveness may engender mistrust on several grounds. First, teachers may perceive their principal's competency to be less than is required to lead a challenging and risky improvement effort. Second, inflated self-ratings may signal an interpersonal climate in which the positional power or personality of the principal has prevented an open flow of information and feedback both up and down the organizational hierarchy. In such schools, teachers and leaders are unlikely to agree on the need for change and mutual resistance rather than team work may predominate. Principals who rate their own effectiveness more highly than do their teachers may adversely influence their organization by setting unrealistic school targets or by being blind to their need to learn and improve (Fleenor et al., 2010). Hence, the identification of principals who possess characteristics associated with a high probability of being an over-rater would be a valuable step toward enhancing their organization's educational goals.

The under-rating principals may also present a risk to teacher-principal trust. If teachers perceive leadership as more positive than does the principal, this could signal that the principal holds higher standards of leadership than the teachers. This also brings risks if teachers do not

10

aspire to the standards that the principal holds and fear they will be generalized to their own performance. Whatever the level of effectiveness and the direction of the discrepancy, large perceptual gaps may signal low organizational coherence, low trust and poor teamwork (Bryk & Schneider, 2002).

While we are aware that much of the above discussion of the consequences of discrepancy for school development is speculative, research in business contexts does include some evidence of positive correlations between SOA and job satisfaction and organizational commitment (Fleenor et al., 2010). It is likely that the organizational consequences of SOA, such as level of trust and quality of problem-solving are cumulative over time and therefore difficult to discern with cross-sectional designs (Halverson et al., 2005).

Predictors of Discrepancy

Predictors of discrepancy have been given much attention in the literature on leadership and management, with studies typically involving leaders from the manufacturing, transport, and finance sectors (Fleenor et al., 2010). More recently, researchers in the field of educational leadership (Atwater & Waldman, 1998; E. Goldring & Goff, 2010) have begun examining convergence and discrepancy in ratings of leadership effectiveness, but not surprisingly, given the limited use of multi-source evaluation in education, there is, as yet, little empirical evidence. While findings about predictors of self-other agreement in business contexts are not entirely consistent, they do suggest some predictors that may be applicable to the school context.

Among the personal characteristics of leaders that predict discrepancy are gender and age. Both male and older leaders are more likely to rate their leadership effectiveness more positively than their subordinates (Avolio, Mhatre, Norman, & Lester, 2009; Brutus, Fleenor,

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

& McCauley, 1999; Fleenor et al., 2010; Ostroff et al., 2004; Vecchio & Anderson, 2009;

Visser, Ashton, & Vernon, 2008).

Predictors of differences between 'self' and 'other' ratings of leadership behaviors were

examined in a study of 4,493 U.S. managers across 654 organizations participating in a

leadership development program (Ostroff et al., 2004). The managers were rated by various

'others' -13,706 subordinates (an average of four per manager), 13,752 peers, and 3,994

supervisors. They found that while female managers rated themselves similarly to the ratings of

subordinates, male managers were over estimators—rating themselves more highly than their

subordinates rated them. The authors suggest that the male overestimation effect is not because

men rated themselves higher than female managers but, rather, because subordinates rated male

managers lower than women. A similar pattern was found for peers and supervisors.

In a study of managers taking part in a leadership development program, managers' self-

ratings were compared with those of their peers, supervisors and subordinates (Brutus et al.,

1999). The authors suggested two possible explanations for their finding that male managers

overrated themselves in comparison to the ratings they received from their subordinates and

peers. The first possibility is that both males and females are equal in their actual leadership

effectiveness. In this scenario, the explanation for males' overestimation could be that the

majority of those in subordinate and peer roles were male (making up the majority of the

'other' ratings) and may be more competitive, therefore rating their male seniors or peers lower

than their actual performance deserves. The second possibility is that female leadership

effectiveness is actually higher than males. In this scenario, the explanation for males'

overestimation, Brutus et al. (1999) suggest, is that female managers face organizational

obstacles and selection hurdles in progressing up the ranks, and thus only better-than-average

female managers are promoted. That, in turn, might explain the relatively higher effectiveness

ratings given to female managers and the reduced likelihood of female managers

overestimating in the way that males did.

In a study of managers in a range of organizations , including education, health care,

government, insurance and manufacturing, (Vecchio & Anderson, 2009) males also tended to

overrate their effectiveness, but only in relation to the ratings of peers and superiors (not

subordinates). This study, however, restricted the sample of subordinates to just one randomly

sampled subordinate for each manager whereas multiple subordinates were used in the other

studies mentioned.

With regard to age, Ostroff et al. (2004) found that managers who were younger, rated

themselves similarly to how their subordinates and supervisors rated them. Older managers,

however, were more likely to rate themselves more highly than their subordinates and

supervisors rated them. The overrating pattern of older managers was magnified because of the

tendency of subordinates and supervisors to rate older managers lower than they rated younger

managers. That tendency contributed to the finding that older managers typically overestimate

their leader behaviors. These findings mirror those described by Brutus et al. (1999) who also

found that managers' age was related to an increase in self-supervisor discrepancy. The authors

propose several possible reasons for this – one is the significant relationship between age and

self-ratings – the tendency to rate one's performance higher as one gets older. This tendency,

they suggest, could be due to the presence of age discrimination in the workplace, leading male

participants to compensate for the discrimination by giving themselves inflated ratings. Or

perhaps males in the study were incorporating age as a positive factor in their own ratings in a

way that others did not.

Characteristics of the situation may also be predictive of discrepancy. A number of

studies indicate that the validity of person perception increases with acquaintanceship (Paulhus

& Bruce, 1992) so that there is likely to be greater perceptual discrepancy when a principal is new to a school. A study of discrepancy in a school context, however, found quite the opposite. In a United States study of 76 school principals and more than 2000 teachers there was less congruence between principal and teachers' ratings when they had spent more time together (Atwater et al., 1998; Atwater & Waldman, 1998). The same study did show the variable of principal experience (the total number of years as a principal at their current and previous schools) to be a modest predictor of teacher-principal congruence. Atwater and Waldman (1998) did not, surprisingly, find a relationship between teacher-principal congruence and the size of the organization.

Given our earlier discussion of the significance of perceptual discrepancy and the paucity of research about leader-teacher discrepancy, our study was designed to describe the magnitude and direction of discrepancy in a national sample of school leaders and to test the ability of selected leader and school characteristics to predict principals who rated their own effectiveness more highly than did their teachers (over-raters). Three empirical questions frame the research: 1) How comparable are principals' self-ratings of their effectiveness with the ratings given to them by their teachers? 2) What is the magnitude and direction of the discrepancy between principal and teacher perceptions of principal effectiveness? 3) What predicts principals who over-rate in relation to their own teachers?

Method

The findings reported in this study were drawn from a wider data set comprising teacher and principal ratings of the effectiveness of both distributed (school-wide) and principal leadership. In this paper we focus exclusively on self and teacher ratings of principal effectiveness. In this section we first describe the sampling procedures, treatment of missing data and the characteristics of our final sample of principal and teacher participants. Then we

outline the properties of our principal effectiveness measure and our approach to the

calculation of school-level discrepancy between principals' self-ratings and those of their

teachers.

Research Context

The research took place in New Zealand, an education system covering approximately

2,500 schools where school leaders and teachers have a great deal of autonomy in relation to

curriculum and pedagogy. Student achievement levels in New Zealand are generally high,

though equity of student achievement across diverse learner groups is low. In PISA, for

example, New Zealand has higher-than-average scores but the impact of socio-economic status

on attainment (17%) was higher than the OECD average of 14% (Pont, Figueroa, Zapata, &

Fraccola, 2013). Performance and completion gaps persist within the student population. And

particularly for Māori (indigenous people) and Pasifika students who represent more than one-

third of the student population. Diversity of the student population in New Zealand is

increasing, but Māori and Pasifika students face lower outcomes and may be less likely to

complete their secondary education.

The student achievement challenge, and the need for school leaders who can effectively

address those challenges is a key concern of the New Zealand education system. Since the

system is characterized by self-governing schools those challenges are core business for

principals and their Board of Trustees, who are elected by the parents of the children enrolled

in that school. A standard Board of Trustees membership includes between three and seven

parent-elected trustees, the principal of the school, on e staff elected trustee, one student

elected trustee I schools with students above Year 9) as well as any co-opted trustees or

proprietors representatives in state integrated religious schools.

Principals act as the Board's chief executive, and Boards are responsible for the

appointment and performance management of their principal, including evaluation of their

leadership effectiveness. There are few regulations about how the principal is to be evaluated,

with the main stipulation being that an annual evaluation occurs. Typically, the school's Boards

of Trustees contract a consultant to carry out the principal evaluation on their behalf—the

approaches used to judge principal effectiveness vary greatly between consultants and are

rarely be based on assessments generated from multi-source evaluations using quantitative

measures or from validated measures. Alternatively, the school Board can carry out the

evaluation using the capabilities of Board members—the risk with this approach is that there is

no requirement for Board members to have educational expertise, and often their experience

and expertise lies in other fields.

The results of principal evaluations are confidential to the principal and Board. Many of

the principals involved in this study used their own data about principal and school-wide

leadership effectiveness from the measures discussed in the following section to inform those

evaluations.

Our Model of Discrepancy

Research on the causes and consequences of discrepancy has shown the importance of

drawing distinctions between the direction and magnitude of discrepancy (Atwater &

Waldman, 1998). The significance of discrepant ratings differs depending on the magnitude

and the direction of any discrepancy (Table 1). Many prior studies have examined the

significance of such differences on measures of effectiveness and outcomes. There has been

much criticism of research that has used a categorization approach to examine the predictors of

effectiveness and outcomes—for example studies of whether leaders' membership in a

category of under- or over-raters predicts important organizational outcomes.

For the purposes of our study, we agree that consideration of both magnitude and direction

of discrepancy is important, but our question is distinctly different from those mentioned above

since we are interested in predicting discrepancy, rather than examining what discrepancy

predicts. Therefore, we argue that in the case of educational leadership in particular, predicting

principals' likelihood of overrating their effectiveness in relation to their teachers' ratings

requires a categorization approach that considers both magnitude and direction of discrepancy

(see Table 1).

Table 1

*A Six Group Categorization of Discrepancy in Principal and Teachers' Ratings of Principal
Effectiveness*

| Principal effectiveness | Discrepancy categorization | | |
	Discrepancy – over-rater	In agreement	Discrepancy – under-rater
High	1. A high performing principal who rates their effectiveness more highly than do the principal's teachers	2. A high performing principal whose self-rating of effectiveness is in agreement with the rating of the principal's teachers	3. A high performing principal who rates their effectiveness lower than do the principal's teachers
Low	4. A low performing principal who rates their effectiveness more highly than do the principal's teachers	5. A low performing principal whose self-rating of effectiveness is in agreement with the rating of the principal's teachers	6. A low performing principal who rates their effectiveness lower than do the principal's teachers

*A Six Group Categorization of Discrepancy in Principal and Teachers' Ratings of Principal
Effectiveness*

As the table above shows, we propose a six categorization model organised around

measures of both principals' effectiveness and levels of discrepancy or agreement. The

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
Assessment, Evaluation and Accountability*

categories of 'in-agreement' refer not to perfect agreement (since that is rare) but rather to

levels of discrepancy that are not markedly different from the levels of discrepancy on average

across all principals (our attention to magnitude). The categories of 'over-rater' and 'under-

rater' refer to those principals who rate themselves markedly higher or lower than teachers (our

attention to direction). We use a measure (detailed later) of individual principal/teachers'

discrepancy compared to the standard deviation (SD) of all principals' discrepancy to

determine inclusion in the over- and under-rater categories—those who exceed the average SD

of discrepancy are treated as discrepant.

For the purposes of this study, our focus has been on predicting the likelihood of

principals over-rating in relation to teachers rather than under-rating or being in agreement

with them. That question presents much complexity and has been an important first step in this

area (predicting discrepancy in educational contexts) that has had little research attention to

date. We did not include consideration in this phase of the research to the actual level of

principal effectiveness (high effective or low effective) as signalled in the left hand column of

the table. We have, however retained that aspect in the table, since it signals an important

consideration for future research.

While the magnitude of discrepancy is high for categories 1 and 3, the direction of

discrepancy is quite different with principals in category 1 overrating their effectiveness

relative to their staff and principals in category 3 doing the opposite. As we shall see in the

next section, the causes and consequences of these two types of discrepancy are likely to be

quite different, including their implications for principal development. While the level of

effectiveness of category 5 principals is problematic, leader and teachers both agree that this is

the case. Public discussion of the information could be a first step to the negotiation of support

and increased appreciation of the principal's openness to leadership learning.

In contrast, principals in category 4, who are over-raters in comparison to their teachers,

may be equally as ineffective and, in addition, unaware that they are perceived to be so. In

contrast to the category 5 situation, there are two potential leadership problems for category 4

principals; low capability and low self-awareness. The implications of this type of discrepancy

for principal and school development are very different from that represented in category 3.

There are at least two possible causes of this 'under-rating', each holding different implications

for leadership evaluation and development. The first is that category 3 principals are overly

modest in their judgment. The second possibility is that these principals hold higher standards

for their own performance than their teachers hold for them. In such cases, the discrepancy

may prove to be productive if teachers and principals develop a shared understanding of and

commitment to the leader's higher standards. We argue in the following sections that, in

general, under-rating is a less problematic type of discrepancy than overrating.

Sampling Procedures and Sample Attrition

Principals (*n* = 159) were recruited from two sources. Ninety-five experienced principals

were recruited from an 18 month national professional development (PD) program. A separate

process of voluntary recruitment into the associated research was run shortly after the

commencement of the development program. A further national sample of 64 principals was

recruited by approaching school principals that had characteristics similar to those in the PD

sample on principal gender, school type (primary, intermediate, secondary, composite), and

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
        Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

school decile[1] group (low, medium, high). Eighty-three principals in the PD program (87.4%

response rate) and fifty-seven of the 64 principals from the national sample (89.1% response

rate) completed the survey on leadership effectiveness. The overall principal response rate was

88.1%.

Teachers (*n* = 3,160) were recruited from the schools of the principals who had already

agreed to participate. Of these, 2,254 were from the schools in which the principal was

undertaking the PD program, with the remaining 906 being from the schools in which the

principal was from the national sample. The response rate for teachers could not be established

as all full time teachers at the participating schools were asked to complete the survey online

and the staffing complement at each school is not known to the researchers.

Prior to participation, principals and teachers gave written consent. Three incentives were

offered to principals who agreed to participate. The first was provision of detailed reports of

the perceived effectiveness of school-wide and principal leadership in their own school. The

second incentive was a book voucher and the third was a professional learning opportunity to

discuss the findings of this research and its possible implications for their leadership.

Missing data. According to Scheffer (2002), when data are either missing completely at

random (MCAR), or missing at random (MAR), acceptable results using imputation

procedures can be obtained for data sets with up to 50% missing data. In this study, the more

conservative strategy of withholding cases with more than 30% missing data was adopted.

---

[1] A school's decile rating indicates the extent to which it draws its students from low socio-economic

communities. Decile 1 schools are the 10% of schools with the highest proportion of students from low socio-

economic communities, whereas decile 10 schools are the 10% of schools with the lowest proportion of these

students.

Table 2 reports the source of sample attrition at each stage of determining our final

sample. The initial step was to consider missing data across eight leadership scales – a

Principal Effectiveness Scale (PE) (which is the focus of this article) and seven additional

school-wide leadership scales that comprise the overall instrument (see measures section for

additional detail on the scales).

Table 2

Number and Percentage of Participants Removed at Each Stage of Sample Attrition

Stage of sample attrition	Principal (*n*)			Teacher (*n*)		
	PD Sample	National Sample	Total	PD Sample	National Sample	Total
Survey responses	83 (100%)	57 (100%)	140 (100%)	2,254 (100%)	906 (100%)	3,160 (100%)
Cases removed due to missing data	0 (0%)	0 (0%)	0 (0%)	160 (7.1%)	190 (21.0%)	350 (11.1%)
Cases removed due to non-corresponding principal or teacher	2 (2.4%)	3 (5.3%)	5 (3.6%)	52 (9.4%)	1 (5.9%)	53 (1.7%)
Remaining cases for analysis (% of responses received)	81 (97.6%)	54 (94.7%)	135 (96.4%)	2,042 (90.6%)	715 (78.9%)	2,757 (87.2%)

Where more than 30% data were missing across the eight leadership scales the participant

was removed from further analysis. This removed eight principals, and 2,612 teachers. The

next step removed cases where there was no response to more than 30% of the items on any

one of the scales (350 teachers, 71 of whom had been removed because of their low response

level to the PE scale). In addition, five principals and 53 teachers were removed because either

a principal completed the survey but no teachers from that principal's school did, or teachers

completed the survey but the principal associated with the school did not. At the conclusion of

data cleaning, response records were available for 135 principals and 2,757 teachers. Table 3

presents demographic information for the final sample.

Table 3

Characteristics of the Principal and Teacher Respondents to the ELP Survey

	Principals Total (*n* = 135)	Teachers Total (*n* = 2,757)
Age Categories		
Less than or equal to 49 years	29 (22%)	
50 - 59 years	84 (62%)	
60+ years	20 (15%)	
Not specified	2 (2%)	
Gender		
Female	75 (56%)	2,040 (74%)
Male	60 (44%)	643 (23%)
Not Specified	-	74 (3%)
Ethnicity		
New Zealand Europeans/Pakeha	122 (90%)	2,021 (73%)
Maori	2 (2%)	97 (4%)
Pasifika	1 (1%)	121 (4%)
Asian	-	164 (6%)
Other	-	81 (3%)
Mixed	9 (6%)	183 (7%)
Not specified	2 (2%)	90 (3%)
Highest educational qualification		
Below Bachelor's degree	20 (14%)	
Bachelor's degree	18 (13%)	
Post-graduate Diploma or Certificate	8 (6%)	
Master's degree or higher	34 (25%)	
Not Specified	55 (41%)	
Years of experience		
0-5 years	18 (13%)	698 (25%)
6-10 years	43 (32%)	535 (19%)
11-15 years	32 (24%)	362 (13%)
More than 15 years	40 (30%)	1,103 (40%)
Not specified	2 (2%)	59 (2%)
Time in role at school		
0-2 years	17 (13%)	824 (30%)
3-5 years	39 (29%)	778 (28%)
6-10 years	42 (31%)	601 (22%)
11-15 years	22 (16%)	228 (8%)

More than 15 years	13 (10%)	264 (10%)
Not specified	2 (2%)	62 (2%)
School type[a]		
Primary (Years 1-6 and Years1-8 schools)	82 (61%)	1,197 (43%)
Intermediate (Years 7-8 schools)	14 (10%)	287 (10%)
Secondary (Years 7-13 and Years 9-13 schools)	28 (21%)	1,027 (37%)
Composite (Years 1-13 schools)	9 (7%)	209 (8%)
Other	2 (2%)	37 (1%)
School Economic Status of School Community		
Low Decile[b]	36 (27%)	701 (25%)
Medium Decile	45 (33%)	846 (31%)
High Decile	53 (39%)	1,201 (44%)
Not specified	1 (1%)	9 (0%)

[a] In the United States the first three New Zealand school group categories are comparable to elementary, middle and high school, respectively.

[b] A school's decile rating indicates the extent to which it draws its students from low socio-economic communities. Decile 1 schools are the 10% of schools with the highest proportion of students from low socio-economic communities, whereas decile 10 schools are the 10% of schools with the lowest proportion of these students.

Supplementary missing values analyses were carried out to ascertain the extent to which the remaining missing data on the items were random or systematic in their occurrence. The SPSS 19.0 MVA module was used to test the patterns of missing data on the PE scale items. Analysis was first carried out separately on the principal and teacher groups. For principals, Little's Missing Completely At Random (MCAR) test indicated items on the PE scale were not MCAR ($\chi^2 = 45.97$, df = 30, $p = .03$) when school and principal characteristics (Table 4) were used as predictors. Further investigation revealed that 41.4% of missing data for the principals' highest qualification resulted from whether the principal was in the PD program or sourced from the national sample, i.e. principals from the national sample group tended to not provide their highest qualification. The MVA was subsequently re-run with the principals separated based on this grouping. Although the items on the PE scale were still not MCAR for these two groups, only four participants were missing data in conjunction with other predictor variables,

no other variables in Table 4 showed significant differences in their patterns of missing data,

and there was less than 5% data missing overall. The principals' missing data were

subsequently treated as random. For teachers, the PE scale was MCAR (χ^2 = 1256.31, df =

1,234, p = .32) when school and teacher characteristics were used as predictors (Table 4).

Table 4

Predictor Variables Used in Missing Values Analysis

School	Principal	Teacher
Socio Economic Status of School Community	Age Group	Gender
School Type (Primary, Intermediate, Secondary, Composite and Other)	Gender	Length of Role as Teacher
Aggregated School Type	Length of Role as Principal	Length of Role as Teacher at School
School Area (Urban, Rural)	Length of Role as Principal at School	Length of Role Teaching with Principal
School Level (Primary, Secondary)	Length of Role as Senior Manager	Employment Status (Fixed Term, Contract etc.)
Percentage of Maori Students	Previous Teaching Experience	Employment Status (Full or Part Time)
	Highest Qualification	
	Completion of PD Program	
	Upgrade Level	

As missing data were either MAR or MCAR multiple imputation was employed

separately for principals and teachers. Imputation was undertaken using the random generator

Mersenne Twister and random seed 234567, with five datasets additional to the original being

generated. The PE scale items were imputed using the predictors in Table 4. For principals,

qualification type and whether the school was primary or secondary were not used.

Measures

The Educational Leadership Practices (ELP) survey was administered by the New Zealand Council for Educational Research. One version was designed for completion by principals, and the other by teachers[2]. Both versions were administered electronically and included seven scales (66 items in total) focused on the effectiveness of school-wide leadership practices (Goal Setting, Strategic Resourcing, Curriculum Quality, Quality of Teaching, Collaborative Teacher Learning and Development, Safe and Orderly Environment, Relationships with Families and Community), and one scale (focused on the effectiveness of the principal (the PE scale). The PE scale is the focus of the study reported here.

The PE scale comprises 16 items (Appendix) which ask teacher respondents to indicate how effective (rated on a 5-point scale - 1 = Ineffective; 2 = Minimally effective; 3 = Satisfactorily effective; 4 = Highly effective; 5 = Outstandingly effective) they consider their principal to be at decision making, problem solving, leading instructional improvement, leading teacher learning and gaining the respect of staff and the parent community. Those items were developed based on a theory of student-centered leadership that outlines leadership capabilities that are required for school leaders to be effective in their practice (V. M. J. Robinson, 2011). Furthermore, the 16 items of the PE scale are intended to measure a single, unidimensional common factor, i.e. principal effectiveness. In addition contextual information, including demographic details, were included in the questionnaire.

---

[2] We wish to thank Andrew Porter (University of Pennsylvania), Joseph Murphy (Vanderbilt University), Ellen Goldring (Vanderbilt University), and Stephen N. Elliott (Arizona State University), the authors of the Vanderbilt Assessment of Leadership in Education, for use of portions of the Vanderbilt Assessment of Leadership in Education (VAL-ED) structure and other items from the VAL-ED.

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

Scale reliability. The reliability of the PE scale score was examined prior to imputation and following imputation (six data sets in total). Cronbach alphas for all data sets were between .968 and .972 and no corrected item-total correlations were less than .709.

Item level means for each imputed data set were compared to the original data set. No item mean for any imputed data set was significantly different from the original data set. On that basis one imputed data set was randomly selected (data set 4) for the remaining analyses.

Scale structure. Confirmatory factor analysis (CFA) was undertaken to validate the theoretical unidimensional structure of the PE scale. Based on the Chi-Square statistic alone, the model and data did not reflect a good fit ((χ^2 (96, $N = 2,892) = 1777.28, p = < .01$)). However, good fit was indicated when Hu and Bentler's Two-Item Index Presentation Strategy (1999) was employed (CFI = .96, SRMR = .03). In this situation error terms between items assessing similar themes (e.g. teaching/learning, and respect/integrity) were allowed to covary. Based on the strength of the reliability and CFA results, subsequent analyses performed on the PE total score are psychometrically justified and consistent with the measurement recommendations of Fleenor et al. (2010).

Principal effectiveness discrepancy analyses. There are both conceptual and methodological issues involved in the choice of a specific difference measure (Atwater et al., 1998; Atwater & Yammarino, 1997). Balancing calls over the past decade for sophisticated statistical models such as response surface modeling with an appreciation that these techniques are "not very useful" at times and that "simple indices" may be recommended as useful indicators of over- and under-ratings (Fleenor et al., 2010, p. 1027), we present and discuss an approach which may be employed at the individual principal level within any school organization. We recognize that there are highly contested research design issues around a) the accuracy of self- and other-ratings, b) the extent of their agreement or discrepancy, c) the

27

choice and modeling of predictors of rating congruence and d) how the ratings are used, either separately or in combination, to predict leader or organizational performance. While we accept that these wider design challenges exist, we focus on how to maximize a school organization's effectiveness through the early identification of principals most likely to over-rate their leadership effectiveness.

In comparing the perceptions of teachers and principals, important information is gained by knowing whether teachers, on average, rate leadership as more positively, the same, or less positively than does the principal. This information may be provided by a calculation of the simple mean difference, or mean discrepancy, based on each individual teacher's discrepancy score: $PE_{Discrepancy} = PE_{Principal} - PE_{Teacher}$. While we recognize the benefits of multivariate regression approaches to prediction when the prediction of both joint and separate self and other variables is useful (Fleenor et al., 2010), for the purposes of our context and research question, only the prediction of a single index of discrepancy was relevant, hence our use of a raw difference score.

The limitation of the simple difference formula we used is that it underestimates perceptual differences because negative and positive differences tend to cancel one another out when a mean is calculated. This approach can create a mean school simple difference score of 0 (suggesting coherence in the views of the principal and their teachers) that masks the considerable differences within the staff about how they see the effectiveness of the principal—for example the mean school simple difference in a school could be zero, but two teachers in that school consider the principal ineffective, two consider the principal outstandingly effective, and half of the teachers have more moderate views (minimally/satisfactorily or highly effective).

When the same data as in the above example are used to calculate a mean based on the

absolute value of the difference, the direction of the difference is ignored and the result is a

larger average discrepancy because this approach calculates the overall disparity of perceptions

in both directions. This difference is calculated as $|\text{Individual PE}_{\text{Discrepancy}}| = |\text{PE}_{\text{Principal}} - \text{PE}_{\text{Teacher}}|$. Both simple and absolute differences were employed in our analyses.

The next step in this process consisted of calculating teacher PE means for each school,

i.e. summing the individual teacher scores and averaging them for that school. Mean simple

and absolute differences were then calculated from the difference between a principal's self-

rating of PE and the principal's teachers' mean to determine the direction and the magnitude of

the discrepancy for that school. These school level simple and absolute difference means are

the data for the discrepancy prediction analyses.

Analyses

An iterative approach to analysis was employed. The first steps employed various t-test,

correlational, and contingency table procedures to identify variables that differentiated between

over- and under-rater response patterns. These steps were followed by logistic regression

procedures that predicted the probability of being an over-rater while simultaneously

minimizing their misclassification. The logistic regressions consisted of full models containing

all predictors of principal and teacher PE discrepancies and then reduced models yielding a

final parsimonious solution.

<div align="center">

Findings

</div>

We first outline our findings about the effectiveness of principals as rated by both

principals and teachers. We then report the magnitude and direction of discrepancy between

principal and teacher ratings and conclude with analyses of discrepancy predictors.

Principal Effectiveness

On average, both principals and teachers rated principals between satisfactorily and highly effective (principals M= 3.69 (SD = 0.47) and teachers M = 3.73, SD = 0.86). An independent samples t-test (with unequal variances) showed no statistically significant difference between the principal and teacher means; t(179.802) = -.880, (n principals = 135, n teachers = 2,757), p = .380, d = .06. Across the two samples, teachers and principals generally agreed in their overall perceptions of the latters' effectiveness.

The distributions of principal and average school level teacher ratings, however, differ in their variability and both reflect a negative skew, i.e., the PE ratings tend to fall at the high end of the scale (Figure 1). Interestingly, at the very highest level, teachers tended to rate their principal higher than the principals rated themselves. Slightly more than 20% of principals rated themselves as being highly or outstandingly effective, whereas teachers were nearly twice as likely to rate their principal at this level. Principals were also less likely (3%) to rate themselves as being ineffective to satisfactorily effective than their teachers, who were six times more likely to rate their principal at these lower levels. Hence, even though the teachers tended to use the highest rating category more than the principals, the teacher and principal mean ratings were not statistically different because the teacher mean was brought down by their tendency to also rate some principals much lower than did the principals themselves.

With respect to the variability in their ratings, the variance of the principal ratings is statistically significantly less than the variance of the teacher mean ratings (SD^2 = 0.22 versus SD^2 = 0.74; Levene's homogeneity of variance test = 54.762, p < .001). One possible reason for the relatively narrow variance of the principals' self-ratings may be that they were aware that they would be provided with a report that compared their own and their teachers' ratings. Even though they had control over the distribution of this report, they may have felt obligated

30

to share it with their teachers. Their anticipation of such disclosure may have led them to avoid

rating themselves as either highly effective or as anything less than satisfactory.

In summary, principals' average self-ratings of effectiveness closely matched those of

teachers, with both groups rating principals as, on average, between satisfactorily and highly

effective. The principals, however, showed much less variability in their ratings of their

effectiveness than did the teachers.
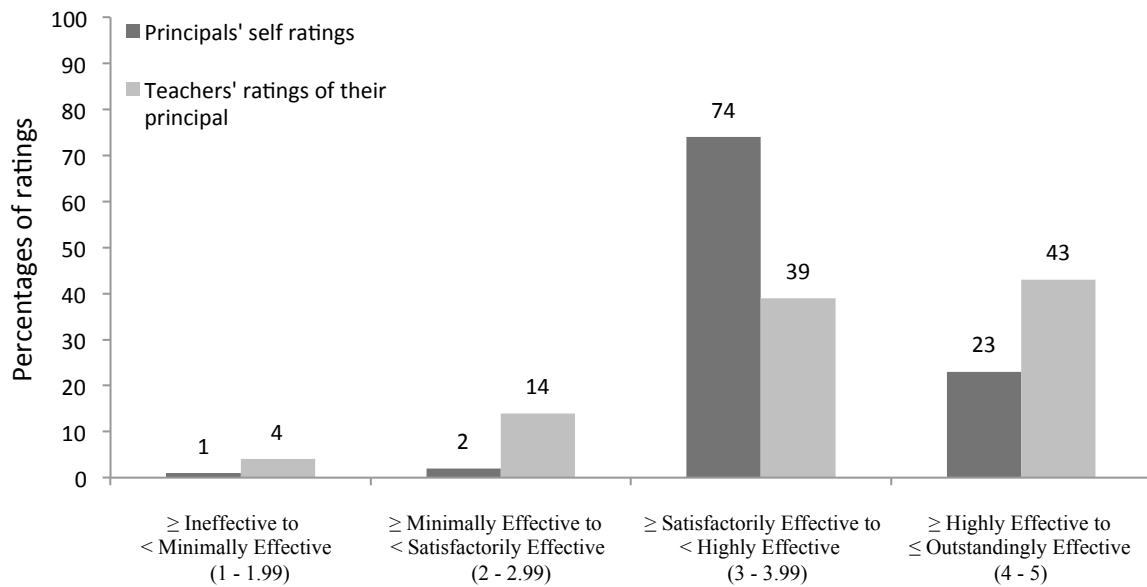
Figure 1. Distribution of average principal effectiveness ratings.

The Magnitude and Direction of PE Discrepancies

The mean simple discrepancy between principal and teacher ratings at the school level

was small, as expected, and negative ($M = -0.10$, $SD = 0.60$). This indicates that principals'

self-ratings were slightly lower than the ratings they received from their teachers and is

consistent with the results of the principal and teacher mean rating comparisons reported

above. The mean absolute discrepancy at the school level ($n = 135$), however, was 0.74 of an

interval ($SD = 0.26$; $SE_{mean} = 0.02$) on the five point scale. Not only is a ¾ point discrepancy

meaningful on a five point scale, it is statistically significant when the sample mean (0.74) and

its standard error (0.02) is tested against a null hypothesis of zero mean absolute discrepancy (z

$= 33.6$, $p<.001$).

Predicting Magnitude of Discrepancy

Although the mean ratings for the teachers and principals were not significantly different,

the fact that the teachers tended to give both higher and lower ratings of PE than did their

respective principals suggests that these discrepancies may be due to specific school and

principal characteristics. Table 5 contains the mean absolute discrepancies for the principal and

school demographic variables. The first column represents the overall sample. The next two

columns separate the principals into "under-raters" and "over-raters". In some instances the

categories of the variables were combined to increase the sample sizes for significance testing.

Table 5

Mean Absolute Differences between Principal and Teacher Ratings of Principal Effectiveness for all Principals, and Under and Over-rating Principals

Participant and School Context Characteristics	All Principals (*N* = 135)		Under-rating Principals (*N* = 31)		Over-rating Principals (*N* = 23)	
	n	Mean *(SD)* Absolute Difference	*n*	Mean *(SD)* Absolute Difference	*n*	Mean *(SD)* Absolute Difference
Gender						
Female	75	0.76 *(0.29)*	20	0.93 *(0.27)*	14	0.97 *(0.35)*
Male	60	0.72 *(0.21)*	11	0.75 *(0.27)*	9	0.86 *(0.17)*
Age						
Less than or equal to 49 Years	29	0.77 *(0.30)*	9	0.82 *(0.33)*	2	1.11 *(0.59)*
50 to 59 years	84	0.74 *(0.26)*	18	0.93 *(0.26)*	13	0.95 *(0.30)*
60+ years	20	0.73 *(0.19)*	4	0.67 *(0.14)*	8	0.85 *(0.21)*
Not specified	2					
Length of role as principal						
0 to 5 years	18	0.70 *(0.21)*	2	0.85 *(0.50)*	2	0.84 *(0.12)*
6 to 10 years	43	0.73 *(0.27)*	13	0.92 *(0.32)*	5	0.90 *(0.22)*
11 to 15 years	32	0.80 *(0.29)*	10	0.86 *(0.26)*	7	0.92 *(0.37)*
16+ years	40	0.73 *(0.24)*	6	0.76 *(0.14)*	9	0.98 *(0.32)*
Not indicated	2					
Length of role as principal at the school						
0 to 2 years	17	0.73 *(0.33)*	5	0.71 *(0.20)*	4	1.07 *(0.47)*
3 to 5 years	39	0.75 *(0.23)*	4	1.01 *(0.23)*	5	0.93 *(0.17)*
6 to 10 years	42	0.77 *(0.31)*	12	0.99 *(0.32)*	7	0.95 *(0.34)*
11 to 15 years	22	0.73 *(0.20)*	6	0.78 *(0.24)*	6	0.85 *(0.21)*
16+ years	13	0.65 *(0.12)*	4	0.70 *(0.13)*	1	0.63 *(-)*
Not indicated	2					
School type						
Primary (including Primary and Intermediate)	124	0.73 *(0.25)*	29	0.85 *(0.27)*	21	0.91 *(0.27)*
Secondary (including Secondary, Composite and Other)	11	0.87 *(0.33)*	2	1.25 *(0.09)*	2	1.12 *(0.56)*
School Economic Status of School Community						
Low decile	36	0.74 *(0.25)*	8	0.82 *(0.35)*	3	1.00 *(0.45)*
Medium decile	45	0.76 *(0.29)*	13	0.90 *(0.26)*	8	0.93 *(0.28)*
High decile	53	0.72 *(0.25)*	10	0.86 *(0.26)*	12	0.91 *(0.29)*

33

Although analysis of variance tests did not reveal statistically significant differences in the means of the absolute discrepancy levels on any of these demographic variables, there are descriptive differences that suggest further investigation. Namely, female principals, younger principals, those with less time as a principal, those with less time as a principal at a given school, principals at secondary schools, and principals at lower decile schools tended to have the greatest discrepancies in their self-versus teacher ratings. While these patterns are revealing about the kinds of principals and schools where there are discrepancies, they tell us nothing about their direction.

Predicting Direction of Discrepancy

To investigate direction of discrepancy, principals were categorized as under-raters, equal-raters, or over-raters. While there has been critique of a categorization approach when discrepancy is treated as a predictor, in our study the categorical variable of leadership discrepancy was the outcome variable to be predicted. Specifically, based on our earlier literature review, and on our developmental purposes, we saw the identification of over-raters as the most important category to understand and predict membership within.

Our three levels of categorization were determined through a conservative approach aimed at identifying extreme under- and over- raters (rather than predicting minor degrees of discrepancy on an interval-like scale). First, the proportion of teachers in each school that the principal over- or under-rated against was calculated. The average proportion of teachers that all principals under-rated against in their school was 55% (*SD* = 29%). A category threshold of 84% was set where any principal who self-rated below at least 84% of the principal's teachers was categorized as an under-rater. For those principals who over-rated compared to their

34

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
        Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
        Assessment, Evaluation and Accountability*

teachers, the average proportion was 42% (*SD* = 29%). A category threshold of 71% was set to

categorize principals who rated themselves at or beyond 71% of their teachers as over-raters.

This process produced three groups of principals: those with patterns of under-ratings (*n* = 31);

in agreement ratings (*n* = 81), or over-ratings (*n* = 23) compared to their teachers.

Referring to Table 5, it is possible to link the magnitudes of absolute differences to

specific discrepancy directions. For over-raters, the youngest age, least time as principal at the

school, and lowest decile categories were all associated with higher degrees of discrepancy

magnitudes in their respective variables. For under-raters, in contrast, the youngest age, least

time as principal at the school and lowest decile categories were associated with lesser degrees

of discrepancy magnitudes. In addition, over-raters show a steady pattern of increasing

discrepancy as their length of time as a principal increases while under-raters show a tendency

towards decreasing discrepancy magnitude as their length of time as principal increases.

A series of binomial logistic regression models were now constructed for the purpose of

understanding how these personal and school characteristics contribute to the direction of

discrepancies (Hosmer & Lemeshow, 1989; Menard, 1995). The criterion in these models was

a two-group under/over grouping variable for the principals. Recognizing the inherent risk of

using difference scores (which may, but not necessarily, exhibit greater degrees of unreliability

than their respective generating variable's scores (Fleenor et al., 2010; Thomas & Zumbo,

2012) to form either 3 or 4-group categories (Atwater et al., 1998), we chose to ignore the

"congruent" principals in order to focus on the two extreme discrepancy groups.

The initial predictors are presented in Table 2. The grouping variable is coded 1=over-

rater, 0=under-rater; this form coding yields a probability associated with each principal being

an over-rater. The focus is to determine which of the possible variables are significant

predictors of being an over-rater, how well the predictors actually classified the known

under/over status of each principal, and what prediction equation and classification strategy

would serve as a useful "at-risk of being an over-rater" indicator in practical applications of

these results.

The analyses started with single predictor logistic regressions wherein the maximum

predictive power of each predictor was tested. Then a forced-entry strategy was employed

where all predictors were included in the full model to test the unique contribution of each

predictor and the overall classification success based on all predictors. This was followed by a

backward selection strategy (using the likelihood ratio as the exclusion criterion) where all

predictors are entered into the model and then statistically non-significant ones are removed.

The results for the final, most parsimonious reduced model are presented in Table 6.

For principals who are older, with more years of experience as a principal, and at schools

with larger rolls but with less time at the school, the odds are increased that they will be an

over-rater principal. As shown in Table 6, the positive B coefficients for age, years as a

principal, and school roll indicate that as the predictor variable values increase, so does the

probability of being an over-rater. The negative B coefficient indicates that principals with less

time at the school have higher probabilities of being over-raters. The summary statistics

indicate an excellent fit for this model: a) model chi-square $= 26.129$, $p < .001$, b) Nagelkerke

$R^2 = .515$, c) Cox & Snell $R^2 = .384$, and d) Hosmer and Lemeshow Test chi-square $= 6.581$, df

$= 8$, $p = .582$.

Table 6

Significant Predictors of Under/Over-rater Status

Variables in the Equation[a]	B	S.E.	Wald	df	Sig.	Exp(B)
Principal age	2.535	.941	7.263	1	.007	12.620
Years as a principal	1.172	.492	5.678	1	.017	3.227
Years as a principal at the school	-1.515	.544	7.746	1	.005	.220
School total roll	.003	.001	5.521	1	.019	1.003
Constant	-6.052	2.031	8.877	1	.003	.002

a. The criterion variable is coded: 1 = Over-rater, 0 = Under-rater.

The column labeled "Exp(B)" converts the B coefficient logits into odds of being an over-rater. For example, for each unit increase in principal age (40-49 years to 50-59 years, and 50-59 years to 60+ years), the odds of being an over-rater versus an under-rater increase by 12.6. For each unit increase in length of role as a principal, the odds of being an over-rater increase by 3.2. For each 1000 unit increase (number of students) in school roll, the odds of being an over-rater increase by 3. For length of role as principal at the school the odds of being an over-rater increase by 4.5 with each unit decrease in time at the school.

From a research perspective the results in Table 6 suggest variables to bear in mind when principal effectiveness is evaluated from either the principal's or teacher's point of view. From a practical perspective, however, it is worthwhile to use the coefficients to estimate the probability that any given principal will be an over-rater. The estimation equation takes the following form:

Estimated probability of being an over-rater = (exp(-6.052 + 2.535(age) + 1.172(years as principal) -1.515(years as principal at school) +.003(school total roll)) / (1+ exp(-6.052 + 2.535(age) + 1.172(years as principal) -1.515(years as principal at school)).

The probability of being an over-rater ("O") for each of the 135 principals, generated from

the above equation, is presented in Figure 2. It is apparent that principals with high over-rater

probabilities (p=>.6) were correctly identified by this procedure. Between p=.3 to .6, there is

over-lap between over- and under-raters. Probabilities of less than .3 correctly identify

principals as largely under-raters. Furthermore, this equation may be employed proactively to

yield the probability of being an over-rater for any principal with characteristics similar to the
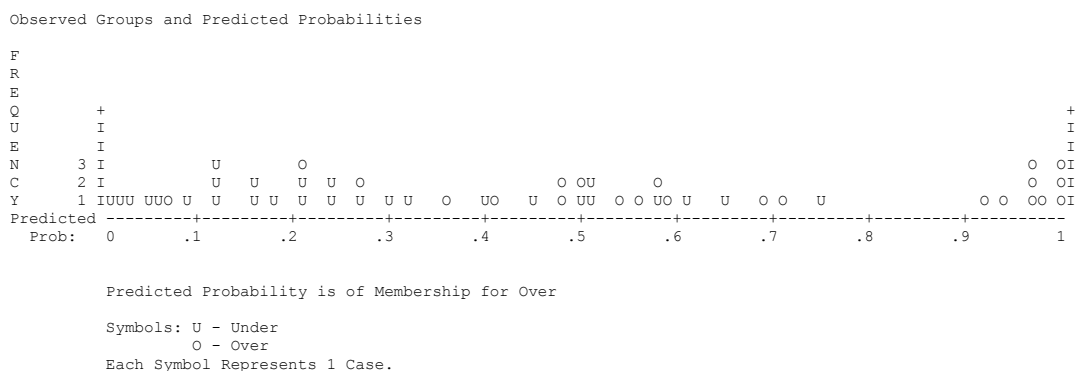
principals in the present study.

```
Observed Groups and Predicted Probabilities

F
R
E
Q       +                                                                                    +
U       I                                                                                    I
E       I                                                                                    I
N    3 I         U       O                                                           O  OI
C    2 I         U   U    U  U  O                 O OU      O                         O  OI
Y    1 IUUU UUO U  U   U U  U  U  U U    O   UO  O O UU O O UO U   U   O O   U        O O  OO OI
Predicted --------+---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:  0      .1        .2        .3        .4        .5        .6        .7        .8        .9        1


          Predicted Probability is of Membership for Over

          Symbols: U - Under
                   O - Over
          Each Symbol Represents 1 Case.
```

Figure 2. Predicted group membership from final model.

Discussion

In the current study both school and principal level characteristics predicted principals

who are over-raters. In particular, a principal's age, years of experience (as a principal and at

their current school) and the size of the school they lead (school roll) was predictive of their

status as over-raters. Principals were more likely to be over-raters when they were older, had

more experience as a principal overall but less time as a principal in their current school, and

when they worked in larger schools. This finding has implications for the identification of

principals who may need support in developing trust, team work and shared teacher-leader

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
        Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational
        Assessment, Evaluation and Accountability*

expectations. These principals, given their greater likelihood of being over-raters, may benefit

most from interventions focused on leadership effectiveness generally and relational trust in

particular.

The potential to identify these principals, and address their development needs is

particularly useful in the New Zealand context described earlier, where there are high

achievement levels but low levels of equity for the achievement of diverse learner groups.

While such issues are of concern for the wider New Zealand system, they are also of concern at

the school level, and might be reduced through improved principal effectiveness. The need to

improve principal effectiveness and particularly their efforts to strengthen relational trust and

team-work in ways that support student achievement is not, however, exclusive to the New

Zealand context. This research seeks to inform the work of those in other contexts where there

are similar calls to improve principal effectiveness and similar calls for school leadership to

contribute to reducing achievement disparity.

Unlike studies from non-educational contexts, gender was not a significant predictor of

over-raters (and our descriptive statistics showed female principals to have greater

discrepancies than males). This is of interest since previous studies of predictors of agreement

indicated that men rate themselves more positively than women, and also that gender

differences disappear when self-ratings are made with a promise of confidentiality (Brutus et

al., 1999; Daubman, Heatherington, & Ahn, 1992; Ellen Goldring et al., 2009). Why then were

gender differences absent in our study, despite our participants not having a promise of

confidentiality?

In fact, participants were likely to treat their own and others' data in a quite public way.

Principals in our study were encouraged and typically willing, to share summaries comparing

their own and others' ratings of their effectiveness with their staff, their mentor, and their

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal?
Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

peers. Their anticipation of publicly sharing their data, we argue, may have served to reduce the tendency for gender differences that might otherwise have been the case. The threat of being seen publicly to over-rate oneself may have led principals to give themselves more moderated ratings than might otherwise have been the case.

Our examination of different cut score probabilities (where we used the default software criterion of .5 and then changed it to .4 and .6) to classify principals as over- or under-raters leads us to recommend .4 for such analyses. While there were more overall incorrect classifications at this level than at other levels (referring to Figure 2 and Table 7), more of the observed over-raters (19/23) were correctly classified. If this classification procedure, using the above equation, is used to identify the most likely participants in need of leadership development work, the cost of incorrectly classifying some under-rating principals as over-raters (e.g., incurring the extra expense of professional development that may not necessarily be needed) would presumably be lower than the cost of incorrectly classifying some over-rating principals as under-raters (e.g., incurring the "cost" of diminished educational achievement and subsequent organizational disruption because they received no extra development work). Furthermore, since under-rating principals are still likely to benefit from interventions aimed at reducing discrepancy, there should be little issue with including those under-raters who are incorrectly classified as over-raters in such interventions.

Table 7

Cut-score Classification Strategies

Cut score	Observed	Predicted	
		Under	Over
Cut score p = .60	Under	28	3
	Over	12	11
Cut score p = .50	Under	25	6
	Over	8	15
Cut score p = .40	Under	23	8
	Over	4	19

Reducing Discrepancy

Findings from empirical studies and theoretical models on self-other perception, such as Funder's (1995) Realistic Accuracy Model (RAM) are useful for considering how to reduce discrepancy. They have identified, for example, visibility of dimensions being judged as a useful predictor of agreement. Greater consensus and self-other agreement has been found for more visible traits (Kenny & West, 2010). Efforts to reduce discrepancies in schools where they exist, might then benefit from research that qualitatively reveals more about how school leaders make their effectiveness visible. What leader behaviors, for example, are observable by teachers in relation to the items in the PE scale? The issue of visibility is germane to the choice of and purpose of self-other comparisons. If the primary purpose of multi-rater evaluation is to gain a more accurate assessment of principal effectiveness, then it might be preferable to obtain ratings from those, such as assistant principals, with rich opportunities to observe and interact with the principal (Grissom & Loeb, 2011). If, on the other hand, the primary purpose is to assess organizational coherence rather than accuracy, then provision of other ratings by the wider teacher group makes sense.

41

Key implications for leadership development policy and practice are the need to attend to the potential for discrepancy at the organizational level and particularly to the possibility that principals may rate themselves more highly than their teachers. The availability of tools that schools can use to evaluate perceptions of the quality of leadership in their schools (both school-wide and individual principal) is a key to such comparisons, as is support for principals in leading staff discussion of their school's results. In particular, the causes and consequences of any discrepancy should be central to the discussion. Different perceptions of effectiveness overall are noteworthy, as are different perceptions on any one item in a scale. But more critical still are the consequences of discrepancy for school performance, since perceptions of the quality of leadership are a key determinant of teacher-leader trust (Halverson et al., 2005). For this reason we recommend further research that examines empirically the causes and consequences of discrepancy especially in schools where it reflects principal over-rating. Attention to levels of effectiveness alongside measures of discrepancy would also make a useful contribution. One fruitful strategy would be a series of case studies of schools with varying magnitude and direction of principal-teacher discrepancy, and varying levels of effectiveness. A case study design would enable the cultural and interpersonal processes that generate, sustain and reduce discrepancy to be identified. Quantitative research with repeated measures of links between discrepancy, relational trust and school improvement would also give us important insights into the ways that different types of discrepancy produce differentially effective school organizations.

Ongoing research attention to discrepancy in principal leadership, and also to school-wide leadership is important from both a system and a school perspective. From a school perspective, the ability to identify and explain the direction and magnitude of principal-teacher discrepancy is important since reducing discrepancy is so clearly linked to desirable conditions

for school improvement. It is also important from a system perspective, since a system served

by schools in which there are high levels of coherence, trust, and shared expectations between

leaders and teachers is more conducive to sustained improvement of both the professional

experience of educators in schools, and most importantly the achievement of learners in those

school

References

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51*(3), 577-598. doi: doi: 10.1111/j.1744-6570.1998.tb00252.x

Atwater, L. E., & Waldman, D. (1998). 360 degree feedback and leadership development. *The Leadership Quarterly, 9*(4), 423-426. doi: doi: 10.1016/S1048-9843(98)90009-1

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. doi: doi: 10.1111/j.1744-6570.1992.tb00848.x

Atwater, L. E., & Yammarino, F. J. (1997). Self–other rating agreement: A review and model. In G. R. Ferris (Ed.), *Research in personnel and human resources management* (pp. 121-174): JAI Press.

Avolio, B. J., Mhatre, K., Norman, S. M., & Lester, P. (2009). The moderating effect of gender on leadership intervention impact: An exploratory review. *Journal of Leadership & Organizational Studies, 15*(4), 325-341. doi: doi: 10.1177/1548051809333194

Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930-942. doi: 10.1037/0021-9010.86.5.930

Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development, 18*, 417-435. doi: doi: 10.1108/02621719910273569#sthash.dkq9ZYtV.dpuf

Bryk, A. S., & Schneider, B. L. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation.

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.

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. doi: doi: 10.1111/j.1559-1816.1997.tb00283.x

Daubman, K. A., Heatherington, L., & Ahn, A. (1992). Gender and the self-presentation of academic achievement. *Sex Roles, 27*, 187-204. doi: doi: 10.1007/BF00290017

Day, C., Sammons, P., Leithwood, K., Hopkins, D., Gu, Q., Brown, E., & Ahtaridou, E. (2011). *Successful School Leadership: Linking with Learning and Achievement*. Maidenhead: McGraw Hill Open University Press.

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology, 46*, 641-665. doi: 10.1111/j.1744-6570.1993.tb00889.x

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*(3), 265-287. doi: 10.1177/109442810143005

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self–other rating agreement in leadership: A review. *The Leadership Quarterly, 21*, 1005-1034. doi: doi: 10.1016/j.leaqua.2010.10.006

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670. doi: 0033-295X/95/S3.00

Goddard, R. D., Tschannen-Moran, M., & Hoy, W. K. (2001). A multilevel examination of the distribution and effects of teacher trust in students and parents in urban elementary schools. *The Elementary School Journal, 102*(1), 3-17. doi: doi: 10.2307/1002166

Goldring, E., Cravens, X. C., Murphy, J., Porter, A. C., Elliott, S. N., & Carson, B. (2009). The evaluation of principals: What and how do states and urban districts assess leadership? *The Elementary School Journal, 110*(1). doi: 10.1086/598841

Goldring, E., & Goff, P. T. (2010). *Understanding feedback from teachers on leadership effectiveness: Do teachers and principals agree?* Paper presented at the American Educational Research Association (AERA), Denver.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal, 48*(5), 1091-1123. doi: doi: 10.3102/0002831211402663

Halverson, S. K., Tonidandel, S., Barlow, C. B., & Dipboye, R. L. (2005). Self–other agreement on a 360-degree leadership evaluation. In S. Reddy (Ed.), *Multi-source performance assessment: Perspective and Insights*. Hyderabad, India: ICFAI University Press.

Heck, R. H. (1992). Principals' instructional leadership and school performance: Implications for policy development. *Educational Evaluation and Policy Analysis, 14*(1), 21-34. doi: 10.3102/01623737014001021

Heck, R. H., & Hallinger, P. (2010). Testing a longitudinal model of distributed leadership effects on school improvement. *The Leadership Quarterly, 21*(5), 867-885. doi: 10.1016/j.leaqua.2010.07.013

Heck, R. H., Marcoulides, G. A., & Lang, P. (1991). Principal instructional leadership and school achievement: The application of discriminant techniques. *School Effectiveness and School Improvement, 2*(2), 115-135. doi: 10.1080/0924345910020204

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley.

Hu, L. t., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi: doi: 10.1080/10705519909540118

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self-and other perception: a meta-analysis. *Personality and Social Psychology Review, 14*(2), 196-213. doi: 10.1177/1088868309353414

Kwan, V. S. Y., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology, 94*, 1062-1077. doi: 10.1037/0022-3514.94.6.1062

Latham, G. P., & Locke, E. A. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*(5), 265-268. doi: 10.1111/j.1467-8721.2006.00449.x

Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal-setting research. *European Psychologist, 12*(4), 290-300. doi: 10.1027/1016-9040.12.4.290

Le Fevre, D., & Robinson, V. M. J. (2014). The interpersonal challenges of instructional leadership: Principals' effectiveness in conversations about performance issues. *Educational Administration Quarterly*, 1-38. doi:10.1177/0013161X13518218

Le Fevre, D., Robinson, V. M. J., & Sinnema, C. E. L. (In press). Genuine inquiry: Widely espoused yet rarely enacted. *Education Management and Leadership*.

Author copy - submitted and accepted version of the following publication:
Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal? Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Menard, S. (1995). *Applied Logistic Regression Analysis*. Thousand Oaks, California: Sage.

Minor, E. C., Porter, A. C., Murphy, J., Goldring, E., Cravens, X., & Elliott, S. N. (2014). A known group analysis validity study of the Vanderbilt Assessment of Leadership in Education in US elementary and secondary schools. *Educational Assessment, Evaluation and Accountability, 26*(1), 29-48. doi: 10.1007/s11092-013-9180-z

Murphy, J. (2013). The architecture of school improvement. *Journal of Educational Administration, 51*(3), 252-263. doi: 10.1108/09578231311311465#sthash.BUnYQyul.dpuf

Ostroff, C., Atwater, L. E., & Feinberg, B. J. (2004). Understanding self-other agreement: A Look at rater and ratee characteristics, context, and outcomes. *Personnel Psychology, 57*, 333-375. doi: doi: 10.1111/j.1744-6570.2004.tb02494.x

Paulhus, D. L., & Bruce, M. N. (1992). The effect of acquaintanceship on the validity of personality impressions: A longitudinal study. *Journal of Personality and Social Psychology, 63*(5), 816-824. doi: 10.1037/0022-3514.63.5.816

Pont, B., Figueroa, D. T., Zapata, J., & Fraccola, S. (2013). Education outlook: New Zealand (pp. 1-24).

Reeves, D. B. (2008). *Assessing educational leaders: evaluating performance for improved individual and organizational results (2nd ed.)*. Thousand Oaks, CA: Corwin Press.

Robinson, V., Lloyd, C., & 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. doi: 10.1177/0013161X08321509

Robinson, V., & Timperley, H. S. (2007). The leadership of the improvement teaching and learning: Lessons from initiatives with positive outcomes for students. *Australian Journal of Education, 51*(3), 247-262. doi: 10.1177/000494410705100303

Robinson, V. M. J. (2011). *Student-centered leadership*. San Francisco: Jossey-Bass.

Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences, 3*, 153-160.

Sinnema, C. E. L., Le Fevre, D., Robinson, V. M. J., & Pope, D. (2013). When others' performance just isn't good enough: Educational leaders' framing of concerns in private and public. *Leadership and Policy in Schools, 12*(4), 301-336. doi: doi: 10.1080/15700763.2013.857419

Sinnema, C. E. L., & Robinson, V. M. J. (2007). The leadership of teaching and learning: Implications for teacher evaluation. *Leadership and Policy in Schools, 6*(4), 319-343. doi: 10.1080/15700760701431603

Sinnema, C. E. L., & Robinson, V. M. J. (2012). Goal setting in principal evaluation: Goal quality and predictors of achievement. *Leadership and Policy in Schools, 11*(2), 135-167. doi: doi: 10.1080/15700763.2011.629767

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. *Personnel Psychology, 58*(1), 33-66. doi: 10.1111/j.1744-6570.2005.514_1.x

Sun, M., Youngs, P., Yang, H., Chu, H., & Zhao, Q. (2012). Association of district principal evaluation with learning-centered leadership practice: Evidence from Michigan and Beijing. *Educational Assessment, Evaluation and Accountability 24*, 189-213. doi: 10.1007/s11092-012-9145-7

Sinnema, C., Robinson, V. M. J., Ludlow, L. H., & Pope, D. (In press). How Effective is the Principal? Discrepancy between Teachers' and Principals' Perceptions of Principal Effectiveness. *Educational Assessment, Evaluation and Accountability*

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and Psychological Measurement, 72*(1), 37-43. doi: 10.1177/0013164411409929

Vecchio, R. P., & Anderson, R. J. (2009). Agreement in Self–Other Ratings of Leader Effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment, 17*, 165-179. doi: 10.1111/j.1468-2389.2009.00460.x

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What makes you think you're so smart? Measured abilities, personality, and sex differences in relation to self-estimates of multiple intelligences. *Journal of Individual Differences, 29*, 35-44. doi: 10.1027/1614-0001.29.1.35

Appendix

Principal and Teacher Versions of PE Scale

Principal: How effective are you in...

Teacher: How effective is the principal of your school in...

1. using research on teaching and learning to inform important school decisions?

2. learning alongside teachers about how to improve teaching and learning?

3. serving the interests of the whole school rather than of particular interest groups?

4. leading useful discussions about the improvement of teaching and learning?

5. identifying and resolving conflict quickly and fairly?

6. promoting and modeling the values of this school?

7. maintaining integrity in difficult situations?

8. showing both personal and professional respect for staff?

9. earning the respect of all of the staff?

10. earning the respect of the wider community?

11. earning the respect of the different ethnic communities served by the school?

12. seeking high quality information about the situation before making a final decision?

13. being open to learning and admitting mistakes?

14. saying what I think and explaining why?

15. actively seeking others' views?

16. making tough decisions when necessary?